BRNO UNIVERSITY OF TECHNOLOGY
VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION
FAKULTA ELEKTROTECHNIKY
A KOMUNIKAČNÍCH TECHNOLOGIÍ

DEPARTMENT OF BIOMEDICAL ENGINEERING
ÚSTAV BIOMEDICÍNSKÉHO INŽENÝRSTVÍ

COMPUTATIONAL ANALYSES AND ANNOTATIONS OF NON-MODEL BACTERIA FOR WHITE BIOTECHNOLOGY
VÝPOČETNÍ ANALÝZY A ANOTACE NEMODELOVÝCH BAKTERIÍ PRO BÍLOU BIOTECHNOLOGII

HABILITATION THESIS
HABILITAČNÍ PRÁCE

AUTHOR **Mgr. Ing. Karel Sedlář, Ph.D.**
AUTOR PRÁCE

BRNO 2023

ABSTRACT

The habilitation thesis presents a commented collection of 16 published papers by the author on the computational processing of sequencing data in order to assemble and analyze genomes and transcriptomes of non-model bacteria that could be used in white biotechnology, i.e., a technology that uses living cells to synthesize the easily degradable products. Bioinformatics analyses of non-model organisms are specific because many computational tools require the use of datasets that are unavailable for novel bacteria due to non-existing microbiological kits to perform desired experiments or simply due to missing knowledge of required input data. This thesis presents the author's personal experience with overcoming these obstacles by using unique combinations of computational tools and techniques in order to maximize information gains from analyses of novel genomes of non-model bacteria.

KEYWORDS

next generation sequencing; third generation sequencing; functional annotation; non-model bacterium; white biotechnology; transcriptomics

ABSTRAKT

Tato habilitační práce představuje komentovaný soubor autorových 16 publikovaných prací o výpočetním zpracování sekvenačních dat za účelem sestavení a analýzy genomů a transkriptomů nemodelových bakterií využitelných v bílé biotechnologii, tj. technologii využívající živé buňky k syntéze snadno rozložitelných produktů. Bioinformatické analýzy nemodelových organismů jsou specifické, protože mnoho výpočetních nástrojů vyžaduje použití datových sad, které jsou pro nové bakterie nedostupné, buď kvůli neexistujícím mikrobiologickým kitům k provádění požadovaných experimentů, nebo jednoduše kvůli chybějícím požadovaným vstupním datům. Tato práce představuje autorovy osobní zkušenosti s překonáváním těchto překážek pomocí unikátních kombinací výpočetních nástrojů a technik s cílem maximalizovat získané informace z analýz nových genomů nemodelových bakterií.

KLÍČOVÁ SLOVA

sekvenování nové generace; sekvenování třetí generace; funkční anotace; nemodelová bakterie; bílá biotechnologie; transkriptomika

ACKNOWLEDGEMENT

Firstly, I would like to thank my wife Lucka for her endless support and patience, I would not be able to reach my goals without her; and my family for their love and understanding. I would also like to thank my colleagues, not only my former supervisors and mentors, Valentýna Provazník, Petra Patáková, and Ralf Zimmer from whom I learned a lot, but also other collaborators and students, especially Stanislav Obruča, Barbora Branská, Helena Škutková, Jana Musilová, Kateřina Jurečková, Markéta Nykrýnová and many others with whom I had a pleasure to work over the past few years and who directly or indirectly contributed to the ideas presented in this thesis.

CONTENTS

1 INTRODUCTION

1.1 My Contribution to the Presented Articles

By the time of writing this thesis, I have published 48 (14 as the first or corresponding author) articles in impacted international journals, 16 conference articles indexed in Web of Science, and one patent. This thesis presents a commented collection of 16 peer-reviewed papers published between the years 2015 and 2022 in collaboration with my colleagues and my students from the Department of Biomedical Engineering, Brno University of Technology and Department of Informatics, Ludwig Maximilian University of Munich, where I worked in this time period, and also with other national and international collaborators. My contribution to the 16 selected studies in terms of research and manuscript preparation, supervision of students, and research direction is summarized in the tables below. Here, the papers are listed chronologically according to the publication date:

[I.] **SEDLAR, Karel**, KOLEK, Jan, SKUTKOVA, Helena, BRANSKA, Barbora, PROVAZNIK, Ivo and PATAKOVA, Petra. Complete genome sequence of Clostridium pasteurianum NRRL B-598, a non-type strain producing butanol. *Journal of Biotechnology*. 2015. 214, p. 113–114. (2015 IF = 2.667, Q2 in BIOTECHNOLOGY & APPLIED MICROBIOLOGY)

Research and manuscript (%)	Supervision (%)	Research direction (%)
51	-	10

[II.] KOLEK, Jan, **SEDLAR, Karel**, PROVAZNIK, Ivo and PATAKOVA, Petra. Dam and Dcm methylations prevent gene transfer into Clostridium pasteurianum NRRL B-598: Development of methods for electrotransformation, conjugation, and sonoporation. *Biotechnology for Biofuels*. 2016. 9(1), p. 14. (2016 IF = 5.203, Q1 in BIOTECHNOLOGY & APPLIED MICROBIOLOGY)

Research and manuscript (%)	Supervision (%)	Research direction (%)
30	-	10

[III.] **SEDLAR, Karel**, KOLEK, Jan, PROVAZNIK, Ivo and PATAKOVA, Petra. Reclassification of non-type strain Clostridium pasteurianum NRRL B-598 as Clostridium beijerinckii NRRL B-598. *Journal of Biotechnology*. 2017. 244, p. 1–3. (2017 IF = 2.533, Q2 in BIOTECHNOLOGY & APPLIED MICROBIOLOGY)

Research and manuscript (%)	Supervision (%)	Research direction (%)
70	-	30

[IV.] SNOPKOVÁ, Kateřina, **SEDLÁŘ, Karel**, BOSÁK, Juraj, CHALOUPKOVÁ, Eva, SEDLÁČEK, Ivo, PROVAZNÍK, Ivo and ŠMAJS, David. Free-Living Enterobacterium Pragia fontium 24613: Complete Genome Sequence and Metabolic Profiling. *Evolutionary Bioinformatics*. 2017. 13, p. 1176934317700863. (2017 IF = 1.877, Q2 in MATHEMATICAL & COMPUTATIONAL BIOLOGY)

Research and manuscript (%)	Supervision (%)	Research direction (%)
30	-	-

[V.] PATAKOVA, Petra, KOLEK, Jan, **SEDLAR, Karel**, KOSCOVA, Pavlina, BRANSKA, Barbora, KUPKOVA, Kristyna, PAULOVA, Leona and PROVAZNIK, Ivo. Comparative analysis of high butanol tolerance and production in clostridia. *Biotechnology Advances*. 2018. 36(3), p. 721–738. (2018 IF = 12.831, D1 in BIOTECHNOLOGY & APPLIED MICROBIOLOGY)

Research and manuscript (%)	Supervision (%)	Research direction (%)
12.5	25	25

[VI.] **SEDLAR, Karel**, KOSCOVA, Pavlina, VASYLKIVSKA, Maryna, BRANSKA, Barbora, KOLEK, Jan, KUPKOVA, Kristyna, PATAKOVA, Petra and PROVAZNIK, Ivo. Transcription profiling of butanol producer Clostridium beijerinckii NRRL B-598 using RNA-Seq. *BMC Genomics*. 2018. 19(1), p. 415. (2018 IF = 3.501, Q2 in BIOTECHNOLOGY & APPLIED MICROBIOLOGY)

Research and manuscript (%)	Supervision (%)	Research direction (%)
45	30	40

[VII.] **SEDLAR, Karel**, KOLEK, Jan, GRUBER, Markus, JURECKOVA, Katerina, BRANSKA, Barbora, CSABA, Gergely, VASYLKIVSKA, Maryna, ZIMMER, Ralf, PATAKOVA, Petra and PROVAZNIK, Ivo. A transcriptional response of Clostridium beijerinckii NRRL B-598 to a butanol shock. *Biotechnology for Biofuels*. 2019. 12(1), p. 243. (2019 IF = 4.815, Q1 in BIOTECHNOLOGY & APPLIED MICROBIOLOGY)

Research and manuscript (%)	Supervision (%)	Research direction (%)
60	10	40

[VIII.] VASYLKIVSKA, Maryna, BRANSKA, Barbora, **SEDLAR, Karel**, JURECKOVA, Katerina, PROVAZNIK, Ivo and PATAKOVA, Petra. Phenotypic and Genomic Analysis of Clostridium beijerinckii NRRL B-598 Mutants With Increased Butanol Tolerance. *Frontiers in Bioengineering and Biotechnology*. 2020. 8, p. 1307. (2020 IF = 5.890, Q1 in MULTIDISCIPLINARY SCIENCES)

Research and manuscript (%)	Supervision (%)	Research direction (%)
20	15	30

[IX.] **SEDLAR, Karel**, VASYLKIVSKA, Maryna, MUSILOVA, Jana, BRANSKA, Barbora, PROVAZNIK, Ivo and PATAKOVA, Petra. Phenotypic and genomic analysis of isopropanol and 1,3-propanediol producer Clostridium diolis DSM 15410. *Genomics*. 2021. 113(1), p. 1109–1119. (published online in 2020, 2020 IF = 5.736, Q1 in BIOTECHNOLOGY & APPLIED MICROBIOLOGY)

Research and manuscript (%)	Supervision (%)	Research direction (%)
50	20	35

[X.] MUSILOVA, Jana, KOURILOVA, Xenie, BEZDICEK, Matej, LENGEROVA, Martina, OBRUCA, Stanislav, SKUTKOVA, Helena, and **SEDLAR, Karel**. First Complete Genome of the Thermophilic Polyhydroxyalkanoates-Producing Bacterium Schlegelella thermodepolymerans DSM 15344. *Genome Biology and Evolution*. 2021. 13(4), p. evab007. (2021 IF = 4.065, Q2 in EVOLUTIONARY BIOLOGY)

Research and manuscript (%)	Supervision (%)	Research direction (%)
20	50	60

[XI.] JURECKOVA, Katerina, RASCHMANOVA, Hana, KOLEK, Jan, VASYLKIVSKA, Maryna, BRANSKA, Barbora, PATAKOVA, Petra, PROVAZNIK, Ivo and **SEDLAR, Karel**. Identification and Validation of Reference Genes in Clostridium beijerinckii NRRL B-598 for RT-qPCR Using RNA-Seq Data. *Frontiers in Microbiology*. 2021. 12, p. 476. (2021 IF = 6.064, Q1 in MICROBIOLOGY)

Research and manuscript (%)	Supervision (%)	Research direction (%)
15	50	50

[XII.] **SEDLAR, Karel**, NYKRYNOVA, Marketa, BEZDICEK, Matej, BRANSKA, Barbora, LENGEROVA, Martina, PATAKOVA, Petra, and SKUTKOVA, Helena. Diversity and Evolution of Clostridium beijerinckii and Complete Genome of the Type Strain DSM 791T. *Processes*. 2021. 9(7), p. 1196. (IF = 3.352, Q2 in ENGINEERING, CHEMICAL)

Research and manuscript (%)	Supervision (%)	Research direction (%)
45	30	50

[XIII.] HERMANKOVA, Kristyna, KOURILOVA, Xenie, PERNICOVA, Iva, BEZDICEK, Matej, LENGEROVA, Martina, OBRUCA, Stanislav, and **SEDLAR, Karel**. Complete Genome Sequence of the Type Strain Tepidimonas taiwanensis LMG 22826T, a Thermophilic Alkaline Protease and Polyhydroxyalkanoate Producer. *Genome Biology and Evolution*. 2021. 13(12), p. evab280. 2021 IF = 4.065, Q2 in EVOLUTIONARY BIOLOGY)

Research and manuscript (%)	Supervision (%)	Research direction (%)
15	50	50

[XIV.] PATAKOVA, Petra, BRANSKA, Barbora, VASYLKIVSKA, Maryna, JURECKOVA, Katerina, MUSILOVA, Jana, PROVAZNIK, Ivo, and **SEDLAR, Karel**. Transcriptomic studies of solventogenic clostridia, Clostridium acetobutylicum and Clostridium beijerinckii. *Biotechnology Advances*. 2022. 58, p. 107889. (published online in 2021, 2021 IF = 17.681, D1 in BIOTECHNOLOGY & APPLIED MICROBIOLOGY)

Research and manuscript (%)	Supervision (%)	Research direction (%)
20	25	35

[XV.] OBRUČA, Stanislav, DVOŘÁK, Pavel, SEDLÁČEK, Petr, KOLLER, Martin, **SEDLÁŘ, Karel**, PERNICOVÁ, Iva, and ŠAFRÁNEK, David. Polyhydroxyalkanoates synthesis by halophiles and thermophiles: towards sustainable production of microbial bioplastics. *Biotechnology Advances*. 2022. 58, p. 107906. (2022 IF not announced yet, 2021 IF = 17.681, D1 in BIOTECHNOLOGY & APPLIED MICROBIOLOGY)

Research and manuscript (%)	Supervision (%)	Research direction (%)
15	-	15

[XVI.] MUSILOVA, Jana, KOURILOVA, Xenie, PERNICOVA, Iva, BEZDICEK, Matej, LENGEROVA, Martina, OBRUCA, Stanislav, and **SEDLAR, Karel**. Novel thermophilic polyhydroxyalkanoates producing strain Aneurinibacillus thermoaerophilus CCM 8960. *Applied Microbiology and Biotechnology*. 2022. 106(12), p. 4669–4681. (2022 IF not announced yet, 2021 IF = 5.560, Q1 in BIOTECHNOLOGY & APPLIED MICROBIOLOGY)

Research and manuscript (%)	Supervision (%)	Research direction (%)
15	50	50

A common topic in these papers is assembling novel bacterial genomes and their structural and functional annotation supplemented by transcriptomic data processing in order to infer new knowledge on regulations within bacterial cell factories that could be utilizable for white biotechnology, i.e., a technology that uses living cells to synthesize the easily degradable products. Thanks to the particular cooperation with biotechnology researchers, this thesis aims primarily at bacteria capable of producing biofuels such as butanol and hydrogen or bioplastics in the form of polyhydroxyalkanoates. Nevertheless, the utilized computational approaches are widely applicable to other non-model bacteria, which is demonstrated by references to additional papers I coauthored not included in the presented collection.

The papers are hard to be sorted into particular sections, as from the bioinformatics point of view, a majority of them uses combinations of several computational approaches to infer biological knowledge, yet together, they present a comprehensive overview of how bacterial data can be computationally processed. For this reason, some of the attached papers are cited repeatedly in different subchapters of the following introductory text. The collection contains also three review papers. While two of them directly discuss the application of bioinformatics to biotechnology and applied microbiology, the other demonstrates how the such application can be used to review biotechnological knowledge.

During the years I worked on bacterial genomics and transcriptomics, I participated in many other studies dealing with various bacteria. These studies, while introducing much new on bacterial biology and biotechnology, used the same or very similar computational approaches as papers within this collection or are not aimed at potential industrial use and, thus, would be redundant to be included on the list. Yet, they are cited in the introductory text, where appropriate, together with other papers describing novel computational procedures I participated in. All 34 articles and one patent I coauthored are highlighted in bold in the References section.

1.2 Bioinformatics for Bacterial Biotechnology

Computational processing of data obtained during experiments with non-model bacteria is quite specific, which only reflects the nature of the bacterial domain of life. Bacteria represent an extremely diverse group of organisms covering a majority of 10^6 to 10^8 separate prokaryotic genospecies present on Earth [1]. Moreover, the majority (>99%) of microorganisms seems to be uncultivable by standard techniques [2], which makes the investigation of non-model bacteria difficult. Novel specialized techniques for their isolations are being continuously developed, e.g. osmoselection [3], proposed for the isolation of polyhydroxyalkanoates (PHAs) producing thermophiles. One such potent PHAs producer obtained using this technique is *Aneurinibacillus thermoaerophilus* CCM 8960 whose genome was recently assembled and analyzed by our group in the study by Musilova et al., 2022 [4] **(Article [XVI.])**. Once a bacterium is isolated and suitable culture conditions are found, another problem arises in obtaining material for the following analyses. There are various sampling and DNA/RNA isolation kits, some are optimized for Gram-positive or Gram-negative bacteria and some can be used for both. Nevertheless, particular sampling and isolation kits affect the quality of obtained DNA as my colleague and I proved in the study by Videnska et al. 2019 [5]. This might be an issue for particular computational analyses comparing results across multiple studies. Finally, when samples are processed by a lab device, primarily a sequencing machine, data are generated and various bioinformatics analyses are performed.

There are, without exaggeration, hundreds of computational tools and packages to perform various kinds of bioinformatics analyses that could be applicable to bacterial data. Many of these tools are able to perform similar tasks, yet with different results. It is usually the job of a bioinformatician to try several approaches and select the most suitable results for particular tasks. Despite the number of various tools, the overall strategy of data processing is the same for all studies of bacteria utilizable for industrial biotechnology. Such a pipeline was described by me and my colleagues in the review by Patakova et al., 2018 [6] **(Article [V.])**. Although the review is related to research on clostridia and their ability to produce butanol, the same steps can be performed with any bacterium producing any valuable compound, see Fig. 1.1. To understand observed phenotype manifestation that would be utilizable in biotechnology, one has to usually begin with genome sequencing which is followed by a range of bioinformatics procedures. Sequencing reads need to be assembled into sequences of genomes that are further structurally annotated to define particular genes [7–11] **(Articles [XIII.], [X.], [IV.], [I.], [IX.])** or mapped to a reference genome for direct comparison [12] **(Article [VIII.])**. Genomes are then usually used for phylogenetic analyses and correct identification of bacteria as previous phenotype identification may be unsuccessful [4] **(Article [XVI.])** or erroneous [13] **(Article [III.])**. Nowadays, even large phylogenomic studies comparing tens or even hundreds of whole genomes are possible [14] **(Article [XII.])**. Revealed relations to already known genomes can help to find orthologues of selected genes [15] **(Article [XV.])** and perform functional annotation. Possibly, sequencing data can be mined for additional, for example epigenetic, information [16] **(Article [II.])**. The functional annotation itself provides only a potential functional capacity of an organism, transcriptomics is needed to reveal which processes are running under different conditions [17, 18] **(Articles [VI.], [VII.])**. A majority of current studies rely on genome-wide transcription measured with RNA-Seq [19] **(Article [XIV.])**, nevertheless, transcription of selected genes can be measured with much cheaper reverse-transcriptase quantitative polymerase chain reaction (RT-qPCR) when suitable reference genes are found by a bioinformatics

analysis of RNA-Seq data [20] **(Article [XI.])**. Genomic and transcriptomic information can be further connected in gene regulatory networks, genome-scale or other models by the means of systems biology to mark suitable parts of genomes for further engineering by the tools of synthetic biology, which is, however, out of the scope of this thesis.

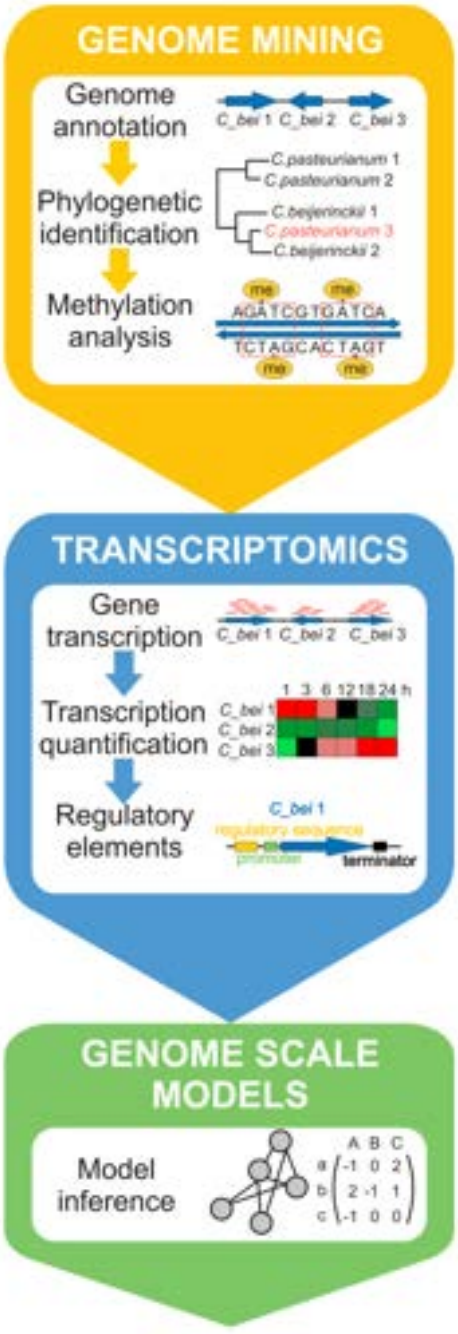


Fig. 1.1 Overview of bioinformatics analysis, taken from [6] (Article [V.])
Different bioinformatics tools can be used to perform desired particular tasks. Since a majority of tools is controlled through the command line interface, simple scripting using bash shell or scripting languages such as R or python is sufficient to prepare in-house pipelines for batch processing of data gathered within particular labs.

2 BACTERIAL GENOME ASSEMBLY, ANNOTATION, AND ANALYSIS

Bacterial genetic information is, in the vast majority of cases, carried on a single circular chromosome that can be supplemented by one or more plasmids. On average, 88% of the genome is formed by coding regions [21], yet the differences between different species can be enormous. For example, the shortest bacterial genome of *Nasuia deltocephalinicola* [22] contains only 112,091 bp, while the longest genome of *Sorangium cellulosum* [23] is more than 100 times longer, containing 14,782,125 bp. There are also other basic differences, e.g., guanine-cytosine (GC) content that can vary from 18 % to 85 %. Other differences lie in the absence or presence of particular genes or are hidden in sequences of particular orthologous genes. As already said, bacteria are a very diverse group. While it is somehow natural to presume their interspecies diversity, bacteria also demonstrate high intraspecies diversity, as described for model organisms *Escherichia coli* [24]. The same applies to non-model bacteria as my colleagues and I demonstrated in the large phylogenomic study of 237 genomes of *Clostridium beijerinckii* species [14] **(Article [XII.])**. Before computational analyses of a bacterial genome can be performed, genome sequencing is necessary. Data processing and possible results vary according to the platform used for sequencing.

2.1 Genome Sequencing

The huge progress in biotechnology and applied microbiology would not be possible without constantly evolving sequencing technologies. They can be divided into three generations that differ considerably in their principle. In fact, it is the combination of different sequencing principles that makes the assembly of bacterial genomes quite easy and cheap nowadays. The advent of genome sequencing during the last two decades is apparent from the number of bacterial genomes available in public databases. For example, the GenBank database under National Centre for Biotechnology Information (NCBI), which covered less than 300 bacterial genomes in 2006, reached in the first half of 2022 almost 1.2 million genome assemblies of almost 45,000 bacterial species [25].

The first generation sequencing covers only one commercially viable method – Sanger sequencing [26]. Despite its parallelization using multiple capillaries of capillary electrophoresis, the overall output of Sanger sequencing platforms is limited to several kilobases (kbp) per sequencing run and is, therefore, unsuitable for sequencing whole bacterial genomes. While the price of sequencing per 1 kbp is high, Sanger sequencing allows to produce a limited number of sequences, so sequencing of a single gene is possible at low costs. This is the reason why this technique is still being used in the research of novel genomes. Newly isolated bacteria are hard to be identified by a simple phenotype observation. On the other hand, sequencing of a marker gene, the most commonly 16S rRNA gene, can suggest if a novel taxon was found and its worth of whole genome sequencing (WGS). Recently, we used such a strategy for describing several novel taxa, *Pedobacter fastidiosus* [27], *Rugamonas violacea* [28], *Massilia antartica* [29], and four novel species of the genus *Corynebacterium* [30]. Bioinformatics processing of these data is trivial, any Basic Local Alignment Search Tool (BLAST) [31] like algorithm can be easily used for the identification of these sequences. Possibly, they can be used

for constructing a phylogenetic tree. Such use of Sanger sequencing will likely diminish as many novel genomes are metagenome-assembled genomes (MAGs). There is a whole new initiative, SeqCode [32], that allows description of novel taxa directly from sequence data without requiring to have a culture. Nevertheless, having a culture with the bacterium is still crucial for biotechnology research, since it allows us to study a bacterium under various conditions in a laboratory.

Next generation sequencing (NGS) [33] was a breaking-through technology that allowed massive sequencing of bacterial genomes. It started with the Roche 454 pyrosequencing platform which is no longer available. Currently used platforms are Illumina and IonTorrent. Their common feature is a massive parallelization so the typical output of the sequencing run is counted in Gbp. Since tens or hundreds of bacterial genomes can be sequenced at once, the price per typical bacterial genome can be around 100 €. NGS platforms are used in any study, not only for DNA sequencing, other specialized techniques based on sequencing as RNA-Seq for transcriptomics [34] or ChIP-Seq for epigenetics [35] are done using NGS. Unfortunately, NGS reads are too short to cover large repetitive segments of bacterial genomes. Thus, these techniques have to be supplemented by long-read sequencing in order to assemble a complete bacterial genome.

Third generation sequencing (TGS) [36] allowed easy completion of bacterial genomes as it provides reads of the length of units or even tens of kbp that are easier to assemble. The generation is represented by two commercially available sequencing platforms: Pacific Biosciences (PacBio) Single Molecule Real Time (SMRT) sequencing and Oxford Nanopore Technologies (ONT) sequencing. Unlike NGS, the accuracy of these techniques is considerably lower. Although their accuracy gets substantially higher over time, NGS is usually needed for polishing assemblies made of long reads. Another advantage of TGS is its ability to capture epigenetic modifications of particular bases. While PacBio newly detects besides 6-methyl adenosine (6mA) and 4-methyl cytosine (4mC) also 5-methyl cytosine (5mC) modifications, ONT remains more experimental and allows only m5C modification detection.

Lowering sequencing costs for NGS and TGS is possible thanks to the parallel sequencing of more genomes. This is done by multiplexing, which is a wet lab procedure. Before fragments of DNA from different genomes are mixed up and sequenced, short oligonucleotide barcodes are added at the ends of a fragment. A unique barcode is used for each sequenced genome. Thanks to the current output of sequencing platforms, multiplexing is used almost in every study nowadays since tens of bacterial genomes are usually sequenced together. Nevertheless, only a few years ago, the situation was the opposite and more sequencing runs were needed to sufficiently sequence a single bacterial genome.

2.2 Genome Assembly

DNA sequencing is usually performed to read complete genetic information that a selected bacterium carries. Since even TGS does not allow to read a complete chromosome, but only its fragments, sequencing reads need to be assembled into original sequences before other bioinformatics analyses are performed. The whole process is quite complicated and it might be uneasy to complete the original sequence. Such incomplete assemblies, divided into contigs or scaffolds, are referred to as draft

assemblies. Although they might be still very informative, complete genome assemblies are necessary for deep studies of bacterial genomes.

2.2.1 Quality Assessment and Trimming

Genome assembly itself is a computationally extensive and complicated process that can be spoiled by contamination in reads and low-quality base readings. Thus, quality trimming of raw reads is always needed before they are further used. Reads themselves are strings over the alphabet $\{A, C, G, T\}$. Particular characters are produced during base-calling, which is a process of assigning these bases to raw sequencing signals. The nature of the signal (fluorescence, electric current, pH, etc.) depends on the sequencing platform and the accuracy of the base-calling can be quantified using the PHRED score [37]:

$$Q = -10\, log_{10} P, \tag{2.1}$$

where P is the base-calling error probability. Logarithmic nature of PHRED score means that Q20 corresponds to the 1% probability of base being called incorrectly. Accuracy requirements are different for particular tasks, but scores >Q30, corresponding to 99.9% read accuracy are highly desirable.

Base-calling is usually accompanied by demultiplexing, which is unlike multiplexing, a computational operation. Reads with the same barcodes are divided into different files and barcodes, as artificial sequences, are trimmed. Similar happens to sequencing adapters that are needed for sequencing but must be removed from text files. Quality assessment can be done by various tools, for example, FastQC can produce informative and interactive html reports that can be summarized over multiple samples with MultiQC [38]. Long reads may require specialized handling, for example with MinIONQC designed for ONT reads [39]. If any contamination or low-quality segments are reported, trimming is needed. Thanks to the internal database of Illumina adapters, Trimmomatic [39] can be easily used for this platform. Cleansed reads can be used for the following analyses.

2.2.2 Genome Assembly Basics

Genome assembly can be done *de novo* when a completely novel bacterium is studied. This is typical when a novel strain is selected for its suitable phenotype. On the contrary, reference-based assembly is performed when a wild-type strain genome is known and needs to be compared to genomes of engineered mutants. Eventually, many current studies combine data from different platforms which require to combine also tools for genome assemblies. Such assemblies are referred to as hybrid assemblies.

De novo assembly is performed with tools using graph algorithms. While the first group utilizes overlap layout consensus (OLC) graphs, the second relies on de Bruijn graphs (DBG) [40]. The OLC approach is more suitable for long reads of uneven length because reads are placed as vertices of the graph and edges represent overlaps between the reads. The original sequence is assembled as a path within the graph, which is an NP-hard problem. Moreover, dynamic programming used to determine overlap between vertices to create edges causes the time complexity of an assembly to be quadratic. Since every new vertex usually brings many new overlaps, the whole graph grows superlinearly with the number of reads. Yet, this is the most suitable strategy to capture longer repetitive elements in bacterial genomes because DBG uses a different strategy that is unable to capture them. Vertices in

DBG are overlaps, while edges represent the sequence of reads and the original sequence is reconstructed by a walk through the graph. To define these overlaps, reads have to be cut into shorter sub-sequences of the selected length, i.e., *k*-mers. The length of the *k*-mer cannot be longer than the shortest read. Thus, this approach is better for shorter reads of even length. Edges of DBG are then defined as overlaps of length *k*-1 and can be found in linear time. Moreover, DBG grows sublinearly with the growing number of reads as new reads usually do not introduce new edges. While in the past, OLC assemblers such as Celera or Newbler were used for NGS data [41], nowadays they are exclusively used for TGS, e.g., Flye, Canu, NECAT, etc. [42].

Reference-based assembly is created by mapping reads to a reference genome. There are various approaches to perform such mappings. Short reads can be mapped for example with BWA [43] using Burrows-Wheeler transformation and long reads with minimap2 [44] that presumes a higher error rate when handling long reads. Nowadays, even the BLAST [31] is so quick that is commonly used for mapping large metagenomic datasets to reference genomes [45]. Mapping usually results in a versatile SAM/BAM file that can be used to create a novel complete (or draft) assembly with SAMtools [46]. Possibly, mapped reads can be used for polishing an original assembly. Short high quality reads can be used to polish assemblies of erroneous long reads with Pilon [47] to produce high-quality hybrid assemblies. Another trick is to use long-reads to self-repair TGS-based assemblies. Unlike NGS, which uses PCR to achieve a better signal-to-noise ratio and, thus, suffers from systematic errors, TGS relies on a single molecule sequencing approach and produces a larger amount of rather random errors. Thus, they can be filtered by re-mapping the reads onto an assembly and performing consensus calling. An example of such a standalone consensus module can be found in racon [48].

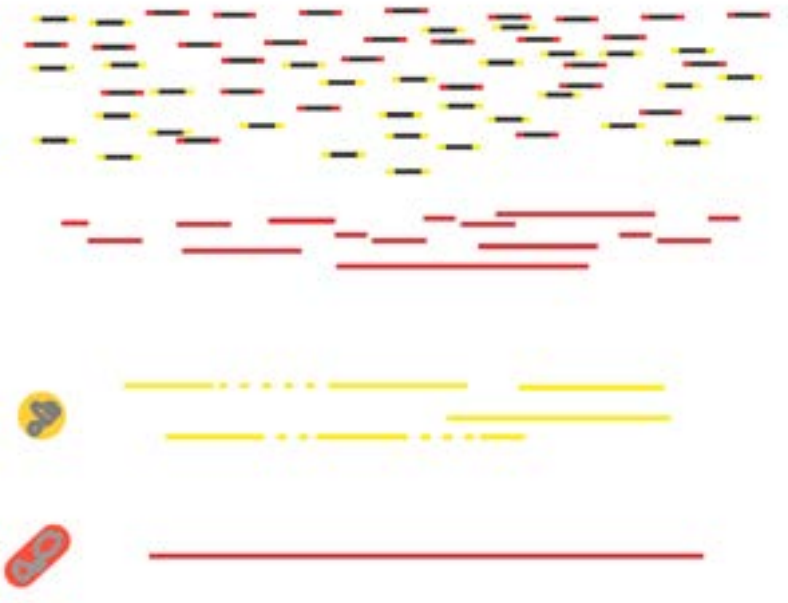


Fig. 2.1 Genome assembly
A schema of assemblies of two organisms. Thanks to the utilization of barcodes (red and yellow), two organisms can be sequenced at once using NGS. The red bacterium is sequenced again using TGS. Short reads assembly (yellow) resulted in draft assembly, yet thanks to the utilization of pair-end sequencing, gaps (dashed lines) were estimated. Long reads used for the red bacterium allow to assemble a complete genome sequence.

Unlike short reads assemblies, hybrid assemblies often result in complete genome sequences, see Fig. 2.1. This is crucial for biotechnology research as draft assemblies may miss some genes that are part of genomes and they are unable to distinguish among chromosomal and plasmid sequences. The presence of a plasmid, or several plasmids, is usually unknown before sequencing. Yet, it plays an important role in the potential biotechnological use of a bacterium since it reflects potential genetic stability. For example, the *sol* operon responsible for the production of valuable solvents is carried by a plasmid in *Clostridium acetobutylicum* ATCC 824, which is inconvenient as the strain frequently loses its ability to produce solvents after repeated subculturing due to the loss of the plasmid [49]. Since bacterial chromosomes and plasmids are, except for some special cases, circular, it is necessary to verify that the final assembly, which is provided in a linear (a long string) form, represents a complete circular molecule. This can be done manually, by concatenating the ends of an assembly and by re-mapping the reads. Possible gaps or overlaps can be solved by adding a missing sequence or by trimming the overlap from one end. Some of the newer TGS assemblers, for example, Flye [50], examine the sequence circularity on their own.

2.2.3 Different Approaches to Assemble Bacterial Genomes

The enormous progress in bacterial genome sequencing is clearly visible from the presented collection of papers. The first genome, we assembled only eight years ago, was those of *Clostridium beijerinckii* NRRL B-598, at that time misidentified as *C. pasteurianum* NRRL B-598. The first study was based on two sequencing runs performed with Roche 454 GS Junior sequencing platform [51]. These two runs, costing more than 4000€, contained less than 180 Mbp (only 28× genome coverage) of 500 bp long reads. Data were processed by Roche software GS De Novo Assembler which performed quality trimming and OLC *de novo* assembly. The resulting draft assembly consisted of 138 contigs. The genome was later resequenced using the TGS PacBio RSII platform. The overall output of the platform was incomparable to the current output of PacBio Sequel IIe which allows to sequence tens of bacterial genomes per a single SMRTcell, and two SMRTcells had to be used for sequencing. Nevertheless, the output of 490 Mbp allowed us to complete the genome [7] **(Article [I.])**. To achieve so, quite a complicated approach had to be used. The longest reads, corresponding to 30× genome coverage, were used to perform initial genome assembly with HGAP [52]. The resulting three contigs were further polished with all long reads, corresponding to 79× genome coverage with Quiver. The genome was then finalized by combining these three contigs with 138 contigs of the original draft genome assembly in Geneious [53] which was also used to verify the circularity of the final sequence and manual trimming of the overlap between ends. Before the genome was uploaded to the GenBank database, the sequence was rearranged so it starts in replication origin *oriC*. This unwritten rule that linear string representing a circular sequence should start in replication origin is, unfortunately, still not being followed by all researchers. Finding a replication origin is easy with Ori-finder [54], an online tool searching for DnaA boxes and calculating four (RY: purine pyrimidine, MK: amino keto, GC: guanine cytosine, and AT: adenine thymine) disparity curves. These curves can be also replaced by a single curve of cumulated phase signal as I and my colleagues showed [55]. Although the hybrid assembly of *C. beijerinckii* NRRL B-598 was based on NGS and TGS data, the error rate of Roche 454 pyrosequencing left many single nucleotide errors that were discovered during a transcriptomic study [17] **(Article [VI.])**. Final resequencing with high-quality Illumina NextSeq platform was needed to polish the assembly using

Pilon [47] and present the final version of the genome sequence [18] **(Article [VII.])**. In summary, to complete the genome of *C. beijerinckii* NRRL B-598 took five years and cost around 10,000€, which is now, less than five years later, hardly imaginable.

We used an almost similar procedure when assembling the genome of *Pragia fontium* 24613 with the only difference of using SOLiD sequencing data for polishing instead of Illumina [56]. The SOLiD platform using ligation for sequencing is no longer available for several years and was never widely used, at least in Czechia. Only one sequencing device was present in Czechia in comparison to more than 20 Roche 454 pyrosequencing devices. Since SOLiD produced very short reads of length only 35 bp, the polishing procedure was different. Firstly, reads were assembled using DBG assembler and resulting contigs were mapped onto the genome assembly to repair single nucleotide errors in Geneious. The resulting high-quality complete genome assembly allowed us to perform annotation and thorough description of the bacterium [8] **(Article [IV.])**.

Every genome is different and a similar procedure as described above was not sufficient to assemble the complete genome sequence of *Clostridium diolis* DSM 15410 [11] **(Article [IX.])**. At that time, a PacBio RSII platform used for long-read sequencing was at the end of its support and was substantially enhanced in comparison to previous studies. It allowed to use continuous long reads (CLR) corresponding to the longest possible reads or to generate circular consensus sequencing (CCS) reads as consensus sequences of particular subreads from the same sequencing well, in PacBio SMRT sequencing referred to as zero-mode waveguide (ZMW) [57]. Although CCS reads were shorter, their accuracy was >99%, which was considerably higher than ~85% of CLR. These reads were more typical for the following PacBio Sequel platform and were later replaced by high fidelity (HiFi) reads that are both, accurate and long. In the presented study, we started directly with hybrid assembly using CLR and high-quality short Illumina NextSeq reads assembled with SPAdes [58]. Gaps in the resulting draft genome assembly were closed with CCS reads using GMcloser [59]. The final contig was then polished, in the first round, by all PacBio subreads using the SMRTlink tool and, in the second round, by BWA mapping of short reads and polishing with Pilon. Finally, overlaps at the end of a linear contig representing a circular molecule were found with MUMmer [60] and the duplicated sequence was manually trimmed from one end of the contig.

In the following years, newly emerging ONT sequencing started to play a major role in assembling complete bacterial genomes. Because of its high error rate, it should be always accompanied by high-quality short-read sequencing. There is other bunch of tools for ONT data processing and, according to my personal experience, their performance may vary among different genomes. During the years, my colleagues and I were very successful with the following pipeline. ONT reads are assembled using OLC assembler Flye and NGS reads with DGB assembler SPAdes. Their comparison performed with MUMmer allows identifying erroneous Flye ONT contigs to be omitted from the following polishing. The first step of polishing is then done by four rounds of mapping long reads with minimap2 [44] and polishing with racon [48]. The second step consists of two rounds of polishing with medaka, again in combination with ONT reads. Finally, the third step of polishing is performed by mapping short reads with BWA and polishing with Pilon. This mapping and polishing is done repeatedly until the sequence of the assembly remains untouched by the additional round of polishing. Usually, at

least two rounds are needed. Such a combination of tools gathered popularity in the bacterial community and was used in the pipeline called Unicycler [61] that allows to call all of these tools and assemble a complete bacterial genome with a single command. Using these tools, we were able to assemble genomes of *Clostridium beijerinckii* DSM 791^T [14] **(Article [XII.])**, *Schlegelella thermodepolymerans* DSM 15344^T [9] **(Article [X.])**, *Tepidimonas taiwanensis* LMG 22826^T [10] **(Article [XIII.])**, and *Aneurinibacillus thermoaerophilus* CCM 8960 [4] **(Article [XVI.])**. The advantage of this pipeline lies in its ability to reveal plasmids as shorter circular sequences with sufficient coverage. Since plasmids are also circular, it is necessary to find their replication origin *oriV* and rearrange the sequence according to it. Searching for *oriV* is more complicated than for *oriC* because plasmids do not show such a disparity between nucleotides and must be searched based on sequence similarities using BLAST. Unfortunately, *oriVs* are diverse and searches against the DoriC database [62] containing replication origins of chromosomes and plasmids may not be successful. In that case, annotation of a plasmid is needed to find the correct start of a sequence in the linear form, see the following chapter.

Thanks to the higher sequencing output of current sequencing devices and the possibility of multiplexing, the real price per completion of a genome dropped to a range between 1000€ and 2000€ and the final assembly can be completed in several days. The higher the coverage, the easier to reach a complete genome assembly and reveal plasmids. When combining NGS and TGS, a coverage of 200× is usually sufficient to assemble the complete genome. Nevertheless, this is not the rule and we experienced many problems when assembling the genome of *C. diolis* DSM 15410 [11] **(Article [IX.])** despite having a total coverage of 700×. It can be taken for granted that bacterial genome assembly will experience another change in the near future as PacBio sequencing was "reborn" last year. Only recently, two new sequencers were introduced, Revio for long-read sequencing and Onso for short read sequencing, both promising so far unmatched accuracy >Q40. Novel bioinformatics tools and novel pipelines compiled with current tools will certainly emerge and the whole process will be again easier than ever before.

2.2.4 Searching for Variants in Bacterial Genomes

Although reference-based assembly seems to be less complicated, it introduces completely new hurdles that need to be overcome as shown in the study of nine mutants strains of *C. beijerinckii* NRRL B-598 with increased butanol tolerance [12] **(Article [VIII.])**. In this case, it is not so important to get a long string representing a mutant's genome as it is almost similar to the wild-type (WT) strain genome, but to highlight changes between these genomes. Beside single nucleotide polymorphisms (SNPs) also larger structural variants such as copy number variations (CNVs). These are quite neglected topics in bacterial bioinformatics. While in eukaryotes, these studies are often performed by single-cell DNA sequencing (scDNA-Seq) [63], in bacterial studies, sequencing of a culture containing many bacteria is still preferred. The reason lies in the difficulties of obtaining genetic material from non-model organisms as available DNA/RNA isolation kits show different efficiency for various bacteria and it may be impossible to get a sufficient amount of DNA from a single cell. Therefore, we will presume standard sequencing of a culture that will show some low, yet non-negligible diversity. The beginning of the procedure is straightforward, cleansed reads are mapped onto the WT strain genome with BWA or any other tool and mapping SAM/BAM files are prepared. The following procedures, however, differ for the prediction of CNVs and the prediction of SNPs.

Prediction of CNVs can be done with Pilon, since it is able, besides polishing an assembly, also to infer SNPs and CNVs. CNVs are copies of a specific segment of DNA while a number of these copies varies among different genomes. Although they are studied mainly in eukaryotes, they can be identified in bacterial genomes, where they can highly influence the final phenotype [64]. In the study by Vasylkivska et al., 2020 [12] **(Article [VIII.])**, we used Pilon for their detection in mutants strains of *C. beijerinckii* NRRL B-598. Nevertheless, there are other tools that can be used, for example, CNproScan presented by our group [65]. In the study by Jugas et al., 2021 [65], we also compared our tool with the other seven tools and showed that results can substantially vary between particular tools. Nevertheless, this study is aimed at healthcare-associated infections (HAIs) causing bacteria, since CNVs plays important role in drug resistance [66]. Analyses of bacteria involved in HAIs are different from those studied for utilization in the industry. Studies usually aim only at highly variable parts of their genome that can be used to distinguish bacterial strains from one species by genotyping. How to find such regions from sequencing data was described by me and my colleagues in the study by Nykrynova et al., 2021 [67].

To capture SNPs, inspiration can be taken from scDNA-Seq data processing that removes PCR duplicates to prevent algorithms from putting more weight on polymorphisms that were artificially multiplied before sequencing. While scDNA-Seq distinguishes duplicates by small oligonucleotides, called unique molecular identifiers (UMIs) [68], ligated to ends of sequenced fragments, in standard sequencing they cannot be precisely identified. However, we are able to remove all duplicated sequences mapping to the same position by simply deleting multiple copies of the same reads. Remember that the two same reads may be still preserved when they map to different loci. This is advantageous as we want to presume roughly even coverage of the genome. However, a mapping tool that maps these multi-mapping reads randomly is necessary, BWA fulfills this presumption. Demultiplication can be done by indexing the mapped reads with SAMtools and sorting them while deleting duplications with Picard tools [46]. Under no circumstances should the demultiplication be done for the detection of CNVs as their prediction is based on discrepancies in the genome coverage. Since reads are coming from many bacteria, captured polymorphisms are usually not present in all reads as they are simply not present in all bacteria and low frequent variants may be even the results of sequencing errors. Therefore, results are somewhat similar to variant calling in polyploid eukaryotes containing different alleles and even the WT strain culture shows many variable positions within its genome as not all sequenced cells are the same. Considering this, it is evident that predictions of SNPs in mutant stains are biased and thresholds of variant frequency have to be set to predict mutations exclusive for cultures of mutant strains. In the study by Vasylkivska et al., 2020 [12] **(Article [VIII.])**, we set up our own approach, combining laboratory work and computational data processing steps, to minimize false detections while capturing SNPs. Besides the WGS of mutant strains, we did resequencing of the WT strain. We used variant calling in the WT strain to set a threshold for mutant stains. Only SNPs with higher frequency than the highest frequency in WT and sufficient coverage were counted for mutant strains. This was possible by using GATK [69] which can report frequencies of particular SNPs. Not all variant calling tools are able to do that, for example, Pilon reports all variants without their frequencies.

Besides genome sequencing, single nucleotide variants can be also captured by a cheaper technique, particularly quantitative, also referred to as real-time, PCR (qPCR). Unfortunately, qPCR is

only able to predict a presence of a mutation in analyzed amplicon (short analyzed part of DNA) by the difference in melting temperature. Therefore, its use for biotechnology research is very limited. Yet, it finds its utilization in the identification of HAIs causing bacteria thanks to its much lower price in comparison to DNA sequencing. Processing of qPCR data also requires specialized computational approaches. One such approach proposed by myself is a part of Czech patent CZ308046 [70].

2.3 Genome Annotation

Genome assembly itself is only the beginning of bioinformatics work in the research of bacteria as it brings the basic structure to be further analyzed. As DNA contains genes that are somehow organized and have different functions, the following steps belong to the genome annotation. While the structural annotation serves to reveal the positions of particular genes, the functional annotation predicts the behavior and purpose of particular genes.

2.3.1 Structural Annotation

Basic structural annotation of genomes is a computationally extensive, yet user-friendly procedure. The majority of novel sequences is annotated automatically when uploaded to public databases. For example, every genome uploaded to the GenBank database can be annotated by the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) [71] by simply selecting this option when submitting new data. PGAP is a pipeline consisting of many other computational tools that are being continuously developed and extended. It can be run online using the computational resources of NCBI or it is available for download and local use. This may be advantageous for preliminary analyses without the need to publish an analyzed sequence, for example, when searching for the *dnaA* gene while trying to predict replication origin. Another possibility for extensive annotation of bacterial genomes, also available online, is the Rapid Annotations using Subsystems Technology (RAST) [72]. Not only are these two tools the most widely used in the research of clostridia [6] **(Article [V.])**, but they find their use in the annotation of any bacteria. Our group prefers using PGAP, therefore it was used in all studies describing novel genomes within this collection [4, 7, 9–11, 14] **(Articles [I.], [IX.], [X.], [XII.], [XIII.], [XVI.])** but the genome description of *Pragia fontium* 24613, where additional gene mining was manually curated based on results of the RAST [8] **(Article [IV.])**. The outputs of an annotation are identified loci, i.e. genomic coordinates, of identified elements. These are primarily coding sequences (CDSs), i.e. mostly protein-coding genes, and rRNA and tRNA genes. Newer versions of PGAP also add other non-coding RNA genes and Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) loci. Each locus is assigned a unique alphanumeric identifier referred to as a locus tag. Locus tags are unique for any genome sequence assigned an accession number but they can differ between various versions of the sequence. Therefore, it is necessary to clearly state the accession number including the version of the genome, when referring to genes using their locus tags. Generally, it is better to use genome sequences and corresponding annotations stored in the GenBank, European Nucleotide Archive (ENA), or DNA Data Bank of Japan (DDBJ) databases that are shared and use the same locus tags that do not differ between particular versions of sequences. Newly annotated genes between versions are given new locus tags and locus tags of deleted genes are no longer used. Unfortunately, many users refer to the NCBI Reference Sequence (RefSeq) database. This database contains copies of complete genomes stored in the GenBank but uses annotation of the latest PGAP version. All locus tags are renamed between

versions which makes the comparison of findings between different versions quite uneasy. This happened to our group when locus tags in the RefSeq were renamed for *C. beijerinckii* NRRL B-598 between versions NZ_CP011966.2 and NZ_CP011966.3 which was problematic for comparing our new results from the study Sedlar et. al, 2019 [18] **(Article [VII.])** to the older ones from the study Sedlar et al., 2018 [17] **(Article [VI.])**. Since RefSeq does not presume further polishing of available genomes, many users are not aware of this limitation. This may be a problem for future bioinformatics data processing that relies on automated text mining and natural language processing while performing extensive data mining from available published papers.

Basic structural annotation can be complemented with the outputs of other tools annotating additional elements. Since the discovery of CRISPR-Cas9 genome editing [73], CRISPR arrays are usually reported for novel genomes. While the latest PGAP reports CRISPRs, higher sensitivity in their detection can be achieved with other tools, e.g. CRISPRDetect [74]. Besides possibilities of future genome editing, their presence can be used to predict a presence of an active bacterial immune system as CRISPR-associated system (Cas) can form an adaptive immune system in bacteria [75]. Thus, a Cas presence may be accompanied by an absence of potentially active prophages as we found in the genomes of four novel *Corynebacterium* species [30] and vice versa their absence may be followed by a presence of active prophages as we reported for the genome of *Pedobacter fastidiosus* [27]. It is also possible to combine more tools and report only consensus of CRISPR arrays detection as we used in the study of *Tepidimonas taiwanensis* [10] **(Article [XIII.])**. Computational prediction of prophages is also quite a common procedure when describing novel genomes. There are several tools that can be used, some of them can even predict if the prophage is active, and combining result from several tools is also possible [9, 10] **(Articles [X.] and [XIII.])**. Possibly, an activity of *in silico* predicted prophages can be further examined by transcriptomic data as we reported for *Clostridium beijerinckii* [17] **(Article [VI.])**. Nevertheless, the prediction of prophages will probably undergo rapid development in the near future as prophages are probably far more abundant in bacterial genomes than expected. A recent study by Piligrimova et al. 2021 [76] showed that whole plasmid prophages are very common in the bacterial domain. Their prediction is cumbersome, as many prophages remain to be undiscovered and computational tools relying on homology searches reporting sequence similarities to already known phages are, therefore, inefficient. Nevertheless, when combined with other computational approaches, for example, with a calculation of plasmid coverage and other lab techniques, novel prophages can be reported. Only recently, our group discovered a novel plasmid prophages pAT1 in the genome of the original isolate *Aneurinibacillus thermoaerophilus* CCM 8960 [4] **(Article [XVI.])**.

Besides monocistronic transcripts that encode only one protein, bacteria produce also an ineligible amount of polycistronic transcripts. Adjacent genes that are transcribed together and share a promotor form an operon. Annotation of operons is not widely applied in the research of non-model bacteria, yet they can be predicted *in silico* from the genome sequence, for example with operon-mapper [77]. This approach requires also functional annotation as genes in an operon belong to the same pathway, i.e. they belong to the same functional category. Annotation is not standardized and operon-mapper output has to be parsed with custom-made scripts to prepare a flat-file with an annotation table that can be uploaded to the GenBank database to update a genome record. Thus, only a minority of records contains information about operons, an example can be found in the genome of *Schlegelella*

thermodepolymerans DSM 15344 published by our group under the genome accession number CP064338.1 [9] **(Article [X.])**. On the other hand, annotation of operons is very often done only to discuss the presence of selected biological pathways and important genes without uploading it to the genome record as we did when describing genes and operons involved in violacein production in *Massilia antarctica* [29]. This strategy can be used to explain unexpected behavior even for publicly available draft genomes as we did with *S. thermodepolymerans* before its resequencing. We found a unique operon in its genome containing genes for xylose utilization together with a gene for xylose transport into the cell. This explained the preference of the bacterium to consume xylose over glucose [78]. Operon can be also identified from transcriptomic data by searching for correlation in transcriptional profiles of adjacent genes. Nevertheless, to my best knowledge, there is currently no specialized tool that would be able to perform an automated identification of operons from RNA-Seq data in non-model bacteria.

There are also other, mostly regulatory, elements, e.g. promoters, untranslated regions (UTRs), or small non-coding RNA (sRNA) genes, that can be identified in a genome sequence. Nevertheless, unlike the loci above that are being predicted directly from the genome sequence, regulatory elements are identified using additional primarily transcriptomic data. Thus, they will be described in the third chapter.

2.3.2 Functional Annotation

Basic structural annotation is always accompanied by basic functional annotation because homology searches used for verification of predicted ORFs very often transfer also a putative function. Thus, identified loci are described by features, e.g. "CDSs", which are assigned other qualifiers with descriptions, e.g. "gene" containing an abbreviation of a gene or "product" containing the basic description of a protein-coding gene. If the function remains unknown, which is not so rare for non-model bacteria, a generic description "hypothetical protein" is added under the product qualifier. Overall, such functional annotation is not very accurate for non-model organisms as orthologous genes may share very low sequence similarities to their counterparts in well-annotated model bacteria. Manual reannotation of the function is, however, possible. This may be extremely time-consuming and requires many manual comparisons using BLAST and even the processing of transcriptomic data. It took our group a couple of years and a thorough analysis of three RNA-Seq datasets to reannotate several genes in the *C. beijerinckii* NRRL B-598 genome for proper visualization of transcriptional changes in selected metabolic processes [79]. Similar reannotations uploaded to public databases are highly desirable as they improve the annotation of future genomes because corrected annotations are further spread for new orthologues.

Some groups of genes with the same function may be even further divided into classes, for example, PHA synthases [15] **(Article [XV.])** that are divided according to their subunit composition. Their correct identification in non-model organisms may be extremely challenging as we experienced during analyses of *Rubrobacter xylanophilus* and *Rubrobacter spartanus* [80]. Although we were able to easily identify PHA synthases in their genomes by similarity searches using BLAST, further annotation of their subunits was not possible that way because comparable and very low sequence similarities of 30% to both PhaE and PhaR subunits were found. Nevertheless, since both of these

subunits have very different molecular weights, we were able to distinguish them by calculating their putative weight directly from their amino acid sequences with the sequence manipulation suite [81].

Since the basic functional annotation added by PGAP or other pipelines may be unreliable and tries to be unnecessarily detailed, more coarse-grained annotations are usually used for the description of novel genomes. One of the simplest, yet the most widely used analysis is annotation dividing genes into Clusters of Orthologous Groups (COGs), also referred to as clusters of orthologous genes. COGs allow the classification of protein-coding genes using their homologous relationships, the whole database is manually curated and its latest update comes from 2020 [82]. Annotation of novel genomes can be done with BLAST against the COG database itself or using other specialized tools, e.g. eggNOG-mapper [83]. Besides the description of the potential functional capacity of a novel organism, COGs can be used to verify the completeness of a genome assembly as genes with housekeeping functions have to be always present in a genome. Particular COGs are divided into functional categories that can be also used for genome-wide visualizations. As our group uses these visualizations quite often [4, 9–11, 14] **(Articles [IX.], [X.], [XII.], [XIII.], [XVI.])**, we designed a python package, COGtools (available with full documentation and tutorials from https://github.com/xpolak37/COGtools under the MIT license.), capable to combine annotations from various sources and help to produce publication quality visualizations, see Fig. 2.2.

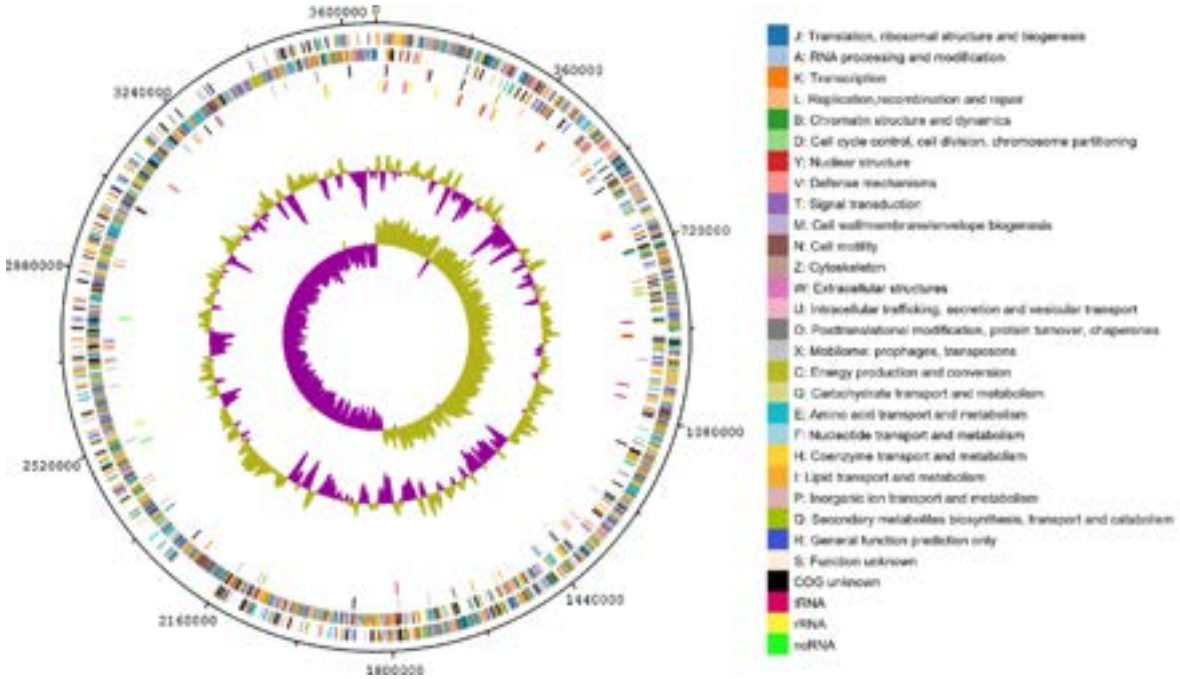


Fig. 2.2 Circular chromosomal map of Aneurinibacillus thermoaerophilus CCM 8960
The first and the second outermost circles represent CDSs on the forward and reverse strands, respectively. The third circle represents pseudogenes. Colors highlight functional categories of particular loci. The fourth outermost circle shows RNA genes, distinguishing between tRNA, rRNA, and ncRNA. The inner area represents the GC content and GC skew.

Our preliminary, yet unpublished, results show that COGtools can improve the sensitivity and overall accuracy of annotations. Since we also discovered some discrepancies between various sources building upon COG vocabulary, more research on this topic is needed in the near future to help improve the annotation of non-model bacteria.

Another possibility for functional annotation that is more typical for model organisms is the use of Gene Ontology (GO) [84]. GO is a kind of controlled vocabulary defining basic terms and relationships among genes. Since the GO is maintained exclusively by the Open Biological and Biomedical Ontology foundry, it does not contain any discrepancies and presents a basis for the interoperability of databases and computational tools. GO dictionary has 3 categories or domains: Biological Process (BP) describing an easily recognizable series of events, Molecular Function (MF) describing basic gene (or gene product) activity, and Cellular Component (CC) describing where a gene product can be found in the cell. Moreover, GO has a hierarchical structure in the form of a directed acyclic graph (DAG). Thus, general terms can be replaced by more specific terms to perform a detailed analysis, and vice versa, specific terms can be replaced by general ones to describe basic functions. GO annotation of a non-model bacterium is not an easy task. It requires manual inspection of various databases of extensive searching for annotated orthologues with BLAST. Even though many databases can be accessed through a single interface of QuickGO tool [85], manual curation of the result to remove duplications and to parse annotation in a correct format that can be further used is needed. So far, we were able to prepare further utilizable GO annotation of only one organism, *C. beijerinckii* NRRL B-598 [18] **(Article [VII.])** using custom-made R scripts. Nevertheless, results show that it is possible to annotate even very specific terms to a non-model bacterium as visible from the bar plot showing the distance of annotated terms from the root of the GO DAG, see Fig. 2.3.

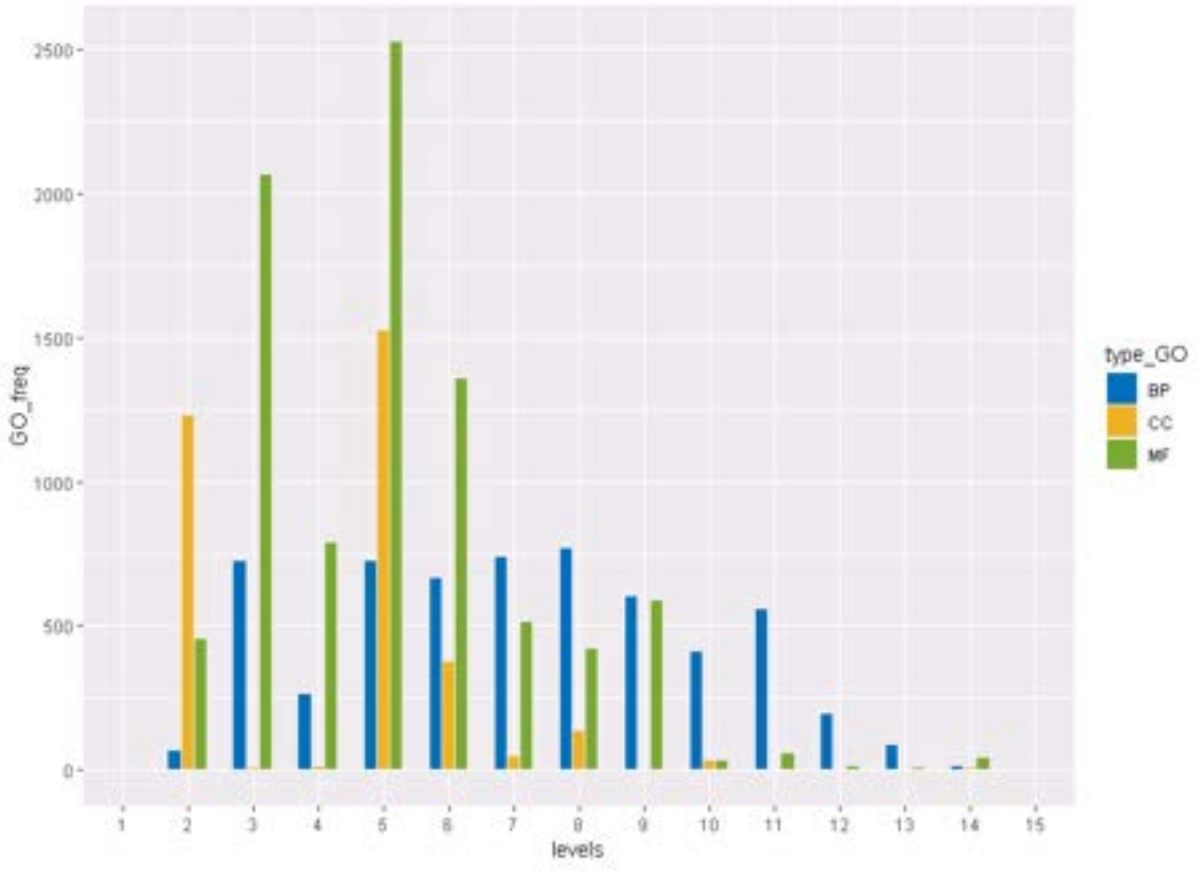


Fig. 2.3 *C. beijerinckii* NRRL B-598 GO annotation, taken from [18] (Article [VII.])
Levels (longest distance from the root) of 18,020 unique GO terms assignments in *C. beijerinckii* NRRL B-598 manually curated GO annotation divided into GO categories: BP = Biological Process, CC = Cellular Component, and MF = molecular function.

Although there is a wide range of other specialized resources for functional annotation, the only other comprehensive enough to be mentioned here is the Kyoto Encyclopedia of Genes and Genomes (KEGG) [86]. KEGG is a knowledge base for a systematic analysis of gene functions. It contains information on reactions and pathways gene products are involved in, i.e. it links genomic information with higher-order functional information. Reference complete genome assemblies uploaded to GenBank/EMBL/DDBJ are automatically annotated and presented in the KEGG databases. Annotation of additional genomes is possible with BlastKOALA [87]. This might be advantageous to select genes whose transcription is worth analyzing as we did for *C. beijerinckii* NRRL B-598 [88] or just to discuss the ability of novel species to produce valuable enzymes as we did, in comparison to the results of metabolic fingerprinting, for *Massilia antarctica* CCM 8941 [29] and *P. fontium* 24613 [8] **(Article [IV.])**, where we identified thiosulfate reductase responsible for H_2S production, which is quite an atypical feature for enterobacteria.

2.4 Identification and Phylogeny

Newly sequenced genomes need to be identified, i.e. their taxonomic name needs to be verified in the case of bacteria from collections or a new name must be assigned in the case of novel isolates. After that, phylogenetic analyses showing evolutionary relationships can be performed.

2.4.1 Taxonomic Identification

A majority of bacteria from the public collections was identified based on their phenotype. Such identification is not so precise and reidentifications of particular strains or reclassifications of whole species are not so rare after genomes of these organisms are sequenced. Correct identification is crucial for biotechnology as it allows to predict the behavior of analyzed organisms from related, already described organisms. The whole concept of identification is based on genome similarities. While in the past, genome similarities were tested in labs by DNA-to-DNA hybridization (DDH), in the genomic era, hybridization can be calculated directly from genome sequences by digital DDH (dDDH). The most powerful tool to identify novel genomes relying on an extensive database of reference genomes and dDDH is the Type (Strain) Genome Server (TYGS) [89]. Since tens of thousands of bacterial genomes are already known, it is nowadays quite easy to confirm correct taxonomic placement of a bacterium as we recently did for *Schegelella thermodepolymerans* DSM 15344 [9] **(Article [X.])** or *Tepidimonas taiwanensis* LMG 22826 [10] **(Article [XIII.])**. Similarly, we were able to assign a species name to formerly unidentified *Aneurinibacillus thermoaerophilus* CCM 8960 [4] **(Article [XVI.])**, an original PHA-producing isolate from Brno. At last, but not least, the inability to identify analyzed bacteria helps to confirm that novel species were found as we recently did for several novel isolates *Pedobacter fastidiosus* [27], *Rugamonas violacea* [28], *Massilia antartica* [29], or four novel species of *Corynebacterium* genus [30].

Massive sequencing of bacterial genomes which filled databases with reference data is, however, the matter of the last five years. Correct identification of analyzed strains was more challenging in the past as we experienced with *C. beijerinckii* NRRL B-598, formerly misidentified as *Clostridium pasteurianum* [13] **(Article [III.])**. Its correct taxonomic identification was challenging because the genome sequence of the type strain that should be used as a reference was not available for

C. beijerinckii back then. Moreover, a draft genome sequence of *Clostridium diolis* showed very high genome similarities to our strain. Therefore, we had to supplement dDDH calculations with other non-type strains by phylogenomic analysis. As we later sequenced genomes of the type strains *C. beijerinckii* DSM 791^T [14] **(Article [XII.])** and *C. diolis* DSM 15410^T [11] **(Article [IX.])**, we were finally able to confirm that *C. beijerinckii* and *C. diolis* are heterotypic synonyms, i.e., they represent the same species.

Besides dDDH, there are other parameters showing genome similarities capable to perform species delineation. Probably the most basic parameter is Average Nucleotide Identity (ANI) that we still use to confirm dDDH values [27–30]. There are several approaches to calculate ANI. They are usually computationally demanding as they require the alignment of long sequences and, therefore, are available as server applications. My colleagues and I proposed a slightly different approach using genomic signal processing that is able to outperform other ANI-based tools and run on a standard desktop computer [90].

2.4.2 Phylogenetic Analysis

Identification is often followed by a more detailed analysis of relationships between various bacteria or between selected orthologous genes. There is a wide range of techniques, tools, mathematical models, and correction techniques to construct and rescale phylogenetic trees. In the most basic way, bacteria are compared using sequences of 16S rRNA genes. This standard procedure in the research of bacteria that we use for taxonomic studies [27–30] is usually supplemented or replaced by other techniques that can bring important knowledge for white biotechnology.

Firstly, phylogeny based on selected protein sequences conserved across a bacterial domain can supplement taxonomic identification as we did for the reidentification of *C. beijerinckii* NRRL B-598 [13] **(Article [III.])** and confirmation of *C. diolis* reclassification [11] **(Article [IX.]).** An efficient way to perform similar analysis can take advantage of PhyloPhlAn [91] which calculates with an internal database of conserved sequences in the bacterial domain and, therefore, can quickly analyze any dataset of bacterial genomes. Recently, we used this pipeline also to identify a novel strain *A. thermoaerophilus* CCM 8960 [4] **(Article [XVI.])** as the original isolate was unable to be taxonomically identified without genome sequencing [3].

Secondly, the most precise phylogeny for a selected group of bacteria, for example, a selected species with the potential to produce a valuable chemical, can be done by maximizing the amount of information that is analyzed. This can be done by defining a core genome, i.e., all genes that are shared by a group of analyzed genomes. Such analysis is computationally extensive as it requires pairwise comparisons of whole genomes. Nevertheless, fast clustering techniques can be used, for example, USEARCH can be selected as a clustering algorithm within BPGA pipeline [92]. This way, we were able to reconstruct the most precise phylogeny of 242 strains of *C. beijerinckii* and define 16 clusters of very similar strains [14] **(Article [XII.])**. Such information is valuable for diverse species like *C. beijerinckii* as only one strain representing the whole cluster can be selected for further analyses to reveal its potential for biotechnology. Other strains of the cluster can be presumed to have the same phenotype manifestation and behavior. On the contrary, different clusters or even well-distinguished strains are worthy to be further examined as they are expected to have different behavior. This is directly

visible from the Fig. 2.4, where strains *C. beijerinckii* NRRL B-598 and *C. diolis* DSM 15410^T present the two closest, yet well-distinguished branches (please note that *C. beijerinckii* and *C. diolis* are heterotypic synonyms, i.e., the same species as stated before). While the strain *C. beijerinckii* NRRL B-598 is capable of acetone-butanol-ethanol fermentation with negligible production of ethanol [7] **(Article [I.])**, the strain *C. diolis* DSM 15410^T is capable of isopropanol-butanol-ethanol fermentation and shows a unique ability to produce isopropanol [11] **(Article [IX.])**. Moreover, the type strain *C. beijerinckii* DSM 791^T is hidden in Cluster 5. Therefore, the use of the type strain is not always desirable for biotechnology research and results gathered using a single strain are not always generalizable on the whole species.

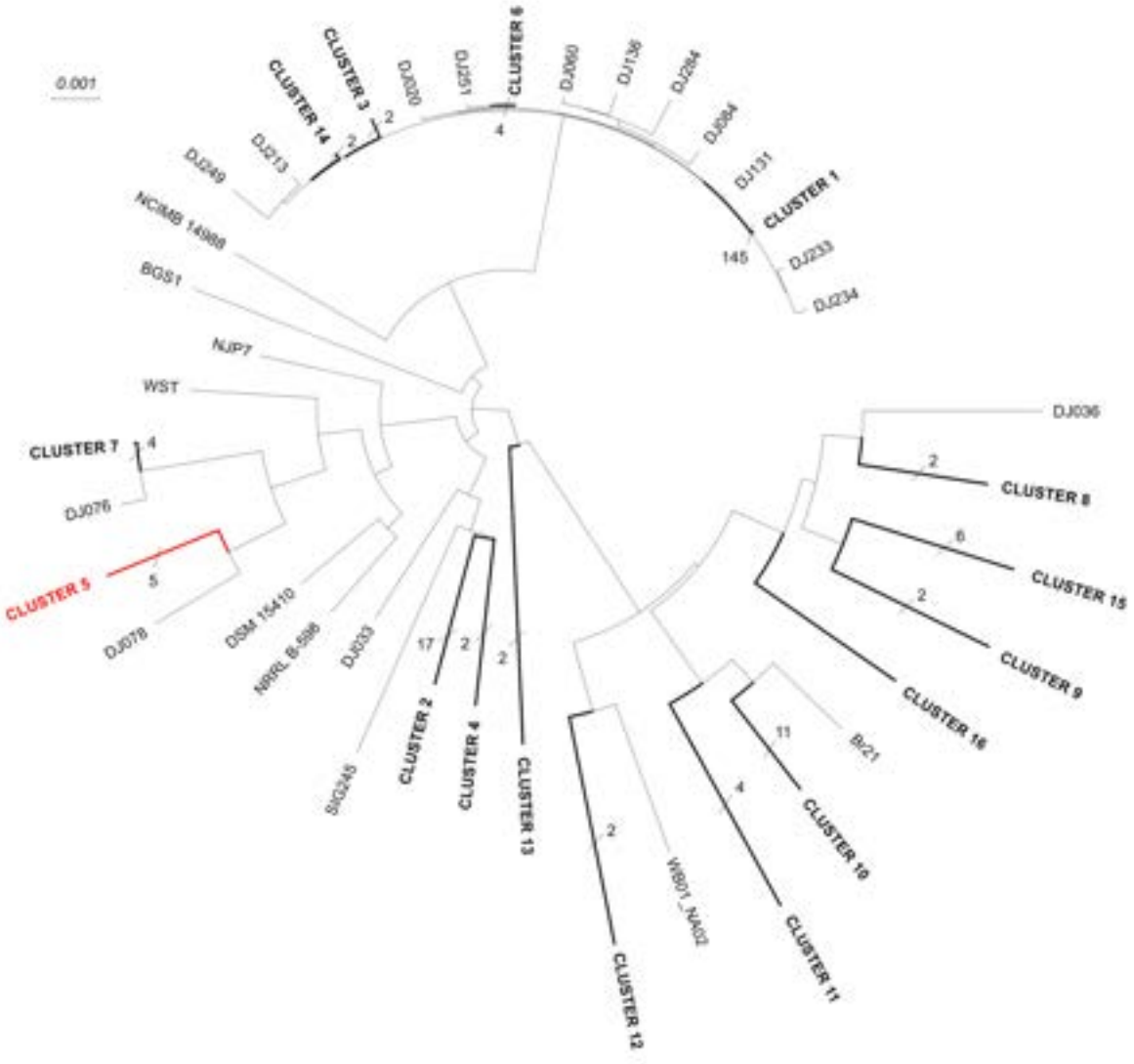


Fig. 2.4 Phylogeny of *C. beijerinckii* species, taken from [14] (Article [XII.])
The phylogeny was reconstructed using 2308 genes of the core genome, present in every strain. Branches that were shorter than 1% of the maximum branch length were collapsed. The number of strains hidden under collapsed branches is written to each collapsed branch. The branch containing the type strain *C. beijerinckii* DSM 791^T is highlighted in red.

Third, phylogenetic relations between selected genes can help to distinguish among various classes of genes that produce enzymes utilizable in white biotechnology. These trees do not represent

the evolution of species or strains but rather classify selected enzymes. In the study by Obruca et al., 2022 [15] **(Article [XV.])**, we compared PHA polymerases from extremophiles. We showed that halophiles and thermophiles carry Class I or Class III PhaC synthases and that there might be some yet undiscovered classes of PhaC synthases. This study presents also another phenomenon. Although the article type is "review", bioinformatics work done in the study is standard research. This demonstrates that bioinformatics serves as a tool for biotechnology that can be used either for novel research or to review current knowledge.

3 REGULATIONS IN BACTERIAL GENOMES

Computational analyses and annotations of genome sequences can only reflect the functional potential of a bacterium. This is the first part of biotechnological research that serves to select suitable organisms for further research. Nevertheless, the real functions and regulations that would be utilizable in white biotechnology have to be revealed with additional epigenetic, transcriptomic, or proteomic experiments.

3.1 Methylation

Although bacteria seem to be simple unicellular organisms, they have several mechanisms working as adaptive immune systems. One of them, the CRISPR-Cas system was already mentioned in the previous chapter. Another, no less important, system is the restriction modification system (RM system). RM systems are formed by methylases and restriction nucleases. The latter enzymes cleave DNA at specific motifs, but the former can protect bacterial own DNA by adding a methyl group. This way, only foreign DNA is destroyed. This is an important feature for biotechnology as foreign DNA is often introduced by an infectious agent, for example, a bacteriophage. Contamination of the biotechnological process is extremely dangerous as it causes huge financial losses and at the same time contamination precautions, for example, sterilization, are costly and prevent the use of many organisms in industrially viable projects. Thus, the targeted utilization of RM systems could bring a completely novel concept to white biotechnology supplementing current trends such as the Next Generation Industrial Biotechnology (NGIB) that relies on the use of extremophiles [15] **(Article [XV.])**. Extreme conditions that are suitable for extremophilic bacteria are usually hostile to other contaminants including phages. This is, however, not true in general and even extremophiles contain foreign DNA and prophages as we discovered in *Tepidimonas taiwanensis* LMG 22826^{T}, a thermophilic PHA and protease producer [15] **(Article [XIII.])**. On the other hand, the incorporation of foreign DNA is desirable during the engineering of various strains to improve their features. In that case, RM systems may present unnecessary hurdles and exclude strains from further research. Thus, revealing genome-wide methylations is highly beneficial for biotechnology research.

Unfortunately, the detection of methylations is possible only with selected TGS platforms and to some limited level with NGS. For NGS, bisulfite sequencing, usually using the Illumina platform, is needed [93]. Bisulfite conversion changes all but methylated cytosines into uracils. Thus, by mapping reads to previously assembled reference genomes the methylated cytosines are revealed. Unfortunately, that way identified 5mC methylations are typical for eukaryotes and their role in bacterial genomes, where 4mC or 6mA methylations are usually more prevalent, is often unimportant. These two modifications are, however, easily detected with the PacBio platform. Moreover, with an improved detection algorithm, PacBio SMRT is able to detect also 5mC modifications from 2022 on with newer sequencing devices. Since methylations are directly identified from raw sequencing fluorescence signal, no additional lab procedures are needed and methylations are labeled directly in the newly assembled genome without the need for any former reference sequence. Methylations can be also identified with ONT platforms in a very similar way. However, the processing of raw current nanopore signals, referred to as squiggles, is not so precise and needs to be further developed for practical use.

Our group took advantage of the PacBio RSII platform used for *C. beijerinckii* NRRL B-598, formerly misidentified as *C. pasteurianum* NRRL B-598, genome sequencing when we were not able to transform the strain according to a protocol working for other clostridia [16] **(Article [II.])**. We described formerly unknown two motifs containing 6mA methylations and matched these recognition sites with two methyltransferases using the REBASE database [94] where we published the results. Finally, we were able to prepare a protocol for a gene transfer into *C. beijerinckii* NRRL B-598 using unmethylated plasmid DNA. Similarly, we revealed thousands of 6mA methylated positions in the *P. fontium* 24613 genome and matched the inferred GATC recognition site to functional well-described DNA adenine methylase responsible for mismatch repair of DNA [8] **(Article [IV.])**.

Unfortunately, it is not always possible to match motifs of newly discovered recognition sites to corresponding methylases. Although we detected nine methylation motifs in *C. diolis* DSM 15410^T genome of which seven were hitherto undescribed, we were not able to couple them with methylases as only a single methyltransferase was annotated in the genome [11] **(Article [IX.])**. Inferred motifs were probably not genuine, caused by errors in motif calling caused by limitations of today obsolete PacBio RSII sequencing platform. Although there must be an active RM system in *C. diolis* DSM 15410^T genome, because the strain requires pre-methylated DNA for transformation [95], its own RM systems must be strain specific and different from RM systems identified in *C. beijerinckii* NRRL B-598, the closest strain within the genus (please note that *C. beijerinckii* and *C. diolis* are heterotypic synonyms), see Fig. 2.4. Similarly, we were not able to couple eight 5mC modification containing motifs to methylases annotated in the *A. thermoaerophilus* CCM 8960 genome [4] **(Article [XVI.])**.

3.2 Transcription

A transcriptomic study, which reflects the involvement of individual genes in all cellular processes within a population, at any given sampling moment, is a valuable tool for gaining a deeper insight into microbial cell factories. Measuring the expression of genetic information is a standard research procedure that reveals microbial behavior by providing information on the regulation of their genes. Transcriptomic techniques can aim at a limited number of preselected genes or capture genome-wide transcription. They can even help to infer novel, hitherto unannotated genes, explain the functions of poorly described genes, or annotate non-coding elements with important regulatory functions as reviewed by Patakova et al., 2022 [19] **(Article [XIV.])**. While the study is aimed at solventogenic clostridia, summarized bioinformatics techniques for transcriptomic data processing are applicable in general to any bacterium utilizable in white biotechnology.

3.2.1 Transcription of Selected Genes

A low number of selected genes can be studied with RT-qPCR to capture changes in their expression. The method is very cheap but processing a larger number of genes is not possible as unique primers have to be proposed for every single analyzed gene. Not only serves RT-qPCR for validation of the data from other genome-wide techniques [19] **(Article [XIV.])**, transcription of key genes in a selected pathway can be studied to compare transformed strains with WT strains. In the study by Kolek et al., 2017 [96], we compared expression profiles of six selected key genes in sporulation, acidogenesis, and solventogenesis as these processes are tightly connected and orchestrated by the master regulator *spo0A*.

To be able to compare results for the WT *C. beijerinckii* NRRL B-598 to the mutant strain OESpo0A with overexpressed *spo0A*, normalization of the data was needed as a standard task of RT-qPCR data quantification. As genome-wide expression of the WT was not available at that time, we selected four candidate genes from the literature search as possible reference genes and examined the stability of their expression with RT-qPCR. Eventually, *rpsJ* was chosen as the reference gene for data normalization. Since the transcription of the gene was tested only under the specific conditions of our study, we calculated only the relative expression of studied genes and rated fold changes in their expression profiles.

Although utilization of universal reference genes, for example, 16S rRNA coding gene, whose expression is believed to be stable, is very popular for normalizing RT-qPCR data, it does not seem to be correct as many studies have shown that the expression of commonly used reference genes is not always stable [20] **(Article [XI.])**. Bacteria are very diverse and their expression profiles are specific even for particular strains. We used several genome-wide transcription studies of *C. beijerinckii* NRRL B-598 to infer its most suitable reference gene [20] **(Article [XI.])**. The selection of candidate reference genes was conducted by a series of bioinformatics steps to filter RNA-Seq data. Besides calculating differential expression between all 66 time-points pairs for all genes in the genome, the coefficient of variation of transcript per million counts was used. Candidate genes were further analyzed by RT-qPCR and *zmp* and *greA* were selected as the most suitable reference genes for *C. beijerinckii* NRRL B-598. These genes are not widely applicable to *C. beijerinckii* species because *pepT* was confirmed to be the most suitable reference gene for the strain *C. beijerinckii* NCIMB 8052 by two studies [19] **(Article [XIV.])**.

3.2.2 Genome-wide Transcription

Genome-wide transcription was formerly examined mostly by microarrays. The term “genome-wide” is not fully appropriate, because microarrays are designed to capture expression of pre-selected genes. Nevertheless, there are hundreds of spots each corresponding to a specific gene on a single microarray and, thus, the technique belongs to the high throughput methods. The disadvantage is that sequence of analyzed genes has to be known to prepare a microarray. Thus, current genome-wide transcription studies of non-model bacteria mostly rely on transcriptome sequencing, i.e., RNA-Seq, because no prior information about analyzed genes is needed and the technique offers an unlimited dynamic range.

The bioinformatics preprocessing of RNA-Seq data is not trivial and consists of numerous steps, e.g., quality assessment and trimming, computational ribodepletion, RNA-Seq specific mapping, and data quantification. The most popular tools to perform these tasks are summarized in Patakova et al., 2022 [19] **(Article [XIV.])**. Data preprocessing depends on the aim of the study, while some quantification techniques, for example, Reads Per Kilobase Million (RPKM), are suitable to compare values within a single sample, other techniques, for example, Transcripts Per Million (TPM), allow to compare values between samples. Eventually, the most reliable result can be achieved with statistical tests to calculate differential expression, i.e., if the expression of the gene between two conditions significantly changed, then can be supplemented by informative visualizations using Venn diagrams, bar plots, or dimensionality reduction techniques [17, 18] **(Articles [VI.], [VII.])**. Expression profiles can be visualized in heatmaps using Z-score to normalize the data. Possibly, moments, where

statistically significant changes were captured, can be highlighted as we did with small arrows in the study by Patakova et al, 2020 [79], see Fig. 3.1. Besides already mentioned comprehensive studies describing the overall response of an organism [17, 18] **(Articles [VI.], [VII.])**, the same data can be reused to describe particular metabolic traits, i.e., acidogenesis, solventogenesis, and metabolic stress [88] or amino acid, metal ion, vitamin and carbohydrate uptake [97].

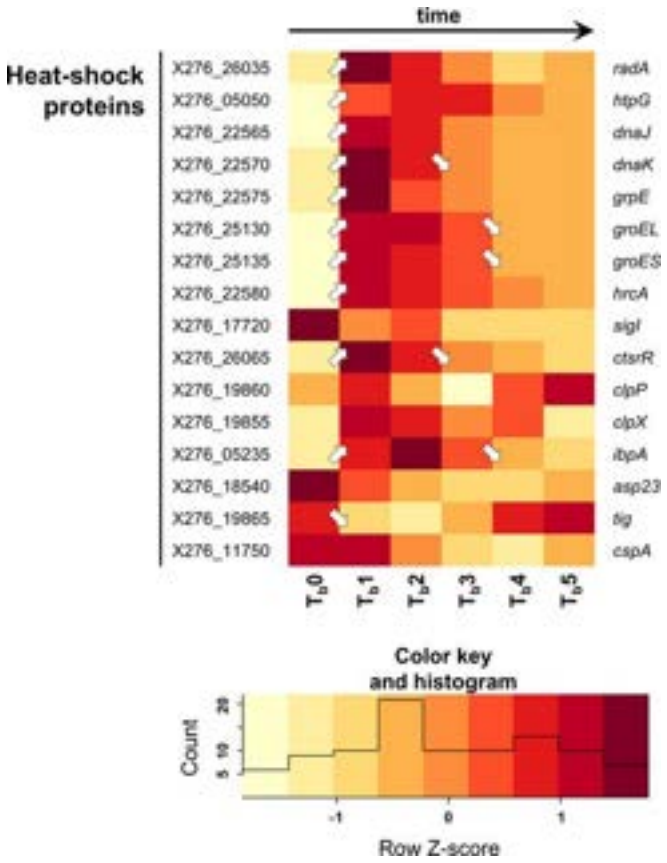


Fig. 3.1 Heatmap of selected genes during shocked ABE fermentation, taken from [79]
Heatmap showing expression of heat-shock protein-coding genes during butanol shocked ABE fermentation in *C. beijerinckii* NRRL B-598 (arrows ↗ and ↘ indicate statistically significant (p-adj <0.001, Benjamini–Hochberg correction) upregulation and downregulation of related genes transcription; if there is no arrow in the figure, transcription was not changed significantly).

Eventually, formerly prepared functional annotation can be used to characterize biological processes in an organism under selected conditions. Simple quantification techniques can be used to estimate the activity of annotated elements as we did for prophages in *C. beijerinckii* NRRL B-598 genome using RPKM values [17] **(Article [VI.])**. More complex techniques, especially differential expression analysis can be used to define regulated genes of interest and a gene universe to summarize biological processes, molecular function, and cellular components describing studied phenomenon in a bacterium with GO enrichment analysis. That way, we described the response of *C. beijerinckii* NRRL B-598 to the butanol shock [18] **(Article [VII.])** and characterized its unregulated processes [20] **(Article [XI.]).**

3.2.3 Annotation with Transcriptomic Data

Some of the regulating genomics elements, e.g., promoters, UTRs, and sRNA genes, remain to be discovered and cannot be directly annotated in a genome sequence by basic structural annotation because their orthologous sequences are unknown. Nevertheless, these elements, when showing some expression, can be inferred from RNA-Seq data. There are not many automated tools for their annotation and they usually work with limited kinds of data. Novel tools will be highly desirable in the near future as transcriptomic studies of non-model bacteria are more and more common nowadays. As adjustments of sequencing protocols are quite common, it is hard to design a tool that can work with any kind of data. Nevertheless, it is possible to propose methods that work with similar sensitivity for the older types of data. Our preliminary results, for example, show that *trans*-encoded sRNA genes can be annotated using old non-stranded RNA-Seq libraries with similar sensitivity to new stranded RNA-Seq data [98].

4 CONCLUSIONS

During the last decade, I contributed to the description and analysis of numerous bacteria. Rather than an austere taxonomic description, I always aimed at functional annotation in order to infer new knowledge that could be further utilized in biotechnology. In this introductory text to the 16 selected articles, I mainly aimed at studies of organisms relevant to white biotechnology, a concept that relies on the utilization of bacterial cell factories for environmentally friendly production of valuable compounds, e.g., biofuels and bioplastics. I believe the biological production of materials of daily use from renewable sources is the only possible future for humankind due to the limited resources of our planet. Nevertheless, biotechnology research that is needed to establish industrially viable projects for a circular economy using bacterial cell factories will not be possible without bioinformatics support. Only thorough bioinformatics description of processes, which were revealed by biotechnology itself, can bring their full understanding needed to adjust these processes and maximize yields. From that point of view, I see bioinformatics being one of the most important tools for current biotechnology.

Nevertheless, bioinformatics analyses of non-model bacteria are not trivial for several reasons. The first obstacle is the fact that many tools and algorithms require input data that are simply unavailable for hitherto undescribed organisms because their measuring would be too costly or is simply impossible due to the missing biological kits to perform desired experiments. The second obstacle, a more general, lies in newly emerging laboratory technologies, mainly constantly changing sequencing technologies that pose new requirements on the computational tools. This is clearly visible from the collection of articles as genomes of presented bacteria needed to be assembled and analyzed by various computational approaches. For that reason, I do not believe it will be possible to propose comprehensive computational pipelines that would be able to automatically annotate and analyze bacterial genomes for biotechnologists in the near future. On the contrary, experienced bioinformaticians with a broad overview of computational tools that are able to combine various tools in order to infer biological knowledge will be needed more than ever before in the multidisciplinary teams working on bacterial biotechnology.

The main aim of this thesis was to present bioinformatics as the modern and constantly changing discipline that offers a wide range of techniques that, when appropriately combined, can offer extremely powerful apparatus to infer biological knowledge for biotechnology and possible industrial utilization of bacteria. Apart from the brief introduction of sequencing technologies and approaches to assemble genomes, I rather aimed at practical aspects of bioinformatics research with as many examples of problems I faced in the research during the last 10 years as possible. I believe this practical illustration could be appealing not only to bioinformaticians but also to biotechnologists that would like to start using bioinformatics as another tool in their research.

REFERENCES

[1] SIMON, Carola and DANIEL, Rolf. Metagenomic Analyses: Past and Future Trends. *Applied and Environmental Microbiology*. 2011. **77**(4), p. 1153–1161.

[2] HUGENHOLTZ, Philip, GOEBEL, Brett M and PACE, Norman R. Impact of Culture-Independent Studies on the Emerging Phylogenetic View of Bacterial Diversity. *Journal of Bacteriology*. 1998. **180**(18), p. 4765–4774.

[3] PERNICOVA, Iva, NOVACKOVA, Ivana, SEDLACEK, Petr, KOURILOVA, Xenie, KOLLER, Martin and OBRUCA, Stanislav. Application of osmotic challenge for enrichment of microbial consortia in polyhydroxyalkanoates producing thermophilic and thermotolerant bacteria and their subsequent isolation. *International Journal of Biological Macromolecules*. 2020. **144**, p. 698–704.

[4] MUSILOVA, Jana, KOURILOVA, Xenie, PERNICOVA, Iva, BEZDICEK, Matej, LENGEROVA, Martina, OBRUCA, Stanislav and SEDLAR, Karel. Novel thermophilic polyhydroxyalkanoates producing strain Aneurinibacillus thermoaerophilus CCM 8960. *Applied Microbiology and Biotechnology*. 2022. 106(12), p. 4669–4681.

[5] VIDENSKA, Petra, SMERKOVA, Kristyna, ZWINSOVA, Barbora, POPOVICI, Vlad, MICENKOVA, Lenka, SEDLAR, Karel and BUDINSKA, Eva. Stool sampling and DNA isolation kits affect DNA quality and bacterial composition following 16S rRNA gene sequencing using MiSeq Illumina platform. *Scientific Reports*. 2019. 9(1), p. 1–14.

[6] PATAKOVA, Petra, KOLEK, Jan, SEDLAR, Karel, KOSCOVA, Pavlina, BRANSKA, Barbora, KUPKOVA, Kristyna, PAULOVA, Leona and PROVAZNIK, Ivo. Comparative analysis of high butanol tolerance and production in clostridia. *Biotechnology Advances*. 2018. 36(3), p. 721–738.

[7] SEDLAR, Karel, KOLEK, Jan, SKUTKOVA, Helena, BRANSKA, Barbora, PROVAZNIK, Ivo and PATAKOVA, Petra. Complete genome sequence of Clostridium pasteurianum NRRL B-598, a non-type strain producing butanol. *Journal of Biotechnology*. 2015. 214, p. 113–114.

[8] SNOPKOVÁ, Kateřina, SEDLÁŘ, Karel, BOSÁK, Juraj, CHALOUPKOVÁ, Eva, SEDLÁČEK, Ivo, PROVAZNÍK, Ivo and ŠMAJS, David. Free-Living Enterobacterium Pragia fontium 24613: Complete Genome Sequence and Metabolic Profiling. *Evolutionary Bioinformatics*. 2017. 13, p. 1176934317700863.

[9] MUSILOVA, Jana, KOURILOVA, Xenie, BEZDICEK, Matej, LENGEROVA, Martina, OBRUCA, Stanislav, SKUTKOVA, Helena and SEDLAR, Karel. First Complete Genome of the Thermophilic Polyhydroxyalkanoates-Producing Bacterium Schlegelella thermodepolymerans DSM 15344. *Genome Biology and Evolution*. 2021. 13(4), p. evab007.

[10] HERMANKOVA, Kristyna, KOURILOVA, Xenie, PERNICOVA, Iva, BEZDICEK, Matej, LENGEROVA, Martina, OBRUCA, Stanislav and SEDLAR, Karel. Complete Genome Sequence of the Type Strain Tepidimonas taiwanensis LMG 22826T, a Thermophilic Alkaline Protease and Polyhydroxyalkanoate Producer. *Genome Biology and Evolution*. 2021. 13(12), p. evab280.

[11] SEDLAR, Karel, VASYLKIVSKA, Maryna, MUSILOVA, Jana, BRANSKA, Barbora, PROVAZNIK, Ivo and PATAKOVA, Petra. Phenotypic and genomic analysis of isopropanol and 1,3-propanediol producer Clostridium diolis DSM 15410. *Genomics*. 2021. 113(1), p. 1109–1119.

[12] VASYLKIVSKA, Maryna, BRANSKA, Barbora, SEDLAR, Karel, JURECKOVA, Katerina, PROVAZNIK, Ivo and PATAKOVA, Petra. Phenotypic and Genomic Analysis of Clostridium beijerinckii NRRL B-598 Mutants With Increased Butanol Tolerance. *Frontiers in Bioengineering and Biotechnology*. 2020. 8, p. 1307.

[13] SEDLAR, Karel, KOLEK, Jan, PROVAZNIK, Ivo and PATAKOVA, Petra. Reclassification of non-type strain Clostridium pasteurianum NRRL B-598 as Clostridium beijerinckii NRRL B-598. *Journal of Biotechnology*. 2017. 244, p. 1–3.

[14] SEDLAR, Karel, NYKRYNOVA, Marketa, BEZDICEK, Matej, BRANSKA, Barbora, LENGEROVA, Martina, PATAKOVA, Petra and SKUTKOVA, Helena. Diversity and Evolution of Clostridium beijerinckii and Complete Genome of the Type Strain DSM 791T. *Processes*. 2021. 9(7), p. 1196.

[15] OBRUČA, Stanislav, DVOŘÁK, Pavel, SEDLÁČEK, Petr, KOLLER, Martin, SEDLÁŘ, Karel, PERNICOVÁ, Iva and ŠAFRÁNEK, David. Polyhydroxyalkanoates synthesis by halophiles and thermophiles: towards sustainable production of microbial bioplastics. *Biotechnology Advances*. 2022. 58, p. 107906.

[16] KOLEK, Jan, SEDLAR, Karel, PROVAZNIK, Ivo and PATAKOVA, Petra. Dam and Dcm methylations prevent gene transfer into Clostridium pasteurianum NRRL B-598: Development of methods for electrotransformation, conjugation, and sonoporation. *Biotechnology for Biofuels*. 2016. 9(1), p. 14.

[17] SEDLAR, Karel, KOSCOVA, Pavlina, VASYLKIVSKA, Maryna, BRANSKA, Barbora, KOLEK, Jan, KUPKOVA, Kristyna, PATAKOVA, Petra and PROVAZNIK, Ivo. Transcription profiling of butanol producer Clostridium beijerinckii NRRL B-598 using RNA-Seq. *BMC Genomics*. 2018. 19(1), p. 415.

[18] SEDLAR, Karel, KOLEK, Jan, GRUBER, Markus, JURECKOVA, Katerina, BRANSKA, Barbora, CSABA, Gergely, VASYLKIVSKA, Maryna, ZIMMER, Ralf, PATAKOVA, Petra and PROVAZNIK, Ivo. A transcriptional response of Clostridium beijerinckii NRRL B-598 to a butanol shock. *Biotechnology for Biofuels*. 2019. 12(1), p. 243.

[19] PATAKOVA, Petra, BRANSKA, Barbora, VASYLKIVSKA, Maryna, JURECKOVA, Katerina, MUSILOVA, Jana, PROVAZNIK, Ivo and SEDLAR, Karel. Transcriptomic studies of solventogenic clostridia, Clostridium acetobutylicum and Clostridium beijerinckii. *Biotechnology Advances*. 2022. 58, p. 107889.

[20] JURECKOVA, Katerina, RASCHMANOVA, Hana, KOLEK, Jan, VASYLKIVSKA, Maryna, BRANSKA, Barbora, PATAKOVA, Petra, PROVAZNIK, Ivo and SEDLAR, Karel. Identification and Validation of Reference Genes in Clostridium beijerinckii NRRL B-598 for RT-qPCR Using RNA-Seq Data. *Frontiers in Microbiology*. 2021. 12, p. 476.

[21] LAND, Miriam, HAUSER, Loren, JUN, Se-Ran, NOOKAEW, Intawat, LEUZE, Michael R., AHN, Tae-Hyuk, KARPINETS, Tatiana, LUND, Ole, KORA, Guruprased, WASSENAAR, Trudy, POUDEL, Suresh and USSERY, David W. Insights from 20 years of bacterial genome sequencing. *Functional & integrative genomics*. 2015. **15**(2), p. 141–61.

[22] BENNETT, Gordon M. and MORAN, Nancy A. Small, smaller, smallest: The origins and evolution of ancient dual symbioses in a phloem-feeding insect. *Genome Biology and Evolution*. 2013. **5**(9), p. 1675–1688.

[23] SCHNEIKER, Susanne, PERLOVA, Olena, KAISER, Olaf, GERTH, Klaus, ALICI, Aysel, ALTMEYER, Matthias O, et al. Complete genome sequence of the myxobacterium Sorangium cellulosum. *Nature biotechnology*. 2007. **25**(11), p. 1281–1289.

[24] BINNEWIES, Tim T., MOTRO, Yair, HALLIN, Peter F., LUND, Ole, DUNN, David, LA, Tom, HAMPSON, David J., BELLGARD, Matthew, WASSENAAR, Trudy M. and USSERY, David W. Ten years of bacterial genome sequencing: Comparative-genomics-based discoveries. *Functional and Integrative Genomics*. 2006. **6**(3), p. 165–185.

[25] SAYERS, Eric W., BOLTON, Evan E., BRISTER, J. Rodney, CANESE, Kathi, CHAN, Jessica, COMEAU, Donald C., et al. Database resources of the national center for biotechnology information. *Nucleic acids research*. 2022. **50**(D1), p. D20–D26.

[26] SANGER, Frederick, NICKLEN, S and COULSON, A R. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*. 1977. **74**(12), p. 5463–5467.

[27] ŠVEC, Pavel, KRÁLOVÁ, Stanislava, STAŇKOVÁ, Eva, HOLOCHOVÁ, Pavla, SEDLÁŘ, Karel, KOUDELKOVÁ, Sylva, et al. Pedobacter fastidiosus sp. nov., isolated from glacial habitats of maritime Antarctica. *International Journal of Systematic and Evolutionary Microbiology*. 2022. 72(4), p. 005309.

[28] SEDLÁČEK, Ivo, HOLOCHOVÁ, Pavla, SOBOTKA, Roman, BUSSE, Hans-Jürgen, ŠVEC, Pavel, KRÁLOVÁ, Stanislava, ŠEDO, Ondrej, PILNÝ, Jan, STAŇKOVÁ, Eva, KOUBLOVÁ, Vendula and SEDLÁŘ, Karel. Classification of a Violacein-Producing Psychrophilic Group of Isolates Associated with Freshwater in Antarctica and Description of Rugamonas violacea sp. nov. *Microbiology spectrum*. 2021. 9(1), p. e0045221.

[29] SEDLÁČEK, Ivo, HOLOCHOVÁ, Pavla, BUSSE, Hans Jürgen, KOUBLOVÁ, Vendula, KRÁLOVÁ, Stanislava, ŠVEC, Pavel, SOBOTKA, Roman, STAŇKOVÁ, Eva, PILNÝ, Jan, ŠEDO, Ondrej, SMOLÍKOVÁ, Jana and SEDLÁŘ, Karel. Characterisation of Waterborne Psychrophilic Massilia Isolates with Violacein Production and Description of Massilia antarctica sp. nov. *Microorganisms*. 2022. 10(4), p. 704.

[30] ŠVEC, Pavel, BUSSE, Hans Jürgen, SEDLÁŘ, Karel, MUSILOVÁ, Jana, KRÁLOVÁ, Stanislava, STAŇKOVÁ, Eva, ŠEDO, Ondrej, KRSEK, Daniel, KOUBLOVÁ, Vendula, KRIŠTOFOVÁ, Lucie and SEDLÁČEK, Ivo. Corynebacterium antarcticum sp. nov., Corynebacterium marambiense sp. nov., Corynebacterium meridianum sp. nov., and Corynebacterium pygosceleis sp. nov., isolated from Adélie penguins (Pygoscelis adeliae). *Systematic and Applied Microbiology*. 2023. 46(1), p. 126390.

[31] ALTSCHUL, Stephen F., GISH, Warren, MILLER, Webb, MYERS, Eugene W. and LIPMAN, David J. Basic local alignment search tool. *Journal of molecular biology*. 1990. **215**(3), p. 403–410.

[32] HEDLUND, Brian P., CHUVOCHINA, Maria, HUGENHOLTZ, Philip, KONSTANTINIDIS, Konstantinos T., MURRAY, Alison E., PALMER, Marike, PARKS, Donovan H., PROBST, Alexander J., REYSENBACH, Anna Louise, RODRIGUEZ-R, Luis M., ROSSELLO-MORA, Ramon, SUTCLIFFE, Iain C., VENTER, Stephanus N. and WHITMAN, William B. SeqCode: a nomenclatural code for prokaryotes described from sequence data. *Nature Microbiology*. 2022. **7**(10), p. 1702–1708.

[33] METZKER, Michael L. Sequencing technologies — the next generation. *Nature Reviews Genetics*. 2009. **11**(1), p. 31–46.

[34] WANG, Zhong, GERSTEIN, Mark and SNYDER, Michael. RNA-Seq: A revolutionary tool for transcriptomics. *Nature Reviews Genetics*. 2009. **10**(1), p. 57–63.

[35] PARK, Peter J. ChIP–seq: advantages and challenges of a maturing technology. *Nature Reviews Genetics*. 2009. **10**(10), p. 669–680.

[36] BLEIDORN, Christoph. Third generation sequencing: technology and its potential impact on evolutionary biodiversity research. *Systematics and Biodiversity*. 2015. **14**(1), p. 1–8.

[37] EWING, Brent, EWING, Brent, HILLIER, Ladeana, HILLIER, Ladeana, WENDL, Michael C, WENDL, Michael C, GREEN, Phil and GREEN, Phil. Base-Calling of Automated Sequencer Traces Using Phred. I. Accuracy Assessment. *Genome Research*. 2005. **8**(206), p. 175–185.

[38] EWELS, Philip, MAGNUSSON, Måns, LUNDIN, Sverker and KÄLLER, Max. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*. 2016. **32**(19), p. 3047–3048.

[39] LANFEAR, R., SCHALAMUN, M., KAINER, D., WANG, W. and SCHWESSINGER, B. MinIONQC: Fast and simple quality control for MinION sequencing data. *Bioinformatics*. 2019. **35**(3), p. 523–525.

[40] LI, Zhenyu, CHEN, Yanxiang, MU, Desheng, YUAN, Jianying, SHI, Yujian, ZHANG, Hao, GAN, Jun, LI, Nan, HU, Xuesong, LIU, Binghang, YANG, Bicheng and FAN, Wei. Comparison of the two major classes of assembly algorithms: Overlap-layout-consensus and de-bruijn-graph. *Briefings in Functional Genomics*. 2012. **11**(1), p. 25–37.

[41] MILLER, Jason R., KOREN, Sergey and SUTTON, Granger. *Assembly algorithms for next-generation sequencing data*. 2010. ISBN 1089-8646 (Electronic)r0888-7543 (Linking).

[42] WICK, Ryan R. and HOLT, Kathryn E. Benchmarking of long-read assemblers for prokaryote whole genome sequencing. *F1000Research*. 2019. **8**, p. 2138.

[43] LI, Heng and DURBIN, Richard. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009. **25**(14), p. 1754–1760.

[44] LI, Heng. Minimap2: pairwise alignment for nucleotide sequences. BIROL, Inanc (ed.), *Bioinformatics*. 2018. **34**(18), p. 3094–3100.

[45] GERHARDT, Kenji, RUIZ-PEREZ, Carlos A., RODRIGUEZ-R, Luis M., CONRAD, Roth E. and KONSTANTINIDIS, Konstantinos T. RecruitPlotEasy: An Advanced Read Recruitment Plot Tool for Assessing Metagenomic Population Abundance and Genetic Diversity. *Frontiers in Bioinformatics*. 2022. **1**, p. 88.

[46] LI, H., HANDSAKER, B., WYSOKER, A., FENNELL, T., RUAN, J., HOMER, N., MARTH, G., ABECASIS, G. and DURBIN, R. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*.

2009. **25**(16), p. 2078–2079.

[47] WALKER, Bruce J., ABEEL, Thomas, SHEA, Terrance, PRIEST, Margaret, ABOUELLIEL, Amr, SAKTHIKUMAR, Sharadha, CUOMO, Christina A., ZENG, Qiandong, WORTMAN, Jennifer, YOUNG, Sarah K. and EARL, Ashlee M. Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. WANG, Junwen (ed.), *PLoS ONE*. 2014. **9**(11), p. e112963.

[48] VASER, Robert, SOVIĆ, Ivan, NAGARAJAN, Niranjan and ŠIKIĆ, Mile. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Research*. 2017. **27**(5), p. 737–746.

[49] CORNILLOT, Emmanuel, NAIR, Ramesh V., PAPOUTSAKIS, Eleftherios T. and SOUCAILLE, Philippe. The genes for butanol and acetone formation in Clostridium acetobutylicum ATCC 824 reside on a large plasmid whose loss leads to degeneration of the strain. *Journal of Bacteriology*. 1997. **179**(17), p. 5442.

[50] KOLMOGOROV, Mikhail, YUAN, Jeffrey, LIN, Yu and PEVZNER, Pavel A. Assembly of long, error-prone reads using repeat graphs. *Nature Biotechnology*. 2019. **37**(5), p. 540–546.

[51] KOLEK, Jan, SEDLÁŘ, Karel, PROVAZNÍK, Ivo and PATÁKOVÁ, Petra. Draft Genome Sequence of Clostridium pasteurianum NRRL B-598, a Potential Butanol or Hydrogen Producer. *Genome Announcements*. 2014. 2(2), p. 192–206.

[52] CHIN, Chen Shan, ALEXANDER, David H., MARKS, Patrick, KLAMMER, Aaron A., DRAKE, James, HEINER, Cheryl, CLUM, Alicia, COPELAND, Alex, HUDDLESTON, John, EICHLER, Evan E., TURNER, Stephen W. and KORLACH, Jonas. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nature Methods*. 2013. **10**(6), p. 563–569.

[53] KEARSE, Matthew, MOIR, Richard, WILSON, Amy, STONES-HAVAS, Steven, CHEUNG, Matthew, STURROCK, Shane, BUXTON, Simon, COOPER, Alex, MARKOWITZ, Sidney, DURAN, Chris, THIERER, Tobias, ASHTON, Bruce, MEINTJES, Peter and DRUMMOND, Alexei. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*. 2012. **28**(12), p. 1647–1649.

[54] GAO, Feng and ZHANG, Chun-Ting. Ori-Finder: a web-based system for finding oriCs in unannotated bacterial genomes. *BMC bioinformatics*. 2008. **9**, p. 79.

[55] MADERANKOVA, Denisa, SEDLAR, Karel, VITEK, Martin and SKUTKOVA, Helena. The identification of replication origin in bacterial genomes by cumulated phase signal. In : *2017 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology, CIBCB 2017*. Institute of Electrical and Electronics Engineers Inc., 2017. p. 267–271. ISBN 9781467389884.

[56] SNOPKOVÁ, Kateřina, SEDLÁŘ, Karel, BOSÁK, Juraj, CHALOUPKOVÁ, Eva, PROVAZNÍK, Ivo and ŠMAJS, David. Complete Genome Sequence of Pragia fontium 24613, an Environmental Bacterium from the Family Enterobacteriaceae. *Genome Announcements*. 2015. 3(4), p. e00740-15.

[57] RHOADS, Anthony and AU, Kin Fai. PacBio Sequencing and Its Applications. *Genomics, Proteomics & Bioinformatics*. 2015. **13**(5), p. 278–289.

[58] BANKEVICH, Anton, NURK, Sergey, ANTIPOV, Dmitry, GUREVICH, Alexey A., DVORKIN, Mikhail, KULIKOV, Alexander S., et al. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*. 2012. **19**(5), p. 455–477.

[59] KOSUGI, Shunichi, HIRAKAWA, Hideki and TABATA, Satoshi. GMcloser: Closing gaps in assemblies accurately with a likelihood-based selection of contig or long-read alignments. *Bioinformatics*. 2015. **31**(23), p. 3733–3741.

[60] KURTZ, Stefan, PHILLIPPY, Adam, DELCHER, Arthur L, SMOOT, Michael, SHUMWAY, Martin, ANTONESCU, Corina and SALZBERG, Steven L. Versatile and open software for comparing large genomes. *Genome biology*. 2004. **5**(2), p. R12.

[61] WICK, Ryan R., JUDD, Louise M., GORRIE, Claire L. and HOLT, Kathryn E. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLOS Computational Biology*. 2017. **13**(6), p. e1005595.

[62] LUO, Hao and GAO, Feng. DoriC 10.0: An updated database of replication origins in prokaryotic genomes including chromosomes and plasmids. *Nucleic Acids Research*. 2019. **47**(D1), p. D74–D77.

[63] NAWY, Tal. Single-cell sequencing. *Nature Methods*. 2013. **11**(1), p. 18.

[64] PERIWAL, Vinita and SCARIA, Vinod. Insights into structural variations and genome rearrangements in prokaryotic genomes. *Bioinformatics*. 2015. **31**(1), p. 1–9.

[65] JUGAS, Robin, SEDLAR, Karel, VITEK, Martin, NYKRYNOVA, Marketa, BARTON, Vojtech, BEZDICEK, Matej, LENGEROVA, Martina and SKUTKOVA, Helena. CNproScan: Hybrid CNV detection for bacterial genomes. *Genomics*. 2021. 113(5), p. 3103–3111.

[66] HUANG, Tzu Wen, CHEN, Te Li, CHEN, Ying Tsong, LAUDERDALE, Tsai Ling, LIAO, Tsai Lien, LEE, Yi Tzu, et al. Copy Number Change of the NDM-1 Sequence in a Multidrug-Resistant Klebsiella pneumoniae Clinical Isolate. *PLOS ONE*. 2013. **8**(4), p. e62774.

[67] NYKRYNOVA, Marketa, BARTON, Vojtech, SEDLAR, Karel, BEZDICEK, Matej, LENGEROVA, Martina and SKUTKOVA, Helena. Word Entropy-Based Approach to Detect Highly Variable Genetic Markers for Bacterial Genotyping. *Frontiers in Microbiology*. 2021. 12, p. 631605.

[68] KIVIOJA, Teemu, VÄHÄRAUTIO, Anna, KARLSSON, Kasper, BONKE, Martin, ENGE, Martin, LINNARSSON, Sten and TAIPALE, Jussi. Counting absolute numbers of molecules using unique molecular identifiers. *Nature Methods*. 2011. **9**(1), p. 72–74.

[69] MCKENNA, Aaron, HANNA, Matthew, BANKS, Eric, SIVACHENKO, Andrey, CIBULSKIS, Kristian, KERNYTSKY, Andrew, GARIMELLA, Kiran, ALTSHULER, David, GABRIEL, Stacey, DALY, Mark and DEPRISTO, Mark A. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*. 2010. **20**(9), p. 1297–1303.

[70] SEDLÁŘ, Karel, MAYER, Jiří, RÁČIL, Zdeněk, KOCMANOVÁ, Iva, BEZDÍČEK, Matěj and LENGEROVÁ, Martina. Způsob typizace bakteriálních kmenů Escherichia coli a oligonukleotidy pro použití při tomto způsobu. 308046. 2019. CZ : Úřad průmyslového vlastnictví.

[71] TATUSOVA, Tatiana, DICUCCIO, Michael, BADRETDIN, Azat, CHETVERNIN, Vyacheslav, NAWROCKI, Eric P., ZASLAVSKY, Leonid, LOMSADZE, Alexandre, PRUITT, Kim D., BORODOVSKY, Mark and OSTELL, James. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Research*. 2016. **44**(14), p. 6614–6624.

[72] AZIZ, Ramy K., BARTELS, Daniela, BEST, Aaron, DEJONGH, Matthew, DISZ, Terrence, EDWARDS, Robert A., et al. The RAST Server: Rapid annotations using subsystems technology. *BMC Genomics*. 2008. **9**(1), p. 75.

[73] DOUDNA, Jennifer A. and CHARPENTIER, Emmanuelle. The new frontier of genome engineering with CRISPR-Cas9. *Science*. 2014. **346**(6213), p. 1258096.

[74] BISWAS, Ambarish, STAALS, Raymond H.J., MORALES, Sergio E., FINERAN, Peter C. and BROWN, Chris M. CRISPRDetect: A flexible algorithm to define CRISPR arrays. *BMC Genomics*. 2016. **17**(1), p. 356.

[75] SOREK, Rotem, LAWRENCE, C Martin and WIEDENHEFT, Blake. CRISPR-mediated adaptive immune systems in bacteria and archaea. *Annual Review of Biochemistry*. 2013. **82**, p. 237–266.

[76] PILIGRIMOVA, Emma G., KAZANTSEVA, Olesya A., KAZANTSEV, Andrey N., NIKULIN, Nikita A., SKORYNINA, Anna V., KOPOSOVA, Olga N. and SHADRIN, Andrey M. Putative plasmid prophages of Bacillus cereus sensu lato may hold the key to undiscovered phage diversity. *Scientific Reports*. 2021. **11**(1), p. 7611.

[77] TABOADA, Blanca, ESTRADA, Karel, CIRIA, Ricardo and MERINO, Enrique. Operon-mapper: a web server for precise operon identification in bacterial and archaeal genomes. *Bioinformatics*. 2018. **34**(23), p. 4118–4120.

[78] KOURILOVA, Xenie, PERNICOVA, Iva, SEDLAR, Karel, MUSILOVA, Jana, SEDLACEK, Petr, KALINA, Michal, KOLLER, Martin and OBRUCA, Stanislav. Production of polyhydroxyalkanoates (PHA) by a thermophilic strain of Schlegelella thermodepolymerans from xylose rich substrates. *Bioresource Technology*. 2020. 315, p. 123885.

[79] PATAKOVA, Petra, KOLEK, Jan, JURECKOVA, Katerina, BRANSKA, Barbora, SEDLAR, Karel, VASYLKIVSKA, Maryna and PROVAZNIK, Ivo. Deeper below the surface—transcriptional changes in selected genes of Clostridium beijerinckii in response to butanol shock. *MicrobiologyOpen*. 2021. 10(1), p. e1146.

[80] KOUŘILOVÁ, Xenie, SCHWARZEROVÁ, Jana, PERNICOVÁ, Iva, SEDLÁŘ, Karel,

MRÁZOVÁ, Kateřina, KRZYŽÁNEK, Vladislav, NEBESÁŘOVÁ, Jana and OBRUČA, Stanislav. The First Insight into Polyhydroxyalkanoates Accumulation in Multi-Extremophilic Rubrobacter xylanophilus and Rubrobacter spartanus. *Microorganisms 2021, Vol. 9, Page 909*. 2021. 9(5), p. 909.

[81] STOTHARD, P. The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *BioTechniques*. 2000. **28**(6).

[82] GALPERIN, Michael Y., WOLF, Yuri I., MAKAROVA, Kira S., ALVAREZ, Roberto Vera, LANDSMAN, David and KOONIN, Eugene V. COG database update: focus on microbial diversity, model organisms, and widespread pathogens. *Nucleic Acids Research*. 2021. **49**(D1), p. D274–D281.

[83] CANTALAPIEDRA, Carlos P., HERŃANDEZ-PLAZA, Ana, LETUNIC, Ivica, BORK, Peer and HUERTA-CEPAS, Jaime. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Molecular Biology and Evolution*. 2021. **38**(12), p. 5825–5829.

[84] ASHBURNER, Michael, BALL, Catherine A., BLAKE, Judith A., BOTSTEIN, David, BUTLER, Heather, CHERRY, J. Michael, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature genetics*. 2000. **25**(1), p. 25–29.

[85] BINNS, David, DIMMER, Emily, HUNTLEY, Rachael, BARRELL, Daniel, O'DONOVAN, Claire and APWEILER, Rolf. QuickGO: a web-based tool for Gene Ontology searching. *Bioinformatics*. 2009. **25**(22), p. 3045.

[86] KANEHISA, Minoru and GOTO, Susumu. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*. 2000. **28**(1), p. 27–30.

[87] KANEHISA, Minoru, SATO, Yoko and MORISHIMA, Kanae. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *Journal of molecular biology*. 2016. **428**(4), p. 726–731.

[88] PATAKOVA, Petra, BRANSKA, Barbora, SEDLAR, Karel, VASYLKIVSKA, Maryna, JURECKOVA, Katerina, KOLEK, Jan, KOSCOVA, Pavlina and PROVAZNIK, Ivo. Acidogenesis, solventogenesis, metabolic stress response and life cycle changes in Clostridium beijerinckii NRRL B-598 at the transcriptomic level. *Scientific Reports*. 2019. 9(1), p. 1371.

[89] MEIER-KOLTHOFF, Jan P. and GÖKER, Markus. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. *Nature Communications*. 2019. **10**(1), p. 2182.

[90] MADERANKOVA, Denisa, JUGAS, Robin, SEDLAR, Karel, VITEK, Martin and SKUTKOVA, Helena. Rapid Bacterial Species Delineation Based on Parameters Derived From Genome Numerical Representations. *Computational and Structural Biotechnology Journal*. 2019. 17, p. 118–126.

[91] SEGATA, Nicola, BÖRNIGEN, Daniela, MORGAN, Xochitl C. and HUTTENHOWER, Curtis. PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. *Nature Communications*. 2013. **4**(1), p. 2304.

[92] CHAUDHARI, Narendrakumar M., GUPTA, Vinod Kumar and DUTTA, Chitra. BPGA- an ultra-fast pan-genome analysis pipeline. *Scientific Reports 2016 6:1*. 2016. **6**(1), p. 24373.

[93] KRUEGER, Felix, KRECK, Benjamin, FRANKE, Andre and ANDREWS, Simon R. DNA methylome analysis using short bisulfite sequencing data. *Nature Methods 2012 9:2*. 2012. **9**(2), p. 145–151.

[94] ROBERTS, Richard J., VINCZE, Tamas, POSFAI, Janos P. and MACELIS, Dana. REBASE: restriction enzymes and methyltransferases. *Nucleic Acids Research*. 2003. **31**(1), p. 418–420.

[95] LI, Ang, WEN, Zhiqiang, FANG, Dahui, LU, Minrui, MA, Yuheng, XIE, Qian and JIN, Mingjie. Developing Clostridium diolis as a biorefinery chassis by genetic manipulation. *Bioresource Technology*. 2020. **305**, p. 123066.

[96] KOLEK, J., DIALLO, M., VASYLKIVSKA, M., BRANSKA, B., SEDLAR, K., LÓPEZ-CONTRERAS, A. M. and PATAKOVA, P. Comparison of expression of key sporulation, solventogenic and acetogenic genes in C. beijerinckii NRRL B-598 and its mutant strain overexpressing spo0A. *Applied Microbiology and Biotechnology*. 2017. 101(22), p. 8279–8291.

[97] VASYLKIVSKA, Maryna, JURECKOVA, Katerina, BRANSKA, Barbora, SEDLAR, Karel, KOLEK, Jan, PROVAZNIK, Ivo and PATAKOVA, Petra. Transcriptional analysis of amino acid, metal ion, vitamin and carbohydrate uptake in butanol-producing Clostridium beijerinckii NRRL

B-598. *PLOS ONE*. 2019. 14(11), p. e0224560.

[98] SEDLAR, Karel and ZIMMER, Ralf. Comparison of Stranded and Non-stranded RNA-Seq in Predicting Small RNAs in a Non-model Bacterium. *Lecture Notes in Computer Science*. 2022. 13347, p. 45–56.

LIST OF ABBREVIATIONS

ANI	Average Nucleotide Identity
BLAST	Basic Local Alignment Search Tool
bp	base pair
BP	Biological Process
Cas	CRISPR-associated system
CC	Cellular Component
CCS	circular consensus sequencing
CDS	coding sequences
CLR	continuous long read
CNV	copy number variation
COG	Clusters of Orthologous Groups
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
DAG	Directed Acyclic Graph
DBG	de Bruijn graph
DDBJ	DNA Data Bank of Japan
dDDH	digital DNA to DNA hybridization
DDH	DNA to DNA hybridization
ENA	European Nucleotide Archive
GC	guanine-cytosine (content)
GO	Gene Ontology
HAI	healthcare-associated infection
MAG	metagenome-assembled genome
MF	Molecular Function
NCBI	National Centre for Biotechnology Information
NGIB	Next Generation Industrial Biotechnology
NGS	next generation sequencing
OLC	Overlap Layout Consensus
ONT	Oxford Nanopore Techniques
PacBio	Pacific Biosciences
PCR	polymerase chain reaction
PHA	polyhydroxyalkanoate
qPCR	quantitative (=real-time) polymerase chain reaction
RM	restriction modification (system)
RPKM	Read Per Kilobase Million
RT-qPCR	quantitative reverse-transcriptase polymerase chain reaction
SMRT	Single Molecule Real Time

SNP	single nucleotide polymorphism
sRNA	small (non-coding) RNA
TGS	third generation sequencing
TPM	Transcripts Per Million
UMI	unique molecular identifier
UTR	untransated region
WGS	whole genome sequencing
WT	wild-type
ZMW	zero-mode waveguide

LIST OF FIGURES

5 PRESENTED JOURNAL ARTICLES

5.1 Article I

SEDLAR, Karel, KOLEK, Jan, SKUTKOVA, Helena, BRANSKA, Barbora, PROVAZNIK, Ivo and PATAKOVA, Petra. Complete genome sequence of Clostridium pasteurianum NRRL B-598, a non-type strain producing butanol. *Journal of Biotechnology*. 2015. 214, p. 113–114. (2015 IF = 2.667, Q2 in BIOTECHNOLOGY & APPLIED MICROBIOLOGY)

DOI: https://doi.org/10.1016/j.jbiotec.2015.09.022

Article I: Complete genome sequence of *Clostridium pasteurianum* NRRL B-598, a non-type strain producing butanol.

Journal of Biotechnology 214 (2015) 113–114

Contents lists available at ScienceDirect

Journal of Biotechnology

journal homepage: www.elsevier.com/locate/jbiotec

Genome announcement

Complete genome sequence of *Clostridium pasteurianum* NRRL B-598, a non-type strain producing butanol

Karel Sedlar[a,*], Jan Kolek[b], Helena Skutkova[a], Barbora Branska[b], Ivo Provaznik[a], Petra Patakova[b]

[a] Department of Biomedical Engineering, Brno University of Technology, Technicka 12, 616 00 Brno, Czech Republic
[b] Department of Biotechnology, University of Chemistry and Technology Prague, Technicka 5, 166 28 Prague, Czech Republic

ARTICLE INFO

Article history:
Received 15 September 2015
Accepted 21 September 2015
Available online 26 September 2015

Keywords:
Clostridium pasteurianum
Genome
Butanol
Oxygen tolerance
Pyrosequencing
Single molecule real-time sequencing

ABSTRACT

The strain *Clostridium pasteurianum* NRRL B-598 is non-type, oxygen tolerant, spore-forming, mesophilic and heterofermentative strain with high hydrogen production and ability of acetone–butanol fermentation (ethanol production being negligible). Here, we present the annotated complete genome sequence of this bacterium, replacing the previous draft genome assembly. The genome consisting of a single circular 6,186,879 bp chromosome with no plasmid was determined using PacBio RSII and Roche 454 sequencing.

Clostridia are a diverse, polyphyletic group of usually rod-shaped, spore-forming anaerobes including a large number of solvent producing species. The strain *Clostridium pasteurianum* NRRL B-598 ferments wide spectrum of substrates containing glucose, xylose, arabinose, mannose, saccharose, lactose, cellobiose or starch by bi-phasic acetone–butanol fermentation. In addition, the strain has exceptional proteolytic abilities which allow alternation of complex nitrogen sources like yeast extract or tryptone in a cultivation medium with cheap protein or peptide containing waste (e.g. whey or feather meal). The start of solventogenic phase in this strain is clearly associated with formation of endospores but no acids re-utilization usually occurs and solventogenesis begins in exponential growth phase. Unlike other *C. pasteurianum* strains, this strain cannot utilize glycerol as a substrate. The strain forms an unknown pale purple pigment, possibly of polyketide origin, the production of which is linked with high solvents yield. For this strain, butanol stress response was studied (Kolek et al., 2015), which resulted in increased content of membrane plasmalogens. Further, *sol* operon and Spo0A global regulator were identified in the genome (Sedlar et al., 2014) and the strain was used for development of flow cytometric analyses enabling cell-level insight to the bacterial population (Patakova et al., 2013; Linhova et al., 2012). Until now, only the draft genome was available (Kolek et al., 2014) which was now replaced with the complete genome sequence. Due to high genetic stability, oxygen tolerance, and versatile sugar-fermenting and proteolytic abilities, this strain has a potential to be a useful platform for further genetic modification.

Total genome DNA of *C. pasteurianum* NRRL B-598 was sequenced using PacBio (*PacBio RSII*; GATC Biotech, Inc., Constance, Germany) and Roche 454 (Roche GS Junior; UCT Prague, Prague, Czech Republic). PacBio single molecule real-time (SMRT) analysis v.2.3 was used for PacBio raw reads (2× SMRT cell, covering ≈490 Mbp) treatment. The HGAP software (Chin et al., 2013) was used for *de novo* genome assembly with ≈30× coverage of self-corrected reads with length over 6626 bp. Contig accuracy was enhanced with the Quiver tool using the entire read set (≈79× coverage). The resulting 3 PacBio contigs were combined with previous Roche assembly containing 138 contigs (Kolek et al., 2014) using Geneious R7 (Kearse et al., 2012). The final contig was circularized while overlaps at its ends were manually trimmed. The genome sequence was arranged to *oriC* predicted by Ori-Finder (Gao and Zhang, 2008). Annotation was added by the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) (http://www.ncbi.nlm.nih.gov/genome/annotation_prok/). ProSplign (Kiryutin et al., 2007) and GeneMarkS+ (Besemer et al., 2001) were used for open reading frame (ORF) detection; tRNAscan-SE (Lowe and Eddy, 1997) was

* Corresponding author.
E-mail addresses: sedlar@feec.vutbr.cz (K. Sedlar), jan.kolek@vscht.cz (J. Kolek), skutkova@feec.vutbr.cz (H. Skutkova), barbora.branska@vscht.cz (B. Branska), provaznik@feec.vutbr.cz (I. Provaznik), petra.patakova@vscht.cz (P. Patakova).

http://dx.doi.org/10.1016/j.jbiotec.2015.09.022
0168-1656/

Article I: Complete genome sequence of *Clostridium pasteurianum* NRRL B-598, a non-type strain producing butanol.

Table 1
Genome features of *Clostridium pasteurianum* NRRL B-598.

Features	Chromosome
Length (bp)	6,186,879
GC content (%)	29.8
Total number of genes	5,365
Protein coding genes (CDSs)	5,002
rRNAs (5S, 16S, 23S)	49
tRNAs	94
ncRNA	1
Pseudogenes	219

used for tRNA prediction, and ribosomal RNAs were predicted by a sequence similarity search using BLAST against an RNA sequence database, and/or using Infernal and Rfam models.

C. pasteurianum NRRL B-598 contains only chromosomal DNA, no plasmids were detected. By filling gaps in the previous draft genome assembly, the length of the complete genome sequence increased to a total of 6,186,879 bases. The G+C content of the sequence was calculated as 29.8%. In total, 5365 genes were predicted by PGAP, including 5,002 protein coding sequences (CDSs), which is less than in previous draft genome. However, 219 sequences were newly annotated as pseudogenes. 49 rRNA, 94 tRNA, and single ncRNA genes were identified in the genome sequence, see summary in Table 1. The genome of *C. pasteurianum* NRRL B-598 is larger than that of type strain *C. pasteurianum* ATCC 6013 (4.35 Mb), which is closely related but probably not identical with the strain DSM 525 (Rotta et al., 2015). The Spo0A master regulator of sporulation is very similar to that of *C. beijerinckii* NCIMB 8052 differing in only a single amino acid (Sedlar et al., 2014).

Nucleotide sequence accession number: the *C. pasteurianum* NRRL B-598 whole genome sequence has been deposited at DDBJ/EMBL/GenBank under accession No. CP011966, replacing the wgs genome AYXR00000000. The strain is available from the Agricultural Research Service (NRRL) Culture Collection.

Acknowledgements

Supported by the grant project GACR P102/11/1068 NanoBioTECell and by the Ministry of Education, Youth and Sports of the Czech Republic, project No. MSM 6046137305. Computational resources were partially provided by the MetaCentrum under the program LM2010005 and the CERIT-SC under the program Centre CERIT Scientific Cloud, part of the Operational Program Research and Development for Innovations, Reg. no. CZ.1.05/3.2.00/08.0144. Special thanks belong to Dr. Heinrich Klefenz from RTM company (Germany) for inspiration and financial support.

References

Besemer, J., Lomsadze, A., Borodovsky, M., 2001. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. Nucleic Acids Res. 29, 2607–2618, http://dx.doi.org/10.1093/nar/29.12.2607.

Chin, C., Alexander, D., Marks, P., Klammer, A., Drake, J., Heiner, C., Clum, A., Copeland, A., Huddleston, J., Eichler, E., Turner, S., Korlach, J., 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat. Methods 10, 563–569, http://dx.doi.org/10.1038/nmeth.2474.

Gao, F., Zhang, C., 2008. Ori-Finder: a web-based system for finding oriCs in unannotated bacterial genomes. BMC Bioinform. 9, 79–85, http://dx.doi.org/10.1186/1471-2105-9-79.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., Drummond, A., 2012. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28, 1647–1649, http://dx.doi.org/10.1093/bioinformatics/bts199.

Kiryutin, B., Souvorov, A., Tatusova, T., 2007. ProSplign—protein to genomic alignment tool. In: In Proc. 11th Annual International Conference in Research in Computational Molecular Biology, San Francisco, USA.

Kolek, J., Patakova, P., Melzoch, K., Sigler, K., Rezanka, T., 2015. Changes in membrane plasmalogens of *Clostridium pasteurianum* during butanol fermentation as determined by lipidomic analysis. PLoS One 10 (3), http://dx.doi.org/10.1371/journal.pone.0122058, e0122058.

Kolek, J., Sedlar, K., Provaznik, I., Patakova, P., 2014. Draft genome sequence of *Clostridium pasteurianum* NRRL B-598, a potential butanol or hydrogen producer. Genome Announc. 2, http://dx.doi.org/10.1128/genomeA.192-14, e00192-14-e00192-14.

Linhova, M., Branska, B., Patakova, P., Lipovsky, J., Fribert, P., Rychtera, M., 2012. Rapid flow cytometric method for viability determination of solventogenic Clostridia. Folia Microbiol. 57, 307–311, http://dx.doi.org/10.1007/s12223-012-0131-8.

Lowe, T.M., Eddy, S.R., 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25, 955–964, http://dx.doi.org/10.1093/nar/25.5.0955.

Patakova, P., Linhova, M., Rychtera, M., Paulova, L., Melzoch, K., 2013. Novel and neglected issues of acetone–butanol–ethanol (ABE) fermentation by clostridia: *Clostridium* metabolic diversity, tools for process mapping and continuous fermentation systems. Biotechnol. Adv. 31, 58–67, http://dx.doi.org/10.1016/j.biotechadv.2012.01.010.

Rotta, C., Poehlein, A., Schwarz, K., McClure, P., Daniel, R., Minton, N.P., 2015. Closed genome Sequence of *Clostridium pasteurianum* ATCC 6013. Genome Announc. 3, e01596-14, http://dx.doi.org/10.1128/genomeA.01596-14.

Sedlar, K., Skutkova, H., Kolek, J., Patakova, P., Provaznik, I., 2014. Identification and characteriazation of sol operon in *Clostridium pasteurianum* NRRL B-598 genome. In: In: Proceedings Of The 2nd International Conference On Chemical Technology. Czech Society of Industrial Chemistry, Mikulov, pp. 435–439, http://dx.doi.org/10.13140/2.1.3688.6402.

5.2 Article II

KOLEK, Jan, **SEDLAR, Karel**, PROVAZNIK, Ivo and PATAKOVA, Petra. Dam and Dcm methylations prevent gene transfer into Clostridium pasteurianum NRRL B-598: Development of methods for electrotransformation, conjugation, and sonoporation. *Biotechnology for Biofuels*. 2016. 9(1), p. 14. (2016 IF = 5.203, Q1 in BIOTECHNOLOGY & APPLIED MICROBIOLOGY)

DOI: https://doi.org/10.1186/s13068-016-0436-y

Article II: Dam and Dcm methylations prevent gene transfer into *Clostridium pasteurianum* NRRL B-598

Kolek et al. Biotechnol Biofuels (2016) 9:14
DOI 10.1186/s13068-016-0436-y

Biotechnology for Biofuels

RESEARCH Open Access

Dam and Dcm methylations prevent gene transfer into *Clostridium pasteurianum* NRRL B-598: development of methods for electrotransformation, conjugation, and sonoporation

Jan Kolek[1*], Karel Sedlar[2], Ivo Provaznik[2] and Petra Patakova[1]

Abstract

Background: Butanol is currently one of the most discussed biofuels. Its use provides many benefits in comparison to bio-ethanol, but the price of its fermentative production is still high. Genetic improvements could help solve many problems associated with butanol production during ABE fermentation, such as its toxicity, low concentration achievable in the cultivation medium, the need for a relatively expensive substrate, and many more. *Clostridium pasteurianum* NRRL B-598 is non-type strain producing butanol, acetone, and a negligible amount of ethanol. Its main benefits are high oxygen tolerance, utilization of a wide range of carbon and nitrogen sources, and the availability of its whole genome sequence. However, there is no established method for the transfer of foreign DNA into this strain; this is the next step necessary for progress in its use for butanol production.

Results: We have described functional protocols for conjugation and transformation of the bio-butanol producer *C. pasteurianum* NRRL B-598 by foreign plasmid DNA. We show that the use of unmethylated plasmid DNA is necessary for efficient transformation or successful conjugation. Genes encoding DNA methylation and those for restriction-modification systems and antibiotic resistance were searched for in the whole genome sequence and their homologies with other clostridial bacteria were determined. Furthermore, activity of described novel type I restriction system was proved experimentally. The described electrotransformation protocol achieved an efficiency 1.2×10^2 cfu/μg DNA after step-by-step optimization and an efficiency of 1.6×10^2 cfu/μg DNA was achieved by the sonoporation technique using a standard laboratory ultrasound bath. The highest transformation efficiency was achieved using a combination of these approaches; sono/electroporation led to an increase in transformation efficiency, to 5.3×10^2 cfu/μg DNA.

Conclusions: Both Dam and Dcm methylations are detrimental for transformation of *C. pasteurianum* NRRL B-598. Methods for conjugation, electroporation, sonoporation, and a combined method for sono/electroporation were established for this strain. The methods described could be used for genetic improvement of this strain, which is suitable for bio-butanol production.

Keywords: *Clostridium*, Butanol, Transformation, Conjugation, Methylation, Dam, Dcm, Electroporation, Sonoporation

*Correspondence: kolekj@vscht.cz
[1] Department of Biotechnology, University of Chemistry and Technology Prague, Technicka 5, 166 28 Prague, Czech Republic
Full list of author information is available at the end of the article

Background

Interest in biofuel production, which could represent a useful substitute for standard fuels derived from fossil resources, has increased significantly over the last decade [1]. Butanol formed during acetone-butanol-ethanol (ABE) fermentation by solventogenic clostridia represents an interesting option for biofuel production, especially taking into account its physico–chemical properties that better suit requirements of gasoline motors compared to ethanol. Although butanol production by ABE has been known for more than 100 years [2], its industrial-scale production is hampered by a low final concentration, lower yield compared to ethanol, and in most species, an association of butanol production with sporulation. In addition, clostridia, including solventogenic species, are a polyphyletic group of bacteria, where transfer of knowledge gathered for one species, or even strain to another species/strain is difficult if not impossible. Most knowledge regarding the ABE process has been obtained from a single strain, *Clostridium acetobutylicum* ATCC 824, which differs in many features from other solventogenic clostridia [3]. Most other species, with the exception of *C. beijerinckii* NCIMB 8052 [4], have been described relatively poorly. These drawbacks have precluded the biotechnological production of bio-butanol on a larger scale [5]. Genetics and metabolic engineering represent new approaches with the possibility of significantly improving the ABE process.

The existence of methods for genetic manipulation of industrial microorganisms is generally essential for improving their properties to be appropriate for biofuel production. However, these methods are also very important for better, quicker and more effective research that could lead to the acquisition of important information useful in industrial processes. The most commonly used method for introducing foreign DNA into bacterial cells is transformation (an exogenous molecule of DNA is introduced directly through the cell membrane), conjugation (mediated by tight contact between donor-recipient cells and pili formation), and transduction (mediated by virus particles). In most cases, transformation of Gram-positive bacteria is more difficult compared to Gram-negatives and the development of transformation protocols is demanding. Gram-positive bacteria possess a thick peptidoglycan layer that is further enveloped by a protein S-layer and these bacteria also have only one cytoplasmic membrane, whose distortion can lead to immediate disruption of cell homeostasis and often death.

Transformation of gram-positive, strictly anaerobic bacteria of the genus *Clostridium*, is also usually accompanied by many drawbacks. For the introduction of foreign DNA into clostridial cells, several protocols have been described, based on conjugation with *Escherichia coli* [6, 7] or *Enterococcus* [8] donors, PEG-induced protoplast transformation [9, 10] and more recently, electroporation [11–14]. In addition, some less frequently used transformation approaches such as chemical treatment by Tris-PEG method [15] or sonoporation [16] have been tested.

Here, we describe the development of methods for genetic modification of *C. pasteurianum* NRRL B-598—a solventogenic bacterium producing butanol, acetone, and ethanol [17]. This strain is unique in its exceptional oxygen resistance, which is much higher than the standard butanol-producing model strains such as *C. pasteurianum* ATCC 6013, *C. beijerinckii* NCIMB 8052 or *C. acetobutylicum* ATCC 824. Also the whole genomic sequence is available for this strain [18, 19]. Moreover, only one system for genetic manipulation of *C. pasteurianum* species (type strain ATCC 6013) has been published [12]. We found that the development of methods for introducing DNA into the non-type, and at first sight untransformable, strain *C. pasteurianum* NRRL B-598, was problematic and completely different from other clostridia. We believe that our contribution to this field will strengthen knowledge on bacterial (especially *Clostridium*) transformation methods and encourage those who tackle similar tasks, trying to apply protocols developed for different species/strains, to their particular microorganisms.

Results

Initial transformation attempts

Initially, we conducted a series of pilot experiments based on previous descriptions of the transfer of foreign DNA to other clostridial species, as described in the literature [6, 8, 20, 21]. First, we tested various conditions for plasmid transfer by conjugation using various growth media (TYA, RCM, CBM, P2, YTG), time of conjugation (5–24 h), donor:recipient ratios (from 1:10 to 10:1) and, when no transformants resulted, electroporation was tested using various growth states of cells (OD 0.4–1.2), electroporation buffers (SMP, PEG, glycerol), cuvettes (0.2 and 0.4 cm gap), and electrical parameters (field strength 2.5–15 kV cm^{-1}, time constant 5–20 ms). We also used plasmids from the pMTL80000 series encoding different replicons and antibiotic resistance markers [21]; this was to minimize the possibility that the plasmids may encode unsuitable origins of replication or antibiotic resistance for our strain. Unfortunately, no conditions that we tested during these pilot experiments led to successful transformation.

During pilot experiments, we discovered that strain *C. pasteurianum* NRRL B-598 was naturally resistant to chloramphenicol and thiamphenicol, therefore plasmids

encoding thiamphenicol resistance, classically used as a selection marker for most clostridial strains, were not applicable. On the other hand, such a marker could be used for counter-selection during conjugation. We also verified that *C. pasteurianum* NRRL B-598 was not resistant to erythromycin or spectinomycin (20 µg/µl, 700 µg/µl resp.) at concentrations previously reported in the literature [21], but when a lower concentration of antibiotic was used, or too many cells were seeded onto agar plates, a very strong background growth was observed. Similarly, almost normal growth of cells was observed after longer periods (2–3 days) in TYA broth supplemented with appropriate concentrations of antibiotics.

Bioinformatics analysis of the *C. pasteurianum* NRRL B-598 genome

Because all attempts at plasmid transformation of our strain failed, we decided to perform a more detailed bioinformatics analysis. The main purpose was to reveal genes encoding putative restriction-modification (R-M) systems that could present a problem during transformation of clostridia, and genes encoding putative DNA methyltransferases that could be connected with these R-M systems for protection of their own DNA [8, 12, 22, 23].

We took advantage of SMRT sequencing data used for the genome assembly [19] to study DNA methylation on a genome-wide scale. We analyzed all base modifications to determine modified sequence motifs. Out of the total, 2033 positions in the *C. pasteurianum* NRRL B-598 genome were detected as being methylated (m4C or m6A) with the majority being m6A methylations (1996 positions). Both detected motifs (GA**A**YNNNNNNNNRTANYC, GAYNNNNNNC**T**AG) demonstrated novel recognition sequences that have not been described previously. Letters in bold denote methylated bases. Highlighted 'T' represents methylation of 'A' in the opposite strand.

The data were deposited in the REBASE PacBio database (http://rebase.neb.com/cgi-bin/pacbiolist) [24] and were connected to the R-M system based on homology searching. The detected methylation motifs, both m6A types, are summarized in Table 1, along with the corresponding methyl transferase (MT)-encoding genes.

In addition to above-mentioned type I R-M systems, three more putative R-M systems were predicted, including two type II R-M systems and a single type IV R-M system. A summary of all five systems is found in Table 2. BLAST results also showed that no genes homologous to *E. coli* Dam and Dcm were present in the *C. pasteurianum* NRRL B-598 genome.

We also searched for antibiotic resistance genes. In total, 28 ORFs with antibiotic resistance functions, divided into nine resistance classes, were identified in the genome. All of these ORFs were assigned GenBank accession numbers for the relevant protein product (Fig. 1). As expected, we verified the presence of a gene for chloramphenicol acetyltransferase (cat, [GenBank: ALB45592]) that encoded resistance to chloramphenicol and thiamphenicol, as observed during our experiments. Moreover, genes encoding erythromycin or spectinomycin resistance were not identified. A substantial part of the antibiotic resistance of *C. pasteurianum* NRRL B-598 is mediated by an antibiotic efflux system.

Investigation of potential restriction barriers

As described previously, nucleases can be located on the surface of cells and in some cases and degradation of DNA can already start after adding DNA to the cells [25]. In other cases, enzymes with nuclease activity are located in the cytoplasm. Hence we examined nuclease activities in both the protoplast crude lysate (without any parts of the cell envelope) as well as in the whole cell extract.

We did not detect any restriction activity when pMTL83253 (plasmid does not contain motifs of predicted type I R-M systems) was incubated with crude extracts and whole cell lysate. In the case of pMTL82254 (contains one of each predicted motifs), plasmid DNA was nearly completely digested in broad spectrum of cultivation conditions. Restriction did not provide separate bands (DNA fragments) like in case of cultivation with crude extract from *C. pasteurianum* DSM 525, but led to one fuzzy smear (see Fig. 2). The same restriction pattern was obtained at 30 and 37 °C.

Influence of methylation and establishment of an electroporation protocol

As a next step, we wanted to test whether plasmid DNA without Dam and Dcm methylation could be used for transformation. We extracted plasmids from *E. coli*

Table 1 Methylated motif detected for *C. pasteurianum* NRRL B-598

R-M system type	Motifs (±strand)	No. in genome	No. detected (±strand)	% detected (±strand)	Locus tag	Nomenclature
I	GRNTAYNNNNNNNRTTC/ GAAYNNNNNNNNRTANYC	406	385/380	94.83/93.60	X276_10630	M.Cpa598I
I	CTAGNNNNNNRTC/ GAYNNNNNNCTAG	606	573/560	94.55/92.41	X276_12360	M.Cpa598II

Table 2 R-M systems in *C. pasteurianum* NRRL B-598 genome

Type	Name	Gene[a]	Meth. type	Recognition	Locus (X276_)	Most similar (% identity)
I	Cpa598IP	R	m6A	GAAYNNNNNNNRTANYC	10620	CspMORF4102P (95 %)
	M.Cpa598I	M			10630	M.CbeG117ORFCP (97 %)
	S.Cpa598I	S			10635	S.CspMORF4102P (49 %)
I	Cpa598IIP	R	m6A	GAYNNNNNNCTAG	12355	Csc25775ORFJP (90 %)
	M.Cpa598II	M			12360	M.Csc25775ORFJP (96 %)
	S.Cpa598II	S			12365	S.Bme201ORFGP (56 %)
II	M1.Cpa598ORF20205	M	m5C	–	–	M1.CboKAPB3ORF12160P (87 %)
	M2.Cpa598ORF20205	M			01545	M2.CboKAPB3ORF12160P (84 %)
	R1.Cpa598ORF20205	R			–	R2.Cce743ORF4007P (46 %)
	R2.Cpa598ORF20205	R			01555	R1.Bce3081ORF2217P (31 %)
II	M.Cpa598ORF2410P	M	m6A	GATC[b]	20735	M.Cbe598ORF1284P (100 %)
IV	Cpa598ORF12465P	R	–	–	12465	Cdi15410ORFAP (93 %)

[a] *R* restriction endonuclease, *M* restriction endonuclease coupled methylation protein, *S* R-M specific protein

[b] Predicted recognition site

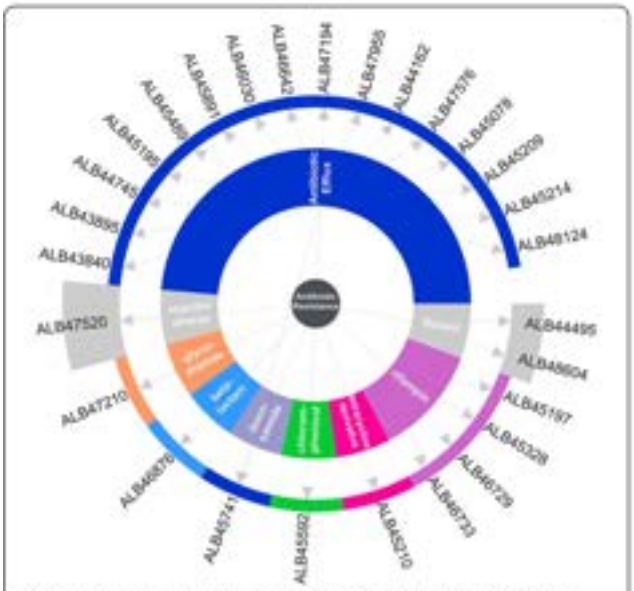
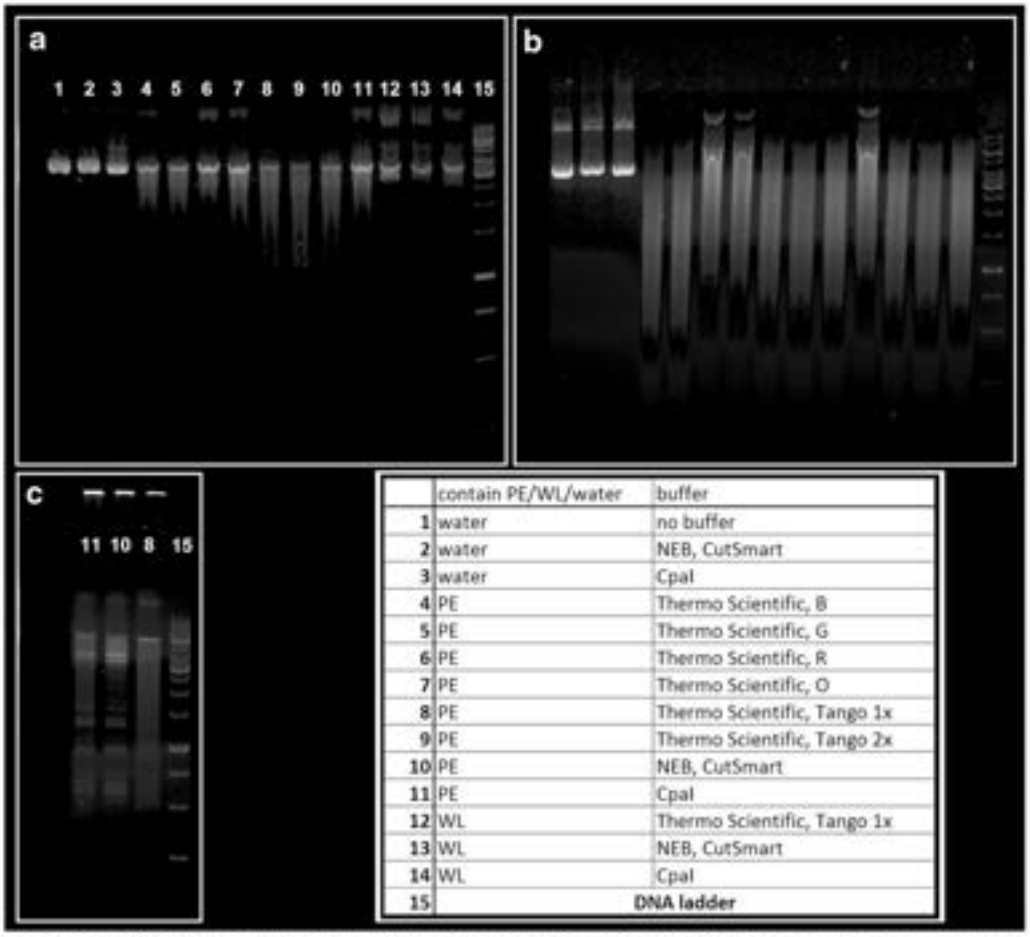


Fig. 1 Antibiotic resistance genes in *C. pasteurianum* NRRL B-598 genome. Overall resistance in the *center*, resistance classes in the *middle*, and individual resistance genes (and their NCBI accession numbers) on the *outer ring*

JM110 (*dam−/dcm−*), a strain used for preparation of unmethylated DNA. After pilot electrotransformation experiments using unmethylated pMTL83253 (containing the pCB102 origin derived from *C. butyricum*) and conditions described previously for *C. beijerinckii* [25], a few erythromycin-resistant colonies (1–12 CFU) were obtained after 48 h of growth on selective agar medium. Also other tested plasmids (pMTL83353-pCB102 replicon and spectinomycin selection marker; pMTL82251-pBP1 replicon; pMTL84251-pCD6 replicon; pMTL85251-pIM13 replicon) were transformed successfully but the CFU yields were much lower (a maximum of 4 CFU). Because of the best transformation efficiency achieved, as well as the fact that the pCB102 origin is the replicon that is used, for example, in standard pMTL007 plasmids (ClosTron system) [7] used for fast and specific knock-outs, we performed all following experiments with pMTL83253. The presence of pMTL83253 in erythromycin-resistant colonies was verified by its isolation and restriction digestion by *Pst*I. Bands of the digested DNA were compared to bands of pMTL83253 isolated from *E. coli* and digested in the same way (Fig. 3). The presence of pMTL83253 was confirmed in all erythromycin-resistant colonies that we tested.

After achieving successful transformation, we aimed to improve transformation efficiency for unmethylated plasmid DNA because the twelve colonies observed (observed maximum) corresponded to a transformation efficiency of only 6 cfu/μg DNA, which is very low and would not be compatible with the use of this method for genetic manipulations.

Initially, we tested different voltages (2500–15,000 V cm^{-1}). A second parameter, investigated and optimized during the first experiments, was the growth state of the cells, represented by culture optical density. For this purpose, we prepared electrocompetent cells from cultures of different OD_{600} (0.6–0.8 and 1.2–1.4), representing the previously used states of culture for electrotransformation of clostridia. When cells at an OD_{600} of around 1.2–1.4 were used, transformation efficiency was significantly improved (Fig. 4). In the following electroporation experiments, time constant, as the main parameter of electroporation, was investigated using the best voltage and cell growth conditions (see



	contain PE/WL/water	buffer
1	water	no buffer
2	water	NEB, CutSmart
3	water	CpaI
4	PE	Thermo Scientific, B
5	PE	Thermo Scientific, G
6	PE	Thermo Scientific, R
7	PE	Thermo Scientific, O
8	PE	Thermo Scientific, Tango 1x
9	PE	Thermo Scientific, Tango 2x
10	PE	NEB, CutSmart
11	PE	CpaI
12	WL	Thermo Scientific, Tango 1x
13	WL	NEB, CutSmart
14	WL	CpaI
15	DNA ladder	

Fig. 2 Testing the presence of potential restriction barriers. Cultivation of pMTL83253 (**a**) and pMTL82254 (**b**) with crude protoplast extract (*PE*) and whole cell lysate (*WL*) prepared from *C. pasteurianum* NRRL B-598 at 37 °C. Positive control (**c**): cultivation of pMTL83253 with *PE* prepared from *C. pasteurianum* DSM 525 by the same method

above). We observed that shorter electric pulses (5 ms) were significantly better for transformation efficiency compared to higher values. CFUs obtained using different time constants are shown in Fig. 4. Square-wave pulse delivery was also tested, but transformation efficiencies were significantly lower than with exponential pulse mode (see Fig. 4).

We also tested a set of various electroporation buffers (30 % PEG 8000 and SMP buffer at different pH values). However, no increase in transformation efficiency was obtained in any other buffers during these experiments. The addition of cell-wall weakening additives (different concentrations of glycine, ampicillin or Tween 80) or treatments with various concentrations of lysozyme prior to electroporation, which have been described previously [12, 26] as methods for significantly increasing transformation efficiency in Gram-positive bacteria, was not successful and no transformants or poor transformation efficiencies were observed (data not shown). Generally, very poor growth was observed in the presence of low concentrations of glycine (more than 0.25 %), even with sucrose or PEG osmotic protection. Equally, addition of osmoprotective agents (various concentrations of sucrose, PEG or lactose) to the recovery medium always had detrimental effects on growth and transformation efficiency, and addition of sucrose to the growth medium at high concentrations (0.2 M and more) led to a significant decrease in growth. Importantly, when culture degeneration [27] was observed (represented mainly by formation of very long, mycelium-like cells in log and late-log phase), transformation efficiency was reduced drastically and only a few colonies grew on the selective medium.

After optimization of electrotransformation steps, we wanted to better understand the influence of Dam and Dcm methylation individually to resolve which one is detrimental or potentially helpful in transformation. We compared electroporation transformation

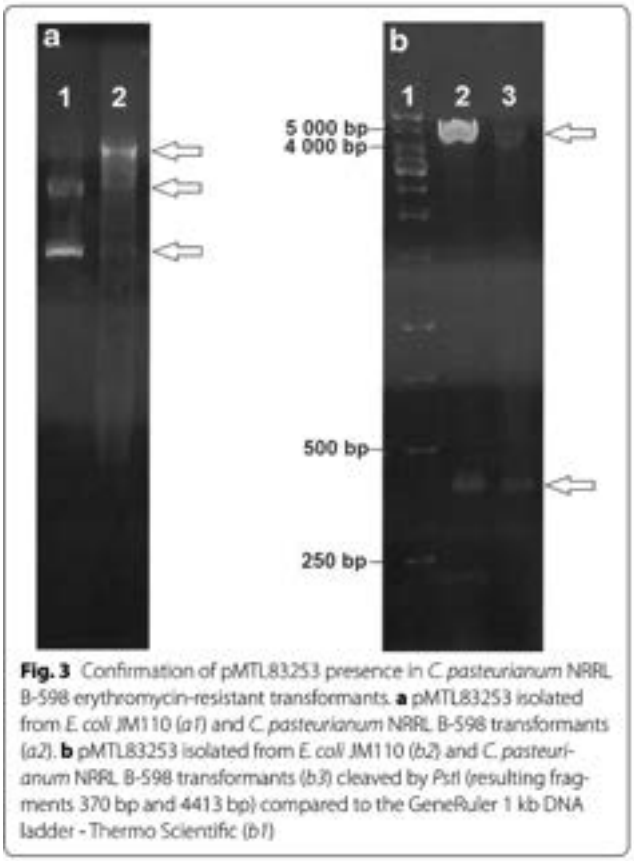


Fig. 3 Confirmation of pMTL83253 presence in *C. pasteurianum* NRRL B-598 erythromycin-resistant transformants. **a** pMTL83253 isolated from *E. coli* JM110 (*a1*) and *C. pasteurianum* NRRL B-598 transformants (*a2*). **b** pMTL83253 isolated from *E. coli* JM110 (*b2*) and *C. pasteurianum* NRRL B-598 transformants (*b3*) cleaved by *Pst*I (resulting fragments 370 bp and 4413 bp) compared to the GeneRuler 1 kb DNA ladder - Thermo Scientific (*b1*)

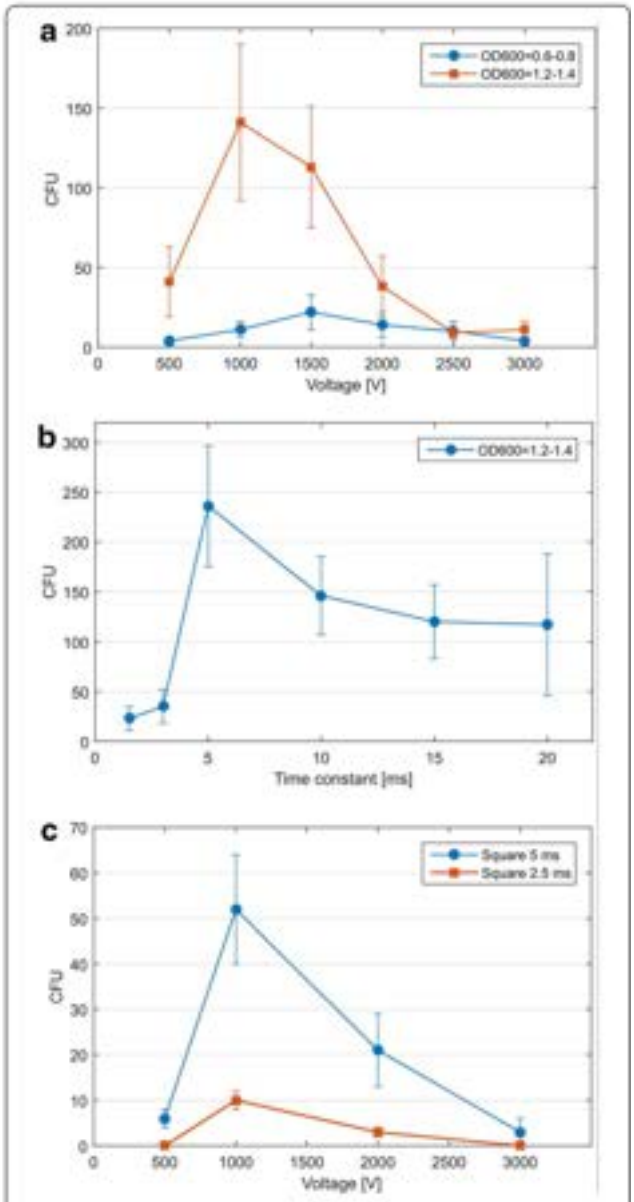


Fig. 4 Optimization of electrotransformation conditions. Influence of various voltages and cell growth state presented by OD_{600} (used conditions: 0.2 cm gap electroporation cuvette, time constant 11 ms) on transformation efficiency (**a**); influence of different time constants (conditions used: 0.2 cm gap electroporation cuvette, voltage 1 000 V) on transformation efficiency (**b**); influence of square-wave pulse delivery (**c**)

efficiencies of experiments where plasmid DNA isolated from the following methylation-deficient *E. coli* strains were used: JM110 (*dam*−/*dcm*−), BL21 (*dam*+/*dcm*−) and GM33 (*dam*−/*dcm*+). DNA extracted from *E. coli* DH5α (*dam*+/*dcm*+) was also used for confirmation that Dam and Dcm methylations represent a real obstacle to transformation, even when the optimized electrotransformation protocol was performed. A few erythromycin-resistant colonies (a maximum 8 of CFU) containing pMTL83253 were sometimes obtained if DNA from DH5α (fully methylated) was transformed. Relatively consistent results were achieved by transformation of hemimethylated plasmid DNA. Both methylations led to a significant reduction in transformation efficiency. The influence of various methylations on electrotransformation efficiencies is summarized in Table 3.

Establishment of conjugational transfer

Conjugation was not observed when an *E. coli* strain supporting Dam or Dcm methylation was used as a donor for transmission of pMTL80000 series plasmids to our strain used in the pilot experiment (see above). Based on our experience from electrotransformation experiments, we constructed a new conjugation donor strain by transmission of RP4 helper plasmid to *E. coli* JM110 (*dam*−/*dcm*−) containing pMTL83253. With this donor ensuring transfer of unmethylated pMTL83253 between donor and recipient cells, we tested for conjugation. Conjugation using a methylation-deficient donor was successful and many erythromycin-resistant colonies were

Table 3 Influence of DNA methylation stage to the electrotransformation efficiency

DNA amount [μg]/*E. coli* strain (designation)	CFU (average count)[a]	Efficiency (CFU per μg DNA)
2 μg/DH5α (*dam* +/*dcm* +)	3	1.5
2 μg/GM33 (*dam* −/*dcm* +)	27	13.5
2 μg/BL21 (*dam* +/*dcm* −)	28	14
2 μg/JM110 (*dam* −/*dcm* −)	236	118

[a] 2 μg of DNA was used for transformation

observed after 48 h. CFUs achieved after various conjugation times (6 or 24 h) are summarized in Table 4.

Use of sonoporation for transmission of plasmid DNA

As described previously, ultrasound could also be a useful technique to use for transformation of Gram-positive bacteria. From a few sonoporation media tested (TYA broth, 0.5 M $CaCl_2$, sterile water, SMP and PEG), only 10 and 30 % PEG 8000 were suitable for relatively high-efficiency transformation. No or only a few transformants were achieved when other sonoporation media were used. An adequate time of ultrasonic pulse was designed according to previous experiences with sonoporation of Gram-positive bacteria, where 20 s was identified as a critical time for ultrasound-mediated plasmid DNA degradation but less time led to a reduction in transformation efficiency [16]. Sonoporation has been proven to be a very effective method of transformation that provides even higher transformation efficiencies than electrotransformation. Efficiencies of transformation achieved by sonoporation are summarized in Table 4.

Combined sono/electroporation for increased transformation efficiency

Because cell-wall weakening approaches were not successful, we compiled a combined method using both sono- and electroporation for improving transformation efficiencies. During the first set of sono/electroporation experiments, we observed that a square-wave pulse provided more consistent results and significantly higher efficiency than the previously used exponential pulse. Also, different amounts of DNA (0.25–2 μg) were used for establishing the most efficient approach. Slightly higher voltage (1250 V) produced the most transformants in the square-wave mode and best transformation efficiency was achieved with 0.5 μg of plasmid DNA (see Fig. 5). By a combination of both techniques, we were able to reach a transformation efficiency of 5.3×10^2 cfu/μg DNA (see Table 4).

Discussion

The development of methods for efficient genetic manipulation of clostridial bacteria is generally very challenging. Protocols for transmission of foreign DNA to many clostridial species have been developed [20], but these transformation procedures use very different conditions and their overall efficiencies vary by orders of magnitude from 10^0 to 10^6 transformants/μg of DNA. Furthermore, transformation conditions are often useful for only one strain and cannot readily be used for other species or even strains. At least, rational step-by-step optimization of the protocol is necessary in order to achieve consistent results. A unique approach to transformation must be developed when the strain expresses a specific restriction

Table 4 Summary of pMTL83353 containing CFU yielded by conjugation, sonoporation, and combined sono/electroporation approaches

Method	CFU (average count)	Efficiency (CFU per μg DNA)
Conjugation (*E. coli* JM110 containing RP4 and pMTL83253 donor)		
6 h of conjugation	12	
24 h of conjugation	37	
Sonoporation		
10 % PEG 8000 buffer, 20 s pulse	225[a]	112.5
30 % PEG 8000 buffer, 20 s pulse	321[a]	160.5
Sono/electroporation		
30 % PEG 8000 buffer, 20 s ultrasound pulse, 5 ms square-wave pulse 5 ms (1250 V)	265[b]	530

[a] 2 μg of DNA was used for transformation

[b] 0.5 μg of DNA was used for transformation

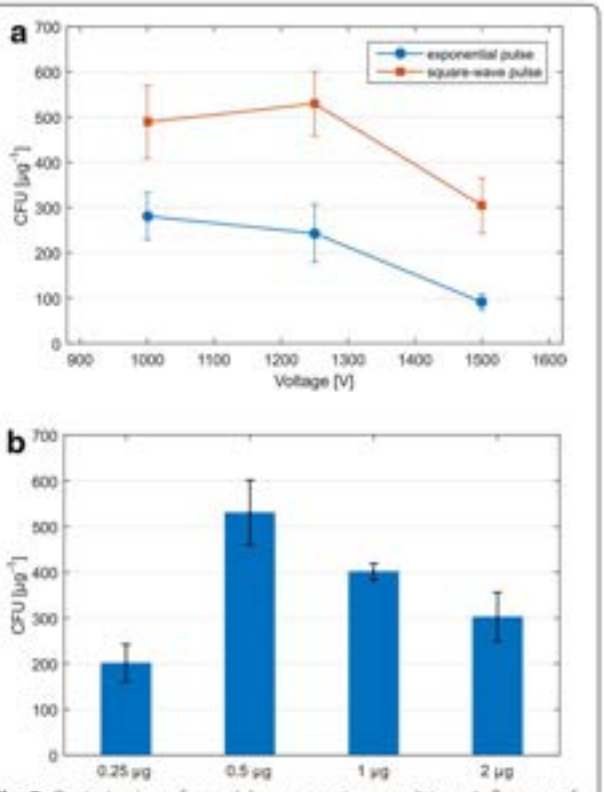


Fig. 5 Optimization of sono/electroporation conditions. Influence of various voltages and exponential and square-wave pulse deliveries (**a**); influence of DNA amount on transformation efficiency (**b**)

barrier that prevents effective transformation, or when conditions from previously published approaches are unsuccessful, as in our case.

C. pasteurianum NRRL B-598 represents a non-type strain of solventogenic clostridium that could be a good candidate for production of organic solvents in an ABE process. This strain excels in very high oxygen resistance and overall robustness that could be helpful for a large-scale ABE process. Moreover, biosynthesis of some nonspecific proteases that allows the use of cheap nitrogen sources in its cultivation (e.g., waste whey products) has been described previously for this strain [28]. During our experiments, we showed that *C. pasteurianum* NRRL B-598 carries a *cat* gene encoding resistance to chloramphenicol and thiamphenicol, the antibiotics that is normally effective against many strains of clostridium bacteria. This finding is a little surprising because chloramphenicol and thiamphenicol resistances have only been observed in solventogenic species such as *C. beijerinckii*, but not *C. pasteurianum*.

The action of various restriction-modification (R-M) systems represents a frequent obstacle in the transformation of clostridia, as well as other Gram-positive species. Type II R-M systems recognize a defined short sequence in the foreign DNA and promote its degradation after transmission to the cytoplasm, or even immediately on the cell surface [29]. R-M II systems were described as a reason preventing transformation of *C. acetobutylicum* ATCC 824 [22], *C. pasteurianum* ATCC 6013 [12] or *C. cellulolyticum* ATCC 35319 [8]. In these cases, special treatment by DNA-methyltransferase, which masks all recognition sequences, was necessary before transformation. Type I R-M systems could also be responsible for a decrease in transformation efficiency like in *C. saccharobutylicum* NCP 262 [23]. Specific protein inhibitors (such as TypeOne restriction inhibitor), protective methylation or heat inactivation could be approaches for overcoming these systems [29]. Equally, reduction of transformation efficiency could be caused by R-M III or IV, but these systems have, so far, been very poorly described in clostridia.

Based on the analysis of PacBio SMRT data, we demonstrated the genomic existence of two type I R-M systems, Cpa598I and Cpa598II. Activity of these systems was also confirmed experimentally by cultivation of pMTL82254 which contained recognition sequences of both R-M systems. Restriction provides probably unspecific cleaving of DNA in the direction from the recognized motifs which is typical for type I R-M systems [30]. Both recognized motifs are included in sequence of pBP1 replication origin module of pMTL80000 plasmids system thus it is better to use other replicon for transformation of this strain. On the other hand, when unmethylated pBP1 replicon-based plasmid (pMTL82251) was transformed by electroporation, we were still able to obtain a few transformants.

Both type II R-M systems are most certainly inactive because no methylated recognition sequence for Cpa598ORF2410 system was found and no m5C methylations assigned to Cpa598ORF20205 system were detected. We note that the kinetic signatures of m5C bases may not have been strong enough to study properly, but in a relatively high sequence coverage (79×) not a single m5C methylation was detected and also no active type II R-M system was obtained during experimental testing of their presence in the protoplast or whole cell lysates. Activity of the remaining type IV R-M system remains unclear, since these systems are poorly described and neither recognition sequence nor the type of methylation was assigned to this system. Nevertheless, because Cpa598ORF12465P is a methyl-directed restriction enzyme, its activity could also be the reason for decreased transformation efficiency. Further studies are required to verify these hypotheses.

The *C. pasteurianum* NRRL B-598 genome contains a relatively large number of antibiotic efflux genes. Antibiotic resistance can be confirmed by almost normal

growth of cells in a medium containing various antibiotics over long periods of time.

The addition of TypeOne restriction inhibitor, which has been described previously as a functional agent for overcoming R-M I systems in *E. coli* or *Salmonella typhimurium* [31], also did not lead to successful transformation. Based on these results, we assumed that a restriction barrier requiring methylation protection of plasmid DNA probably did not constitute a relevant obstacle during transformation of DNA extracted from *E. coli* or its conjugal transfer to *C. pasteurianum* NRRL B-598.

Methylation of transmitted DNA can also clearly affect the efficiency of bacterial transformation. Significant reductions in transformation efficiencies when methylated DNA was used were described for many bacterial species such as *Streptomyces* or *Lactobacillus*. Methyl-specific restriction systems probably play a major role in these observations [32, 33], but the fact that methylated *ori* sequences on a plasmid may not associate with a specific replication protein could also play an important role in transformation efficiency [34]. Fully methylated DNA isolated from *Escherichia coli* (*dam*+/*dcm*+) was, in most cases, referred to as the best template for clostridial transformation because Dam and Dcm methylation could protect DNA from degradation by nucleases and could increase clostridial transformation efficiencies. Reported cases of detrimental influences of *E. coli* methylation were observed in *C. thermocellum* DSM1313 and *C. ljungdahlii* DSM 13528, but eventually only Dcm methylation was identified as the origin of transformation problems in both experiments [13, 34]. Surprisingly, when unmethylated plasmid DNA was used for electrotransformation of *C. pasteurianum* NRRL B-598, we suddenly obtained a few transformants. For electrotransformation, a previously published protocol for *C. beijerinckii* NCIMB 8052 [25] was used and the maximum transformation efficiency, achieved with pMTL82353, was 6 cfu/μg DNA. The transformation efficiency achieved was very low compared to other clostridia or Gram-positive bacteria and could not be used for effective genetic manipulations or research on this strain. Because a previously published protocol for other species was used without changes, we wanted to optimize it directly for *C. pasteurianum* NRRL B-598, hopefully leading to an improved transformation efficiency.

The efficiency of electrotransformation may be affected by many parameters such as growth medium, cell growth phase, composition of electroporation buffer, voltage of electric pulse, or its length (influenced mainly by capacitance and resistance of the electroporator). For electrotransformation of clostridial species, cells in early-log to late-log growth phase, different electroporation buffers with low conductivity containing osmostabilizing agents (sucrose, PEG, etc.), and a relatively low electric field (around 5 kV cm^{-1}) are usually used [20]. We found that the best growth phase of *C. pasteurianum* NRRL B-598 for electrotransformation was between late logarithmic and early-stationary phase (OD_{600} 1.2–1.4), which is not typical for most solventogenic strains. Similarly, the best transformation efficiency was obtained when electroporation was conducted in 10 % PEG 8000 and decreased when the SMP electroporation buffer (at various pH values) was used. Through step-by-step optimization, we were able to achieve an average electrotransformation efficiency of 1.2×10^2 cfu/μg DNA when unmethylated DNA was used. This was much lower than for the type strains *C. acetobutylicum* or *C. beijerinckii*, where the electrotransformation efficiencies reached 10^4–10^5 transformants per μg of DNA [22, 25]. Nevertheless, this efficiency is sufficient to use this method for some genetic improvements and basic research on this intractable strain.

Achieved transformation efficiency showed clearly that with a decreasing number of any *E. coli* methylations, transformation efficiency significantly increased. Thus, both Dam and Dcm methylations were shown to be detrimental to transformation, a fact that has not been described previously in transformation of other clostridia. Previously, Pyne et al. [20] described similar effect of CpG methylation which presence led to obtain no transformants even though CpG provided good protection against digestion by described R-M system. If we take into account the number of Dam- and Dcm-specific methylation sites on pMTL83253 (10 and 18 resp.), we can postulate that Dam methylation could be a little more detrimental than Dcm, which is at variance with findings obtained previously [13, 35]. Decreased efficiency could be caused by a reduction in replication efficiency or some methyl-specific restriction system that may be present in cells as a protection against foreign DNA, e.g., bacteriophage, exhibiting a foreign methylation pattern. The best described similar systems are, for example, the *Dpn*I system in *Streptococcus pneumoniae* [36] or model methylation-dependent systems *Mcr*A, *Mcr*BC, and *Mrr* as described in *E. coli* [30]. If some methyl-specific type IV restriction system occurs in our strain (see above), it would be quite interesting because no restrictions were obtained when we conducted an examination of restriction systems with Dam and Dcm methylated pMTL83253. However, we focused mainly on R-M I and II systems, so some putative R-M IV (methyl-specific) systems may not be active under these in vitro conditions.

The influence of *E. coli* methylation was also verified in conjugation experiments, where pMTL82353 transmission was only successful in the methylation-deficient

donor strain (JM110 containing RP4). The existence of effective conjugal transfer could be very useful because it represents an effective way to transfer large plasmids to *C. pasteurianum* NRRL B-598, which is poorly transformable by electroporation and sonoporation techniques. No evidence concerning the use of a conjugation donor mediating transfer of unmethylated DNA between *E. coli* and clostridia has been published previously and this method could represent a fast and relatively easy method for an initial examination of the influence of methylation on transmission efficiency because this IncP-based conjugation method is applicable for many clostridial species in a similar arrangements of experiments.

Sonoporation is a relatively new method that is not used frequently for bacterial transformation. It is probably based on the cavitation of the cell wall and membrane, mediated by ultrasound pulse delivery that results in transmission of DNA into the cell [37]. Historically, a few transformations of thermophilic clostridia were conducted successfully using ultrasound-mediated transfer [16]. We were able to transform *C. pasteurianum* NRRL B-598 by sonoporation using a simple 20-s ultrasound pulse. Surprisingly, the average efficiency of pMTL82353 transfer was 1.6×10^2 cfu/μg DNA, which was even more efficient than electrotransformation. Moreover, sonoporation is a method that does not require any special or expensive equipment and is fast and reliable. On the other hand, it is likely that ultrasound-mediated transformation is limited by the size of the transferred plasmid because larger plasmids can be more rapidly destroyed by sonication. Polyethylene glycol probably plays an important role in transformation of *C. pasteurianum* NRRL B-598 because it may act as an osmostabilizer and also as an agent ensuring easier transmission through the bacterial membrane. Sonoporation of unmethylated DNA was the necessary condition and when DNA extracted from DH5α was used, no or only a few transformants were obtained.

Ultrasound pre-treatment prior to electrotransformation was used previously e.g., for *Saccharopolyspora erythraea* [38] or *Streptomyces spp.* [39]. Ultrasound can effectively disorganize the cell wall; therefore, it may be useful to increase the efficiency of transformation. Because we were not successful using standard cell-wall weakening procedures (glycine addition or lysozyme treatment), we attempted to enhance the uptake of DNA into bacterial cells by sonication prior to electroporation, especially in this case where sonication was proven to be the best approach for transformation. Sono/electroporation proved to be the best method for transformation of *C. pasteurianum* NRRL B-598, producing relatively consistent results over many replicates. It was shown to be important to use a square-wave pulse during sono/electroporation because when a standard exponential pulse was delivered, transformation efficiency decreased. This was mainly result of higher cell mortality probably due to ultrasound-caused cell wall disturbances. By the combination of both methods, we were able to achieve transformation efficiency of 5.3×10^2 cfu/μg DNA, which was about four times higher than using sonoporation or electroporation alone.

The transformation efficiency that was achieved is sufficient for effective plasmid DNA delivery to *C. pasteurianum* NRRL B-598 and could be used, for example, for simple gene over-expression or knock-out experiments. Due to restrictions with equipment, all transformation steps (electroporation, sonoporation, and partial culture manipulation) were performed outside of the anaerobic chamber. We assume that strict anaerobic conditions may improve the efficiency of DNA transmission however even under the described conditions, we were able to achieve usable and repeatable results for this oxygen resistant strain. It is also possible that less well-described *E. coli* methylases (e.g., genomic orphan methylases) could be responsible for the relatively low efficiency of DNA transmission and could be the subject of further research.

Conclusions

We have described methods for transmission of foreign DNA to *C. pasteurianum* NRRL B-598 for future potential genetic manipulation. Using PacBio kinetic data, we described 2 previously unknown recognition motifs for type I R-M systems in the *C. pasteurianum* NRRL B-598 genome as well as demonstrated the inactivity of 2 type II R-M systems. We also discovered a putative type IV methyl-directed R-M system that could be responsible for low transformation efficiency. Transformation or conjugal transfer of non-methylated DNA was necessary for high-efficiency transmission by all methods tested, which is unusual for clostridial transformation methods described to date. Methods for conjugation, electrotransformation, not frequently used sonoporation, and even their combination (sono/electroporation) were described and a maximum transformation efficiency of 5.3×10^2 cfu/μg DNA was achieved. In this paper, we also demonstrated that development of genetic methods for a non-type strain could be challenging and be completely different to the type strain or even other clostridia. All described methods could lead to more effective research that would make this strain useful in biofuel production. This work also reveals new knowledge about the diversity of defense mechanisms against foreign DNA in solventogenic clostridia and shows the possibility of using sono/electroporation for efficient transformation of Gram-positive bacteria.

Methods

Bacterial strains and growth conditions

All strains described in this paper are summarized in Table 5. *C. pasteurianum* NRRL B-598 was maintained as a spore suspension in sterile distilled water and grown in TYA medium [40] containing in g/l: 20 glucose; 2 yeast extract (Merck); 6 tryptone (Sigma); 0.5 KH_2PO_4; 3 ammonium acetate; 0.3 $MgSO_4.7H_2O$; 0.01 $FeSO_4$. TYA plates (solidified by 1.5 % agar) were supplemented with erythromycin (20 µg/ml), spectinomycin (700 µg/ml), chloramphenicol (25 µg/ml), or thiamphenicol (15 µg/ml) as required. *C. pasteurianum* DSM 525 was cryopreserved in 30 % glycerol solution (maintained in −80 °C) and grown in RCM broth (Merck) supplemented by glucose to a final concentration of 20 g/l. Cultivation of both strains was performed in an anaerobic chamber (Concept 400; Ruskinn Technology, UK) in a stable atmosphere of 95 % N_2/5 % H_2 and at 37 °C. Clostridium basal medium (CBM) [41], semi-defined P2 medium [42], and YTG [43] media were also used during this study.

All *E. coli* strains were cryopreserved in 20 % glycerol solution (maintained in −80 °C) and grown on LB medium (containing in g/l: 10 tryptone; 5 yeast extract; 5 NaCl) in 37 °C. LB broth or plates (1.5 % agar) were supplemented with erythromycin (500 µg/ml), spectinomycin (100 µg/ml), ampicillin (100 µg/ml), or streptomycin (30 µg/ml) as necessary.

Plasmids, oligonucleotides, and DNA manipulation

All plasmids used in this paper are summarized in Table 5. Plasmid DNA was transmitted to *E. coli* strains by standard $CaCl_2$ treatment; transmission of RP4 helper plasmid between *E. coli* strains was performed by conjugation. For isolation of plasmid DNA, a High Pure Plasmid Isolation Kit miniprep (Roche, Switzerland) was used. Plasmid DNA from *C. pasteurianum* NRRL B-598 was extracted by the method described previously for *C. pasteurianum* ATCC 6013 [12] with modifications. For isolation, 8 ml of culture (OD_{600} ca. 1.3–1.5) was harvested by centrifugation (10,000×*g*, 2 min.), washed once in 1.5 ml KET buffer (0.5 M KCl; 0.1 M EDTA; and 0.05 M Tris–HCl; pH 8.0) and SET buffer (25 % sucrose, 0.05 M EDTA, and 0.05 M Tris–HCl, pH 8.0) and resuspended in 250 µl of SET buffer containing 5 mg/ml of lysozyme. The mixture was incubated for 10 min at 37 °C. Lysis and purification were completed using the High Pure Plasmid Isolation Kit miniprep (Roche, Switzerland) where the first step was addition of 250 µl of lysis buffer. The original protocol was followed after this step.

Detection of restriction systems

For identification of putative restriction systems in *C. pasteurianum* NRRL B-598, a protoplast crude extract and whole cell lysate were tested for restriction activity. The whole cell lysate was prepared by sonication (30 min) of the bacterial cells, which were harvested from 30 ml of culture (OD_{600} 0.6–0.8) and resuspended in 5 ml of nuclease-free distilled water. For protoplast preparation, 50 ml of culture (OD_{600} 0.6–0.8) was centrifuged (10,000×*g*, 2 min.), washed with lactose-containing protoplast buffer (25 mM potassium phosphate, 6 mM $MgSO_4$, 15 % lactose, pH 7.0) [12, 44] and resuspended in 2–4 ml of protoplast buffer containing 10 mg/ml of lysozyme. The mixture was incubated at 37 °C in the anaerobic chamber for 45–60 min (at least 90 % of

Table 5 Summary of bacterial strains and plasmid DNA used in this thesis

Bacterial strains	Genotype	Source
Clostridium pasteurianum NRRL B-598		ARL collection (NRRL)
Clostridium pasteurianum DSM 525 (ATCC 6013)	Coding type II restriction system CpaAI	DSMZ
Escherichia coli DH5α (DSM 6897)	*dam* +/*dcm*+	DSMZ
Escherichia coli JM110 (DSM 11539)	*dam*−/*dcm*−	DSMZ
Escherichia coli BL21(DE3)	*dam*+/*dcm*−	CGSC
Escherichia coli GM33	*dam*−/*dcm*+	CGSC
Escherichia coli HB101 (DSM 1607)	*dam*+/*dcm*+	DSMZ
E. coli plasmids		
RP4 (RK2)	Coding IncP-based conjugation function	DSMZ
E. coli/*Clostridium* shuttle plasmids		
pMTL83353	*aad9*, pCB102 origin of replication	[21]
pMTL82251	*ermB*, pBP1 origin of replication	[21]
pMTL83253	*ermB*, pCB102 origin of replication	[21]
pMTL84251	*ermB*, pCD6 origin of replication	[21]
pMTL85251	*ermB*, pIM13 origin of replication	[21]

cells were transformed to protoplasts). Protoplasts were collected by centrifugation (1200×*g*, 10 min) and lysed in 20 ml TEMK buffer [22] at 37 °C for 1 h after which, cell debris were removed by additional centrifugation (20,000×g, 20 min., 4 °C). The *C. pasteurianum* DSM 525 protoplast crude extract was prepared in the same way as above (15–20 min. cultivation with lysozyme-containing buffer was enough in this case) and used as a positive control in the restriction-system detection assay. Protoplasts and whole cell crude extracts were used immediately for reactions with plasmid DNA.

The reaction mixture composition was the following: 5 µl of protoplast crude extract or whole cell lysate; 0.5 µg of plasmid DNA (pMTL83253 and pMTL82254); reaction buffer added to a final 1× concentration; deionized water was added to a final volume of 20 µl. Reactions were performed at 30 and 37 °C for at least 8 h (4 h in the case of the positive control). After incubation, reactions were analyzed by standard 1 % agarose-gel electrophoresis. Reaction buffers that were tested were the following: commercial R, O, G, B, and Tango buffers for restriction enzymes (Thermo Scientific, USA), a commercial CutSmart buffer for restriction enzymes (NEB, UK) and CpaAI reaction buffer [45].

Bioinformatics

Bioinformatics analysis was focused on revealing genes for antibiotic resistance, putative restriction barriers and methylation enzymes and motifs in the *C. pasteurianum* NRRL B-598 whole genome sequence.

The methylome was characterized using PacBio Single Molecule Real-Time sequencing (2× SMRT cell) kinetic data collected during the genome sequencing process [46]. SMRT Analysis v.2.3 using "RS_Modification_and_Motif_Analysis.1" protocol was used for genome-wide base modification and detection of the affected motifs. The default quality value (QV) score of 30 was used for motif determination. The detected motifs were uploaded and further analyzed using the REBASE database [24]. The complete genome was also scanned for homologs of R-M system genes using BLAST searching against REBASE and GenBank databases.

Identification of antibiotic resistance genes was carried out with RGI (Resistance Gene Identifier) version 2 [47]. The predicted ORFs were manually compared to genes in the *C. pasteurianum* NRRL B-598 complete genome [19] predicted by NCBI Prokaryotic Genome Annotation Pipeline (PGAP) (http://www.ncbi.nlm.nih.gov/genome/annotation_prok/) and GenBank accession numbers of protein products of relevant genes were assigned.

Statistics analyses describing transformation efficiency were calculated and visualized using Matlab 2014b.

Preparation of electrocompetent cells and electroporation conditions

For all electroporation experiments, a GenePulser Xcell™ electroporator including both CE and PC module (Bio-Rad, USA) was used. For preparation of electrocompetent cells, 100 ml of TYA medium was inoculated with different proportions of spores and grown overnight. Following a day's culture, the competent cells were prepared from cells in late-log to early-stationary growth phase (OD_{600} 1.2–1.4). Bacterial cells were centrifuged (10,000×*g*, 3 min, 4 °C), washed once with an equal volume of chilled electroporation buffer (10 % PEG 8000) and gently resuspended in 1/20 volume of the same buffer. Electrocompetent cells were maintained on ice and used for electroporation immediately.

Into a 0.2-cm gap electroporation cuvette (BioRad, USA), 480 µl of competent cells and 2 µg of plasmid DNA dissolved in 20 µl of demineralized water were mixed and transferred to the electroporator. During optimization of electroporation parameters a Time Constant mode was used. The most successful parameters were the following: 5 ms time constant, 1000 V (corresponding to 50 µF capacitance and 100 Ω resistance). Electroporated cells were incubated for 10 min in the anaerobic chamber on ice and 100 µl of shocked cells were then inoculated into 2 ml of prewarmed and prereduced TYA broth. After 6 h of culture, all cells were harvested by centrifugation, resuspended in 100–500 µl of TYA and plated onto TYA agars with appropriate antibiotic selection, or directly seeded onto plates in different volumes. Growth of antibiotic-resistant colonies was observed after 24–48 h.

All centrifugation and electroporation steps were conducted out of the anaerobic chamber because chamber construction did not allow them to be performed inside.

Gene transfer by conjugation

Escherichia coli HB101 and JM110 both containing helper plasmid RP4 were used as conjugation donors. The donor was transformed by pMTL83253 as described above and conjugation was conducted as described previously [21]. An overnight culture of donor (1 ml) was washed twice with LB broth and 200 µl of overnight recipient culture were added. The mixture was spotted in small drops onto TYA agar medium without antibiotic selection and incubated for 6–24 h. Cells were scraped and washed from the agar with 600 µl of PBS, twice, and plated onto TYA with appropriate antibiotic selection and chloramphenicol or thiamphenicol counter-selection for suppression of *E. coli* donor growth.

Gene transfer by sonoporation

Sonoporation was performed using a standard laboratory ultrasonic bath (Elmasonic E120H, Elma Schmidbauer

GmbH, Switzerland). Competent cells were prepared in the same way as electrocompetent cells (see above) but were finally resuspended in 1/20 volume of sonoporation buffer (30 % PEG 8 000). Into a flat-bottom glass vial, 480 µl of competent cells and 2 µg of plasmid DNA were mixed and immediately sonoporated in the middle of the ultrasound bath for 20 s. Recovery of the mixture was conducted in the same way as during electrotransformation. Growth of antibiotic-resistant colonies was observed after 24–48 h.

Combined technique for higher transformation efficiency

For the best transformation efficiency, a combination of sonoporation and electroporation was performed. Competent cells and transformation mix were prepared in the same way as during the standard sonoporation procedure; however only 0.25–2 µg of plasmid DNA was used for transformation. Immediately after sonoporation, cells were transferred to the 0.2-cm gap electroporation cuvette and electroporated using a square-wave pulse (5 ms, 1250 V). For recovery of the cells, the standard method was used (see above).

Statistical and control approaches

All transformation experiments were performed at least three times. Transfer efficiencies of foreign DNA were calculated as an average value derived from three independent experiments. Negative controls (transformation mixture without DNA added or conjugation with donor strain without appropriate a pMTL80000 series plasmid) were used in all transformation experiments.

Abbreviations

ABE: acetone-butanol-ethanol fermentation; BLAST: basic local alignment search tool; CFU: colony forming units; Dam: DNA adenine methyltransferase; Dcm: DNA cytosine methyltransferase; DNA: deoxyribonucleic acid; m4C: 4-methylcytosine; m5C: 5-methylcytosine; m6A: 6-methylcytosine; MT: methyltransferase; OD_{600}: optical density at 600 nm; ORF: open reading frame; PBS: phosphate buffered saline; PEG: polyethylene glycol; R-M system: restriction-modification system; SMRT: single molecule real-time sequencing.

Authors' contributions

JK carried out the experimental work on putative restriction-system detection, establishment of transformation and conjugation protocols, preparation of the draft manuscript and study design. KS carried out the bioinformatics analysis and drafted the manuscript. IP participated in the study design and helped to draft the manuscript. PP conceived the study and helped to draft the manuscript. All authors read and approved the final manuscript. We have read Biotechnology for Biofuels policy on data and material release, and the data within this manuscript meet those requirements.

Author details

[1] Department of Biotechnology, University of Chemistry and Technology Prague, Technicka 5, 166 28 Prague, Czech Republic. [2] Department of Biomedical Engineering, Brno University of Technology, Technicka 12, 616 00 Brno, Czech Republic.

Acknowledgements

This study was supported by the Ministry of Education, Youth and Sports of the Czech Republic, project No. MSM 6046137305, TACR BIORAF No. TE01020080 and by the grant project GACR P102/11/1068 NanoBioTECell. Set of the p80000 modular plasmids were kindly provided by prof. Nigel Minton, University of Nottingham. Special thanks belong to BioRad company (namely ing. Jana Nováková) for providing us with an option to use the GenePulser XCell™ electroporator.

Competing interests

The authors declare that they have no competing interests.

Received: 21 October 2015 Accepted: 8 January 2016
Published online: 20 January 2016

References

1. Panagiotopoulos IA, Bakker RR, de Vrije T, Claassen PA, Koukios EG. Integration of first and second generation biofuels: fermentative hydrogen production from wheat grain and straw. Bioresource Technol. 2013. doi:10.1016/j.biortech.2012.09.083.
2. Jones DT, Woods DR. Acetone-butanol fermentation revisited. Microbiol Rev. 1986;50:484–524.
3. Patakova P, Linhova M, Rychtera M, Paulova L, Melzoch K. Novel and neglected issues of acetone-butanol-ethanol (ABE) fermentation by clostridia: *Clostridium* metabolic diversity, tools for process mapping and continuous fermentation systems. Biotechnol Adv. 2013. doi:10.1016/j.biotechadv.2012.01.010.
4. Wang Y, Li XZ, Mao YJ, Blaschek HP. Genome-wide dynamic transcriptional profiling in *Clostridium beijerinckii* NCIMB 8052 using single-nucleotide resolution RNA-Seq. BMC Genomics. 2012. doi:10.1186/1471-2164-13-102.
5. Zheng YN, Li LZ, Xian M, Ma YJ, Yang JM, Xu X, et al. Problems with the microbial production of butanol. J Ind Microbiol Biot. 2009. doi:10.1007/s10295-009-0609-9.
6. Purdy D, O'Keeffe TAT, Elmore M, Herbert M, McLeod A, Bokori-Brown M, et al. Conjugative transfer of clostridial shuttle vectors from *Escherichia coli* to *Clostridium difficile* through circumvention of the restriction barrier. Mol Microbiol. 2002. doi:10.1046/j.1365-2958.2002.03134.x.
7. Heap JT, Kuehne SA, Ehsaan M, Cartman ST, Cooksley CM, Scott JC, et al. The ClosTron: mutagenesis in *Clostridium* refined and streamlined. J Microbiol Meth. 2010. doi:10.1016/j.mimet.2009.10.018.
8. Jennert KC, Tardif C, Young DI, Young M. Gene transfer to *Clostridium cellulolyticum* ATCC 35319. Microbiology. 2000. doi:10.1099/00221287-146-12-3071.
9. Reysset G, Hubert J, Podvin L, Sebald M. Transfection and transformation of *Clostridium acetobutylicum* strain N1-4081 protoplasts. Biotechnol Tech. 1988;2:199–204.
10. Lin YL, Blaschek HP. Transformation of heat-treated *Clostridium acetobutylicum* protoplasts with pUB110 plasmid DNA. Appl Environ Microb. 1984;48:737–42.
11. Allen SP, Blaschek HP. Electroporation-induced transformation of intact cells of *Clostridium perfringens*. Appl Environ Microb. 1988;54:2322–4.
12. Pyne ME, Moo-Young M, Chung DA, Chou CP. Development of an electrotransformation protocol for genetic manipulation of *Clostridium pasteurianum*. Biotechnol Biofuels. 2013. doi:10.1186/1754-6834-6-50.
13. Leang C, Ueki T, Nevin KP, Lovley DR. A genetic system for *Clostridium ljungdahlii*: a chassis for autotrophic production of biocommodities and a model homoacetogen. Appl Environ Microb. 2013. doi:10.1128/AEM.02891-12.
14. Oultram JD, Loughlin M, Swinfield TJ, Brehm JK, Thompson DE, Minton NP. Introduction of plasmids into whole cells of *Clostridium acetobutylicum* by electroporation. FEMS Microb Lett. 1988. doi:10.1111/j.1574-6968.1988.tb03154.x.
15. Yoshino S, Yoshino T, Hara S, Ogata S, Hayashida S. Construction of shuttle vector plasmid between *Clostridium acetobutylicum* and *Escherichia coli*. Agr Biol Chem Tokyo. 1990. doi:10.1271/bbb1961.54.437.
16. Lin L, Song H, Ji Y, He Z, Pu Y, Zhou J, et al. Ultrasound-mediated DNA transformation in thermophilic gram-positive anaerobes. PloS One. 2010. doi:10.1371/journal.pone.0012582.
17. Kolek J, Patakova P, Melzoch K, Sigler K, Rezanka T. Changes in membrane plasmalogens of *Clostridium pasteurianum* during butanol fermentation

as determined by lipidomic analysis. PloS One. 2015. doi:10.1371/journal.pone.0122058.
18. Kolek J, Sedlar K, Provaznik I, Patakova P. Draft genome sequence of *Clostridium pasteurianum* NRRL B-598, a potential butanol or hydrogen producer. Genome Announc. 2014. doi:10.1128/genomeA.00192-14.
19. Sedlar K, Kolek J, Skutkova H, Branska B, Provaznik I, Patakova P. Complete genome sequence of *Clostridium pasteurianum* NRRL B-598, a non-type strain producing butanol. J Biotechnol. 2015. doi:10.1016/j.jbiotec.2015.09.022.
20. Pyne ME, Bruder M, Moo-Young M, Chung DA, Chou CP. Technical guide for genetic advancement of underdeveloped and intractable *Clostridium*. Biotechnol Adv. 2014. doi:10.1016/j.biotechadv.2014.04.003.
21. Heap JT, Pennington OJ, Cartman ST, Minton NP. A modular system for *Clostridium* shuttle plasmids. J Microbiol Meth. 2009. doi:10.1016/j.mimet.2009.05.004.
22. Mermelstein LD, Welker NE, Bennett GN, Papoutsakis ET. Expression of cloned homologous fermentative genes in *Clostridium acetobutylicum* ATCC 824. Nat Biotechnol. 1992. doi:10.1038/nbt0292-190.
23. Lesiak JM, Liebl W, Ehrenreich A. development of an *in vivo* methylation system for the solventogen *Clostridium saccharobutylicum* NCP 262 and analysis of two endonuclease mutants. J Biotechnol. 2014. doi:10.1016/j.jbiotec.2014.07.005.
24. Roberts RJ, Belfort M, Bestor T, Bhagwat AS, Bickle TA, Bitinaite J, et al. A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. Nucleic Acids Res. 2003. doi:10.1007/978-3-642-18851-0_1.
25. Blaschek HP, Formanek J, Chen CK. *Clostridium* in biotechnology and food technology. In: Eynard N, Teissie J, editors. Electrotransformation of Bacteria. New York: Springer; 2000. p. 49–55.
26. McDonald IR, Riley PW, Sharp RJ, McCarthy AJ. Factors affecting the electroporation of *Bacillus subtilis*. J Appl Bacteriol. 1995. doi:10.1111/j.1365-2672.1995.tb00937.x.
27. Kashket ER, Cao ZY. Clostridial strain degeneration. FEMS Microbiol Rev. 1995. doi:10.1111/j.1574-6976.1995.tb00214.x.
28. Patakova P, Lipovsky J, Cizkova H, Fortova J, Rychtera M, Melzoch K. Exploitation of food feedstock and waste for production of biobutanol. Czech J Food Sci. 2009;27:276–83.
29. Vasu K, Nagaraja V. Diverse functions of restriction-modification systems in addition to cellular defense. Microbiol Mol Biol R. 2013. doi:10.1128/MMBR.00044-12.
30. Blumenthal RM, Xiaodong C. Restriction-modification systems. In: Streips UN, Yasbin RE, editors. Modern microbial genetics. 2nd ed. New York: Wiley; 2002. p. 178–225.
31. Hoffman LM, Haskins DJ, Jendrisak J. TypeOne™ Inhibitor improves transformation efficiencies by blocking type I restriction and modification systems *in vivo*. Epicentre Forum. 2002;9:8.
32. Gonzalez-Ceron G, Miranda-Olivares OJ, Servin-Gonzalez L. Characterization of the methyl-specific restriction system of *Streptomyces coelicolor* A3(2) and of the role played by laterally acquired nucleases. FEMS Microbiol Lett. 2009. doi:10.1111/j.1574-6968.2009.01790.x.
33. Spath K, Heinl S, Grabherr R. Direct cloning in *Lactobacillus plantarum* electroporation with non-methylated plasmid DNA enhances transformation efficiency and makes shuttle vectors obsolete. Microb Cell Fact. 2012;11:141. doi:10.1186/1475-2859-11-141.
34. Zhang G, Wang W, Deng A, Sun Z, Zhang Y, Liang Y, et al. A mimicking-of-DNA-methylation-patterns pipeline for overcoming the restriction barrier of bacteria. PLoS Genetics. 2012. doi:10.1371/journal.pgen.1002987.
35. Guss AM, Olson DG, Caiazza NC, Lynd LR. Dcm methylation is detrimental to plasmid transformation in *Clostridium thermocellum*. Biotechnol Biofuels. 2012. doi:10.1186/1754-6834-5-30.
36. de la Campa AG, Springhorn SS, Kale P, Lacks SA. Proteins encoded by the *Dpn*I restriction gene cassette. Hyperproduction and characterization of the *Dpn*I endonuclease. J Biol Chem. 1988;263:14696–702.
37. Song Y, Hahn T, Thompson IP, Mason TJ, Preston GM, Li G, et al. Ultrasound-mediated DNA transfer for bacteria. Nucleic acids res. 2007. doi:10.1093/nar/gkm710.
38. Fitzgerald NB, English RS, Lampel JS, Vanden Boom TJ. Sonication-dependent electroporation of the erythromycin-producing bacterium *Saccharopolyspora erythraea*. Appl Environ Microb. 1998;64:1580–3.
39. Mazy-Servais C, Baczkowski D, Dusart J. Electroporation of intact cells of *Streptomyces parvulus* and *Streptomyces vinaceus*. FEMS Microb Lett. 1997. doi:10.1111/j.1574-6968.1997.tb12561.x.
40. Hongo M, Murata A, Kono K, Kato F. Lysogeny and bacteriocinogeny in strains of *Clostridium* species. Agr Biol Chem Tokyo. 1968. doi:10.1080/00021369.1968.10859021.
41. O'Brien RW, Morris JG. Oxygen and the growth and metabolism of *Clostridium acetobutylicum*. J Gen Microbiol. 1971. doi:10.1099/00221287-68-3-307.
42. Baer SH, Blaschek HP, Smith TL. Effect of butanol challenge and temperature on lipid Composition and membrane fluidity of butanol-tolerant *Clostridium acetobutylicum*. Appl Environ Microb. 1987;53:2854–61.
43. Li YH, Mandelco L, Wiegel J. Isolation and characterization of a moderately thermophilic anaerobic alkaliphile, *Clostridium paradoxum* Sp Nov. Int J Syst Bacteriol. 1993. doi:10.1099/00207713-43-3-450.
44. Minton NP, Morris JG. Regeneration of protoplasts of *Clostridium pasteurianum* ATCC 6013. J Bacteriol. 1983;155:432–4.
45. Richards DF, Linnett PE, Oultram JD, Young M. Restriction endonucleases in *Clostridium pasteurianum* ATCC 6013 and *C. thermohydrosulfuricum* DSM 568. J Gen Microbiol. 1988. doi:10.1099/00221287-134-12-3151.
46. Flusberg BA, Webster DR, Lee JH, Travers KJ, Olivares EC, Clark TA, et al. Direct detection of DNA methylation during single-molecule, real-time sequencing. Nat Methods. 2010. doi:10.1038/nmeth.1459.
47. McArthur AG, Waglechner N, Nizam F, Yan A, Azad MA, Baylay AJ, et al. The comprehensive antibiotic resistance database. Antimicrob Agents Ch. 2013. doi:10.1128/AAC.00419-13.

5.3 Article III

SEDLAR, Karel, KOLEK, Jan, PROVAZNIK, Ivo and PATAKOVA, Petra. Reclassification of non-type strain Clostridium pasteurianum NRRL B-598 as Clostridium beijerinckii NRRL B-598. *Journal of Biotechnology*. 2017. 244, p. 1–3. (2017 IF = 2.533, Q2 in BIOTECHNOLOGY & APPLIED MICROBIOLOGY)

DOI: https://doi.org/10.1016/j.jbiotec.2017.01.003

Journal of Biotechnology 244 (2017) 1–3

Contents lists available at ScienceDirect

Journal of Biotechnology

journal homepage: www.elsevier.com/locate/jbiotec

Short communication

Reclassification of non-type strain *Clostridium pasteurianum* NRRL B-598 as *Clostridium beijerinckii* NRRL B-598

Karel Sedlar[a,*], Jan Kolek[b], Ivo Provaznik[a], Petra Patakova[b]

[a] Department of Biomedical Engineering, Brno University of Technology, Technicka 12, 616 00 Brno, Czechia
[b] Department of Biotechnology, University of Chemistry and Technology Prague, Technicka 5, 166 28 Prague, Czechia

ARTICLE INFO

Article history:
Received 7 October 2016
Received in revised form 9 January 2017
Accepted 15 January 2017
Available online 19 January 2017

Keywords:
Clostridium pasteurianum
Clostridium beijerinckii
Reclassification
Phylogenomic analysis

ABSTRACT

The complete genome sequence of non-type strain *Clostridium pasteurianum* NRRL B-598 was introduced last year; it is an oxygen tolerant, spore-forming, mesophilic heterofermentative bacterium with high hydrogen production and acetone-butanol fermentation ability. The basic genome statistics have shown its similarity to *C. beijerinckii* rather than the *C. pasteurianum* species. Here, we present a comparative analysis of the strain with several other complete clostridial genome sequences. Besides a 16S rRNA gene sequence comparison, digital DNA–DNA hybridization (dDDH) and phylogenomic analysis confirmed an inaccuracy of the taxonomic status of strain *Clostridium pasteurianum* NRRL B-598. Therefore, we suggest its reclassification to be *Clostridium beijerinckii* NRRL B-598. This is a specific strain and is not identical to other *C. beijerinckii* strains. This misclassification explains its unexpected behavior, different from other *C. pasteurianum* strains; it also permits better understanding of the bacterium for a future genetic manipulation that might increase its biofuel production potential.

Clostridia forms a large and diverse group of typically rod-shape, spore-forming anaerobes. The present study, as well as previous molecular studies, suggests that the genus *Clostridium* represents a polyphyletic group with uncertain phylogenetic affinities. Therefore, misclassification of its representatives, including type strains, is not unusual (Rainey and Lawson, 2016; Moon et al., 2008). Although taxonomic dissection for *Clostridia* based on the 16S rRNA gene sequence was proposed, it has also shown that the genus is not phylogenetically coherent and taxonomic placing of newly sequenced genomes can be questionable (Collins et al., 1994; Stackebrandt et al., 1999). For distantly related prokaryotes, this technique can predict genome-wide levels of similarity very well; but for closely related ones, other marker genes are needed (Lan et al., 2016). In particular, for strains with ambiguous properties, genome-wide studies like DNA–DNA hybridization are needed for the definitive assignment (Janda and Abbott, 2007).

The strain *C. pasteurianum* NRRL B-598 is a spore-forming, mesophilic heterofermentative bacterium with acetone-butanol fermentation ability (Lipovsky et al., 2016). It is available from the Agricultural Research Service Culture Collection (NRRL). The bacterium is unable to utilize glycerol, but it ferments a wide range of substrates containing glucose, xylose, arabinose, mannose, saccharose, lactose, cellobiose or starch. Furthermore, it can liquefy gelatin, decompose casein, produce polysaccharide capsules when grown on starch, and form stock polysaccharide granulose prior to the sporulation. The inability to grow on glycerol supports the misclassification of the strain. This is because this ability is generally considered to be one of the main distinguishing phenotypic traits for *C. pasteurianum* strains (Jensen et al., 2012). The strain is also naturally chloramphenicol/thiamphenicol resistant which has never been observed for the *C. pasteurianum* species, but it sometimes occurs at *C. beijerinckii* (Kolek et al., 2016a). First analyses of the genome showed a similarity of the genes involved in solvetogenesis to *C. beijerinckii* (Sedlar et al., 2014). However, a previous draft genome assembly did not contain any 16S rRNA gene sequence (Kolek et al., 2014); its correct taxonomic identification was therefore impossible. Last year, we presented its complete genome (Sedlar et al., 2015). This is available at DDBJ/EMBL/GenBank under accession No. CP011966. The complete genome sequence allowed us to perform a 16S rRNA gene sequence analysis, in addition to genome-wide analyses that proved misclassification of the strain. Therefore, we suggested its reclassification to be *C. beijerinckii* NRRL B-598.

Analysis of the 16S rRNA gene sequences showed that the strain *C. pasteurianum* (*beijerinckii*) NRRL B-598 shares only 92% similarity with the type strain *C. pasteurianum* ATCC 6013 (Rotta et al.,

* Corresponding author.
E-mail addresses: sedlar@feec.vutbr.cz (K. Sedlar), jan.kolek@vscht.cz (J. Kolek), provaznik@feec.vutbr.cz (I. Provaznik), petra.patakova@vscht.cz (P. Patakova).

http://dx.doi.org/10.1016/j.jbiotec.2017.01.003
0168-1656/

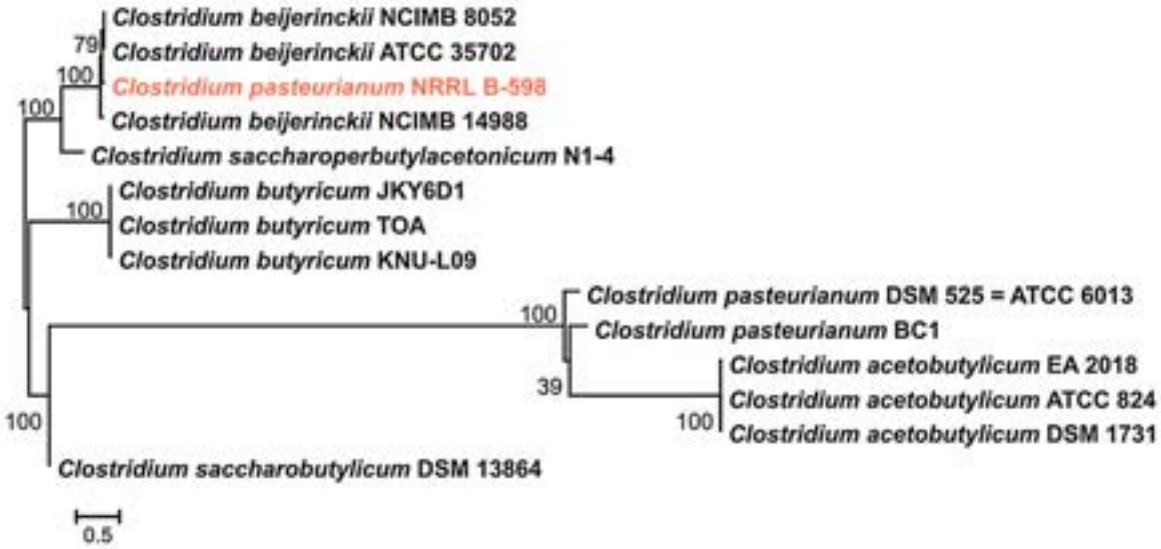


Fig. 1. Phylogenetic position of the strain *C. pasteurianum* (*beijerinckii*) NRRL B-598 based on genome-wide sequence tree. Branches determining the position of the strain are supported with high bootstrap values. The tree was drawn in scale and the scale bar represents the estimated number of amino acid changes per site for a unit of branch length utilizing CAT model.

Table 1
dDDH values among *C. pasteurianum* and *C. beijerinckii* strains. Values DDH >70% indicates that the strains belongs to the same species, values DDH >79% indicates the same subspecies.

	C. pasteurianum NRRL B-598	*C. pasteurianum* ATCC 6013 (=DSM525)	*C. pasteurianum* BC1	*C. beijerinckii* NCIMB 8052	*C. beijerinckii* ATCC 35702	*C. beijerinckii* NCIMB 14988
C. pasteurianum NRRL B-598	–	24.90	26.70	78.40	78.40	75.40
C. pasteurianum ATCC 6013 (=DSM525)		–	28.00	25.30	25.30	23.30
C. pasteurianum BC1			–	26.20	26.20	24.40
C. beijerinckii NCIMB 8052				–	100	74.50
C. beijerinckii ATCC 35702					–	74.50
C. beijerinckii NCIMB 14988						–

2015), while similarity to the well described strain *C. beijerinckii* NCIMB 8052 is as high as 99% (sequence of the type strain is not available). Beside the fact that it has a 99% similarity to other *C. beijerinckii* strains (e.g. ATCC 35702 or NCIMB 14988), and therefore nothing about strain identity can be revealed, this high similarity (98–99%) is also shared with other clostridial species. Examples include: *C. saccharoperbutylacetonicum* N1-4(HMT), *C. butyricum* subsp. convexa JCM 7840, *C. saccharobutylicum* DSM 13864, and others. Therefore, for reaching maximum accuracy, we decided to infer taxonomy of most related strains on a genome-wide scale using complete genomes rather than using only 16S rRNA gene sequences. The analysis was built by a PhyloPhlAn 0.99 (Segata et al., 2013) comparing >400 selected protein sequences conserved across a bacterial domain. The genes were identified using an internal PhyloPhlAn database by translated mapping with USEARCH 8.1 (Edgar, 2010). The final tree was reconstructed using FastTree 2.1 (Price et al., 2010) from protein subsequences of the genes concatenating their most informative amino-acid positions, each aligned using Muscle 3.8 (Edgar, 2004). The topology was computed using neighbor-joining algorithm with utilization of Jukes-Cantor evolution model. Moreover, CAT model and gamma correction were used to optimize and rescale the tree. The resulting tree, visualized using MEGA 6.06 (Tamura et al., 2013), is shown in Fig 1.

The results of a PhyloPhlAn analysis show that *C. pasteurianum* NRRL B-598 is closely related to *C. beijerinckii* strains. To determine whether the strain is identical to another *C. beijerinckii* strain, we performed another genome-wide analysis utilizing dDDH (digital DNA–DNA hybridization) (Auch et al., 2010). This technique replaces the wet-lab DDH by *in silico* comparison using complete genome sequences. The dDDH was computed using GGDC (Genome-to-Genome Distance Calculator) (Meier-Kolthoff et al., 2013). The results, comparing the studied strain with other *C. pasteurianum* and *C. beijerinckii* strains, are summarized in Table 1.

The values indicate that *C. pasteurianum* NRRL B-598 belongs to the *C. beijerinckii* species. This is because similarities to other *C. beijerinckii* species are above the 70% cutoff value for species delineation, while similarities to *C. pasteurianum* are far below this value. The analysis also suggests that strains *C. beijerinckii* NCIMB 8052 and *C. beijerinckii* ATCC 35702 are identical, while *C. pasteurianum* NRRL B-598 represents a standalone *C. beijerinckii* strain, because none of its similarities to other *C. beijerinckii* strains reached the cutoff value of 79%. Although *C. beijerinckii* and *C. pasteurianum* genomes have similar GC content, *C. beijerinckii* genomes are larger with a higher number of genes, as shown in Table 2.

Another substantial difference between *C. beijerinckii* and *C. pasteurianum* genomes can be found in *sol* operon, which forms

Table 2
Basic genome statistics for the three selected strains.

	C. pasteurianum NRRL B-598	*C. beijerinckii* NCIMB 8052	*C. pasteurianum* ATCC 6013
Length (bp)	6,186,879	6,000,632	4,352,101
GC content (%)	29.8	29.9	29.9
Total number of genes	5365	5231	3980
Protein coding genes (CDS)	5002	5023	3791

an important part of the pathway involved in solventogenesis. In *C. pasteurianum*, *sol* operon consists of the *adhE* gene (alcohol/acetaldehyde dehydrogenase), *ctfA* (CoA transferase subunit A), *ctfB* (CoA transferase subunit B), and *adc* (acetoacetate decarboxylase). Moreover, the *adc* gene is carried by the opposite strand from the other genes in operon. *Sol* operon in *C. beijerinckii* strains and *C. pasteurianum* NRRL B-598 genomes lack a *adhE* gene and contain a *ald* (aldehyde dehydrogenase) gene instead. Moreover, all genes in operon are carried by the same strand. Besides mentioned differences, the strain NRRL B-598 also exhibits some unique deviations from *C. beijerinckii* NCIMB 8052 (and other clostridia). For example, a specific DNA methylation pattern and restriction system requires Dam and Dcm methylation free DNA molecules for its transformation (Kolek et al., 2016a). Furthermore, both strains exhibit slightly different behavior in solvent production, sporulation and other cultivation characteristics (Kolek et al., 2016b).

The analyses demonstrate undisputable evidence that *C. pasteurianum* NRRL B-598 is misclassified. Therefore, we suggest its reclassification to be *C. beijerinckii* NRRL B-598. This correction helps to explain why development of the novel transformation method – totally different from the method for *C. pasteurianum* strains – was needed. It also helps us to better understand the behavior of the strain and allows us to properly plan for its future genetic manipulation, in order to increase its biofuel production.

Acknowledgements

Computational resources were partially provided by the CESNET LM2015042 and the CERIT Scientific Cloud LM2015085, provided under the programme "Projects of Large Research, Development, and Innovations Infrastructures". The study was also partly financed by specific university research (MSMT No 21/2012 and MSMT No 20-SVV/2016).

References

Auch, A.F., von Jan, M., Klenk, H.-P., Göker, M., 2010. Digital DNA–DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. Stand. Genomic Sci. 2, 117–134, http://dx.doi.org/10.4056/sigs.531120.

Collins, M., Lawson, P., Willems, A., Cordoba, J., Fernandez-Garayzabal, J., Garcia, P., Cai, J., Hippe, H., Farrow, J., 1994. The phylogeny of the genus clostridium: proposal of five new genera and eleven new species combinations. Int. J. Syst. Evol. Microbiol. 44, 812–826, http://dx.doi.org/10.1099/00207713-44-4-812.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797, http://dx.doi.org/10.1093/nar/gkh340.

Edgar, R.C., 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26, 2460–2461, http://dx.doi.org/10.1093/bioinformatics/btq461.

Janda, J., Abbott, S., 2007. 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. J. Clin. Microbiol. 45, 2761–2764, http://dx.doi.org/10.1128/JCM.01228-07.

Jensen, T.Ø., Kvist, T., Mikkelsen, M.J., Christensen, P.V., Westermann, P., 2012. Fermentation of crude glycerol from biodiesel production by Clostridium pasteurianum. J. Ind. Microbiol. Biotechnol. 39, 709–717, http://dx.doi.org/10.1007/s10295-011-1077-6.

Kolek, J., Sedlar, K., Provaznik, I., Patakova, P., 2014. Draft genome sequence of clostridium pasteurianum NRRL B-598, a potential butanol or hydrogen producer. Genome Announc. 2, e00192–e00214, http://dx.doi.org/10.1128/genomeA.00192-14.

Kolek, J., Branska, B., Drahokoupil, M., Patakova, P., Melzoch, K., 2016a. Evaluation of viability, metabolic activity and spore quantity in clostridial cultures during ABE fermentation. FEMS Microbiol. Lett. 6, fnw031, http://dx.doi.org/10.1093/femsle/fnw031.

Kolek, J., Sedlar, K., Provaznik, I., Patakova, P., 2016b. Dam and Dcm methylations prevent gene transfer into Clostridium pasteurianum NRRL B-598: development of methods for electrotransformation, conjugation, and sonoporation. Biotechnol. Biofuels 9, 1–14, http://dx.doi.org/10.1186/s13068-016-0436-y.

Lan, Y., Rosen, G., Hershberg, R., 2016. Marker genes that are less conserved in their sequences are useful for predicting genome-wide similarity levels between closely related prokaryotic strains. Microbiome 4, 1–35, http://dx.doi.org/10.1186/s40168-016-0162-5.

Lipovsky, J., Patakova, P., Paulova, L., Pokorny, T., Rychtera, M., Melzoch, K., 2016. Butanol production by *Clostridium pasteurianum* NRRL B-598 in continuous culture compared to batch and fed-batch systems. Fuel Process. Technol. 144, 139–144, http://dx.doi.org/10.1016/j.fuproc.2015.12.020.

Meier-Kolthoff, J.P., Auch, A.F., Klenk, H.-P., Göker, M., 2013. Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinf. 14, 1–14, http://dx.doi.org/10.1186/1471-2105-14-60.

Moon, C., Pacheco, D., Kelly, W., Leahy, S., Li, D., Kopecny, J., Attwood, G., 2008. Reclassification of Clostridium proteoclasticum as Butyrivibrio proteoclasticus comb. nov., a butyrate-producing ruminal bacterium. Int. J. Syst. Evol. Microbiol. 58, 2041–2045, http://dx.doi.org/10.1099/ijs.0.65845-0.

Price, M.N., Dehal, P.S., Arkin, A.P., 2010. FastTree 2 – Approximately maximum-likelihood trees for large alignments. PLoS One 5, e9490, http://dx.doi.org/10.1371/journal.pone.0009490.

Rainey, F., Lawson, P., 2016. Proposal to restrict the genus Clostridium Prazmowski to Clostridium butyricum and related species. Int. J. Syst. Evol. Microbiol. 60, 1009–1016, http://dx.doi.org/10.1099/ijsem.0.000824.

Rotta, C., Poehlein, A., Schwarz, K., McClure, P., Daniel, R., Minton, N., 2015. Closed genome sequence of Clostridium pasteurianum ATCC 6013. Genome Announc. 3, e01596–e01614, http://dx.doi.org/10.1128/genomeA.01596-14.

Sedlar, K., Skutkova, H., Kolek, J., Patakova, P., Provaznik, I., 2014. Identification and characterization of sol operon in clostridium pasteurianum NRRL B-598 genome. In: Proceedings of the 2nd International Conference on Chemical Technology. Czech Society of Industrial Chemistry, Mikulov, pp. 435–439, http://dx.doi.org/10.13140/2.1.3688.6402.

Sedlar, K., Kolek, J., Skutkova, H., Branska, B., Provaznik, I., Patakova, P., 2015. Complete genome sequence of Clostridium pasteurianum NRRL B-598, a non-type strain producing butanol. J. Biotechnol. 214, 113–114, http://dx.doi.org/10.1016/j.jbiotec.2015.09.022.

Segata, N., Börnigen, D., Morgan, X.C., Huttenhower, C., 2013. PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. Nat. Commun. 4, 2304, http://dx.doi.org/10.1038/ncomms3304.

Stackebrandt, E., Kramer, I., Swiderski, J., Hippe, H., 1999. Phylogenetic basis for a taxonomic dissection of the genus Clostridium. FEMS Immunol. Med. Microbiol. 24, 253–258, http://dx.doi.org/10.1111/j.1574-695X.1999.tb01291.x.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: Molecular evolutionary genetics analysis version 6. 0. Mol. Biol. Evol. 30, 2725–2729, http://dx.doi.org/10.1093/molbev/mst197.

5.4 Article IV

SNOPKOVÁ, Kateřina, **SEDLÁŘ, Karel**, BOSÁK, Juraj, CHALOUPKOVÁ, Eva, SEDLÁČEK, Ivo, PROVAZNÍK, Ivo and ŠMAJS, David. Free-Living Enterobacterium Pragia fontium 24613: Complete Genome Sequence and Metabolic Profiling. *Evolutionary Bioinformatics*. 2017. 13, p. 1176934317700863. (2017 IF = 1.877, Q2 in MATHEMATICAL & COMPUTATIONAL BIOLOGY)

DOI: https://doi.org/10.1177/1176934317700863

Free-Living Enterobacterium *Pragia fontium* 24613: Complete Genome Sequence and Metabolic Profiling

Evolutionary Bioinformatics
Volume 13: 1–7

DOI: 10.1177/1176934317700863
SAGE

Kateřina Snopková[1], Karel Sedlář[2], Juraj Bosák[1], Eva Chaloupková[1], Ivo Sedláček[3], Ivo Provazník[2] and David Šmajs[1]

[1]Department of Biology, Masaryk University, Brno, Czech Republic. [2]Department of Biomedical Engineering, Brno University of Technology, Brno, Czech Republic. [3]Czech Collection of Microorganisms, Department of Experimental Biology, Masaryk University, Brno, Czech Republic.

ABSTRACT: *Pragia fontium* is one of the few species that belongs to the group of atypical hydrogen sulfide-producing enterobacteria. Unlike other members of this closely related group, *P. fontium* is not associated with any known host and has been reported as a free-living bacterium. Whole genome sequencing and metabolic fingerprinting confirmed the phylogenetic position of *P. fontium* inside the group of atypical H_2S producers. Genomic data have revealed that *P. fontium* 24613 has limited pathogenic potential, although there are signs of genome decay. Although the lack of specific virulence factors and no association with a host species suggest a free-living style, the signs of genome decay suggest a process of adaptation to an as-yet-unknown host.

KEYWORDS: *Pragia fontium*, *Enterobacteriaceae*, whole genome sequence, phylogeny, free-living bacteria

RECEIVED: November 21, 2016. **ACCEPTED:** February 12, 2017.

PEER REVIEW: Seven peer reviewers contributed to the peer review report. Reviewers' reports totaled 1906 words, excluding any confidential comments to the academic editor.

TYPE: Original Research

FUNDING: The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Grant Agency of the Czech Republic (16-21649S) and also by internal grant MUNI/11/InGA04/2014.

DECLARATION OF CONFLICTING INTERESTS: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHOR: David Šmajs, Department of Biology, Masaryk University, Kamenice 753/5, 625 00 Brno, Czech Republic. Email: dsmajs@med.muni.cz

Introduction

To date, the *Enterobacteriaceae* family contains 55 genera and 248 species (www.bacterio.net, September 1, 2016). Most of the enterobacteria live in the vertebrate intestine, whereas several other enterobacterial genera/species represent plant pathogens or invertebrate endosymbionts.[1] Other enterobacteria are believed to live only in the environment, eg, *Pragia*, *Saccharobacter*, *Obesumbacterium*, *Shimwellia*,[1] *Mangrovibacter*,[2] and *Biostraticola*.[3] However, it is possible that their pathogenic/symbiotic potential will be revealed in the future, as it was for *Budvicia*.[4,5]

Pragia fontium is a gram-negative, mesophilic, rod-shaped, motile bacterium. The genus *Pragia* contains only 1 species, *P. fontium*, which was described in 1988.[6] A total of 18 strains were isolated in Czechoslovakia between 1982 and 1986. All strains, except 1, were isolated from water wells and water pipes, whereas 1 strain was obtained from the stool of a healthy woman. Another set of *Pragia* strains was isolated in Ukraine between 1996 and 1997.[7] They were mostly isolated from water (9 strains) and other environmental material (5 strains), although 2 strains came from human clinical material; their relatedness to the Czechoslovakia strains varied from 84% to 95% (based on DNA-DNA hybridization). To date, only strains from these 2 locations have been characterized, and the exact ecological niche and pathogenic potential of *Pragia* remains unclear.

Pragia fontium, as well as *Budvicia* spp. and *Leminorella* spp., is a closely related atypical enterobacterial species. Their common feature is hydrogen sulfide production, with *Budvicia diplopodorum* being the only known exception.[5] These H_2S-producing enterobacteria share several metabolic features including reduced metabolic activity that results in utilization of a limited set of substrates. The optimal growth temperature for *Pragia* and *Budvicia* is 25°C, whereas *Leminorella* is capable of growing at temperatures up to 42°C.[6] *Pragia fontium* can be differentiated from *Budvicia* spp. based on a positive (Simmons) citrate utilization test and from *Leminorella* spp. by its motility, tartrate utilization, tyrosine clearing, and inability to grow at 42°C.[6] In addition, a whole-cell protein pattern analysis of *P. fontium*, *B. aquatica*, and *Leminorella* spp. was determined and the data supported the delineation of these genera.[8] On the DNA level, *Pragia* strains were most closely related to *Budvicia* (based on DNA-DNA hybridization, relatedness 20%-37%) but barely related to other genera, eg, relatedness to *Escherichia coli* K12 was about 3%.[9]

To date, 485 completed enterobacterial genome sequences, covering 21 genera and 47 species, have been deposited in the Genomes OnLine Database (GOLD, https://gold.jgi.doe.gov/). Attention has been focused mainly on clinically and agriculturally important bacteria (eg, *Escherichia*, *Salmonella*, *Klebsiella*, and *Yersinia*), leaving the remaining genera relatively unexplored.

The whole genome sequence and the pilot assembly of *P. fontium* 24613 were published in 2015.[10] In this study, we characterized *P. fontium* based on genomic data, including the relationship of *Pragia* to other genera, and compared metabolic pathways with the results of phenotypic metabolic fingerprinting.

Materials and Methods

Bacterial strains and cultivation conditions

The strains used in this study came from the collection of the Department of Biology, Masaryk University, Brno, Czech Republic (*P. fontium* 24613, originally stored at the National Institute of Public Health, Prague, Czech Republic); from the Czech National Collection of Type Cultures, Prague, Czech Republic (*Budvicia aquatica* CNCTC 6285^T); and from the Czech Collection of Microorganisms, Brno, Czech Republic (*Leminorella grimontii* CCM 4003^T). *Pragia fontium* 24613 came from the same set of strains as *P. fontium* DSM 5563^{T6}. Strains were cultivated in TY medium (8 g casein, 5 g yeast extract, 5 g sodium chloride, pH 7.5; HiMedia, Mumbai, India) at 30°C for 24 hours.

Pragia fontium 24613 genome sequencing and annotation

In our previous study, protocols for DNA extraction, whole genome sequencing, and annotation of *P. fontium* 24613 were described in detail.[10] For additional gene mining and genome comparisons, annotation was manually curated based on results of a RAST (Rapid Annotation using Subsystem Technology) pipeline[11] and DOE-JGI (US Department of Energy-Joint Genome Institute) Microbial Genome Annotation Pipeline.[12] Detected proteins were assigned to Clusters of Orthologous Group (COG) categories based on DOE-JGI results. Methylome was characterized using PacBio single-molecule real-time sequencing (1× SMRT cell) of kinetic data collected during the genome sequencing process.[13] SMRT analysis version 2.3, using the "RS_Modification_and_Motif_Analysis.1" protocol, was used for genome-wide base modification and detection of the affected motifs. Regarding sequencing coverage, a default quality score value of 30 (corresponding to a *P* value of .001) was used for motif determination. The detected motifs were uploaded and further analyzed using the REBASE database.[14] The complete genome was also scanned for homologues of restriction-modification system genes (using a Basic Local Alignment Search Tool [BLAST] search, with the BLASTX algorithm) against the REBASE and GenBank databases.

Phylogenetic position of P. fontium

The genome sequence of *P. fontium* 24613 was compared with other enterobacterial genera on a genome-wide level. Whole genome sequences were downloaded from the GOLD (https://gold.jgi.doe.gov/); their accession numbers are listed in Table S1. Each genus was represented by 1 sequence (except for *Pragia* where both the type strain DSM 5563^T and strain 24613 were used). If available, the sequence of the type strain was used. For genera *Biostraticola*, *Cosenzaea*, *Gibbsiella*, *Mangrovibacter*, *Obesumbacterium*, *Saccharobacter*, and *Samsonia*, no sequences were available. A whole genome phylogenetic analysis was built using PhyloPhlAn 0.99,[15] which compared more than 400 selected protein sequences conserved across bacterial domains. The genes were identified using an internal PhyloPhlAn database by translated mapping with USEARCH 8.1.[16] The topology was computed using the neighbor-joining algorithm in conjunction with the Jukes-Cantor evolution model. Moreover, the CAT model, with gamma correction, was used to optimize and rescale the tree. The final tree was reconstructed, using FastTree 2.1,[17] from protein subsequences of the genes concatenating their most informative amino acid positions, and each was aligned using MUSCLE 3.8.[18] The tree was visualized in MEGA 6.06.[19] Dot plot diagrams between genomes were constructed using the Integrated Microbial Genome platform.[12] The core genome of *P. fontium*, *B. aquatica*, and *L. grimontii* was determined based on orthologous clusters produced by OrthoVenn[20] using a modified OrthoMCL heuristic approach. Default parameters (E-value 1e−5 and inflation value 1.5) were used. Metabolic pathway analysis of *P. fontium* 24613, *Wigglesworthia glossinidia* (acc. no. CP003315), and *Buchnera aphidicola* G002 (acc. no. CP002701) was performed using the KEGG PATHWAY database,[21] which is part of KEGG Web services (http://www.genome.jp/kegg/).

Analyses of metagenomics data

Data from the Human Microbiome Project database (http://hmpdacc.org) and EBI Metagenomics database (https://www.ebi.ac.uk/metagenomics/) were searched with BLASTN 2.2.22[22] using a consensus sequence of 7 16S ribosomal RNA (rRNA) genes of *P. fontium* 24613. The first database contained a complete set of human microbiome data (associating data from several human sites), and the latter database covered data from different environmental sources.

Substrate diversity studies

The Biolog GN2 MicroPlate analysis platform (Biolog, Inc., Hayward, CA, USA) was used for determination of the biochemical profiles of *P. fontium* 24613, *B. aquatica* CNCTC 6285^T, and *L. grimontii* CCM 4003^T cultivated on Biolog Universal Growth (BUG) agar at 30°C for 24 hours. Utilization of 95 carbon sources was tested[23] (Table S2). Media and all reagents were supplied by Biolog and used according to the manufacturer's protocol. Plates were incubated in parallel under aerobic and anaerobic conditions and tests were read after 24 hours of incubation.

Results

Genome analyses of P. fontium 24613

Complete genome sequence of P. fontium 24613. A complete genome sequence for *P. fontium* 24613 represents a single circular chromosome with a length of 4 094 629 bp.[10] The *P. fontium* 24613 genome was compared with 3 draft genomes

Table 1. Genome features of *Pragia fontium* 24613 in comparison with the draft genome of the type strain and the draft genomes of closely related hydrogen sulfide producers.

FEATURE	*P. FONTIUM* 24613	*P. FONTIUM* DSM 5563^T	*B. AQUATICA* DSM 5075^T	*L. GRIMONTII* DSM 5078^T
Genome status	Complete	Draft	Draft	Draft
Genome size	4 094 629 bp	3 950 845 bp	5 670 930 bp	4 222 128 bp
GC content	45.38%	45.23%	45.68%	53.86%
No. of CDS	3579	3464	5130	3878
No. of rRNA genes	22 (8–7–7)	10 (2–5–3)	7 (5–2–0)	16 (8–6–2)
No. of tRNA genes	72	58	57	57
No. of pseudogenes	146 (4.1%)	NA	144 (2.8%)	62 (1.6%)
No. of genes with predicted function	2809 (78.49%)	2862 (82.62%)	3896 (75.95%)	3083 (79.50%)
No. of genes assigned to COG	2601 (72.67%)	2613 (75.43%)	2601 (72.67%)	2804 (72.31%)
No. of genes assigned to KEGG pathways	1160 (32.41%)	1172 (33.83%)	1419 (27.66%)	1217 (31.38%)

Abbreviations: CDS, coding sequences; COG, Clusters of Orthologous Group; KEGG, Kyoto Encyclopedia of Genes and Genomes; rRNA, ribosomal RNA; tRNA, transfer RNA.
Accession numbers of the whole genome sequences of the type strains are listed in Table S1. Order of the rRNA genes in parentheses: 5S-16S-23S. NA—data not available in the GenBank and Genomes OnLine databases.

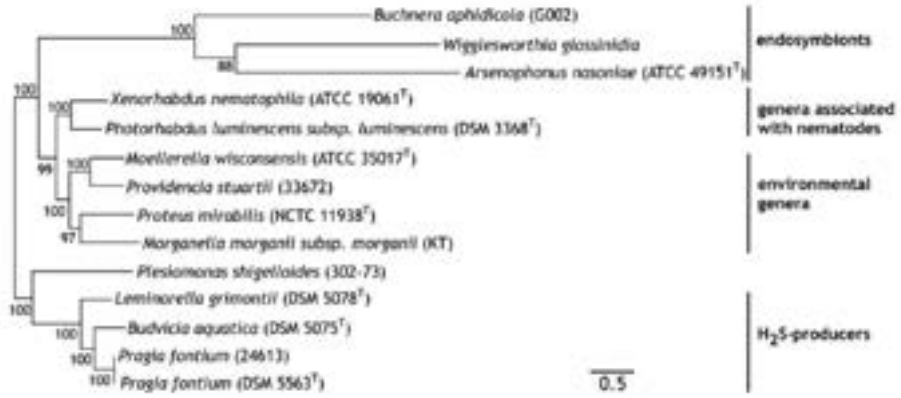


Figure 1. Phylogenetic position of the genus *Pragia* based on a whole genome sequence tree. Only the relevant part of the *Enterobacteriaceae* tree is shown (the whole tree is depicted in Figure S2). All branches are supported with high bootstrap values. The tree was drawn to scale; the scale bar represents the estimated number of amino acid changes per site per unit of branch length.

of related bacteria, including the draft genome of *P. fontium* DSM 5563^T (Table 1). Both the *B. aquatica* DSM 5075^T genome and the *L. grimontii* DSM 5078^{T}genome were larger in size and gene count compared with the complete genome sequence of *P. fontium* 24613. Moreover, the proportion of pseudogenes was larger in *P. fontium* (4.1%) than in the draft genomes of other H_2S-producing enterobacteria (ie, 2.8% and 1.6% for *B. aquatica* and *L. grimontii*, respectively), suggesting genome decay in *P. fontium*. In addition, a clearly higher GC content was found in the *L. grimontii* DSM 5078^T (*L. grimontii*) genome. The draft status was likely responsible for the lower number of predicted rRNA and transfer RNA genes in the *P. fontium* DSM 5563^T, *B. aquatica*, and *L. grimontii* genomes.

Phylogenetic position of P. fontium. A whole genome phylogenetic approach was used to compare the genome sequence of *P. fontium* 24613 with genome sequences of other enterobacterial genera. The relevant part of the *Enterobacteriaceae* tree is shown in Figure 1. Strong support was found for a close relationship among *Pragia* and other atypical H_2S producers, including *Budvicia* and *Leminorella*. The high similarity among genomes was also supported by a dot plot analysis of H_2S producer genomes (Figure S1). Another related genus was *Plesiomonas*, an oxidase-positive genus recently reclassified into the *Enterobacteriaceae* family.[24] A sister clade contains a cluster of genera occurring frequently in the (1) environment (*Providencia*, *Moellerella*, *Proteus*, and *Morganella*), (2) genera associated with nematodes (*Xenorhabdus*, *Photorhabdus*), and (3) endosymbionts (*Arsenophonus*, *Buchnera*, and *Wigglesworthia*). Except for the delineation of *Proteus* vs *Morganella* and endosymbionts *Buchnera* vs *Wigglesworthia*, all other branches were supported by bootstrap values higher than 99%.

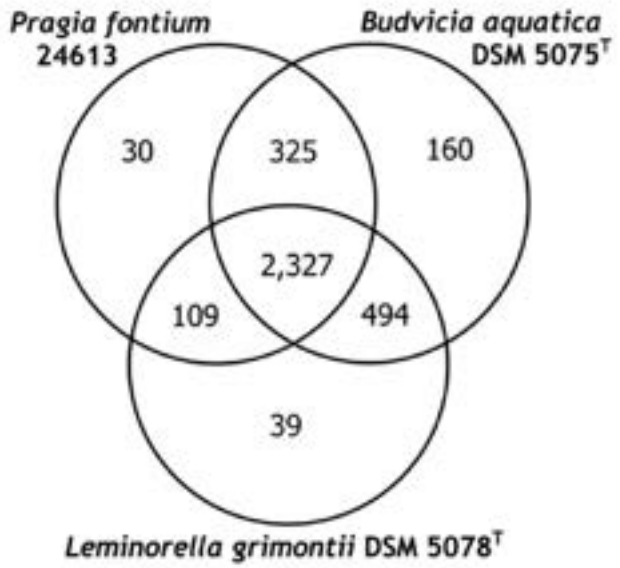


Figure 2. The Venn diagram represents the core genome and pangenome of *Pragia* and the closely related atypical H_2S producers. The numbers represent the gene clusters shared by corresponding group of genera. The diagram shows the close relationships among those inside the group of atypical H_2S producers.

The core genome of enterobacterial hydrogen sulfide producers. The core genomes of *P. fontium*, *B. aquatica*, and *L. grimontii* contain 2327 gene clusters (ie, at least 1 gene from each cluster was found in each genome; Figure 2). The number of gene clusters exclusively shared by 2 genomes was higher for the *P. fontium* and *B. aquatica* genomes (325) compared with the *P. fontium* and *L. grimontii* genomes (109), whereas there were 494 clusters shared by the *L. grimontii* and *B. aquatica* genomes. These data indicate a higher degree of relatedness between *P. fontium* and *B. aquatica* compared with *P. fontium* and *L. grimontii*. A set of 30 gene clusters was unique for the *P. fontium* genome; these clusters encoded homologues to fimbrial genes found in *Serratia* spp. and *Proteus* spp. and also homologues to pyocin S3 and its immunity protein–encoding genes. In total, 10 clusters encoded genes for hypothetical proteins.

Genome-based metabolic and virulence analyses

Analysis of metabolic pathways in the P. fontium genome. Based on the genomic data analysis from KEGG PATHWAY and DOE-JGI, aerobic and facultative anaerobic metabolism of *P. fontium* 24613 was predicted. Oxidized nitrogen and sulfur compounds were capable of serving as alternative terminal electron acceptors under anaerobic conditions. Identification of thiosulfate reductase, responsible for H_2S production, corresponded to previously detected enzyme activity.[6] The genes involved in glycolysis, citrate cycle, and pentose phosphate pathway could also be found in the *P. fontium* genome in addition to genes responsible for amino acid, fatty acid synthesis, lipid, and nucleotide metabolism. *Pragia* was found to be auxotrophic for L-tryptophan, L-histidine, and L-leucine and deficient in biotin synthesis. Compared with *Budvicia* and *Leminorella*, *Pragia* was able to synthetize L-arginine but lacked the genes for fatty acid degradation. In addition, the *P. fontium* genome contained fewer genes involved in carbohydrate metabolism compared with the *L. grimontii* and *B. aquatica* genomes (Table S3).

Genome methylation pattern. Analysis of PacBio sequencing data revealed 24 814 methylated positions of the m6A type, but only a single sequence motif (GATC) was found in all these modifications. More than 80% (21 735 of 26 606) of the GATC positions in the genome were methylated. Methylation type m4C was not found. Kinetic signatures of m5C were subtler than signatures of m6A and m4C and harder to detect using PacBio SMRT sequencing[25]; therefore, they were not assessed. The results of *P. fontium* genome methylation were deposited in the REBASE PacBio database (http://rebase.neb.com/cgi-bin/pacbiolist).[14] In total, 8 different putative restriction-modification systems, all of them type II, were predicted in the genome (Table S4). Seven of them consisted of only methyltransferases, whereas the last one modifying m5C consisted of methyltransferase, mismatch repair endonuclease, and restriction endonuclease.

Virulence and antimicrobial genes in the P. fontium genome. In silico analysis of virulence determinants of the *P. fontium* genome revealed genes involved in iron acquisition (encoding Fe^{2+} and Fe^{3+} transport systems), adhesion (encoding P pili and type I pili), secretion systems (T1SS and T6SS), and antibiotic resistance (encoding AmpC β-lactamase and several efflux pump) (see Table S5).

Production of tailocins, ie, R-type and F-type bacteriocins resembling phage tails, was previously detected in several *Pragia* strains.[26] Gene clusters similar to the phage genes were described as being responsible for production of these antimicrobial compounds.[27] A total of 6 clusters homologous to phage genes were predicted in the *P. fontium* genome, and one of them was likely responsible for tailocin production (see Table S5). The genome search also detected a gene encoding a colicin-like bacteriocin, a homologue of pyocin S3.

Metabolic profiling of P. fontium 24613

The carbohydrate utilization pattern resulting from the testing of various saccharides, carboxylic acids, alcohols, amino acids, aromatic compounds, and their derivatives was determined for *P. fontium* 24613, *B. aquatica* CNCTC 6285T, and *L. grimontii* CCM 4003T. In general, the data obtained from the Biolog assay revealed low levels of metabolic activity in all tested strains. Substrate utilization profiles differed for the 3 tested H_2S producers in 17 substrates (Table S2). *Pragia fontium* 24613 was able to utilize 15 substrates (out of 95; 16%) under aerobic conditions and 22 (out of 95; 23%) under anaerobic conditions. *Pragia* utilized monosaccharides and their derivatives (α-D-glucose, α-D-glucose-1-phosphate, D-glucose-6-phosphate, *N*-acetyl-D-glucosamine, and β-methyl-D-glucoside), monocarboxylic acids (D,L-lactic acid, and D-gluconic acid), dicarboxylic acids (α-keto-glutaric acid, and L-glutamic acid), alcohols and their derivatives (glycerol, D,L-α-glycerol phosphate, myo-inositol, and xylitol), amino acids (D-serine), and aromatic compounds (uridine and thymidine). In addition to substrates

utilized under aerobic conditions, anaerobically cultivated *Pragia* utilized L-arabinose, pyruvic acid methyl ester, D-glucuronic acid, bromosuccinic acid, L-aspartic acid, glycyl-L-aspartic acid, and L-serine. Although *Budvicia* utilized 16 substrates (17%) aerobically and 24 (25%) anaerobically, *Leminorella* utilized only 13 substrates (14%) aerobically and 18 (19%) anaerobically. *Budvicia* and *Leminorella* were able to metabolize several amino acids and their derivatives (L-asparagine, L-aspartic acid, and glycyl-L-aspartic acid) as well as derivatives of organic acids from Krebs cycle (pyruvic acid methyl ester, bromosuccinic acid), which were not utilized by *Pragia*. The complete results of this assay are shown in Table S2. In most of the substrate tests, which differed among H_2S producers, the genes encoding corresponding catabolic enzymes or enzymes possibly involved in metabolism of these compounds were found (Table S6). The only exception was the *B. aquatica* genome, where some of the genes responsible for catabolism of uridine were not found.

Discussion

Pragia belongs to a relatively small group of H_2S-producing enterobacteria containing *P. fontium*, *Budvicia* spp., and *Leminorella* spp. Although all members of this small group are closely related and have a relatively similar biochemical profile, they occupy quite different ecological niches. Although *Budvicia* was originally isolated from freshwater,[28] several other isolates have been described from the intestinal microflora of insects,[26,29] *Diplopoda*,[5] and salmonids.[30] A possible clinical relevance for *B. aquatica* was reported by Corbin et al[4] when this bacterium was isolated from a human clinical sample. *Leminorella* spp. have been exclusively isolated from human clinical specimens and no environmental sources have been reported. Although its clinical significance is unclear,[1] *Leminorella* spp. appear to be associated with urinary tract infections and other human nosocomial infections.[31] In contrast to *Budvicia* and *Leminorella*, *Pragia* has been isolated almost exclusively from environmental sources. Only 3 isolates originated from human clinical samples; there is no information on the role of these strains in infection or disease.[6,7] Because the prevalent habitat of other *Pragia* strains is drinking water, these cases likely reflect accidental isolations. Inspection of metagenomics data revealed the absence of *Pragia* 16S ribosomal DNA (rDNA) in both environmental and host-associated data sets (data not shown). From all the available data, *Pragia* appears to be the only H_2S producer occupying environmental niches with no association with humans or other hosts.

A possible interaction between *Pragia* and a host species was examined by identification and analysis of genes encoding virulence factors. Several common virulence factors shared by most enterobacterial species (even saprophytic ones) were detected. Genes for adhesion, antibiotic resistance, iron uptake, and 2 secretion systems were found. Adhesion and the ability to acquire iron are key factors required for colonization and survival in a host (animal or plant).[32-34] These findings indirectly support an association between *Pragia* and an as-yet-unknown host. We can speculate that if a host organism exists, it will likely be similar to those of the closely related genus *Budvicia*, ie, nonvertebrate hosts such as insects or nematodes. Although the presence of *Pragia* has been detected in the intestines of freshwater salmon,[35] the much more frequent isolation from deepwater wells[6] tends to support a free-living lifestyle of *Pragia*. Both detected secretion systems, T1SS and T6SS, are widely distributed in gram-negative bacteria[36,37] and could mediate interaction with a host or with another bacterium.[38] Although the contribution of T6SS to pathogenesis has been described for several bacteria, eg, *Pseudomonas*[39] and *E coli*,[40] T6SS has also been found in saprophytic bacteria, where it was involved in interactions across the microbial community.[38] Several bacteriocin types have been suggested as putative virulence factors, whereas the importance of others was demonstrated in interactions across microbial community.[41,42] Although the function of *P. fontium* bacteriocins remains unknown, both tailocins and colicin-like homologues were found in the *Pragia* genome. The GATC methylation motif was found in the *P. fontium* genome, and because the corresponding gene for the restriction enzyme recognizing this motif was not found, methylation appears to be more connected to gene expression regulation[43] and not to degradation of foreign nucleic acid molecules.

Metabolic profiling revealed a metabolic pattern for *Pragia*, *Budvicia*, and *Leminorella*, which was quite distinct from other enterobacteria,[44] supporting the distinctness of enterobacterial H_2S producers and also the close relationship of these bacteria within this group. Despite their overall similarity, H_2S-producing enterobacteria revealed several differences in their ability to utilize substrates. Analyses of genomic data supported the metabolic findings, with only 1 case in which some of the genes encoding expected enzymatic activity were not found. This is likely a result of an incomplete genomic sequence in *Budvicia*. Surprisingly, all species were able to degrade multiple substrates under anaerobic conditions suggesting that alternative electron acceptors (nitrate, reduced sulfur compounds) could be used under anaerobic conditions. Nitrogen oxidation could be carried out using the "nitrite reduction to ammonium pathway" for which the corresponding genes were found in the *P. fontium* genome. This pathway is preferred for respiration under anaerobic conditions, and it is common across *Enterobacteriaceae* and in other facultatively anaerobic bacteria.[45]

A comparative genomics approach revealed that almost 80% of the gene clusters were shared by H_2S-producing enterobacteria, whereas only 49% were shared when *E coli* K12 was added to the analysis. Analysis of the complete genome sequence of *Pragia* revealed that the genome contains genes involved in essential metabolic pathways, in nutrient metabolism, and also in the synthesis of most of the amino acids.

However, the "fatty acid degradation pathway" is missing from the *P. fontium* 24613 genome. This pathway is present in most enterobacterial genomes but not in invertebrate endosymbionts with a reduced genome, such as *Wigglesworthia* and *Buchnera*. Nevertheless, when compared with these endosymbionts, the *P. fontium* genome is relatively large and also contains an additional set of genes, eg, those responsible for degradation of more complex polysaccharides. However, *P. fontium* 24613 has a relatively small genome in comparison with other enterobacteria, even in comparison with the genus *Budvicia*. In addition, the proportion of pseudogenes was larger in *Pragia* compared with other closely related bacteria (despite their draft status, which is prone to assembly errors). Larger proportions of pseudogenes have also been observed in bacteria that were associated with or dependent on eukaryotic hosts.[46] Nevertheless, this analysis comes from a limited number of genome sequences per species and it is known that the prevalence of pseudogenes is quite variable even among closely related strains.[47] A reduction in genome size and an increased number of pseudogenes are common signs of bacterial adaptation to a eukaryotic host. In addition, the *P. fontium* genome contains fewer genes involved in carbohydrate utilization compared with other H_2S producers; a large battery of degradation enzymes is important mainly for free-living bacteria. The traces of genome decay (ie, small genome, absence of fatty acid degradation pathways, the small number of genes associated with carbohydrate utilization, and a larger proportion of pseudogenes) suggest an ongoing process of adaptation to a particular host organism. Although no such host has been identified for *P. fontium*, the recent progress in metagenome studies could help to answer this question in the near future.

Conclusions

Analysis of the complete genome sequence of *P. fontium* 24613 and metabolic profiling confirmed the close relatedness of this bacterium to other H_2S-producing enterobacteria, *Budvicia* spp. and *Leminorella* spp., although for each genus a different environmental niche has been described. Virulence gene mining and the absence of *Pragia* 16S rDNA sequences in the human metagenomics data suggest limited pathogenic potential for *Pragia*, consistent with the previously described free-living lifestyle of this bacterium. On the contrary, reduced genome size, limited number of encoded enzymes for carbohydrate and fatty acid degradation, and frequent presence of pseudogenes suggest a process of adaptation to an as-yet-unknown host.

Acknowledgements

The authors thank Jan Šmarda for providing the *P. fontium* 24613 strain.

Author Contributions

KSn and DS conceived, designed, and performed the experiments (genome sequencing). IS performed phenotypical characterization. KSn, KSe, and IP analyzed the data. KSn wrote the first draft of the manuscript. DS and JB contributed to the writing of the manuscript. All authors reviewed and approved the final manuscript.

Disclosures and Ethics

As a requirement of publication, author(s) have provided to the publisher signed confirmation of compliance with legal and ethical obligations including, but not limited to, the following: authorship and contributorship, conflicts of interest, privacy and confidentiality, and (where applicable) protection of human and animal research subjects. The authors have read and confirmed their agreement with the ICMJE authorship and conflict of interest criteria. The authors have also confirmed that this article is unique and not under consideration or published in any other publication, and that they have permission from rights holders to reproduce any copyrighted material. Any disclosures are made in this section. The external blind peer reviewers report no conflicts of interest.

REFERENCES

1. Octavia S, Lan R. The family *Enterobacteriaceae*. In: Rosenberg E, DeLong EF, Lory S, Stackebrandt E, Thompson F, eds. *The Prokaryotes*. 4th ed. Berlin, Germany: Springer-Verlag; 2014:225–275.
2. Rameshkumar N, Lang E, Nair S. *Mangrovibacter plantisponsor* gen. nov., sp. nov., a nitrogen-fixing bacterium isolated from a mangrove-associated wild rice (*Porteresia coarctata* Tateoka). *Int J Syst Evol Microbiol*. 2010;60:179–186.
3. Verbarg S, Frühling A, Cousin S, et al. *Biostraticola tofi* gen. nov., spec. nov., a novel member of the family *Enterobacteriaceae*. *Curr Microbiol*. 2008;56:603–608.
4. Corbin A, Delette C, Besson S, et al. *Budvicia aquatica* sepsis in an immunocompromised patient following exposure to the aftermath of Hurricane Katrina. *J Med Microbiol*. 2007;56:1124–1125.
5. Lang E, Schumann P, Knapp BA, Kumar R, Spröer C, Insam H. *Budvicia diplopodorum* sp. nov. and emended description of the genus *Budvicia*. *Int J Syst Evol Microbiol*. 2013;63:260–267.
6. Aldová E, Hausner O, Brenner DJ, et al. *Pragia fontium* gen. nov., sp. nov. of the family *Enterobacteriaceae*, isolated from water. *Int J Syst Evol Microbiol*. 1988;38:183–189.
7. Pokhyl SI. The properties of *Pragia fontium* bacteria isolated on the territories of Ukraine and the Czech Republic. *Mikrobiol Z*. 2000;62:3–10.
8. Schindler J, Potužníková B, Aldová E. Classification of strains of *Pragia fontium*, *Budvicia aquatica* and of *Leminorella* by whole-cell protein pattern. *J Hyg Epidemiol Microbiol Immunol*. 1992;36:207–216.
9. Hickman-Brenner FW, Vohra MP, Huntley-Carter GP, et al. *Leminorella*, a new genus of *Enterobacteriaceae*: identification of *Leminorella grimontii* sp. nov. and *Leminorella richardii* sp. nov. found in clinical specimens. *J Clin Microbiol*. 1985;21:234–239.
10. Snopková K, Sedlář K, Bosák J, Chaloupková E, Provazník I, Šmajs D. Complete genome sequence of *Pragia fontium* 24613, an environmental bacterium from the family *Enterobacteriaceae*. *Genome Announc*. 2015;3:e00740–15.
11. Aziz RK, Bartels D, Best AA, et al. The RAST Server: rapid annotations using subsystems technology. *BMC Genomics*. 2008;9:75.
12. Markowitz VM, Chen I-MA, Palaniappan K, et al. IMG 4 version of the integrated microbial genomes comparative analysis system. *Nucleic Acids Res*. 2014;42:D560–D567.
13. Flusberg BA, Webster DR, Lee JH, et al. Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat Methods*. 2010;7:461–465.
14. Roberts RJ, Belfort M, Bestor T, et al. A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Res*. 2003;31:1805–1812.
15. Segata N, Börnigen D, Morgan XC, Huttenhower C. PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. *Nat Commun*. 2013;4:2304.
16. Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinforma Oxf Engl*. 2010;26:2460–2461.
17. Price MN, Dehal PS, Arkin AP. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS ONE*. 2010;5:e9490.

18. Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*. 2004;5:113.
19. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol*. 2013;30:2725–2729.
20. Wang Y, Coleman-Derr D, Chen G, Gu YQ. OrthoVenn: a web server for genome wide comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Res*. 2015;43:W78–W84.
21. Zhou T. Computational reconstruction of metabolic networks from KEGG. *Methods Mol Biol Clifton NJ*. 2013;930:235–249.
22. Altschul SF, Madden TL, Schäffer AA, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 1997;25:3389–3402.
23. Garland JL, Mills AL. Classification and characterization of heterotrophic microbial communities on the basis of patterns of community-level sole-carbon-source utilization. *Appl Environ Microbiol*. 1991;57:2351–2359.
24. Janda JM. Genus XXVII. *Plesiomonas*, Habs and Schubert 1962, 324AL. In: Brenner DJ, Krieg NR, Staley JT, Garrity GM, eds. *Bergey's Manual of Systematic Bacteriology*. Vol 2. 2nd ed, New York, NY: Springer; 2005;740–744.
25. Davis BM, Chao MC, Waldor MK. Entering the era of bacterial epigenomics with single molecule real time DNA sequencing. *Curr Opin Microbiol*. 2013;16:192–198.
26. Tagliavia M, Messina E, Manachini B, Cappello S, Quatrini P. The gut microbiota of larvae of *Rhynchophorus ferrugineus* Oliver (Coleoptera: Curculionidae). *BMC Microbiol*. 2014;14:136.
27. Ghequire MGK, De Mot R. Ribosomally encoded antibacterial proteins and peptides from *Pseudomonas*. *FEMS Microbiol Rev*. 2014;38:523–568.
28. Aldová E, Hausner O, Gabrhelová M, Schindler J, Petrášl P, Braná H. A hydrogen sulphide producing Gram-negative rod from water. *Zentralblatt Für Bakteriol Mikrobiol Hyg A*. 1983;254:95–108.
29. Gupta AK, Rastogi G, Nayduch D, Sawant SS, Bhonde RR, Shouche YS. Molecular phylogenetic profiling of gut-associated bacteria in larvae and adults of flesh flies. *Med Vet Entomol*. 2014;28:345–354.
30. Skrodenyte-Arbaciauskiene V, Sruoga A, Butkauskas D. Assessment of microbial diversity in the river trout *Salmo trutta fario* L. intestinal tract identified by partial 16S rRNA gene sequence analysis. *Fish Sci*. 2006;72:597–602.
31. Blekher L, Siegman-Igra Y, Schwartz D, Berger SA, Carmeli Y. Clinical significance and antibiotic resistance patterns of *Leminorella* spp., an emerging nosocomial pathogen. *J Clin Microbiol*. 2000;38:3036–3038.
32. Kline KA, Dodson KW, Caparon MG, Hultgren SJ. A tale of two pili: assembly and function of pili in bacteria. *Trends Microbiol*. 2010;18:224–232.
33. Rossez Y, Holmes A, Wolfson EB, et al. Flagella interact with ionic plant lipids to mediate adherence of pathogenic *Escherichia coli* to fresh produce plants. *Environ Microbiol*. 2014;16:2181–2195.
34. Boyer E, Bergevin I, Malo D, Gros P, Cellier MF. Acquisition of Mn(II) in addition to Fe(II) is required for full virulence of *Salmonella enterica* serovar Typhimurium. *Infect Immun*. 2002;70:6032–6042.
35. Skrodenyte-Arbaciauskiene V, Sruoga A, Butkauskas D, Skrupskelis K. Phylogenetic analysis of intestinal bacteria of freshwater salmon *Salmo salar* and sea trout *Salmo trutta trutta* and diet. *Fisheries Sci*. 2008;74:1307–1314.
36. Costa TRD, Felisberto-Rodrigues C, Meir A, et al. Secretion systems in Gram-negative bacteria: structural and mechanistic insights. *Nat Rev Microbiol*. 2015;13:343–359.
37. Thomas S, Holland IB, Schmitt L. The Type 1 secretion pathway—the hemolysin system and beyond. *Biochim Biophys Acta*. 2014;1843:1629–1641.
38. Basler M. Type VI secretion system: secretion by a contractile nanomachine. *Philos Trans R Soc Lond B Biol Sci*. 2015;370:20150021.
39. Mougous JD, Cuff ME, Raunser S, et al. A virulence locus of *Pseudomonas aeruginosa* encodes a protein secretion apparatus. *Science*. 2006;312:1526–1530.
40. Zhou Y, Tao J, Yu H, et al. Hcp family proteins secreted via the type VI secretion system coordinately regulate *Escherichia coli* K1 interaction with human brain microvascular endothelial cells. *Infect Immun*. 2012;80:1243–1251.
41. Riley MA, Gordon DM. The ecological role of bacteriocins in bacterial competition. *Trends Microbiol*. 1999;7:129–133.
42. Šmajs D, Micenková L, Šmarda J, et al. Bacteriocin synthesis in uropathogenic and commensal *Escherichia coli*: colicin E1 is a potential virulence factor. *BMC Microbiol*. 2010;10:288.
43. Sánchez-Romero MA, Cota I, Casadesús J. DNA methylation in bacteria: from the methyl group to the methylome. *Curr Opin Microbiol*. 2015;25:9–16.
44. Janda JM. New members of the family *Enterobacteriaceae*. In: Dworkin M, Falkow S, Rosenberg E, Schleifer K-H, Stackebrandt E, eds. *The Prokaryotes: Volume 6: Proteobacteria: Gamma Subclass*. 3rd ed. New York, NY: Springer; 2006:5–40.
45. Tiedje JM. Ecology of denitrification and dissimilatory nitrate reduction to ammonium. In: Zehnder AJB, ed. *Environmental Microbiology of Anaerobes*. New York, NY: John Wiley & Sons; 1988:179–244.
46. Lerat E, Ochman H. Psi-Phi: exploring the outer limits of bacterial pseudogenes. *Genome Res*. 2004;14:2273–2278.
47. Kuo C-H, Ochman H. The extinction dynamics of bacterial pseudogenes. *PLoS Genet*. 2010;6:e1001050.

5.5 Article V

PATAKOVA, Petra, KOLEK, Jan, **SEDLAR, Karel**, KOSCOVA, Pavlina, BRANSKA, Barbora, KUPKOVA, Kristyna, PAULOVA, Leona and PROVAZNIK, Ivo. Comparative analysis of high butanol tolerance and production in clostridia. *Biotechnology Advances*. 2018. 36(3), p. 721–738. (2018 IF = 12.831, D1 in BIOTECHNOLOGY & APPLIED MICROBIOLOGY)

DOI: https://doi.org/10.1016/j.biotechadv.2017.12.004

Article V: Comparative analysis of high butanol tolerance and production in clostridia

Biotechnology Advances 36 (2018) 721–738

Contents lists available at ScienceDirect

Biotechnology Advances

journal homepage: www.elsevier.com/locate/biotechadv

Research review paper

Comparative analysis of high butanol tolerance and production in clostridia

Petra Patakova[a,*], Jan Kolek[a], Karel Sedlar[b], Pavlina Koscova[b], Barbora Branska[a], Kristyna Kupkova[b], Leona Paulova[a], Ivo Provaznik[b,c]

[a] Department of Biotechnology, University of Chemistry and Technology Prague, Technicka 5, CZ 16628 Prague 6, Czech Republic
[b] Department of Biomedical Engineering, Brno University of Technology, Technicka 12, CZ 61600, Brno, Czech Republic
[c] Department of Physiology, Masaryk University, Kamenice 5, CZ 625 00, Brno, Czech Republic

ARTICLE INFO

Keywords:
Clostridium
ABE fermentation
Butanol tolerance
Butanol production
Genome mining
Transcriptomics

ABSTRACT

2016, was the 100 years anniversary from launching of the first industrial acetone-butanol-ethanol (ABE) microbial production process. Despite this long period and also revival of scientific interest in this fermentative process over the last 20 years, solventogenic clostridia, mainly *Clostridium acetobutylicum*, *Clostridium beijerinckii*, *Clostridium saccharoperbutylacetonicum* and *Clostridium pasteurianum*, still have most of their secrets. One such poorly understood mechanism is butanol tolerance, which seems to be one of the most significant bottlenecks obstructing industrial exploitation of the process because the maximum achievable butanol concentration is only about 21 g/L. This review describes all the known cellular responses elicited by butanol, such as modifications of cell membrane and cell wall, formation of stress proteins, extrusion of butanol by efflux pumps, response of regulatory pathways, and also maps both random and targeted mutations resulting in high butanol production phenotypes. As progress in the field is inseparably associated with emerging methods, enabling a deeper understanding of butanol tolerance and production, progress in these methods, including genome mining, RNA sequencing and constructing of genome scale models are also reviewed. In conclusion, a comparative analysis of both phenomena is presented and a theoretical relationship is described between butanol tolerance/high production and common features including efflux pump formation/activity, stress protein production, membrane modifications and biofilm growth.

1. Introduction

Acetone-butanol-ethanol (ABE) fermentation is a peculiar but interesting metabolic pathway in some non-pathogenic and non-toxinogenic clostridia that probably relieves metabolic stress caused by the accumulation of acids both inside and outside of cells. ABE fermentation is usually a bi-phasic process, at least for *Clostridium acetobutylicum*, *Clostridium beijerinckii* and *Clostridium saccharoperbutylacetonicum* species, in which mainly butyric and acetic acid formation is superseded by the formation of solvents, n-butanol, acetone/2-propanol and ethanol in later phases of the cell cycle. Solvent formation is usually accompanied by sporulation but currently it seems that regulation of both processes is not tightly linked in all strains. Past industrial fermentative acetone and butanol production processes worldwide used this process (Jones and Woods, 1986; Chiao and Sun, 2007) including the former Czechoslovakia (Patakova et al., 2009) until it was replaced by cheaper chemical production from oil. In China, several plants produced butanol from first generation (food) feedstocks till 2009 but in 2013 only one plant processing mixed first/second generation feedstocks stayed open (Jiang et al., 2015); biobutanol production in China is now unknown. Nowadays, the process has experienced a renewal of interest coupled with fluctuating prices of oil and the largest oil reserves being located in politically unstable countries. In December 2016, the Green Biologics company announced the first commercial shipments of bio-based butanol and acetone from its plant located in Little Falls, Minnesota, U.S.A. (Green Biologics web, 2016).

The main issues that obstruct industrial butanol production are lack of cheap and abundant feedstocks and low values of significant process parameters such as butanol productivity, yield and concentration. Particularly, a low maximum achievable butanol concentration (about

Abbreviations: ABE, aceton-butanol-ethanol; ANI, average nucleotide identity; dDDH, digital DNA-DNA hybridization; EB, ethidium bromide; EMS, ethyl ester of methanesulfonic acid; FBA, flux balance analysis; FC, flow cytometry; GENRE, genome-scale network reconstruction; GGDC, genome to genome distance calculator; GSM, genome scale model; HSP, heat shock protein; m6A, 6-methyl adenosine; m4C, 4-methyl cytosine; m5C, 5-methyl cytosine; MTase, methyl transferase; NGS, next generation sequencing; NTG, *N*-methyl-*N*'-nitro-*N*-nitrosoguanidine; PI, propidium iodide; Rex, redox responsive repressor; RM, restriction modification; SMN, stoichiometric metabolic network; SMRT, single molecule real-time technology; TU, transcription unit; TF, transcription factor

* Corresponding author.
E-mail address: petra.patakova@vscht.cz (P. Patakova).

https://doi.org/10.1016/j.biotechadv.2017.12.004
Received 25 August 2017; Received in revised form 5 December 2017; Accepted 12 December 2017
Available online 15 December 2017
0734-9750/

20 g/L) represents a critical bottleneck that is apparently associated with the poor ability of clostridia to tolerate high butanol concentrations. Since some bacteria are able to tolerate much higher butanol concentrations, such as *Pseudomonas putida* (up to 60 g/L – Ruhl et al., 2009) or *Lactobacillus amylovorus* (40 g/L – Liu et al., 2012) it does not seem unreasonable to also strive for improvement in solventogenic clostridia. However, a general understanding of butanol tolerance in clostridia is still limited and it seems that butanol tolerance may not be in total accord with the butanol stress response and/or high levels of production of butanol. On the one hand, the butanol stress response studied in *Clostridium acetobutylicum* ATCC 824 (pGROE1), a mutant with over-expressed *groESL* operon genes (coding for heat shock proteins) and its control plasmid strain, elicited differential expression of about 200 genes (Tomas et al., 2004) but on the other hand, a butanol tolerant mutant SA-1 derived from parental *Clostridium beijerinckii* NCIMB 8052, differed from the original strain in only a few genes (Sandoval-Espinosa et al., 2013). It might seem that butanol tolerance and butanol production are two distinct issues that are not tightly associated and some studies (Zhao et al., 2003; Liu et al., 2013), in which very tolerant but poor butanol producers were obtained, support this idea. However, from random mutagenesis studies, including gene shuffling, butanol tolerance represents a simple, irreplaceable selection criterion and many high butanol producing strains (Xue et al., 2012; Li et al., 2016c) were obtained based on butanol tolerance.

Although, some reviews have been published (Liu et al., 2017; Peabody and Kao, 2016) dealing with a similar topic, we believe our review brings a new, intriguing insight into the problem. Most studies dealing with solventogenic clostridia were performed using the model strain *C. acetobutylicum* ATCC 824, although this situation is changing now and other strains are being studied (Poehlein et al., 2017). In this comparative work, mainly species *C. acetobutylicum*, *C. beijerinckii*, *Clostridium pasteurianum* and *C. saccharoperbutylacetonicum* are referenced as examples of solventogenic clostridia. The main goals of the review are to put together currently fragmented knowledge, to stress the significance of bioinformatics tools for deeper understanding of the topic, and to revive forgotten attainments and to link them with new information. The review is intended to search for common features of butanol tolerant and high producing clostridial phenotypes which might suggest a way to obtain a "win-win" solution i.e. a highly productive, highly tolerant clostridial strain.

2. Mechanisms increasing butanol tolerance in clostridia

Throughout the years of ABE fermentation studies, many mechanisms were observed, described or hypothesized about how solventogenic clostridia cope with butanol stress elicited either by formation or by deliberate addition of butanol. The complex response includes modification of the cell membrane and cell wall, changes in transport and cell motility/chemotaxis, induction of efflux, formation of stress proteins, accumulation of protective compounds in cells, formation of capsules and preference of growth in the form of a biofilm, and probably others. These phenomena (for their overview see Fig.1), which were studied at different levels and to different depths, are discussed in this section. Results of tolerance assays may reflect a culture history and the design of the particular assay and, therefore, they are often comparable only when this is taken into account (Zingaro et al., 2013). In fact, they may be planned as survival or growth studies, but even within these two categories, different results can be obtained depending on culture density, age, previous adaptation or shocks, stage of life cycle, etc. Comparison of the results set testing different types of solvent tolerance assays was provided for *E. coli* (Zingaro and Papoutsakis, 2012).

2.1. Cell wall and cell membrane modifications

The cell membrane represents a primary target especially for external butanol action. In sub-lethal concentrations (usually from 5 to 15 g/L), butanol initiates a complex stress response that includes membrane modifications with the aim of stabilizing the cell membrane and to prevent its damage. This stress response may include an increase in membrane fluidity, a change in the ratio of saturated/unsaturated fatty acids in membrane lipids, the incorporation of cyclopropane ring containing fatty acids and an increase in plasmalogen content in membrane lipids. Lethal, strain specific, butanol concentrations (in general above 15 g/L) result in impairment of membrane transport functions, inability to generate ATP at the membrane level and dissolution of membrane lipids, which together, cause cell death. Most of these mechanisms were observed in solventogenic clostridia. An increase in the ratio of saturated/unsaturated fatty acids was detected during butanol production or after butanol challenge in *Clostridium acetobutylicum* (Vollherbst-Schneck et al., 1984; Lepage et al., 1987). In contrast, no significant increase in the ratio occurred in *Clostridium beijerinckii* NRRL B-598 (formerly *C. pasteurianum* NRRL B-598 – see Sedlar et al., 2017, 2015) under similar production or addition experiments (Kolek et al., 2015) and surprisingly, production of unsaturated fatty acids was completely stopped by *C. beijerinckii* ATCC 51743 (which should be identical with *C. beijerinckii* NCIMB 8052 based on ATCC information) under conditions of butanol stress (Huffer et al., 2011). Cyclopropanation of unsaturated membrane fatty acids, which may be another mechanism to decrease unsaturated fatty acids in the cell membrane, was found in both *C. acetobutylicum* (Lepage et al., 1987) and *C. beijerinckii* NRRL B-598 (Kolek et al., 2015). An increase in membrane fluidity during butanol production or after butanol challenge was detected in *C. acetobutylicum* (Baer et al., 1987). Complex homeoviscous response of the cell membrane was described during butanol formation by *C. pasteurianum* ATCC 6013 (Venkataramanan et al., 2014) which included an increase in lipid tail length and a decrease in unsaturated fatty acids in the membrane. A probable protective effect of membrane plasmalogens against butanol toxicity was shown for the first time in *C. beijerinckii* NRRL B-598 (Kolek et al., 2015), despite the fact that plasmalogens, diacyl phospholipids containing an alk-1′-enyl ether linked hydrocarbon chain in position sn-1 of glycerol, or glycerol glycolipids with the same ether bond, were found in strictly anaerobic bacteria, including clostridia, tens of years ago (e.g. Kamio et al., 1969; Johnson and Goldfine, 1983). Cardiolipin synthesis in the cell membrane was also found to be involved in butanol resistance in *Escherichia coli* (Reyes et al., 2011). In solventogenic, butanol-producing bacteria, it is necessary to distinguish between the effect of internal butanol occurring during the production phase and external butanol i.e. extracellular butanol contained in the cultivation medium. It seems that butanol added to the cultivation medium, i.e. originally extracellular butanol, induces a much faster response compared to internal butanol i.e. butanol produced by a cell (Kolek et al., 2015).

Although it seems that the lipidic part of the cell membrane is the most sensitive component due to its solubility in butanol, there are other studies (Bowles and Ellefson, 1985; Wang et al., 2005) that emphasize the effect of butanol on membrane proteins involved in the maintenance of internal pH; these are important for butanol tolerance in *C. acetobutylicum* and *C. beijerinckii*. Butanol caused an inhibition of H^+-ATPase in *C. acetobutylicum* (Bowles and Ellefson, 1985) while the Na^+/H^+ antiporter was probably affected by butanol in *C. beijerinckii* (Wang et al., 2005). Moreover, the basis of high butanol tolerance exhibited by a mutant of *C. beijerinckii* BR54 consisted of a higher level of glycation occurring in membrane proteins of the mutant compared with its parental strain *C. beijerinckii* NCIMB 8052 (Wang et al., 2005). Differential analysis of membrane proteins performed for *C. acetobutylicum* DSM 1731 and its butanol-tolerant mutant revealed the main changes in transporter proteins including ion transporters, ATP-ases and structural proteins involved in coat and flagella formation (Mao et al., 2011). Similarly, in other microorganisms, altered antiporter proteins were found to be involved in butanol resistance such as the K^+/H^+

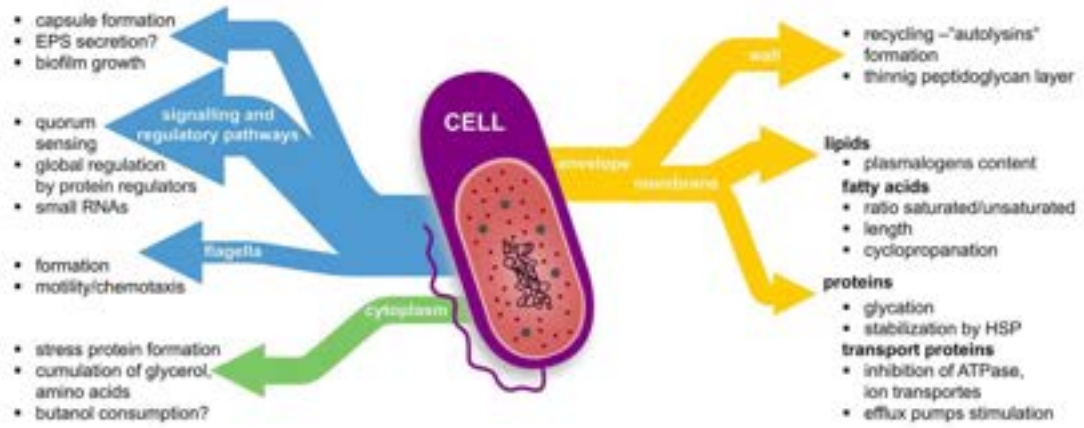


Fig. 1. Overview of butanol response in solventogenic clostridia.
(Overall cell response is highlighted in blue, cell envelope response in yellow and cytoplasmic response in green.) (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

antiporter in *Methylobacterium extorquens* (Hu et al., 2016).

Some of the above described mechanisms of cell membrane modification were used to obtain a butanol tolerant clostridial phenotype. A simple experiment in which elaidic and oleic acids were added to the special cultivation medium and incorporated into the cell membrane of *C. acetobutylicum* ATCC 824 (Baer et al., 1987) did not result in increased butanol tolerance. However, insertion of COE1-5C (a five-ringed oligo-polyphenylene-vinylene conjugation oligoelectrolyte) into the cell membrane of *E. coli* K12 prevented membrane fluidization and depolarization, resulting in increased butanol tolerance (Hinks et al., 2015). Over-expression of cyclopropane fatty acid synthase gene (*cfa*) in *C. acetobutylicum* ATCC 824 resulted in a very tolerant but almost butanol non-productive, mutant phenotype (Zhao et al., 2003). Furthermore, butanol tolerance was studied and engineered by different methods in other, better understood bacteria. In *E. coli*, different genes of the fatty acid operon were overexpressed and *fabD* overexpression was found to be best fitted for the purpose (Bui et al., 2015).

In many bacteria, the cell wall is not a rigid structure but an ever changing element that undergoes constant turnover and recycling. This phenomenon is better described and understood in Gram-negative bacteria, especially *E. coli*, but probably exists also in *C. acetobutylicum* (Reith and Mayer, 2011a). Enzymes, involved in the cell wall recycling activity in *C. acetobutylicum*, such as *N*-acetylmuramic acid/*N*-acetylglucosamine kinase (Reith et al., 2011) or glucosamine/glucosaminide *N*-acetyltransferase (Reith and Mayer, 2011b) were originally referred to as autolysins. Autolysis of part of the *C. acetobutylicum* ATCC 824 population, which has frequently been described during stationary phase in bioreactor batch cultivation, might actually be a controlled process focused on generation of higher spore numbers (Liu et al., 2015). If part of the population is sacrificed and autolysed in a controlled way, the chance of the remaining cells, which may profit from nutrients released during the autolysis, to finish sporulation is higher compared to a non-autolytic population (Liu et al., 2015). Nevertheless, activities of enzymes involved in cell wall recycling and/or autolysis in clostridia were also associated with butanol tolerance (Webster et al., 1981; Van Der Westhuizen et al., 1982; Croux et al., 1992). During industrial fermentations carried out using molasses as a substrate, *C. acetobutylicum* P262 (later reclassified as *Clostridium* sp., see Johnson et al., 1997) cells having a typical swollen clostridial shape were found to be susceptible to degeneration (loss of solvent forming ability) and autolysis under higher butanol concentrations (in the range 7–16 g/L). On the contrary, the *lyt-1* mutant of the same strain (autolysis deficient mutant) never exhibited degeneration and autolysis under the same conditions (Van Der Westhuizen et al., 1982). Moreover it seems that a thinning peptidoglycan layer of *C. beijerinckii* NRRL B-598 (formerly *C. pasteurianum*) might be associated with increased butanol tolerance (Linhova et al., 2010).

2.1.1. Methods for studying the effect of butanol on the cell envelope

A powerful tool for imaging/analysis of a whole range of functional membrane properties is the application of specific fluorescent probes, which, in connection with flow cytometry (FC) or fluorescence microscopy, enables studies of trans-membrane transport, membrane potential or permeability changes at the single cell level (Müller and Nebe-von-Caron, 2010; Tracy et al., 2010). The ability of cells to actively pump various substances from the intra- to the extracellular environment could be measured by ethidium bromide (EB) (Arioli et al., 2014; Czechowska and van der Meer, 2012), a fluorescent probe carrying only a single delocalized positive charge that enables EB to partly penetrate the intact cell membrane. Once inside the cells, it is immediately pumped out (Shapiro, 2005). A decrease or loss in unspecific pump activity leads to accumulation of EB within the cells (Jernaes and Steen, 1994), intercalation into nucleic acid and an intense increase in its fluorescence emission. More highly charged propidium iodide (PI) should not penetrate intact cell membranes and accumulate within the cells. Only where there is an overall loss of membrane integrity and functionality, PI can cross the cell membrane, resulting in increased red fluorescence. Because cells with such compromised membranes are considered dead, PI is utilized much more frequently in viability studies (Gallazzi et al., 2015; González-Peñas et al., 2015; Patakova et al., 2014) than specific investigations of membrane damage. Energetic changes in a membrane caused by butanol or other membrane altering agents can be detected using cationic and anionic slow response potentiometric probes. Cationic rhodamine 123 or carbocyanine dyes and anionic dyes represented by oxonols are used for the detection of membrane hyperpolarization or depolarization respectively (Shapiro, 2000). Application of both types of potentiometric dyes in analysis of bacteria from the genus *Clostridium* was introduced by Linhova et al. (2012), where anionic bis-oxonol was found to be the most convenient.

Despite the wide variety of options for fluorescence analysis, there are a very limited number of studies dealing with solventogenic clostridia. The first attempts to apply fluorescent probes were published by Jones et al. (2008) and Tracy et al. (2008) but the fluorescence staining pattern of different physiological states did not correspond to theoretical expectations. On the other hand, it was found to be a valuable tool for the recognition of different morphological states. FC-based methodology for the functional analysis of cell viability and particular cell features in cultures of solventogenic clostridia was presented very recently (Linhova et al., 2012; Kolek et al., 2016a). A combination of enzymatic and fluorescence assays was used for studying the effect of butanol on inner and outer membranes of *E. coli* and, in both cases, a concentration specific membrane leakiness was found (Fletcher et al.,

2016).

Alterations in cell walls under butanol stress have been studied much less compared to the cytoplasmic membrane. Direct changes in cell shape were observed microscopically e.g. by Fletcher et al. (2016) who described elongation and filamentous growth of *E. coli* under butanol stress. A specific fluorescent alternative to Gram staining suggested the possibility of peptidoglycan thinning during solvent production (Linhova et al., 2010). Electron microscopy offers a greater insight into the synthesis and composition of the peptidoglycan complex and a detailed analysis of the peptidoglycan sacculus. Similarly to studies of the butanol effect on cell membranes, genetic approaches can be utilized e.g. overexpression of genes related to peptidoglycan biosynthesis (Yuan et al., 2014) or transcriptomic analysis (Tracy et al., 2010; Winkler and Kao, 2011). A complex description of the engineering of cell membranes and cell walls in a selection of bacteria in order to obtain solvent tolerant strains was presented by Sandoval and Papoutsakis (2016).

2.2. Stress proteins production

Stress generated by butanol challenge or even butanol production is similar in many ways to that caused by heat shock, and thus it may be also associated with the production of various heat shock proteins (HSP) involved in different housekeeping functions such as protein folding, membrane transport, synthesis and degradation. Increasing expression of heat-shock proteins during butanol challenge experiments has been shown previously in an *E. coli* model (Rutherford et al., 2010) as well as over-expression of several HSPs variants directly leading to overall higher butanol tolerance. A strain over-expressing the GroESL system of chaperones showed faster growth in the presence of 0.75% (v/v) butanol and nearly a 4-fold increase in the number of surviving colonies when cultured in the presence of 1% (v/v) butanol (Zingaro and Papoutsakis, 2013). Heterologous expression of the *C. acetobutylicum* GroESL system in *E. coli* also led to significantly higher butanol tolerance (Abdelaal et al., 2015).

Pich et al. (1990) found that the transition from acidogenesis to solventogenesis in *C. acetobutylicum* DSM 792 (ATCC 824) was accompanied by the production of specific HSPs (hsp 73, hsp 72 [DnaK], hsp 67 [GroESL], hsp 17 and hsp 14). Upregulation of groES, dnaKJ, hsp 18 and hsp 90 was observed in *C. acetobutylicum* ATCC 824 (Tomas et al., 2004) after butanol challenge. Overexpression of the autologic groESL operon in *C. acetobutylicum* resulted in both higher butanol tolerance and production (Tomas et al., 2003) but also revealed the complexity of the cell's response both in the overexpression of the operon and in plasmid insertion. This complexity complicated finding an unambiguous answer as to which chaperon(s) played the most important role in butanol tolerance. The major changes in gene expression included upregulation of motility and chemotaxis genes and downregulation of major stress response genes while host-plasmid interactions, in this case, resulted in different growth and sporulation patterns. Recently, it has been found (Jones et al., 2016) that overexpression of the GroESL chaperon system, resulting in higher butanol tolerance and production, and can be triggered by overexpression of regulatory 6S RNA in *C. acetobutylicum* ATCC 824. Over-expression of HSPs, GrpE and HtpG also led to significantly higher butanol tolerance in *C. acetobutylicum* ATCC 824 (Mann et al., 2012).

GroESL and DnaK, originating from an extremophilic bacterium *Deinococcus wulumuqiensis*, were also successfully over-expressed in *C. acetobutylicum*; these manipulations led to higher butanol tolerance and up to a 2-fold increase in butanol production, with nearly no acid production (Liao et al., 2017).

2.3. Efflux pumps

Butanol excretion by efflux pumps is a topic that has received attention recently. Small uncharged molecules, such as alcohols, were assumed to be excreted from microbial cells through a diffusion process. However, recent transcriptome-based methods revealed that this is probably not the case and surprisingly many genes, probably involved in both influx and efflux from/to microbial cells, were found (Kell et al., 2015). It is likely that engineering of membrane transporters could increase yields and productivity of desired small molecules and that bioengineers should not be focused only on optimization of biosynthetic routes but should also be interested in the problem of how to transport a product from inside the cell to the extracellular space (Boyarskiy and Tullman-Ercek, 2015). Currently, several families of efflux systems have been recognized based on their different molecular architectures, energy sources as well as target molecules: ABC (ATP-binding-cassette), MFS (major-facilitator-superfamily), SMR (small-multidrug-resistance), MATE (multidrug and toxic compound extrusion) and RND (resistance-nodulation-division). RND membrane pumps seem to be of special importance for solvent efflux and tolerance (Mukhopadhyay, 2015).

Significantly increased expression of several efflux related genes were identified in *E. coli* during butanol challenge, confirming that they likely participate in butanol resistance. Formate transporter *focA* was even able to directly increase butanol resistance when over-expressed (Reyes et al., 2011), indicating that it could recognize butanol as a substrate. In contrast, when 43 efflux pumps with described capabilities for solvent extrusion were heterologously expressed in *E. coli*, they were found to be inefficient with short chain alcohols, including butanol (Dunlop et al., 2011). More recently, Fisher et al. (2014) proved that the multidrug efflux pump AcrB, belonging to the RND family, was able to excrete butanol efficiently when a single amino acid was changed in this protein. Furthermore, Reyes et al. (2013) showed that there was probably an antagonistic relationship between butanol resistance systems and resistance to ethanol or isobutanol. This could be caused mainly by a limited maximum number of different efflux pumps that can be incorporated into the membrane, because the more efflux pumps the poorer the growth. To overcome stress generated by overexpression of the AcrB efflux pump (a variant Acrbv2 secreting butanol was developed by Fisher et al., 2014), a negatively responsive promoter was used in *E. coli* to create a controlling mechanism useful for balancing growth, production and butanol efflux (Boyarskiy et al., 2016). In silico, a model of efflux pump expression was constructed, controlled by a feedback loop with a biosensor sensing the level of compound to be excreted (Harrison and Dunlop, 2012). This model showed that feedback regulation is more favourable for growth and production than constant expression of efflux pumps under any condition.

Over the last few years, it has been shown that multidrug resistance efflux systems, belonging to all known families including RND, are also widespread in Gram-positive bacteria including clostridia (Schindler and Kaatz, 2016). Homologues of AcrB transporter, which probably has a slightly different composition because of dissimilar architectures of the cell wall and membrane, could be found in clostridia and other Gram-positive bacteria (Dehoux et al., 2016; Murakami et al., 2002). Thus it is possible that these proteins play a role in butanol resistance and production.

2.4. Other mechanisms involved in butanol tolerance

Although this is certainly not a desirable feature in solventogenic, butanol producing clostridia, in *Clostridium ljungdahlii*, two butanol dehydrogenases were described that could enable butanol consumption by this bacterium through its oxidation to butyric acid (Tan et al., 2014). Regarding the close relatedness between butanol dehydrogenase from *C. ljungdahlii* and alcohol dehydrogenase from *C. acetobutylicum*, coded by the *adh1* gene (Youngleson et al., 1989), and multiple alcohol dehydrogenases with incompletely understood functions described in solventogenic clostridia (Dai et al., 2016; Chen, 1995), it cannot be excluded that in specific cases solventogenic clostridia may consume butanol as a substrate. In addition, the ability to grow on butanol as a sole carbon substrate was found in *Bacillus butanolivorans* and

Enterobacter sp. (Kuisiene et al., 2008; Veeranagouda et al., 2006).

C. acetobutylicum ATCC 824 cells tended to accumulate glycerol and certain amino acids inside their cells during butanol production, which correlated with middle exponential and early stationary growth phases during batch cultivation (Wang et al., 2016). Glycerol accumulation might be explained by the need to maintain osmotic and redox balances in the cells while an increase in the content of amino acids (namely threonine, glycine, alanine, phenylalanine, tyrosine, tryptophan, aspartate and glutamate) might be caused by re-direction of metabolic flux from glycolysis to the TCA cycle. Inhibition of certain glycolytic genes by butanol has been described previously (Alsaker et al., 2004; Tomas et al., 2004) and increased activity of the TCA cycle might correspond to a greater need for NADH for glycerol formation when glycolysis was partly inhibited. Moreover, Lianage et al. (2000) observed that a decrease in glycerol dehydrogenase activity (which may be reversely correlated with glycerol formation) caused by down-regulation of the *gldA* gene by antisense RNA in a mutant of *C. beijerinckii* resulted in a butanol tolerant phenotype.

3. Regulatory pathways involved in butanol stress/tolerance/production and methods of their analysis

Solventogenic clostridia use the Embden-Mayerhof-Parnas glycolytic pathway for processing glucose to pyruvate, which is the starting compound for both acids and solvents formed during acidogenesis and solventogenesis. From the point of view of energy gain, acidogenesis is more profitable because it results in a gain of 4 ATP per mole of glucose in comparison with 2 ATP per mole of glucose if solvents are formed from glucose in solventogenesis. Since part of solvents are generated from acids and the related pathways do not generate ATP, the actual energy profit of solventogenesis is even lower.

The scheme showing ATP and NAD(P)H generation by ABE fermentation was published by Ventura et al. (2013) and a detailed description of all enzymes in *C. acetobutylicum* involved in ABE fermentation was described by Gheslaghi et al. (2009). Changes in intracellular concentrations of ATP/ADP and/or NAD(P)H/NAD(P) during whole ABE fermentation were determined by several research groups (Bowles and Ellefson, 1985; Meyer and Papoutsakis, 1989; Grupe and Gottschalk, 1992; Zhao et al., 2016) under different conditions. On one hand, acidogenesis results in generation of surplus ATP; on the other hand, an increased concentration of acids and the resulting pH drop require increased activity of H^+-ATpase to maintain constant internal pH, which may alter the ATP/ADP ratio. However, it seems that the NAD(P)H/NAD(P) ratio is equally important as that of ATP/ADP for the shift from acidogenesis to solventogenesis. In the past, different models of the metabolic shift based on these regulators were constructed (Grupe and Gottschalk, 1992; Girbal and Soucaille, 1994; Girbal et al., 1995) and current studies (Ventura et al., 2013; Wietzke and Bahl, 2012) bring new evidence for this. Double overexpression of *pfkA* and *pykA* genes of *C. acetobutylicum* (Ventura et al., 2013) resulted in elevation of both ATP and NADH pools together with higher butanol production and tolerance. Redox responsive repressor (Rex) seems to be responsible for the switch to solventogenesis based on a dependence on the NADH/NAD ratio (Wietzke and Bahl, 2012). There is also an increased demand for energy in the form of ATP for many processes involved in butanol stress response, such as re-modelation of cell membrane, biosynthesis and functioning of HSPs, or synthesis of efflux pumps (for a detailed description, see Nicolaou et al., 2010).

Complete regulatory and signalling pathways involved in metabolism, sporulation and stress responses of solventogenic clostridia have not been fully described. By genome mining and studies of individual solventogenic strains, four global transcription regulators, Spo0A, CodY, CcpA and Rex, which are known in many other bacterial genera, were confirmed in 44 solventogenic strains (Poehlein et al., 2017). Spo0A is a master regulator for sporulation and is often associated with the trigger for solventogenesis, although this is now in question. In *Bacillus subtilis*, Spo0A is a global regulator (Molle et al., 2003) and is involved not only in sporulation but also in transcriptional regulation of genes for chemotaxis/motility, efflux pumps, enzymes and DNA replication machinery (in total, 103 genes). In solventogenic clostridia, it remains to be unambiguously clarified how Spo0A is involved in transcriptional regulation. CodY is a stationary phase regulator, identified in clostridia but studied in other bacteria such as *Bacillus* (Belitsky and Soneneshein, 2013), having mostly a repressor effect on the transcription of many genes. In *C. acetobutylicum* ATCC, catabolite control protein A (CcpA) regulator plays a role in the transport and metabolism of saccharides and other substrates, sporulation and solvent production (Ren et al., 2012). Rex seems to be involved in hydrogen production, oxygen tolerance, NAD synthesis, fermentative metabolism and others (Zhang et al., 2014); moreover it probably functions as butanol production repressor, see Section 4.2. Recent GC–MS analysis of intracellular metabolite concentrations during ABE fermentation in *C. acetobutylicum* confirmed tight metabolic regulation and identified key metabolites and metabolic pathways responsible for growth and solvent production (Liu et al., 2016b). Two novel potential regulatory proteins, encoded by genes SMB_G1518 and SMB_G1519, were identified in *C. acetobutylicum* DSM 1791 (Jia et al., 2012). Disruption of these genes resulted in an increase in butanol tolerance, cell motility and increased activity of pyruvate: ferredoxin oxidoreductase, which may imply potential regulatory functions of these genes in butanol tolerance/sensitivity.

Generally, sporulation is considered to be a specific type of stress response. In solventogenic clostridia, sporulation is probably an answer to acid stress, which paradoxically results in a different type of stress – butanol stress because it is often accompanied by a metabolic switch i.e. solvent production. Although a proportion of the acids formed is consumed during solvent formation, the remainder contribute to the creation of overall metabolic stress (Alsaker et al., 2010; Wang et al., 2013). The *Clostridium* sporulation cascade consists of sporulation-specific transcription and sigma factors (Al-Hinai et al., 2015), but their relationship to butanol production is in question and is currently under investigation. Microarrays still maintain their key role in the analysis of sporulation regulators and software are available for microarray data analysis, including identification of differentially expressed genes, cluster analysis for finding co-regulated genes and identification of transcription factor binding sites. Instructions for the whole microarray process along with a list of available software can be found in Olson (2006) and Selvaraj and Natarajan (2011) respectively. Microarray technology was used for transcriptional analysis of *C. acetobutylicum* 824 (pMSPOA), a strain with overexpressed *spo0A* factor (Alsaker et al., 2004). The transcriptional analysis was performed against a plasmid control strain and against the *spo0A* knockout strain (SKO1) with no solventogenesis or sporulation activity. The results showed that even though *spo0A* overexpression led to increased butanol tolerance and to rather major changes in the transcriptional program, solvent production stayed essentially unchanged. Another more recent study (Zhang et al., 2017) utilizing microarrays compared the transcriptional profiles of two *C. beijerinckii* NCIMB 8052 strains, where one of the strains (*C. beijerinckii* DG-8052) was degenerated, i.e. without the ability to produce solvents or sporulate. The results showed that morphological and physiological changes in the degenerated strain DG-8052 were related to altered expression of sigma factors and it seems that they may be involved not only in sporulation but also in the stress response of the cells to acetate and butyrate, with a potentially major role of *sigH* (coding for RNA polymerase sigma factor −70). However, a great number of genes responsible for different functions, such as phosphotransferase transport system, sugar metabolism, motility, sporulation, and solventogenesis in the degenerated strain were differentially expressed and thus unambiguous evaluation is difficult. A different approach, confirming a crucial role of regulatory elements in butanol production by clostridia, was carried out by Sandoval et al. (2015), who performed random mutagenesis and directed evolution selection on *C.*

pasteurianum ATCC 6013. This led to the creation of a mutant (M150B) with substantially increased butanol production. After further analysis, it was discovered that the mutant was *spo0A* deficient, which led to higher tolerance of crude glycerol (which was used as carbon source) and also higher solvent production. This reinforcing influence of *spo0A* deletion on butanol production in *C. pasteurianum* was further verified by Schwarz et al. (2017).

Cell to cell communication probably also plays a role in butanol production and tolerance. Quorum sensing is the production of small, diffusible molecules that are recognized by cells in culture and at specific concentrations, these signalling molecules are able to change gene expression through transcription factors. A pioneer study in solventogenic clostridia described the potential induction of the *sol* operon and initiation of sporulation in *C. saccharoperbutylacetonicum* N1-4 (Kosaka et al., 2007) by a non-peptide compound, probably involved in regulation by quorum sensing. Further, it was found that the agr quorum sensing system actively modulates expression of sporulation factors in *C. acetobutylicum* (Steiner et al., 2012). These agr-like quorum systems could be found across the firmicutes and is probably a widespread way of controlling gene expression in clostridia (Wuster and Babu, 2008).

The intergenic regions filling the space between coding regions in the genome were previously considered to be without purpose. This presumption changed rapidly with the advancement of techniques like microarrays (Yin and Zhao, 2007) or more recently, RNA-seq (Soutourina et al., 2013), which enable RNome studies. It has become evident that the intergenic regions actually play critical roles in cell regulation through small non-coding RNA (sRNA). Apart from experimental studies, it is also possible to predict sRNA in silico using algorithms such as sRNAscanner (Sridhar et al., 2010), sTarPicker (Ying et al., 2011), or SIPHT (Livny, 2012). Both approaches have their drawbacks; ideally a combination of the two should be used in sRNA prediction studies. A large research study on sRNA in clostridia has been carried out by Chen et al. (2011), who predicted sRNA in 21 clostridial genomes including *C. acetobutylicum* and *C. beijerinckii*. This study has brought new insights into regulatory mechanisms in clostridia, including the role of sRNA in *cis*-regulation of efflux pumps in *C. acetobutylicum* and *C. beijerinckii*. Another study, focused strictly on sRNA regulation in *C. acetobutylicum* ATCC 824 under metabolic stress induced by butanol and butyrate, was carried out by Venkataramanan et al. (2013). Here they identified sRNAs that were expressed or differentially expressed in relation to metabolic stresses, and affect transcriptional (6S, S-box, *solB*) and translational (tmRNA, SRP-RNA) processes. The role of 6S and tmRNA in butanol tolerance was confirmed in the follow-up study by Jones et al. (2016), where regulatory sRNAs were overexpressed in *C. acetobutylicum* ATCC 824. This resulted in increased butanol tolerance, as well as butanol production in the case of 6S overexpression, however, a slight decrease in butanol production accompanied tmRNA overexpression.

4. High butanol production by clostridial strains

In general there are two proven approaches to increase butanol production by clostridia; strain improvement using random or targeted mutagenesis, or a change in cultivation conditions. A short overview of the best clostridial producers obtained by mutagenesis is presented in Table 1. From Table 1, it is clear that the maximum achievable butanol concentration (about 20 g/L) is more or less the same for all ABE fermentation-performing *Clostridium* strains and random mutagenesis was equally as successful as targeted mutagenesis. The latter phenomenon probably reflects the low level of understanding of ABE fermentation and its regulation. For comparison, the best heterologous butanol producers, *E. coli* (Shen et al., 2011) and *Clostridium tyrobutyricum* (Yu et al., 2011), can form up to 15 and 16 g/L, respectively, which are considered to be impressive results.

Table 1
The highest butanol titres achieved by mutants of solventogenic clostridia.

Parental strain	Mutant	Mutagenesis method	Butanol titre	Remark	Reference
Clostridium beijerinckii NCIMB 8052 (originally considered to be *C. acetobutylicum* NCIMB 8052 but later re-identified)	BA 101	*N*-methyl-*N*′-nitro-*N*-nitroso-guanidine (NTG) treatment	19 g/L 21 g/L[a]	Selected for high amylase production [a]After acetate addition to the medium	Formanek et al., 1997 [a]Chen and Blaschek, 1999
Clostridium acetobutylicum ATCC 55025 (Jain et al., 1993)[**]	JB200[#]	Adaptive engineering using cultivation in fibrous bed bioreactor in the presence of butanol	19 g/L 20 g/L[a]	[#]Strain protected by US patent (Yang and Zhao, 2013). [a]	Xue et al., 2012 [a]Jiang et al., 2014
Clostridium acetobutylicum ATCC 55025 (Jain et al., 1993)[**]	HMKO	Deletion in cac3319 gene coding for histidine kinase	18 g/L	–	Xu et al., 2015
Asporogenic strain derived from *C. acetobutylicum* ATCC 4259	GS4-3	NTG treatment plus genome shuffling	20 g/L	Cassava used as substrate	Li et al., 2016c
Clostridium acetobutylicum GX01, natural isolate from Guangxi Province, China					
C. saccharoperbutylacetonicum N1-4 (ATCC 13564)	Delta4abk	CRISPR-Cas9 editing used for *pta* and *buk* double deletion	19 g/L	Slight increase of butanol production, yield and selectivity compared to parent strain	Wang et al., 2017
C. acetobutylicum PJC4BK	BKM19	NTG treatment	18 g/L	–	Jang et al., 2013
A butyrate kinase inactivated mutant of *C. acetobutylicum* ATCC 824 (Green et al., 1996)	PJC4BK (pPta-Bfq[for])	Synthetic small regulatory RNA system for knock-down of *pta* gene expression	17 g/L[a]	[a]	[a]Cho and Lee, 2017
C. pasteurianum ATCC 6103	MBEL_GLY2	NTG treatment	18 g/L	Glycerol was used as substrate	Malaviya et al., 2012

** Patent protected strain, currently deposited in ATCC collection as *C. beijerinckii* ATCC 55025.

4.1. Random mutagenesis for obtaining high butanol producers

Random chemical mutagenesis can be performed by mutagenic agents acting directly because repair mechanisms for damaged DNA are lacking in clostridia (Bowring and Morris, 1985). The most common case is the use of *N*-methyl-*N'*-nitro-*N*-nitroso-guanidine (NTG), an alkylating agent, causing mostly unidirectional random transition mutations between GC and AT base pairs in DNA, see Table 1. Another example of a mutagenic agent acting in a similar way as NTG, is the ethyl ester of methanesulfonic acid (EMS) (Bowring and Morris, 1985; Lemmel, 1985; Jain et al., 1993). EMS use for random mutagenesis achieved a high butanol producing stable asporogenic mutant, *C. beijerinckii* ATCC 55025, formerly *C. acetobutylicum* ATCC 55025 (Jain et al., 1993), which was used as a starting strain for further targeted mutagenesis (Yang and Zhao, 2013; Xu et al., 2015). After random mutagenesis, the strains with desired phenotypes may be selected by a readily recognizable property such as growth on a specific substrate or butanol tolerance, but can also be directly tested for high butanol production.

Transcription of selected genes of the oldest butanol high-producing mutant, *C. beijerinckii* BA101, was compared with its parental strain, *C. beijerinckii* NCIMB 8052 (Shi and Blaschek, 2008) and differences were observed mainly in transcription of genes from the sporulation cassette, genes for PTS transport systems (downregulation in both cases) and genes for motility/chemotaxis (upregulation). All changes in genes transcription were in accordance with described phenotypes.

Chemical mutagenesis can be used to generate a library of mutants to be used for genome shuffling strain improvement (Li et al., 2016c; Gao et al., 2012). Genome shuffling is a whole genome engineering method inspired by natural evolution, which includes construction of a parental library, protoplasting, cell fusion, regeneration of fusants and selection of mutants with desired phenotypes (Zhang et al., 2002; Gong et al., 2009). These steps can be repeated several times provided that the mutants obtained after the first round of genome shuffling are included in the parental library before initiation of a second round and so on. The success of the method depends on careful selection of strains that will be included in the parental library. These strains can be obtained by random mutagenesis, targeted mutagenesis, selection using different criteria etc. The parental library usually also includes the original wild-type strain and its members featuring different, desirable properties such as high butanol tolerance (for butanol production), rapid growth on a selected substrate or high selectivity for butanol production. Genome shuffling together with random mutagenesis is comparable with other strain-improving methods, (see Table 1) and resulted in a high butanol concentration 20 g/L after cultivation of a mutant strain (Li et al., 2016c).

The strain *C. acetobutylicum* JB200 (Yang and Zhao, 2013), which was obtained as a random mutant during cultivation in a fibrous bed bioreactor after butanol shock, was further analysed and compared with a parental strain, an asporogenic mutant *Clostridium acetobutylicum* ATCC 55025. A single-base deletion in the *cac3319* gene, encoding histidine kinase, was found (Xu et al., 2015). The product of the *cac3319* gene, an orphan histidine kinase, was already proven to be able to phosphorylate Spo0A protein during in vitro studies (Steiner et al., 2011), however Xu et al. (2015) hypothesized that the increase in both butanol tolerance and production observed in the HKKO mutant (see Table 1), might be caused by an unknown regulatory mechanism with an important role for this histidine kinase because sporulation and solventogenesis did not seem to be tightly connected in the mutant strain.

It was questionable whether all previous "mutants" obtained during continuous culture under phosphate limitation (Meinecke et al., 1984) or during ABE fermentation (Jones et al., 1982) are true mutants or cells exhibiting different phenotypes based on transcriptional or translational rather than genomic changes. A potential proof that use of traditional evolutionary culture methods may result in true mutants was provided recently (Sandoval-Espinosa et al., 2013). Comparative genome analysis was performed for SA-1, a butanol tolerant and high-production mutant (Lin and Blaschek, 1983) derived from *C. beijerinckii* NCIMB 8052 by serial enrichment in culture medium containing increasing concentrations of butanol. The SA-1 strain was re-sequenced and its genome, together with phenotypic parameters, were compared with the parental strain *C. beijerinckii* NCIMB 8052 (Sandoval-Espinosa et al., 2013). Eight single nucleotide changes were found, one small deletion and one large insertion, however with apparently no direct influence on butanol production and tolerance. Surprisingly, the PTS transport system for fructose was deregulated in the SA-1 mutant, which agreed with a similar downregulation of PTS genes observed for another mutant, BA101, derived from the same strain by NTG mutagenesis (Shi and Blaschek, 2008). Therefore Sandoval-Espinosa et al. (2013) hypothesized a potential connection between the PTS system and other regulatory mechanisms for sporulation and solvent production.

4.2. Targeted mutagenesis for improved butanol production

Tools for targeted mutagenesis and metabolic engineering of clostridia have been developed during the last decade and recent and progressive methods, including CRISPR-Cas genome editing, were successfully applied to clostridial research (Pyne et al., 2016a; Wang et al., 2015, 2017; Li et al., 2016b). Currently, methylome studies (see Section 5.1.1) may improve transformation protocols significantly. *C. acetobutylicum* strain improvement has been recently reviewed by Xue et al. (2017).

The first class of engineered strains with improved butanol production are strains with a directly reinforced butanol biosynthetic pathway. This could be done by over-expression of genes directly involved in butanol synthesis, knock-out of genes involved in synthesis of by-products, or by a combination of these approaches. Unfortunately, it is not easy to completely disrupt formation of by-products by single gene down-regulation as reported previously for acetone (Tummala et al., 2003a; Jiang et al., 2009; Wang et al., 2017) or butyric acid (Green et al., 1996; Harris et al., 2000; Wang et al., 2017). Similarly, increased expression of *adhE* or *ald*, which are directly involved in the last steps of butanol formation in *C. acetobutylicum*, *C. pasteurianum* and *C. beijerinckii* respectively, does not automatically lead to significantly higher butanol production (Harris et al., 2000; Tummala et al., 2003b). All studies have already shown that regulation of butanol production is very tight and that the concentration of precursors, by-products, or values of redox potentials inside cells must be tuned precisely. Cooksley et al. (2012) contributed to an understanding of the ABE biosynthetic pathway in *C. acetobutylicum* by constructing different deletion mutants in this pathway. Even though no high-producing strain was obtained, it was shown that it was possible to decrease acetone production significantly by *ctfA*, *ctfB* and *adc* knock-outs; they also clarified the role of essential genes (*thl*, *hydA*) in which knock-out was lethal. Jang et al. (2012) tested the influence of several single knock-outs of genes involved in by-product synthesis (*pta*, *ack*, *buk*, *ptb*) and their combination in *C. acetobutylicum* ATCC 824. A combination of *buk* and *pta* knock-outs led to the best production of butanol, around 16 g/L. Together with over-expression of *adhE*, the strain was able to produce nearly 19 g/L of butanol. Single and combined *pta* and *buk* mutants were also prepared for *C. saccharoperbutylacetonicum* N1-4 (Wang et al., 2017). Production of 19 g/L of butanol was obtained in the case of a double mutant, but production from a single mutant was nearly same as the parental strain.

The second approach is metabolic engineering of solventogenic regulators. Harris et al. (2001) performed knock-out of the *solR* gene, a putative negative regulator of the solventogenic pathway in *C. acetobutylicum* ATCC 824. Inactivation of *solR* led to the production of 14.6 g/L of butanol compared to 11.7 g/L observed for the parental strain. When this manipulation was accompanied by additional *adhE*

over-expression, butanol production of 17.6 g/L was achieved. Factor Rex also represses expression of butanol synthetic genes based on the intracellular NADH/NAD^+ ratio, and by its deletion, it was possible to increase butanol production in *C. acetobutylicum* ATCC 824 (Wietzke and Bahl, 2012) as well as *C. pasteurianum* DSM 525 (Schwarz et al., 2017). Solventogenesis is most likely also under direct regulation of some sporulation factors and quorum sensing mechanisms (see Section 3) and targeted mutagenesis of these mechanisms could also contribute to finding overproducing phenotypes. An example may be the butanol over-expressing *C. pasteurianum* mutant with *spoA* deletion (Sandoval et al., 2015). Another example is the mutant strain HKKO (Xu et al., 2015) derived from *C. acetobutylicum* ATCC 55025 with deletion of histidine-kinase *cac3319*, which is probably involved in Spo0A activation; the mutant strain produced nearly 45% more butanol with 90% higher productivity.

Over-expression of some factors involved in butanol tolerance can also directly lead to higher production, as reported in the case of over-expressing the gene encoding heat-shock protein *groESL* in *C. acetobutylicum* ATCC 824 (Tomas et al., 2003). Maximal butanol concentration achieved with this strain was 17.8 g/L, compared to 12.5 g/L in the control strain. An overall increase in ATP and NADP levels in cells is the next possible way to improve butanol production. Ventura et al. (2013) performed over-expression of 6-phosphofructokinase and pyruvate kinase. Only a small increase in butanol production occurred in the case of a single over-expressing mutant, however a 29% increase was observed in a double over-expression mutant.

4.3. Culture methods for improvement of butanol production

Many reviews (Jones and Woods, 1986; Tashiro and Sonomoto, 2010; Lee et al., 2008d; Patakova et al., 2013) show that the influence of culture conditions on butanol production is highly significant and that parameters such as medium pH, supplementation with acids, overall composition of culture medium and cultivation methods are probably the most important factors affecting concentration, yield, selectivity and productivity of butanol. In this connection, it is necessary to stress that a change in culture medium may result in an increase in butanol production comparable with both targeted and random mutagenesis. An example is cultivation of *C. saccharoperbutylacetonicum* N1-4 in which butanol concentration was usually a maximum of 16 g/L (Tashiro et al., 2004) but after a change of cultivation medium and conditions (Wang et al., 2017), 18 g/L of butanol was reached without any strain engineering. A similar example maybe acid supplementation of the cultivation medium, which usually also improves the butanol concentration (e.g. Chen and Blaschek, 1999; Wang et al., 2013a). Our poor understanding of how different external factors can influence butanol production can be demonstrated by recent studies dealing with the effect of calcium and zinc on solvent production (Wang et al., 2016; Wu et al., 2015, 2016b; Han et al., 2013) in both *C. acetobutylicum* and *C. beijerinckii*. Calcium in the form of calcium carbonate not only has a buffering effect but improves glucose transport, increases butanol tolerance and has a positive effect on solventogenesis (Han et al., 2013).These effects are synergistic with those observed for sole addition of zinc (Wang et al., 2016). In addition, genome-wide transcriptional analysis for *C. beijerinckii* NCIMB 8052 (Wang et al., 2013a) revealed that addition of sodium butyrate (40 mM i.e. 4.4 g/L) to the cultivation medium resulted in an earlier onset of transcription of solventogenic genes, delayed and down-regulated transcription of sporulation genes and prolonged transcription of motility genes. All transcriptional events were in accord with observed phenotypic traits. Wang et al. (2013a) speculated about the potential role of butyryl phosphate in phosphorylation of Spo0A, which might be responsible for improvements in fermentation characteristics after butyrate addition.

Currently, different process optimization methods such as response surface methodology are being used to identify optimal conditions for butanol production using a specific clostridial strain and selected feedstock. This approach may result in process improvement, e.g. 17 g/L of butanol was achieved from corncob hydrolysate using *C. acetobutylicum* NCIM 2877 (Kumar et al., 2014) or 18 g/L using synthetic medium and a mixed *Clostridium* culture (Cheng et al., 2012).

4.3.1. Attached growth and biofilm formation

There are many reports describing positive effects of "attached growth" or even biofilm formation by solventogenic clostridia on butanol production parameters and butanol tolerance, for a review, see Qureshi et al. (2005). Capsule formation by extracellular polysaccharides was described during industrial butanol production in South Africa in the 1980s (Long et al., 1984) and was sporadically studied later (Junelles et al., 1989), but because this feature is not typical for currently studied solventogenic strains, the topic has not developed much further. Recently, extracellular polymer substance (EPS) and biofilm formation have been described in *C. acetobutylicum* CGMCC 5234 during growth on cotton towels and the phenomenon increased tolerance of the strain to acidic conditions and the presence of butanol, as well as production characteristics (Zhuang et al., 2016). The authors (Zhuang et al., 2016) hypothesized that in submersed liquid cultivation of planktonic cells, acid formation during ABE fermentation induces sporulation while growth in a biofilm associated with EPS formation results in a different type of stress response with limited sporulation. Transcriptional analysis of planktonic and biofilm cells of the same strain (*C. acetobutylicum* CGMCC 5234) revealed different levels of expression for 16% of genes (Liu et al., 2016a) and provided indirect proof for the hypothesis of Zhuang et al. (2016), i.e. that up-regulation of sporulation genes in biofilm cells was delayed compared to planktonic cells and occurred in late stationary phase.

Improvements in butanol production were described several times using packed-bed, biofilm or similar types of bioreactors compared to submersed liquid cultivation. However, both types of cultivation, i.e. using submersed liquid cultivation with planktonic cells and surface (biofilm) immobilization cultivation, using the same strain and the same/similar cultivation conditions were only rarely compared (see Table 2). Lipovsky et al. (2016) compared production of butanol by *C. beijerinckii* NRRL B-598 (formerly *C. pasteurianum* NRRL B-598) in batch, fed-batch and continuous cultivations with surface immobilized and planktonic cells. From this comparison (Lipovsky et al., 2016), it is clear that it is not possible to optimize all fermentation parameters, such as butanol concentration, yield and productivity, at once.

Table 2
Immobilization as a factor improving butanol production by clostridia.

Strain	Immobilization support	Improvement in butanol production compared to planktonic cells	Reference
C. acetobutylicum CGMCC 5234	Cotton towels	12%	Liu et al., 2016a
C. acetobutylicum CGMCC 5234	Cotton towels	27%	Chen et al., 2013
Clostridium pasteurianum DSM 525	Corn stover pieces	25%	Gallazzi et al., 2015
Clostridium beijerinckii NCIMB 8052	Porous polyvinyl alcohol	45%	Lee et al., 2008c
Clostridium beijerinckii opinoii	Rashig ceramic rings	25%	Yang et al., 2016

5. Methods to obtain butanol tolerant and high-producing phenotypes

Thorough understanding of molecular processes in organisms used for biorefinery, including their formal description, is required to engineer butanol tolerant and high-producing phenotypes. Thus, interdisciplinary cooperation, utilization of molecular biological methods, and advanced bioinformatics strategies seems to be irreplaceable in current biotechnological research. Current deep sequencing technologies provide us with far more information than only genome sequence composition. Technologies such as RNA-Seq (Wang et al., 2009) or ChIP-Seq (Park, 2009) are replacing older, less sophisticated technologies of microarrays and ChIP-chip, respectively. Their main advantage is the possibility of performing an experiment without prior knowledge of the genome sequence. Therefore, new genes or DNA-binding sites can be discovered. Moreover, third generation sequencing technologies, specifically Pacific Biosciences RS long-read sequencing (PacBio) utilizing the single molecule real-time technology (SMRT) (Roberts et al., 2013), also sequence methylations on a genome-wide scale, providing epigenome along with whole genome information.

An analysis of cellular responses usually requires a comparative study between closely related strains or species with different levels of production and tolerance. RNA-Seq gene expression profiles of two strains can be easily quantified by reference-free methods such as Kallisto (Bray et al., 2016) or Salmon (Patro et al., 2017). However, even more distant strains can exhibit highly similar expression patterns when placed under the same conditions, where the external stimulus boosts transcription of similar genes and leaves the rest unexpressed. Identification of the slightest, yet possibly important, differences is therefore essential for the correct phylogenetic recognition of the strains.

The advantageous bioinformatics strategy to study butanol tolerance and production in clostridia is shown in Fig. 2 and contains three main blocks. The whole strategy starts with genome sequencing and data mining which provide basic information about the studied strain. This is followed by RNA sequencing and data quantification to reveal relation between various genes and to find important regulatory elements. This also helps to clarify the basic behaviour of the strain. Combining knowledge from both of these blocks then finally leads to genome scale model inference. Such a model allows to identify functional elements and their relations, which can be further used in strain engineering and prediction of the strain's behaviour.

5.1. Genome mining in clostridia

Nowadays, novel phenotypic discoveries are always compared and related to genotype. Accurate genome mining is preferably performed using high quality complete genome assemblies. For bacteria, gaps in draft genome assemblies caused by the inability of next generation sequencing (NGS) to produce long sequencing reads can be relatively easily covered by long reads using the PacBio sequencing platform (English et al., 2012). While some genome assemblies of butanol producing clostridial strains remain as drafts, such as *C. acetobutylicum* GXAS18-1 (Mo et al., 2015), *C. diolis* DSM 15410 (Wang et al., 2013b), there is also a range of complete genomes based on NGS platforms data, such as for *C. beijerinckii* SA-1 (Noar et al., 2014) and *C. pasteurianum* DSM 525 (Poehlein et al., 2015). Most recently, a collection of complete genomes was assembled utilizing NGS together with the PacBio platform on *C. carboxidivorans* P7^T (Li et al., 2015), *C. beijerinckii* NRRL B-598 (Sedlar et al., 2015), and *C. butyricum* JKY6D1 (Li et al., 2016a).

5.1.1. Genome annotation

After a genome is assembled, complete annotation is required to predict genes that can be later identified as involved in butanol production and tolerance. One of the first annotations among butanol producing clostridia was performed using the *Clostridium acetobutylicum*

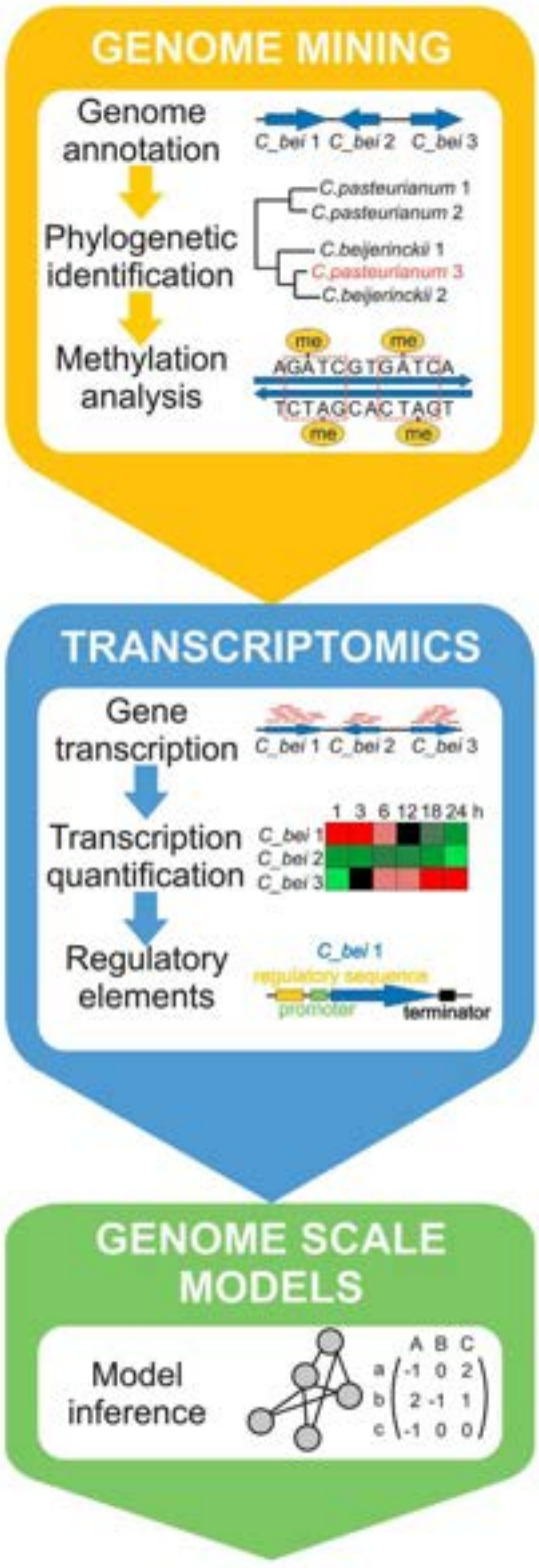


Fig. 2. Bioinformatics analysis scheme.

ATCC 824 genome (Nölling et al., 2001). Although at that time this was a challenging task requiring utilization of custom-made Perl scripts and much manual effort, many genes involved in solvent production were identified in the genome, mainly on a megaplasmid. This gradual enlargement of databases of known genes has led to an expansion of novel automatic tools for annotation that are essential for current research in which bacterial genomes are routinely sequenced. Novel genomes of

butanol producers are therefore processed by a considerable variety of tools for annotating different elements. An example can be the annotation of *C. pasteurianum* DSM 525 which required the use of genomic tools such as Prodigal (Hyatt et al., 2010) for gene prediction, RNAmmer (Lagesen et al., 2007) and tRNAscan-SE (Lowe and Eddy, 1997) for rRNA and tRNA prediction. Preferably, automated pipelines associating particular tools are used for complete annotations of prokaryotic genomes. Among newly published clostridial genomes, NCBI prokaryotic genome annotation pipeline (PGAP) (Tatusova et al., 2016) and RAST (Aziz et al., 2008) are the most popular and were used for annotation of the latest draft as well as the complete genome assemblies described above. Unfortunately, systematic errors can be introduced into novel homology-based annotations. Possibly, pipelines utilizing transcriptomic data, e.g. BRAKER (Hoff et al., 2016), can be used to correct these issues.

5.1.2. Phylogenetic identification

Another step in thorough research of solventogenic clostridia is confirmation of their phylogenetic origins, or its adjustment, respectively. Verified taxonomy helps to plan transformations leading to win-win phenotypes while searching for already described experiments using related strains or species. Unfortunately, many genomes in the GenBank database are misidentified (Federhen et al., 2016), therefore great caution must be taken during phylogenetic analyses. Moreover, clostridia itself represents an extremely heterogeneous genus comprising species that are phylogenetically intermixed with other sporeforming and non-sporeforming genera (Collins et al., 1994). Although the division of the genus based on 16S rRNA gene is given (Stackebrandt et al., 1999), the method has its pitfalls and needs to be supplemented with other, more detailed analyses (Janda and Abbott, 2007). Rather than only 16S rRNA genes, whole genomes can be used for phylogeny inference by phylogenomic analyses. During a GenBank taxonomy workshop (Federhen et al., 2016), average nucleotide identity (ANI) (Richter and Rossello-Mora, 2009) was selected as a simple statistical method for correct taxonomic identification. It can be easily computed by various online and stand-alone tools as ANI calculator (http://enve-omics.ce.gatech.edu/ani/index) and JSpecies (http://www.imedea.uib.es/jspecies) respectively. However, this method also has its caveats. Firstly, a cut-off value for species delineation is species-specific and cannot be generalized and secondly, gaps in draft genomes are not penalized when comparing complete and draft genomes. During re-identification of *C. beijerinckii* NRRL B-598, formerly misidentified as *C. pasteurianum* NRRL B-598 (Sedlar et al., 2017), ANI was supplemented with digital DNA–DNA hybridization (dDDH) (Auch et al., 2010), a technique that replaces the original wet-lab DDH by in silico comparisons. The value was computed using Genome-to-Genome Distance Calculator (GGDC) (Meier-Kolthoff et al., 2013) that allows comparison of complete as well as draft genomes while considering the statistical significance of the output. Additionally, PhyloPhlAn (Segata et al., 2013) was used to construct a phylogenomic tree based on 400 selected protein sequences conserved across the bacterial domain. The final result differed from automatic ANI analysis by NCBI that proposed re-identification as *C. diolis*. Other misidentifications among butanol producing clostridial strains can be expected as they are common across the whole bacterial domain. On the other hand, solventogenic species are usually identified correctly based on their phenotypic traits. Yet, some genomic analyses have led to partitioning of the whole species and re-classification of strains into novel species such as *C. butyricum* (Lawson and Rainey, 2016). Recently, two clades covering most solventogenic strains were proposed by Poehlein et al. (2017). These clades differed mainly in their ability to synthesize the Rnf complex necessary for ferredoxin reduction and formation of pyruvate decarboxylase. Members of clade I included *C. acetobutylicum*, *C pasteurianum* and other related species. Clade II is formed by *C. beijerinckii*, *C. saccharoperbutylacetonicum* and related species. It is peculiar that the division of species and strains based on genome mining follows old, intuitive division based mainly on substrate utilization i.e. clade I - clostridia utilize starch-containing substrates while clade II - clostridia utilize saccharose containing feedstock.

Comparative studies focused on searching specific genes/operons/promotors/DNA binding sites of interest can be performed by genome mining and the results may reveal surprising similarities, differences, rare occurrences or absences in specific traits, which may explain standard or uncommon behaviours of particular strains. For solventogenic clostridia, such comparative studies were performed to compare CRISPR systems (Brown et al., 2014), to envisage the formation of secondary metabolites by cryptic biosynthetic pathways (Behnken and Herweck, 2012), to compare overall features of known, sequenced solventogenic clostridia (Poehlein et al., 2017) and to compare spore germination in clostridia and bacilli (Xiao et al., 2011).

5.1.3. Methylation analysis

Methylation patterns in prokaryotic genomes are important factors for keeping their integrity and are involved in regulation. Methyltransferases (MTases) recognizing these patterns are often connected with restriction-modification (RM) systems protecting the genome. Their detailed analysis helps to design suitable plasmids, which are not destroyed by the host. There are three different types of methylations observable in bacterial and archaeal genomes: 6-methyladenosine (m6A), 4-methylcytosine (m4C), and 5-methylcytosine (m5C), all of which are products of DNA MTase activity (Korlach and Turner, 2012). DNA methylation represents one option for epigenetic gene regulation. The influence of methylation on gene expression has been described extensively for Eukaryotes (Siegfried and Simon, 2010), where methylations are interconnected, especially with chromatin state and repressed gene expression. In bacteria, DNA methylation could be involved in the regulation of many cellular processes including regulation of the cell cycle and initiation of replication, bacterial virulence, DNA repair mechanisms or regulation of gene expression (Casadesús and Low, 2006).

MTases connected with restriction-modification (RM) systems protect host genome from destruction. The drawback of RM systems is that they make engineering of new strains much more difficult, as they recognize plasmids and foreign DNA as a thread and initiate its degradation almost immediately (Murray et al., 2012). Knowing the methylome of the studied organism is therefore essential for developing new techniques for genomic engineering in clostridia, leading to industrially profitable strains. Historically, only m5C methylations were detectable by bisulfite sequencing, and tools were lacking for m6A and m4C methylations, which are dominant in prokaryotic DNA. This changed with the advent of SMRT sequencing, which is able to directly detect all three methylations, although with lower sensitivity to m5C (Clark et al., 2013). In a relatively recent study by Blow et al. (2016) a catalogue of methylations in 230 diverse bacterial and archaeal species was constructed by use of SMRT sequencing, making a rich resource for further research. In relation to *Clostridium* species, SMRT sequencing was used in *C. pasteurianum* ATCC 6013 for detection of methylation patterns to overcome the RM system and allow plasmid transformation to make more tolerant and solvent producing strains (Pyne et al., 2016b). Similarly, Kolek et al. (2016b) developed a new protocol for genetic transformation of *C. beijerinckii* NRRL B-598, formerly misidentified as *C. pasteurianum* NRRL B-598, based on the identified methylation patterns.

5.2. Transcriptomics

After genome mining tasks are done, assumptions for the following genomic analyses can be set up. These usually aim to further explore genes, especially their transcription, under different conditions, e.g. under stress caused by high butanol concentration in cultivation medium. Such an analysis may help to identify genes involved in butanol tolerance.

5.2.1. Transcription of genes

The first high throughput technology utilizable in butanol tolerance research was a technology greatly exceeding the capacity of northern blots - microarrays (Schulze and Downward, 2001). This technology allowed a further description of *C. acetobutylicum* tolerance to butanol by analysis of transcription under butanol stress (Tomas et al., 2004). The main advantage was relatively easy data processing by simple statistical clustering and visualization using tools such as Cluster and TreeView (Eisen et al., 1998). However, to compare gene expression in samples under different conditions, differential expression analysis, a normalization of data prior to comparison, is needed. One such method was proposed using *C. acetobutylicum* strains (Yang et al., 2003). Over the years, many specialized tools for microarray data processing were presented, some of them associated under the TM4 microarray software suite (Saeed et al., 2006) that was used for transcriptional analysis of butanol producing *C. beijerinckii* strains (Shi and Blaschek, 2008). This not only helped to highlight the differences between *C. beijerinckii* NCIMB 8052 and butanol hyper producing strain *C. beijerinckii* BA101 in the expression of primary metabolic as well as chemotaxis genes, but also revealed differences in sporulation between *C. beijerinckii* and *C. acetobutylicum* through an increased rate of transcription of the *sigE*- and *sigG*-regulated genes in the *C. beijerinckii* genome. Although microarrays were recently used in *E. coli* butanol tolerance research to identify membrane-related proteins (Si et al., 2016), the future of this technique in butanol resistance research is questionable for several reasons. The main one is the fact that a microarray has to be designed for a specified set of genes. This is not an issue for the well described *E. coli* genome, but many genomes of solvent producing clostridia are described poorly or are not known completely. Microarrays for them are therefore designed using orthologues that were identified to be important in the *C. acetobutylicum* genome (Shi and Blaschek, 2008). This makes previously undescribed relationships between genes impossible to be detected. Also the dynamic range of the method is limited, therefore the ability to capture a proper level of expression for highly transcribed genes is limited.

The future research of butanol tolerance lies in complete transcriptome sequencing, RNA-Seq (Wang et al., 2009). However, the ability to study transcription of all genes at once over an infinite dynamic range is computationally much more demanding for data preprocessing. Steps typical for sequencing data processing, such as quality trimming and demultiplexing, are needed. Custom made Python or Shell scrips using regular expressions can be easily used as well as specialized tools such as Trimmomatic (Bolger et al., 2014). The main step of data pre-processing lies in mapping reads. STAR (Dobin et al., 2013) or segemehl (Hoffmann et al., 2014) show satisfactory results for non-model organisms. Files with aligned reads are usually large and have to be processed by specialized tools such as samtools (Li et al., 2009). Also, after each step, quality assessment should be done. Fastqc can generate quality reports that can be summarized across all samples by multiqc (Ewels et al., 2016). Although these tools are relatively easy to use, they require a basic knowledge of the command line interface. Possibly, software with a user-friendly environment such as open source Rockhopper (McClure et al., 2013) or commercial CLC Genomics Workbench can be used instead of standalone tools. Once the coordinates of aligned reads are compared to coordinates of genes, a count table summarizing the expression of particular genes can be constructed. Other steps include statistical evaluations similar to those used for microarrays. As a by-product, improvement of previous in silico annotations is possible. Such an example for butanol producing clostridia was presented during transcription profiling of *C. beijerinckii* NCIMB 8052 (Wang et al., 2011).

5.2.2. Quantification of transcription

While RNA-Seq have been already used to identify targets for increasing tolerance of *E. coli* (Rau et al., 2016) to butanol or to describe complex stress responses of *C. acetobutylicum* (Venkataramanan et al., 2015), a question of correct data normalization has arisen. Although the dynamic range of RNA-Seq is theoretically infinite, the higher the sequencing depth, the higher the noise, which leads to false positives (Tarazona et al., 2011). Therefore, increasing the number of sample replicates is advised to improve detection (Rapaport et al., 2013). Special attention should be paid to data normalization for differential expression analysis (Bullard et al., 2010). During transcription profiling of *C. beijerinckii* NCIMB 8052 on a genome-wide scale, a gene encoding peptidase T (Cbei_2428) was selected as a housekeeping gene (Wang et al., 2012). Such genes have a constant level of expression over time and can be used for data normalization. Although the transcriptional level of the gene was verified by qRT-PCR, this kind of normalization is being questioned and use of a spike-in control is advised (Chen et al., 2016). In terms of computational tools, there are a variety of packages and applications for RNA-Seq data normalization and qualification, such as DESeq2 (Love et al., 2014) or edgeR (Robinson et al., 2010). While up to date, RNA-Seq helped to identify genes and pathways involved in butanol production and tolerance, such as in adding butyrate to the cultivation medium of *C. beijerinckii* during fermentation (Wang et al., 2013a), future experiments and novel computational tools will allow refinement of these findings, inference of complete or at least partial gene regulatory networks and the creation of novel genome-scale models.

5.2.3. Transcription units and transcription factors

Transcription units (TUs), as the name implies, are units consisting of a promoter, a transcription start site, a coding region and a transcription terminator. Finding TUs in a cell under different conditions can elucidate the regulatory mechanism of the organism. But because TUs are condition-dependent, it is a challenging task to predict these computationally. For this purpose a novel computational method - SeqTU has been developed, based on the availability of RNA-seq data (Chou et al., 2015; Chen et al., 2017) and its use could advance research on regulatory elements. The utility of computational methods for finding regulatory regions has been proven e.g. by Yang et al. (2017), who studied an important master regulator, catabolite control protein A (CcpA) and its binding sites in several *C. acetobutylicum* and *E. coli* strains and in plasmids. Identification of the binding sites was performed by use of RegPredict web server (Novichkov et al., 2010). The relevance of CcpA in *C. acetobutylicum* in regards to solvent production had been previously proven by Ren et al. (2012), where *ccpA* inactivation affected the three important genes and gene clusters (*sol* operon, *bukII*, and *abrB*) related to solvent production. Finding the motif of CcpA-binding sites and its variability (Yang et al., 2017) revealed that the importance of CcpA could be far greater than previously expected. Yet another study focusing on the importance of transcription factors (TFs) and their binding sites has been published by Wang et al. (2013), where a large microarray experiment was conducted on 96 samples to ascertain the role of TFs and their binding sites in *C. acetobutylicum* ATCC 824 under different levels of butyrate and butanol stress. The identification of differentially expressed TFs under metabolite stress then allowed reconstruction of a stress response network model differentiating between general and metabolic response stresses. These novel findings can be utilized not only to elucidate the behaviour of transcription units under different conditions, but also to synthesize new, more resistant strains by incorporating regulatory-network data in the framework of genome-scale models.

5.3. Genome scale models

Preferably, bioinformatic studies of butanol tolerance and resistance should be conducted with metabolic models based on genome annotation that address the problem of systems complexity and provide us with the formal description of mechanisms standing behind a win-win phenotype. The models, known also as genome-scale metabolic (GSM) models or stoichiometric metabolic network (SMN) models, provide a

global picture of gene-protein relationships and allow investigations of an organism of interest at the robust systems-level by integrating different omics data (Feist et al., 2008; Perez-Garcia et al., 2016). Although there are some limitations to this approach (Cardoso, 2015; Blazeck and Alper, 2010), there are many examples of successful applications of GSM models in research. For instance, they have been used to examine the consequences of deletions, predict the different phenotypes of genetic variants, analyse biological networks and interspecies interactions, and study evolutionary processes (Oberhardt et al., 2009; Liu et al., 2010; Kim et al., 2017). With implementation of this advanced technology, it is also possible to design new metabolic engineering strategies for microorganisms (Jang et al., 2012). However, the main disadvantage of GSM models is that they ignore dynamic behaviours and provide information only about the steady state of balanced metabolites (Santos et al., 2011). In contrast with GSM models, kinetic models are able to predict a rate of all enzymes in the time course. Nevertheless, kinetic models belong to bottom-up approaches of systems biology (Bruggeman and Westerhoff, 2007) and explicitly focus only on a specific isolated metabolic pathway. Since the biomass composition of a metabolic system changes significantly under varying conditions, GSM models incorporating the genome-scale information compute only simplified predictions (Dikicioglu et al., 2015). Most enzymes are not sufficiently explored yet and the required kinetic data on a large scale are not available. Gaining this information can be costly and challenging (Cardoso, 2015). In addition, specific GSM models are only suitable for analysis of a specific organism and cannot be generalized.

Integrating experimental data within a GSM model is a challenging task that involves the compilation of genome-scale network and its transformation to a genome-scale model (Thiele and Palsson, 2010). Genome-scale network reconstruction (GENRE) requires summarizing all known biochemical reactions, represented as a flow of balanced metabolites in a stoichiometric matrix (Feist et al., 2008). The bridge between genome annotation data and the associated reaction fluxes can be compiled using a variety of reaction network databases, e.g. KEGG (Kanehisa et al., 2016). GENRE, computationally transferred in a mathematical format, can then be turned into a GSM model (Feist et al., 2008). Draft model building must be followed by manual curation, validation, and other additional steps (Fondi and Lió, 2015). Various GSM model building and analysis methods and tools have been already proposed for metabolic engineering and creation of in silico mutants (Simeonidis and Price, 2015). Software tools and methods for bacterial metabolic model reconstructions have been reviewed by Thiele and Palsson (2010) or Fondi and Lió (2015).

GSM models can be used in a variety of biotechnology applications, including improvement of biofuel production by microorganisms (Milne et al., 2009; Lee et al., 2008b). Several well-curated GSM models focusing on clostridia have been reconstructed in the last decade. Most of the clostridial genome-scale models have been built for *C. acetobutylicum* ATCC 824 (Lee et al., 2008a; Senger and Papoutsakis, 2008a,b; McAnulty et al., 2012; Dash et al., 2014; Yoo et al., 2015; Salimi et al., 2010; Wallenius et al., 2013), but there have also been published models for *C. beijerinckii* (Milne et al., 2011) and in the case of *C. pasteurianum*, a basic step for formation of a GSM model, a metabolic reconstruction, has recently been performed (Pyne et al., 2016b).

The first genome-scale model reconstructed for clostridia resulted in the formation of an in silico strain with stopped production of acetone or butanol and reduced production of ethanol (Senger and Papoutsakis, 2008a). The model was then coupled with the concept of specific proton flux states to predict medium pH during acidogenesis (Senger and Papoutsakis, 2008b). Genome-scale models serve as a framework allowing the incorporation of various information from different sources – transcriptional regulatory data, associations of genes, proteins and reactions, metabolic flux analysis data, thermodynamic information, and others. Such upgraded models with multi-omics datasets have more promising predictive power. GSM simulations involving the application of flux balance analysis (FBA) can be used to increase understanding of clostridial metabolism and predict the formation of particular metabolites during ABE fermentation. For instance, FBA was used to investigate the relationship between hydrogen and solvent production in the in silico model of *C. acetobutylicum* (Lee et al., 2008a). The simulations showed potentially enhanced butanol production caused by decreased flux through hydrogenase. The model has also been used for in silico gene deletion analysis to study essential genes that could possibly be important as knockout targets. Likewise, the relationship between hydrogen and butanol in production mechanisms was observed through the GSM model of *C. beijerinckii*, where limited hydrogen formation rate confirmed its role in enhancing butanol production (Milne et al., 2011). Evaluating fermentation experiments of this model suggested selective pressures other than optimal growth that effect the resulting phenotype. In silico genome-scale metabolic flux modelling is often employed to investigate if a particular gene overexpression or knockout will lead to a desired feature before mutant strain engineering is attempted in a wet laboratory. Application of FBA linked with flux ratio constraints allowed the design of a simulated knockdown of acetoacetyl-CoA transferase, forcing a greater production of butanol (McAnulty et al., 2012). These results were achieved by implementation of a newly developed flux balance analysis with a flux ratios algorithm that considers a nonlinear relationship between the metabolite at a critical node and the phenotypic expression. It was further applied in metabolic engineering studies of isopropanol, butanol, and ethanol production in *C. acetobutylicum* (Yen et al., 2013). Moreover, GSM models can also help to understand responses to environmental stressors. Core sets for butanol and butyrate stress reactions were revealed through GSM, associated with information regarding changes in the metabolism of *C. acetobutylicum* under different butanol stress conditions (Dash et al., 2014). The reactions were predicted by a CoreReg method integrating transcriptomic data into the model. Similar findings may lead to the development of specific mutant strains. In another model, metabolic flux analysis was conducted for functional characterization of numerous key enzymes involved in butanol production of *C. acetobutylicum* (Yoo et al., 2016). Their role was further explored in the experimental study of in-frame deletion mutant strains (Yoo et al., 2016). However, there are phenomena that are very difficult to include in a GSM model, for example if a specific metabolic reaction may be catalysed by multiple enzymes, coded by multiple genes, which, for example, is the case for butanol formation in *C. acetobutylicum* DSM 1731 (Dai et al., 2016) mediated by up to six different aldehyde/alcohol dehydrogenases. In such a case, modeling returns multiple results. A recent effort in GSM model development for clostridia was mapped by Dash et al. (2016) and Millat and Winzer (2017).

Although currently, predictions on the ability of clostridia to tolerate or produce butanol based on its metabolic behaviour on a genome scale are still limited, the systems biology based GSM models may serve as a valuable tool for further studies and optimization of this process. Integration of kinetic parameters to GSM models would certainly improve their predictive power and facilitate studies of the organism in a broader context. The idea to construct such genome-scale kinetic models has been considered previously (e.g. Jamshidi and Palsson, 2008; Stanford et al., 2013). Efforts to integrate kinetic models with GSM models have already been conducted for *E. coli* (e.g. Chakrabarti et al., 2013; Mannan et al., 2015; Khodayari and Maranas, 2016). In light of these efforts to make GSM models dynamic, we expect that such a supplement will expand the development of new engineering strategies. However, lack of knowledge about biological systems at all levels is the main stumbling block in the proper use of systems approaches to analyse genome-scale data and compute accurate predictions about microbial behaviour. Both bottom-up and top-down approaches still face serious limitations in metabolic engineering - their combination is so far the best way to improve the predictive powers of models (Blazeck

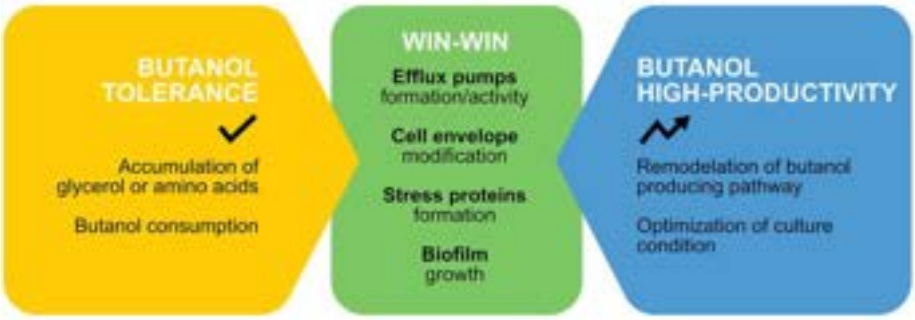


Fig. 3. Searching for common features of butanol tolerant and high-productive phenotypes in solventogenic clostridia.

and Alper, 2010). This goes hand in hand with increasing amounts of information that we can gain about microbes and progress in the development of computational techniques that remains a challenging issue for further studies.

6. Conclusion and future perspectives

Butanol tolerant and butanol high producing phenotypes of solventogenic clostridia share common features that include formation of stress (especially heat shock) proteins, induction of efflux pump formation/activity, modification of cell membranes/walls and probably also the preference to create biofilms or to grow on the surface of a solid carrier (see Fig. 3). The growth in biofilm seems to be advantageous regarding lower specific butanol toxicity (i.e. butanol toxicity expressed per a single cell), which is naturally achieved in biofilm populations having higher cell density compared to planktonic population. Thus it seems that use of butanol tolerance as a simple criterion for selection of random mutants has a rational core. As our understanding of signal pathways and other regulatory mechanisms in solventogenic clostridia is limited, random mutagenesis, directed evolution and gene shuffling are equally successful as targeted interventions that confirm the importance of butanol tolerance as a selection pressure factor.

Widespread availability and accelerated development of molecular biological methods such as next generation sequencing, RNA sequencing and proteomics, together with new technology for generating targeted mutants, such as CRISPR-Cas, generate complex data that may be processed by advanced bioinformatics tools into models improving our knowledge of functioning solventogenic clostridia. Future progress can be envisaged in neglected research fields such as construction of new expression regulators using prophages (Feiner et al., 2015). There is a solid background for this in solventogenic clostridia, in which many different phages and prophages have been identified (Jones et al., 2000; Brown et al., 2014; Pyne et al., 2016a,b). Alternatively, synthetic feedback loops or circuits controlling transcription/translation of selected genes might be designed (Bradley et al., 2016); the first such study regulating efflux pump expression for butanol extrusion in *E. coli* (Boyarskiy et al., 2016) has already been tested. Microbial butanol-producing clostridia profiting from different favourable properties of two or more members can also be designed (Xiaoqiang et al., 2016) and such butanol producing consortia have been developed (Wu et al., 2016a).

Conflict of interest

None declared.

Acknowledgement

This work has been supported by the Czech Science Foundation [grant number 17-00551S]. Many thanks belong to John Brooker for English revision of the manuscript.

References

Abdelaal, A.S., Ageez, A.M., Abd El-Hadi, A.E., Abdallah, N.A., 2015. Genetic improvement of n-butanol tolerance in *Escherichia coli* by heterologous overexpression of *groESL* operon from *Clostridium acetobutylicum*. 3 Biotech. 5, 401–410.

Al-Hinai, M.A., Jones, S.W., Papoutsakis, E.T., 2015. The *Clostridium* sporulation programs: diversity and preservation of endospore differentiation. *Microbiol. Mol. Biol. Rev.* 79, 19–37.

Alsaker, K.V., Spitzer, T.R., Papoutsakis, E.T., 2004. Transcriptional analysis of *spo0A* overexpression in *Clostridium acetobutylicum* and its effect on the cell's response to butanol stress. *J. Bacteriol.* 186, 1959–1971.

Alsaker, K.V., Paredes, C., Papoutsakis, E.T., 2010. Metabolite stress and tolerance in the production of biofuels and chemicals: gene-expression-based systems analysis of butanol, butyrate, and acetate stresses in the anaerobe *Clostridium acetobutylicum*. *Biotechnol. Bioeng.* 105, 1131–1147.

Arioli, S., Guglielmetti, S., Amalfitano, S., Viti, C., Marchi, E., Decorosi, F., Giovannetti, L., Mora, D., 2014. Characterization of tetA-like gene encoding for a major facilitator superfamily efflux pump in *Streptococcus thermophilus*. FEMS Microbiol. Lett. 355, 61–70.

Auch, A.F., von Jan, M., Klenk, H.-P., Göker, M., 2010. Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. *Stand. Genomic Sci.* 2, 117–134.

Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., Formsma, K., Gerdes, S., Glass, E.M., Kubal, M., Meyer, F., Olsen, G.J., Olson, R., Osterman, A.L., Overbeek, R.A., McNeil, L.K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G.D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A., Zagnitko, O., 2008. The RAST Server: rapid annotations using subsystems technology. BMC Genomics 9, 75.

Baer, S.H., Blaschek, H.P., Smith, T.L., 1987. Effect of butanol challenge and temperature on lipid composition and membrane fluidity of butanol-tolerant *Clostridium acetobutylicum*. Appl. Environ. Microbiol. 53, 2854–2861.

Behncken, S., Herweck, C., 2012. Anaerobic bacteria as producers of antibiotics. Appl. Microbiol. Biotechnol. 96, 61–67.

Belitsky, B.R., Sonenshein, A.L., 2013. Genome-wide identification of *Bacillus subtilis* CodY-binding sites at single-nucleotide resolution. PNAS 110, 7026–7031.

Blazeck, J., Alper, H., 2010. Systems metabolic engineering: genome-scale models and beyond. *Biotechnol. J.* 5, 647–659.

Blow, M.J., Clark, T.A., Daum, C.G., Deutschbauer, A.M., Fomenkov, A., Fries, R., Froula, J., Kang, D.D., Malmstrom, R.R., Morgan, R.D., Posfai, J., Singh, K., Visel, A., Wetmore, K., Zhao, Z., Rubin, E.M., Korlach, J., Pennacchio, L.A., Roberts, R.J., 2016. The epigenomic landscape of prokaryotes. PLoS Genet. 12, e1006064.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120.

Bowles, L.K., Ellefson, W.L., 1985. Effects on butanol on *Clostridium acetobutylicum*. Appl. Environ. Microbiol. 50, 1165–1170.

Bowring, S.N., Morris, J.G., 1985. Mutagenesis of *Clostridium acetobutylicum*. *J. Appl. Bacteriol.* 58, 577–584.

Boyarskiy, S., Tullman-Ercek, D., 2015. Getting pumped: membrane efflux transporters for enhanced biomolecule production. *Curr. Opin. Chem. Biol.* 28, 15–19.

Boyarskiy, S., López, S.D., Kong, N., Tullman-Ercek, D., 2016. Transcriptional feedback regulation of efflux protein expression for increased tolerance to and production of n-butanol. *Metab. Eng.* 33, 130–137.

Bradley, R.W., Buck, M., Wang, B., 2016. Tools and principles for microbial gene circuit engineering. *J. Mol. Biol.* 428, 862–888.

Bray, N.L., Pimentel, H., Melsted, P., Pachter, L., 2016. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* 34, 525–527.

Brown, S.D., Nagaraju, S., Utturkar, S., Tissera, S.D., Segovia, S., Mitchell, W., Land, M.L., Dassanayake, A., Kopke, M., 2014. Comparison of single-molecule sequencing and hybrid approaches for finishing the genome of *Clostridium autoethanogenum* and analysis of CRISPR systems in industrial relevant Clostridia. Biotechnol. Biofuels 7, 40.

Bruggeman, F.J., Westerhoff, H.V., 2007. The nature of systems biology. *Trends Microbiol.* 15, 45–50.

Bui, L.M., Lee, J.Y., Geraldi, A., Rahman, Z., Lee, J.H., Kim, S.C., 2015. Improved n-butanol tolerance in *Escherichia coli* by controlling membrane related functions. *J. Biotechnol.* 204, 33–44.

Bullard, J.H., Purdom, E., Hansen, K.D., Dudoit, S., 2010. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments evaluation of statistical methods for normalization and differential expression in

mRNA-Seq experiments. *U.C. Berkeley Div. Biostat. Pap. Ser.* 11, 94.
Cardoso, J., 2015. Analysis of genetic variation and potential applications in genome-scale metabolic modeling. *Front. Bioeng. Biotechnol.* 3, 13.
Casadesús, J., Low, D., 2006. Epigenetic gene regulation in the bacterial world. Microbiol. Mol. Biol. Rev. 70, 830–856.
Chakrabarti, A., Miskovic, L., Soh, K.C., Hatzimanikatis, V., 2013. Towards kinetic modeling of genome-scale metabolic networks without sacrificing stoichiometric, thermodynamic and physiological constraints. *Biotechnol. J.* 8, 1043–1057.
Chen, J.S., 1995. Alcohol dehydrogenase: multiplicity and relatedness in the solvent-producing clostridia. *FEMS Microbiol. Rev.* 17, 263–273.
Chen, C.K., Blaschek, H.P., 1999. Acetate enhances solvent production and prevents degeneration in *Clostridium beijerinckii* BA101. *Appl. Microbiol. Biotechnol.* 52, 170–173.
Chen, Y., Indurthi, D.C., Jones, S.W., Papoutsakis, E.T., 2011. Small RNAs in the genus *Clostridium*. MBio 2.
Chen, Y., Zhou, T., Liu, D., Li, A. Xu, Liu, Q., Li, B., Ying, H., 2013. Production of butanol from glucose and xylose with immobilized cells of *Clostridium acetobutylicum*. Biotechnol. Bioprocess Eng. 18, 234–241.
Chen, K., Hu, Z., Xia, Z., Zhao, D., Li, W., Tyler, J.K., 2016. The overlooked fact: fundamental need for spike-in control for virtually all genome-wide analyses. *Mol. Cell. Biol.* 36, 662–667.
Chen, X., Chou, W., Ma, Q., Xu, Y., 2017. SeqTU: a web server for identification of bacterial transcription units. *Sci. Rep.* 7, 43925.
Cheng, C.-L., Che, P.Y., Chen, B.-Y., Lee, W.-J., Chien, L.-J., Chang, J.-S., 2012. High yield bio-butanol production by solvent-producing bacterial microflora. Bioresour. Technol. 113, 58–64.
Chiao, J.-S., Sun, Z.-H., 2007. History of the acetone-butanol-ethanol fermentation industry in China: development of continuous production technology. *J. Mol. Microbiol. Biotechnol.* 13, 12–14.
Cho, C., Lee, S.Y., 2017. Efficient gene knockdown in *Clostridium acetobutylicum* by synthetic small regulatory RNAs. *Biotechnol. Bioeng.* 114, 374–383.
Chou, W.C., Ma, Q., Yang, S., Cao, S., Klingeman, D.M., Brown, S.D., Xu, Y., 2015. Analysis of strand-specific RNA-seq data using machine learning reveals the structures of transcription units in *Clostridium thermocellum*. *Nucleic Acids Res.* 43, e67.
Clark, T.A., Lu, X., Luong, K., Dai, Q., Boitano, M., Turner, S.W., He, C., Korlach, J., 2013. Enhanced 5-methylcytosine detection in single-molecule, real-time sequencing via Tet1 oxidation. *BMC Biol.* 11, 4.
Collins, M.D., Lawson, P.A., Willems, A., Cordoba, J.J., Fernandez-Garayzabal, J., Garcia, P., Cai, J., Hippe, H., Farrow, J.A., 1994. The phylogeny of the genus *Clostridium*: proposal of five new genera and eleven new species combinations. Int. J. Syst. Bacteriol. 44, 812–826.
Cooksley, C.M., Zhang, Y., Wang, H., Redl, S., Winzer, K., Minton, N.P., 2012. Targeted mutagenesis of the *Clostridium acetobutylicum* acetone-butanol-ethanol fermentation pathway. *Metab. Eng.* 14, 630–641.
Croux, C., Canard, B., Goma, G., Soucaille, P., 1992. Autolysis of *Clostridium acetobutylicum* ATCC 824. *J. Gen. Microbiol.* 138, 861–869.
Czechowska, K., van der Meer, J.R., 2012. Reversible and irreversible pollutant-induced bacterial cellular stress effects measured by ethidium bromide uptake and efflux. Environ. Sci. Technol. 46, 1201–1208.
Dai, Z., Dong, H., Zhang, Y., Li, Y., 2016. Elucidating the contributions of multiple aldehyde/alcohol dehydrogenases to butanol and ethanol production in *Clostridium acetobutylicum*. Sci. Rep. 6, 28189.
Dash, S., Mueller, T.J., Venkataramanan, K.P., Papoutsakis, E.T., Maranas, C.D., 2014. Capturing the response of *Clostridium acetobutylicum* to chemical stresses using a regulated genome-scale metabolic model. Biotechnol. Biofuels 7, 144.
Dash, S., Ng, C.Y., Maranas, C.D., 2016. Metabolic modeling of clostridia: current developments and applications. FEMS Microbiol. Lett. 363, fnw004.
Deboux, P., Marvaud, J.C., Abouelleil, A., Earl, A.M., Lambert, T., Dauga, C., 2016. Comparative genomics of *Clostridium bolteae* and *Clostridium clostridioforme* reveals species-specific genomic properties and numerous putative antibiotic resistance determinants. BMC Genomics 17, 819.
Dikicioglu, D., Kirdar, B., Oliver, S.G., 2015. Biomass composition: the "elephant in the room" of metabolic modelling. Metabolomics 11, 1690–1701.
Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., Gingeras, T.R., 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21.
Dunlop, M.J., Dossani, Z.Y., Szmidt, H.L., Chu, H.C., Lee, T.S., Keasling, J.D., Hadi, M.Z., Mukhopadhyay, A., 2011. Engineering microbial biofuel tolerance and export using efflux pumps. Mol. Syst. Biol. 7, 487.
Eisen, M.B., Spellman, P.T., Browndagger, P.O., Botstein, D., 1998. Cluster analysis and display of genome-wide expression patterns. Proc. Natl. Acad. Sci. U. S. A. 95, 14863–14868.
English, A.C., Richards, S., Han, Y., Wang, M., Vee, V., Qu, J., Qin, X., Muzny, D.M., Reid, J.G., Worley, K.C., Gibbs, R.A., 2012. Mind the gap: upgrading genomes with pacific biosciences RS long-read sequencing technology. PLoS One 7, e47768.
Ewels, P., Magnusson, M., Lundin, S., Käller, M., 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics 32, 3047–3048.
Federhen, S., Rossello-Mora, R., Klenk, H.-P., Tindall, B.J., Konstantinidis, K.T., Whitman, W.B., Brown, D., Labeda, D., Ussery, D., Garrity, G.M., Colwell, R.R., Hasan, N., Graf, J., Parte, A., Yarza, P., Goldberg, B., Sichtig, H., Karsch-Mizrachi, I., Clark, K., McVeigh, R., Pruitt, K.D., Tatusova, T., Falk, R., Turner, S., Madden, T., Kitts, P., Kimchi, A., Klimke, W., Agarwala, R., DiCuccio, M., Ostell, J., 2016. Meeting report: GenBank microbial genomic taxonomy workshop (12–13 May 2015). Stand. Genomic Sci. 11, 15.
Feiner, R., Argov, T., Rabinovich, L., Sigal, N., Borovok, I., Herskovits, A.A., 2015. A new perspective on lysogeny prophages as active regulatory switches of bacteria. *Nat. Rev. Microbiol.* 13, 641–650.
Feist, A.M., Herrgård, M.J., Thiele, I., Reed, J.L., Palsson, B.Ø., 2008. Reconstruction of biochemical networks in microorganisms. *Nat. Rev. Microbiol.* 7, 129–143.
Fisher, M.A., Boyarskiy, S., Yamada, M.R., Kong, N.W., Bauer, S., Tullman-Ercek, D., 2014. Enhancing tolerance to short-chain alcohols by engineering the *Escherichia coli* AcrB efflux pump to secrete the non-native substrate n-butanol. ACS Synth. Biol. 3, 30–40.
Fletcher, E., Pilizota, T., Davies, P.R., McVey, A., French, C.E., 2016. Characterization of the effects of n-butanol on the cell envelope of *E.coli*. *Appl. Microbiol. Biotechnol.* 100, 9563–9659.
Fondi, M., Lió, P., 2015. Genome-scale metabolic network reconstruction. In: Mengoni, A., Galardini, M., Fondi, M. (Eds.), Bacterial pangenomics: methods and protocols. Humana Press Springer Science + Business Media, New York, pp. 233–256.
Formanek, J., Mackie, R., Blaschek, H.P., 1997. Enhanced butanol production by *Clostridium beijerinckii* BA101 grown in semidefined P2 medium containing 6 percent maltodextrin or glucose. *Appl. Environ. Microbiol.* 63, 2306–2310.
Gallazzi, A., Branska, B., Marinelli, F., Patakova, P., 2015. Continuous production of n-butanol by *Clostridium pasteurianum* DSM 525 using suspended and surface-immobilized cells. *J. Biotechnol.* 216, 29–35.
Gao, X., Zhao, H., Zhang, G., He, K., Jin, Y., 2012. Genome shuffling of *Clostridium acetobutylicum* CICC 8012 for improved production of acetone-butanol-ethanol (ABE). *Curr. Microbiol.* 65, 128–132.
Ghoslaghi, R., Scharer, J.M., Moo-Young, M., Chou, C.P., 2009. Metabolic pathways of clostridia for producing butanol. Biotechnol. Adv. 27, 764–781.
Girbal, L., Soucaille, P., 1994. Regulation of *Clostridium acetobutylicum* metabolism as revealed by mixed-substrate steady-state continuous cultures: role of NADH/NAD ratio and ATP pool. *J. Bacteriol.* 176, 6433–6438.
Girbal, L., Croux, C., Vasconcelos, I., Soucaille, P., 1995. Regulation of metabolic shifts in *Clostridium acetobutylicum* ATCC 824. *FEMS Microbiol. Rev.* 17, 287–297.
Gong, J., Zheng, H., Wu, Z., Chen, T., Zhao, X., 2009. Genome shuffling: progress and applications for phenotype improvement. *Biotechnol. Adv.* 27, 996–1005.
González-Peñas, H., Lu-Chau, T.A., Moreira, M.T., Lema, J.M., 2015. Assessment of morphological changes of *Clostridium acetobutylicum* by flow cytometry during acetone/butanol/ethanol extractive fermentation. Biotechnol. Lett. 37, 577–584.
Green Biologics press releases, 2016. http://www.greenbiologics.com/news.php, **Accessed date: 13 December 2016.**
Green, E.M., Boynton, Z.L., Harris, L.M., Rudolph, F.B., Papoutsakis, E.T., Bennett, G.N., 1996. Genetic manipulation of acid formation pathways by gene inactivation in *Clostridium acetobutylicum* ATCC 824. Microbiology 142, 2079–2086.
Grupe, H., Gottschalk, G., 1992. Physiological events in *Clostridium acetobutylicum* during the shift from acidogenesis to solventogenesis in continuous culture and presentation of a model for shift induction. *Appl. Environ. Microbiol.* 58, 3896–3902.
Han, B., Ujor, V., Lai, L.B., Gopalan, V., Ezeji, T.C., 2013. Use of proteomic analysis to elucidate the role of calcium in acetone-butanol-ethanol fermentation by *Clostridium beijerinckii* NCIMB 8052. Appl. Environ. Microbiol. 79, 282–293.
Harris, L.M., Desai, R.P., Welker, N.E., Papoutsakis, E.T., 2000. Characterization of recombinant strains of the *Clostridium acetobutylicum* butyrate kinase inactivation mutant: need for new phenomenological models for solventogenesis and butanol inhibition? Biotechnol. Bioeng. 67, 1–11.
Harris, L.M., Blank, L., Desai, R.P., Welker, N.E., Papoutsakis, E.T., 2001. Fermentation characterization and flux analysis of recombinant strains of *Clostridium acetobutylicum* with an inactivated solR gene. *J. Ind. Microbiol. Biotechnol.* 27, 322–328.
Harrison, M.E., Dunlop, M.J., 2012. Synthetic feedback loop model for increasing microbial biofuel production using a biosensor. *Front. Microbiol.* 3, 360.
Hinks, J., Wang, Y., Matysik, A., Kraut, R., Kjelleberg, S., Mu, Y., Bazan, G.C., Wuertz, S., Seviour, T., 2015. Increased microbial butanol tolerance by exogenous membrane insertion molecules. Chem. Sus. Chem. 8, 3718–3726.
Hoff, K.J., Lange, S., Borodovsky, M., Stanke, M., 2016. BRAKER1: unsupervised RNA-seq-based genome annotation with GeneMark-ET and AUGUSTUS. Bioinformatics 32 (5), 767–769.
Hoffmann, S., Otto, C., Doose, G., Tanzer, A., Langenberger, D., Christ, S., Kunz, M., Holdt, L.M., Teupser, D., Hackermüller, J., Stadler, P.F., 2014. A multi-split mapping algorithm for circular RNA, splicing, trans-splicing and fusion detection. Genome Biol. 15, R34.
Hu, B., Yang, Y.-M., Beck, D.A.C., Wang, Q.-W., Chen, W.-J., Yang, J., Lidstrom, M.E., Yang, S., 2016. Comprehensive molecular characterization of *Methylobacterium extorquens* AM1 adapted for 1-butanol tolerance. Biotechnol. Biofuels 9, 84.
Huffer, S., Clark, M.E., Ning, J.C., Blanch, H.W., Clark, D.S., 2011. Role of alcohols in growth, lipid composition and membrane fluidity of yeasts, bacteria and archea. *Appl. Environ. Microbiol.* 77, 6400–6408.
Hyatt, D., Chen, G.-L., LoCascio, P.F., Land, M.L., Larimer, F.W., Hauser, L.J., 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinform. 11, 119.
Jain, M.K., Beacom, D., Datta, R. 1993. Mutant strain of *C.acetobutylicum* and process for making butanol. United States Patent US5192673A.
Jamshidi, N., Palsson, B.Ø., 2006. Formulating genome-scale kinetic models in the post-genome era. *Mol. Syst. Biol.* 4, 171.
Janda, J.M., Abbott, S.L., 2007. 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. *J. Clin. Microbiol.* 45, 2761–2764.
Jang, Y.S., Lee, J.Y., Lee, J., Park, J.H., Im, J.A., Eom, M.H., Lee, J., Lee, S.H., Song, H., Cho, J.H., Seung, D.Y., Lee, S.Y., 2012. Enhanced butanol production obtained by reinforcing the direct butanol-forming route in *Clostridium acetobutylicum*. MBio e00314–12.
Jang, Y.-S., Malaviya, A., Lee, S.Y., 2013. Acetone-butanol-ethanol production with high productivity using *Clostridium acetobutylicum* BKM19. *Biotechnol. Bioeng.* 110,

1646–1653.
Jernaes, M.W., Steen, H.B., 1994. Staining of *Escherichia coli* for flow cytometry: influx and efflux of ethidium bromide. Cytometry 17, 302–309.
Jia, K., Zhang, Y., Li, Y., 2012. Identification and characterization of two functionally unknown genes involved in butanol tolerance of *Clostridium acetobutylicum*. PLoS One 7, e38815.
Jiang, Y., Xu, C.M., Dong, F., Yang, Y.L., Jiang, W.H., Yang, S., 2009. Disruption of the acetoacetate decarboxylase gene in solvent-geducing *Clostridium acetobutylicum* increases the butanol ratio. Metab. Eng. 11, 284–291.
Jiang, W., Zhao, J., Wang, Z., Yang, S.-T., 2014. Stable high-titer n-butanol production from sucrose and sugarcane juice by *Clostridium acetobutylicum* JB200 in repeated batch fermentations. Bioresour. Technol. 163, 172–179.
Jiang, Y., Liu, J., Jiang, W., Yang, Y., Yang, S., 2015. Current status and prospects of industrial bio-production of n-butanol in China. Biotechnol. Adv. 33, 1493–1501.
Johnson, N., Goldfine, H., 1983. Lipid composition in the classification of the butyric acid-producing clostridia. *J. Gen. Microbiol.* 129, 1075–1081.
Johnson, J.L., Toth, J., Santiwatanakul, S., Chen, J.-S., 1997. Cultures of "*Clostridium acetobutylicum*" from various collections comprise *Clostridium acetobutylicum*, *Clostridium beijerinckii* and two other distinct types based on DNA-DNA reassociation. *Int. J. Syst. Bacteriol.* 47, 420–424.
Jones, D.T., Woods, D.R., 1986. Acetone-butanol fermentation revisited. *Microbiol. Rev.* 50, 484–524.
Jones, D.T., van der Westhuizen, A., Long, S., Allcock, E.R., Reid, S.J., Woods, D.R., 1982. Solvent production and morphological changes in *Clostridium acetobutylicum*. *Appl. Environ. Microbiol.* 43, 1434–1439.
Jones, D.T., Shirley, M., Wu, X., Keis, S., 2000. Bacteriophage infections in the industrial acetone-butanol (AB) fermentation process. *J. Mol. Microbiol.* 2, 21–26.
Jones, S.W., Paredes, C.J., Tracy, B., Cheng, N., Sillers, R., Senger, R.S., Papoutsakis, E.T., 2008. The transcriptional program underlying the physiology of clostridial sporulation. Genome Biol. 9, R114.
Jones, A.J., Venkataramanan, K.P., Papoutsakis, T., 2016. Overexpression of two stress-responsive, small non-coding RNAs, 6S and tmRNA, imparts butanol tolerance in *Clostridium acetobutylicum*. *FEMS Microbiol. Lett.* 363, fnw063.
Junelles, A.M., El Kanouni, A., Petitdemange, H., Gay, R., 1989. Influence of acetic and butyric acid addition on polysaccharide formation by *Clostridium acetobutylicum*. *J. Ind. Microbiol.* 4, 121–125.
Kamio, Y., Kanegasaki, S., Takahashi, H., 1969. Occurance of plasmalogens in anaerobic bacteria. *J. Gen. Appl. Microbiol.* 15, 439–451.
Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., Tanabe, M., 2016. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* 44, D457–D462.
Keli, D.B., Swainston, N., Pir, P., Oliver, S.G., 2015. Membrane transporter engineering in industrial biotechnology and whole cell biocatalysis. Trends Microbiol. 33, 237–246.
Khodayari, A., Maranas, C.D., 2016. A genome-scale *Escherichia coli* kinetic metabolic model k-ecoli457 satisfying flux data for multiple mutant strains. *Nat. Commun.* 7, 13806.
Kim, W.J., Kim, H.U., Lee, S.Y., 2017. Current state and applications of microbial genome-scale metabolic models. *Curr. Opin. Syst. Biol.* 2, 10–18.
Kolek, J., Patakova, P., Melzoch, K., Sigler, K., Rezanka, T., 2015. Changes in membrane plasmalogens of *Clostridium pasteurianum* during butanol fermentation as determined by lipidomic analysis. PLoS One 10, e0122058.
Kolek, J., Branska, B., Drahokoupil, M., Patakova, P., Melzoch, K., 2016a. Evaluation of viability, metabolic activity and spore quantity in clostridial cultures during ABE fermentation. FEMS Microbiol. Lett. 363, fnw031.
Kolek, J., Sedlar, K., Provaznik, I., Patakova, P., 2016b. Dam and Dcm methylations prevent gene transfer into *Clostridium pasteurianum* NRRL B-598: development of methods for electrotransformation, conjugation, and sonoporation. Biotechnol. Biofuels 9, 14.
Korlach, J., Turner, S., 2012. Going beyond five bases in DNA sequencing. Curr. Opin. Struct. Biol. 22, 251–261.
Kosaka, T., Nakayama, S., Nakaya, K., Yoshino, S., Furukawa, K., 2007. Characterization of the sol operon in butanol-hyperproducing *Clostridium saccharoperbutylacetonicum* strain N1-4 and its degeneration mechanism. *Biosci. Biotechnol. Biochem.* 71, 58–68.
Kuisiene, N., Raugalas, J., Sproer, C., Kroppenstedt, R.M., Chitavichius, D., 2008. *Bacillus butanolivorans* sp. nov., a species with industrial application for the remediation of n-butanol. Int. J. Syst. Evol. Microbiol. 58, 505–509.
Kumar, M., Kumar, D., Singh, B., 2014. Utilization of agro residue corncob for production of acetone-butanol using *Clostridium acetobutylicum* and process optimization through RSM. J. Microbiol. Biochem. Technol. S8.
Lagesen, K., Hallin, P., Rødland, E.A., Stærfeldt, H.H., Rognes, T., Ussery, D.W., 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* 35, 3100–3108.
Lawson, P.A., Rainey, F.A., 2016. Proposal to restrict the genus *Clostridium prazmowski* to *Clostridium butyricum* and related species. *Int. J. Syst. Evol. Microbiol.* 66, 1009–1016.
Lee, J., Yun, H., Feist, A.M., Palsson, B.Ø., Lee, S.Y., 2008a. Genome-scale reconstruction and in silico analysis of the *Clostridium acetobutylicum* ATCC 824 metabolic network. *Appl. Microbiol. Biotechnol.* 80, 849–862.
Lee, S.K., Chou, H., Ham, T.S., Lee, T.S., Keasling, J.D., 2008b. Metabolic engineering of microorganisms for biofuels production: from bugs to synthetic biology to fuels. *Curr. Opin. Biotechnol.* 19, 556–563.
Lee, S.-M., Cho, M.O., Park, C.H., Chung, Y.-C., Kim, J.H., Sang, B.-I., Um, Y., 2008c. Continuous butanol production using suspended and immobilized *Clostridium beijerinckii* NCIMB 8052 with supplementary butyrate. Energy Fuel 22, 3459–3464.
Lee, S.Y., Park, J.H., Jang, S.H., Nielsen, L.K., Kim, J., Jung, K.S., 2008d. Fermentative butanol production by clostridia. *Biotechnol. Bioeng.* 101, 209–228.
Lemmel, S.A., 1985. Mutagenesis in *Clostridium acetobutylicum*. *Biotechnol. Lett.* 7, 711–716.
Lepage, C., Fayolle, F., Hermann, M., Vandecasteele, J.P., 1987. Changes in membrane lipid composition of *Clostridium acetobutylicum* during acetone-butanol fermentation: effects of solvents, growth temperature and pH. Microbiology 133, 103–110.
Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078–2079.
Li, N., Yang, J., Chai, C., Yang, S., Jiang, W., Gu, Y., 2015. Complete genome sequence of *Clostridium carboxidivorans* P7T, a syngas-fermenting bacterium capable of producing long-chain alcohols. J. Biotechnol. 211, 44–45.
Li, C., Wang, Y., Xie, G., Peng, B., Zhang, B., Chen, W., Huang, X., Wu, H., Zhang, B., 2016a. Complete genome sequence of *Clostridium butyricum* JKY6D1 isolated from the pit mud of a Chinese flavor liquor-making factory. *J. Biotechnol.* 220, 23–24.
Li, Q., Chen, J., Minton, N.P., Zhang, Y., Wen, Z., Liu, J., Yang, H., Zeng, Z., Ren, X., Yang, J., Gu, Y., Jiang, W., Jiang, Y., Yang, S., 2016b. CRISPR-based genome editing and expression control systems in *Clostridium acetobutylicum* and *Clostridium beijerinckii*. Biotechnol. J. 11, 961–972.
Li, S.-B., Qian, Y., Liang, Z.-W., Guo, Y., Zhao, M.-M., Pang, Z.-W., 2016c. Enhanced butanol production from cassava with *Clostridium acetobutylicum* by genome shuffling. *World J. Microbiol. Biotechnol.* 32, 53.
Lianage, H., Young, M., Kashket, E.R., 2000. Butanol tolerance of *Clostridium beijerinckii* NCIMB 8052 associated with down-regulation of gldA antisense RNA. *J. Mol. Microbiol. Biotechnol.* 2, 87–93.
Liao, Z., Zhang, Y., Luo, S., Suo, Y., Zhang, S., Wang, J., 2017. Improving cellular robustness and butanol titers of *Clostridium acetobutylicum* ATCC824 by introducing heat shock proteins from an extremophilic bacterium. J. Biotechnol. 252, 1–10.
Lin, Y.L., Blaschek, H.P., 1983. Butanol production by a butanol-tolerant strain of *Clostridium acetobutylicum* in extruded corn broth. Appl. Environ. Microbiol. 45, 966–973.
Linhova, M., Patakova, P., Lipovsky, J., Fribert, P., Paulova, L., Rychtera, M., Melzoch, K., 2010. Development of flow cytometry technique for detection of thinning of peptidoglycan layer as a result of solvent production by *Clostridium pasteurianum*. *Folia Microbiol.* 55, 340–344.
Linhova, M., Branska, B., Patakova, P., Lipovsky, J., Fribert, P., Rychtera, M., Melzoch, K., 2012. Rapid flow cytometric method for viability determination of solventogenic clostridia. *Folia Microbiol.* 57, 307–311.
Lipovsky, J., Patakova, P., Paulova, L., Pokorny, T., Rychtera, M., Melzoch, K., 2016. Butanol production by *Clostridium pasteurianum* NRRL B-598 in continuous culture compared to batch and fed-batch systems. Fuel Process. Technol. 144, 139–144.
Liu, L., Agren, R., Bordel, S., Nielsen, J., 2010. Use of genome-scale metabolic models for understanding microbial physiology. FEBS Lett. 584, 2556–2564.
Liu, S., Bischoff, K.M., Leathers, T.D., Qureshi, N., Rich, J.O., Hughes, S.R., 2012. Adaptation of lactic acid bacteria to butanol. *Biocatal. Agric. Biotechnol.* 1, 57–61.
Liu, X.-B., Gu, Q.-Y., Yu, X.-B., 2013. Repetitive domestication to enhance butanol tolerance and production in *Clostridium acetobutylicum* through artificial simulation of bio-evolution. Bioresour. Technol. 130, 638–643.
Liu, Z., Qiao, K., Tian, Q., Liu, Z.-Y., Li, F.-L., 2015. Spontaneous large-scale autolysis in *Clostridium acetobutylicum* contributes to generation of more spores. Front. Microbiol. 6, 950.
Liu, D., Xu, J., Wang, Y., Chen, Y., Shen, X., Niu, H., Guo, T., Ying, H., 2016a. Comparative transcriptomic analysis of *Clostridium acetobutylicum* biofilm and planktonic cells. *J. Biotechnol.* 218, 1–12.
Liu, H., Huang, D., Wen, J., 2016b. Integrated intracellular metabolic profiling and pathway analysis approaches reveal complex metabolic regulation by *Clostridium acetobutylicum*. Microb. Cell Factories 15, 36.
Liu, S., Qureshi, N., Hughes, S.R., 2017. Progress and perspectives on improving butanol tolerance. *World J. Microbiol. Biotechnol.* 33, 51.
Livny, J., 2012. Bioinformatic discovery of bacterial regulatory RNAs using SIPHT. *Methods Mol. Biol.* 905, 3–14.
Long, S., Jones, D.T., Woods, D.R., 1984. The relationship between sporulation and solvent production in *Clostridium acetobutylicum* P262. *Biotechnol. Lett.* 6, 529–534.
Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550.
Lowe, T.M., Eddy, S.R., 1997. TRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25, 955–964.
Malaviya, A., Jang, Y.-S., Lee, S.Y., 2012. Continuous butanol production with reduced byproducts formation from glycerol by a hyper producing mutant of *Clostridium pasteurianum*. *Appl. Microbiol. Biotechnol.* 93, 1485–1494.
Mann, M.S., Dragovic, Z., Schirrmacher, G., Lutke-Eversloh, T., 2012. Over-expression of stress protein-encoding genes helps *Clostridium acetobutylicum* to rapidly adapt to butanol stress. Biotechnol. Lett. 34, 1643–1649.
Mannan, A.A., Toya, Y., Shimizu, K., McFadden, J., Kierzek, A.M., Rocco, A., 2015. Integrating kinetic model of *E. coli* with genome scale metabolic fluxes overcomes its open system problem and reveals bistability in central metabolism. PLoS One 10, e0139507.
Mao, S., Luo, Y., Bao, G., Zhang, Y., Li, Y., Ma, Y., 2011. Comparative analysis of the membrane proteome of *Clostridium acetobutylicum* wild type strain and its butanol tolerant mutant. *Mol. BioSyst.* 7, 1660–1677.
McAnulty, M.J., Yen, J.Y., Freedman, B.G., Senger, R.S., 2012. Genome-scale modeling using flux ratio constraints to enable metabolic engineering of clostridial metabolism in silico. *BMC Syst. Biol.* 6, 42.
McClure, R., Balasubramanian, D., Sun, Y., Bobrovskyy, M., Sumby, P., Genco, C.A., Vanderpool, C.K., Tjaden, B., 2013. Computational analysis of bacterial RNA-Seq data. *Nucleic Acids Res.* 41, e140.
Meier-Kolthoff, J.P., Auch, A.F., Klenk, H.-P., Göker, M., 2013. Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC

Bioinform. 14, 1–14.
Meinecke, B., Bahl, H., Gottschalk, G., 1984. Selection of an asporogenous strain of *Clostridium acetobutylicum* in continuous culture under phosphate limitation. *Appl. Environ. Microbiol.* 48, 1064–1065.
Meyer, C.L., Papoutsakis, E.T., 1989. Increased levels of ATP and NADH are associated with increased solvent production in continuous cultures of *Clostridium acetobutylicum*. *Appl. Microbiol. Biotechnol.* 30, 450–459.
Millat, T., Winzer, K., 2017. Mathematical modelling of clostridial acetone-butanol-ethanol fermentation. *Appl. Microbiol. Biotechnol.* 101, 2251–2271.
Milne, C.B., Kim, P.-J., Eddy, J.A., Price, N.D., 2009. Accomplishments in genome-scale in silico modeling for industrial and medical biotechnology. Biotechnol. J. 4, 1653–1670.
Milne, C.B., Eddy, J.A., Raju, R., Ardekani, S., Kim, P.-J., Senger, R.S., Jin, Y.-S., Blaschek, H.P., Price, N.D., 2011. Metabolic network reconstruction and genome-scale model of butanol-producing strain *Clostridium beijerinckii* NCIMB 8052. *BMC Syst. Biol.* 5, 130.
Mo, X., Pei, J., Guo, Y., Lin, L., Peng, L., Kou, C., Fan, D., Pang, H., 2015. Genome sequence of *Clostridium acetobutylicum* GXAS18-1, a novel biobutanol production strain. *Genome Announc.* 3, e00033–15.
Molle, V., Fujita, M., Jensen, S.T., Eichenberger, P., González-Pastor, J.E., Liu, J.S., Losick, R., 2003. The Spo0A regulon of *Bacillus subtilis*. *Mol. Microbiol.* 50, 1683–1701.
Mukhopadhyay, A., 2015. Tolerance engineering in bacteria for the production of advanced biofuels and chemicals. *Trends Microbiol.* 23, 498–508.
Müller, S., Nebe-von-Caron, G., 2010. Functional single-cell analyses: flow cytometry and cell sorting of microbial populations and communities. FEMS Microbiol. Rev. 34, 554–587.
Murakami, S., Nakashima, R., Yamashita, E., Yamaguchi, A., 2002. Crystal structure of bacterial multidrug efflux transporter AcrB. Nature 419, 587–593.
Murray, I.A., Clark, T.A., Morgan, R.D., Boitano, M., Anton, B.P., Luong, K., Fomenkov, A., Turner, S.W., Korlach, J., Roberts, R.J., 2012. The methylomes of six bacteria. *Nucleic Acids Res.* 40, 11450–11462.
Nicolaou, S.A., Gaida, S.M., Papoutsakis, E.T., 2010. A comparative view of metabolite and substrate stress and tolerance in microbial processing: from biofuels and chemicals to biocatalysis and bioremediation. *Metab. Eng.* 12, 307–331.
Noar, J., Makwana, S.T., Bruno-Bárcena, J.M., 2014. Complete genome sequence of solvent-tolerant *Clostridium beijerinckii* strain SA-1. *Genome Announc.* 2, 6–7.
Nölling, J., Breton, G., Omelchenko, M.V., Makarova, K.S., Zeng, Q., Gibson, G., Lee, Hong Mei, Dubois, J., Qiu, D., Hitti, J., Aldredge, T., Ayers, M., Bashirzadeh, R., Bochner, H., Boivin, M., Bross, S., Bush, D., Butler, C., Caron, A., Caruso, A., Cook, R., Daggett, P., Deloughery, C., Egan, J., Elston, D., Engelstein, M., Ezedi, J., Gilbert, K., Goyal, A., Guerin, J., Ho, T., Holtham, K., Joseph, P., Krugle, P., Kozlovsky, J., LaPlante, M., LeBlanc, G., Lumm, W., Majeski, A., McDougall, S., Mank, P., Mao, J.I., Nocco, D., Patwell, D., Phillips, J., Pothier, B., Prabhakar, S., Richterich, P., Rice, P., Rosetti, D., Rousetti, M., Rubenfield, M., Sachdeva, M., Snell, P., Spadafora, R., Spitzer, L., Shimer, G., Thomann, H.U., Vicaire, R., Wall, K., Wang, Y., Weinstock, K., Wong, Lai Peng, Wonsey, A., Xu, Q., Zhang, L., Wolf, Y.I., Tatusov, R.L., Sabathe, F., Doucette-Stamm, L., Soucaille, P., Daly, M.J., Bennett, G.N., Koonin, E.V., Smith, D.R., 2001. Genome sequence and comparative analysis of the solvent-producing bacterium *Clostridium acetobutylicum*. *J. Bacteriol.* 183, 4823–4838.
Novichkov, P., Rodionov, D., Stavrovskaya, E., Novichkova, E., Kazakov, A., Gelfand, M., Arkin, A., Mironov, A., Dubchak, I., 2010. RegPredict: an integrated system for regulon inference in prokaryotes by comparative genomics approach. Nucleic Acids Res. 38, W299–W307.
Oberhardt, M.A., Palsson, B.Ø., Papin, J.A., 2009. Applications of genome-scale metabolic reconstructions. *Mol. Syst. Biol.* 5, 320.
Olson, N.E., 2006. The microarray data analysis process: from raw data to biological significance. NeuroRx 3, 373–383.
Park, P.J., 2009. ChIP–seq: advantages and challenges of a maturing technology. *Nat. Rev. Genet.* 10, 669–680.
Patakova, P., Lipovsky, J., Cizkova, H., Fortova, J., Rychtera, M., Melzoch, K., 2009. Exploitation of food feedstock and waste for production of biobutanol. *Czech J. Food Sci.* 27, 276–283.
Patakova, P., Linhova, M., Rychtera, M., Paulova, L., Melzoch, K., 2013. Novel and neglected issues of acetone-butanol-ethanol (ABE) fermentation by clostridia: *Clostridium* metabolic diversity, tools for proces mapping and continuous fermentation systems. Biotechnol. Adv. 31, 58–67.
Patakova, P., Linhova, M., Vykydalova, P., Branska, B., Rychtera, M., Melzoch, K., 2014. Use of fluorescent staining and flow cytometry for monitoring physiological changes in solventogenic clostridia. Anaerobe 29, 113–117.
Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., Kingsford, C., 2017. Salmon provides fast and bias-aware quantification of transcript expression. Nat. Methods 14, 417–419.
Peabody, G.L., Kao, K.C., 2016. Recent progress in biobutanol tolerance in microbial systems with an emphasis on *Clostridium*. *FEMS Microbiol. Lett.* 363, fnw017.
Perez-Garcia, O., Lear, G., Singhal, N., 2016. Metabolic network modeling of microbial interactions in natural and engineered environmental systems. *Front. Microbiol.* 7, 673.
Pich, A., Narberhaus, F., Bahl, H., 1990. Induction of heat shock proteins during initiation of solvent formation in *Clostridium acetobutylicum*. *Appl. Microbiol. Biotechnol.* 33, 697–704.
Poehlein, A., Grosse-Honebrink, A., Zhang, Y., Minton, N.P., Daniel, R., 2015. Complete genome sequence of the nitrogen-fixing and solvent-producing *Clostridium pasteurianum* DSM 525. *Genome Announc.* 3, e01591–14.
Poehlein, A., Solano, J.D.M., Flitsch, S.K., Krabben, P., Winzer, K., Reid, S.J., Jones, D.T., Green, E., Minton, N.P., Daniel, R., Dürre, P., 2017. Microbial solvent formation revisited by comparative genome analysis. Biotechnol. Biofuels 10, 58.
Pyne, M.E., Bruder, M.R., Moo-Young, M., Chung, D.A., Chou, C.P., 2016a. Harnessing heterologous and endogenous CRISPR-Cas machineries for efficient markerless genome editing in *Clostridium*. Sci. Rep. 6, 25666.
Pyne, M.E., Liu, X., Moo-Young, M., Chung, D.A., Chou, C.P., 2016b. Genome-directed analysis of prophage excision, host defence systems, and central fermentative metabolism in *Clostridium pasteurianum*. *Sci. Rep.* 6, 26228.
Qureshi, N., Annous, B.A., Ezeji, T.C., Karcher, P., Maddox, I.S., 2005. Biofilm reactors for industrial bioconversion processes: employing potential of enhanced reaction rates. Microb. Cell Factories 4, 24.
Rapaport, F., Khanin, R., Liang, Y., Pirun, M., Krek, A., Zumbo, P., Mason, C.E., Socci, N.D., Betel, D., 2013. Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. Genome Biol. 14, R95.
Rau, M.H., Calero, P., Lennen, R.M., Long, K.S., Nielsen, A.T., 2016. Genome-wide *Escherichia coli* stress response and improved tolerance towards industrially relevant chemicals. Microb. Cell Factories 15, 176.
Reith, J., Mayer, C., 2011a. Peptidoglycan turnover and recycling in Gram-positive bacteria. *Appl. Microbiol. Biotechnol.* 92, 1–11.
Reith, J., Mayer, C., 2011b. Characterization of a glucosamine/glucosaminide *N*-acetyltransferase of *Clostridium acetobutylicum*. *J. Bacteriol.* 193, 5393–5399.
Reith, J., Berking, A., Mayer, C., 2011. Characterization of an *N*-acetylmuramic acid/*N*-acetylglucosamine kinase of *Clostridium acetobutylicum*. *J. Bacteriol.* 193, 5386–5392.
Ren, C., Gu, Y., Wu, Y., Zhang, W., Yang, C., Yang, S., Jiang, W., 2012. Pleiotropic functions of catabolite control protein CcpA in butanol-producing *Clostridium acetobutylicum*. BMC Genomics 13, 349.
Reyes, L.H., Almario, M.P., Kao, K.C., 2011. Genomic library screens for genes involved in n-butanol tolerance in *Escherichia coli*. PLoS One e17678.
Reyes, L.H., Abdelaal, A.S., Kao, K.C., 2013. Genetic determinants for n-butanol tolerance in evolved *Escherichia coli* mutants: cross adaptation and antagonistic pleiotropy between n-butanol and other stressors. *Appl. Environ. Microbiol.* 79, 5313–5320.
Richter, M., Rosselló-Mora, R., 2009. Shifting the genomic gold standard for the prokaryotic species definition. Proc. Natl. Acad. Sci. 106, 19126–19131.
Roberts, R.J., Carneiro, M.O., Schatz, M.C., 2013. The advantages of SMRT sequencing. Genome Biol. 14, 405.
Robinson, M.D., McCarthy, D.J., Smyth, G.K., 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140.
Ruhl, J., Schmid, A., Blank, L.M., 2009. Selected *Pseudomonas putida* strains able to grow in the presence of high butanol concentrations. Appl. Environ. Microbiol. 75, 4653–4656.
Rutherford, B.J., Dahl, R.H., Price, R.E., Szmidt, H.L., Benke, P.I., Mukhopadhyay, A., Keasling, J.D., 2010. Functional genomic study of exogenous n-butanol stress in *Escherichia coli*. Appl. Environ. Microbiol. 76, 1935–1945.
Saeed, A.I., Bhagabati, N.K., Braisted, J.C., Liang, W., Sharov, V., Howe, E.A., Li, J., Thiagarajan, M., White, J.A., Quackenbush, J., 2006. TM4 microarray software suite. *Methods Enzymol.* 411, 134–193.
Salimi, F., Zhuang, K., Mahadevan, R., 2010. Genome-scale metabolic modeling of a clostridial co-culture for consolidated bioprocessing. Biotechnol. J. 5, 726–738.
Sandoval, N.R., Papoutsakis, E.T., 2016. Engineering membrane and cell-wall programs for tolerance to toxic chemicals: beyond solo genes. *Curr. Opin. Microbiol.* 33, 56–66.
Sandoval, N.R., Venkataramanan, K.P., Groth, T.S., Papoutsakis, E.T., 2015. Whole-genome sequence of an evolved *Clostridium pasteurianum* strain reveals Spo0A deficiency responsible for increased butanol production and superior growth. Biotechnol. Biofuels 8, 227.
Sandoval-Espinosa, W.J., Makwana, S.T., Chinn, M.S., Thon, M.R., Azcarate-Peril, M.A., Bruno-Bárcena, J.M., 2013. Comparative phenotypic analysis and genome sequence of *Clostridium beijerinckii* SA-1, an offspring of NCIMB 8052. Microbiology 62, 2558–2570.
Santos, F., Boele, J., Teusink, B., 2011. A practical guide to genome-scale metabolic models and their analysis. *Methods Enzymol.* 500, 509–532.
Schindler, B.D., Kaatz, G.W., 2016. Multidrug efflux pumps of Gram-positive bacteria. Drug Resist. 27, 1–13 Update.
Schulze, A., Downward, J., 2001. Navigating gene expression using microarrays — a technology review. *Nat. Cell Biol.* 3, E190–E195.
Schwarz, K.M., Grosse-Honebrink, A., Derecka, K., Rotta, C., Zhang, Y., Minton, N.P., 2017. Towards improved butanol production through targeted genetic modification of *Clostridium pasteurianum*. Metab. Eng. 40, 124–137.
Sedlar, K., Kolek, J., Skutkova, H., Branska, B., Provaznik, I., Patakova, P., 2015. Complete genome sequence of *Clostridium pasteurianum* NRRL B-598, a non-type strain producing butanol. *J. Biotechnol.* 214, 113–114.
Sedlar, K., Kolek, J., Provaznik, I., Patakova, P., 2017. Reclassification of non-type strain *Clostridium pasteurianum* NRRL B-598 as *Clostridium beijerinckii* NRRL B-598. *J. Biotechnol.* 244, 1–3.
Segata, N., Börnigen, D., Morgan, X.C., Huttenhower, C., 2013. PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. *Nat. Commun.* 4, 2304.
Selvaraj, S., Natarajan, J., 2011. Microarray data analysis and mining tools. Bioinformation 6, 95–99.
Senger, R.S., Papoutsakis, E.T., 2008a. Genome-scale model for *Clostridium acetobutylicum*: part I. Metabolic network resolution and analysis. *Biotechnol. Bioeng.* 101, 1036–1052.
Senger, R.S., Papoutsakis, E.T., 2008b. Genome-scale model for *Clostridium acetobutylicum*: part II. Development of specific proton flux states and numerically determined sub-systems. Biotechnol. Bioeng. 101, 1053–1071.
Shapiro, H.M., 2000. Membrane potential estimation by flow cytometry. Methods 21, 271–279.
Shapiro, H.M., 2005. Parameters and probes. In: Shapiro, H.M. (Ed.), Practical Flow Cytometry. John Wiley & Sons, Inc., pp. 273–410.

Shen, C.R., Lan, E.I., Dekishima, Y., Baez, A., Cho, K.M., Liao, J.C., 2011. Driving forces enable high-titer anaerobic 1-butanol synthesis in *Escherichia coli*. *Appl. Environ. Microbiol.* 77, 2905–2915.

Shi, Z., Blaschek, H.P., 2008. Transcriptional analysis of *Clostridium beijerinckii* NCIMB 8052 and the hyper-butanol-producing mutant BA101 during the shift from acidogenesis to solventogenesis. Appl. Environ. Microbiol. 74, 7709–7714.

Si, H.-M., Zhang, F., Wu, A.-N., Han, R.-Z., Xu, G.-C., Ni, Y., 2016. DNA microarray of global transcription factor mutant reveals membrane-related proteins involved in n-butanol tolerance in *Escherichia coli*. Biotechnol. Biofuels 9, 114.

Siegfried, Z., Simon, I., 2010. DNA methylation and gene expression. *Wires. Syst. Biol. Med.* 2, 362–371.

Simeonidis, E., Price, N.D., 2015. Genome-scale modeling for metabolic engineering. *J. Ind. Microbiol. Biotechnol.* 42, 327–338.

Soutourina, O.A., Monot, M., Boudry, P., Saujet, L., Pichon, C., Sismeiro, O., Semenova, E., Severinov, K., Le Bouguenec, C., Coppee, J.Y., Dupuy, B., Martin-Verstraete, I., 2013. Genome-wide identification of regulatory RNAs in the human pathogen *Clostridium difficile*. PLoS Genet. 9, e1003493.

Sridhar, J., Sambaturu, N., Narmada, S.R., Sabarinathan, R., Ou, H.-Y., Deng, Z., Sekar, K., Rafi, Z.A., Rajakumar, K., 2010. sRNAscanner: a computational tool for intergenic small RNA detection in bacterial genomes. PLoS One 5, e11970.

Stackebrandt, E., Kramer, I., Swiderski, J., Hippe, H., 1999. Phylogenetic basis for a taxonomic dissection of the genus *Clostridium*. FEMS Immunol. Med. Microbiol. 24, 253–258.

Stanford, N.J., Lubitz, T., Smallbone, K., Klipp, E., Mendes, P., Liebermeister, W., 2013. Systematic construction of kinetic models from genome-scale metabolic networks. PLoS One 8, e79195.

Steiner, E., Dago, A.E., Young, D.I., Heap, J.T., Minton, N.P., Hoch, J.A., Young, M., 2011. Multiple orphan histidine kinases interact directly with Spo0A to control the initiation of endospore formation in *Clostridium acetobutylicum*. *Mol. Microbiol.* 80, 641–654.

Steiner, E., Scott, J., Minton, N.P., Winzer, K., 2012. An agr quorum sensing system that regulates granulose formation and sporulation in *Clostridium acetobutylicum*. Appl. Environ. Microbiol. 78, 1113–1122.

Tan, Y., Liu, J., Liu, Z., Li, F., 2014. Characterization of two novel butanol dehydrogenases involved in butanol degradation in syngas-utilizing bacterium *Clostridium ljungdahlii* DSM 13528. J. Basic Microbiol. 54, 996–1004.

Taranova, S., García-Alcalde, F., Dopazo, J., Ferrer, A., Conesa, A., 2011. Differential expression in RNA-seq: a matter of depth. Genome Res. 21, 2213–2223.

Tashiro, Y., Sonomoto, K., 2010. Advances in butanol production by clostridia. In: Méndez-Vilas, A. (Ed.), Current Research, Technology and Education Topics in Applied Microbiology and Microbial Biotechnology. Vol 2. Formatex Research Center, Badajoz, Spain, pp. 1383–1394.

Tashiro, Y., Takeda, K., Kobayashi, G., Sonomoto, K., Ishizaki, A., Yoshino, S., 2004. High butanol production by *Clostridium saccharoperbutylacetonicum* N1-4 in fed-batch culture with pH-stat continuous butyric acid and glucose feeding method. *J. Biosci. Bioeng.* 98, 263–268.

Tatusova, T., Dicuccio, M., Badretdin, A., Chetvernin, V., Nawrocki, E.P., Zaslavsky, L., Lomsadze, A., Pruitt, K.D., Borodovsky, M., Ostell, J., 2016. NCBI prokaryotic genome annotation pipeline. Nucleic Acids Res. 44, 6614–6624.

Thiele, I., Palsson, B., 2010. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat. Protoc.* 5, 93–121.

Tomas, C.A., Welker, N.E., Papoutsakis, E.T., 2003. Overexpression of groESL in *Clostridium acetobutylicum* results in increased solvent production and tolerance, prolonged metabolism, and changes in the cell's transcriptional program. *Appl. Environ. Microbiol.* 69, 4951–4965.

Tomas, C.A., Beamish, J., Papoutsakis, E.T., 2004. Transcriptional analysis of butanol stress and tolerance in *Clostridium acetobutylicum*. *J. Bacteriol.* 186, 2006–2018.

Tracy, B.P., Gaida, S.M., Papoutsakis, E.T., 2008. Development and application of flow-cytometric techniques for analyzing and sorting endospore-forming clostridia. Appl. Environ. Microbiol. 74, 7497–7506.

Tracy, B.P., Gaida, S.M., Papoutsakis, E.T., 2010. Flow cytometry for bacteria: enabling metabolic engineering, synthetic biology and the elucidation of complex phenotypes. Curr. Opin. Biotechnol. 21, 85–99.

Tummala, S.B., Welker, N.E., Papoutsakis, E.T., 2003a. Design of antisense RNA constructs for downregulation of the acetone formation pathway of *Clostridium acetobutylicum*. J. Bacteriol. 185, 2973.

Tummala, S.B., Junne, S.G., Paredes, C.J., Papoutsakis, E.T., 2003b. Transcriptional analysis of product-concentration driven changes in cellular programs of recombinant *Clostridium acetobutylicum* strains. Biotechnol. Bioeng. 84, 842–854.

Van Der Westhuizen, A.V.D., Jones, D.T., Woods, D.R., 1982. Autolytic activity and butanol tolerance of *Clostridium acetobutylicum*. *Appl. Environ. Microbiol.* 44, 1277–1281.

Veeranagouda, Y., Karegoudar, T.B., Neumann, G., Heipieper, H.J., 2006. *Enterobacter* sp. VKGH12 growing with n-butanol as the sole carbon source and cells to which the alcohol is added as pure toxin show considerable differences in their adaptive responses. FEMS Microbiol. Lett. 254, 48–54.

Venkataramanan, K.P., Jones, S.W., McCormick, K.P., Kunjeti, S.G., Ralston, M.T., Meyers, B.C., Papoutsakis, E.T., 2013. The *Clostridium* small RNome that responds to stress: the paradigm and importance of toxic metabolite stress in *C. acetobutylicum*. BMC Genomics 14, 849.

Venkataramanan, K.P., Kunniswan, Y., Boatman, J.J., Haynes, C.H., Taconi, K.A., Martin, L., Bothun, G.D., Scholz, C., 2014. Homeoviscous response of *Clostridium pasteurianum* to butanol toxicity during glycerol fermentation. *J. Biotechnol.* 179, 8–14.

Venkataramanan, K.P., Min, L., Hou, S., Jones, S.W., Ralston, M.T., Lee, K.H., Papoutsakis, E.T., 2015. Complex and extensive post-transcriptional regulation revealed by integrative proteomic and transcriptomic analysis of metabolite stress response in *Clostridium acetobutylicum*. Biotechnol. Biofuels 8, 81.

Ventura, J.R.S., Hu, H., Jahng, D., 2013. Enhanced butanol production in *Clostridium acetobutylicum* ATCC 824 by double overexpression of 6-phosphofructokinase and pyruvate kinase genes. Appl. Microbiol. Biotechnol. 97, 7505–7516.

Vollherbst-Schneck, K., Sands, J.A., Montenecourt, B.S., 1984. Effect of butanol on lipid composition and fluidity of *Clostridium acetobutylicum* ATCC 824. Appl. Environ. Microbiol. 47, 193–194.

Wallenius, J., Vilkilä, M., Survase, S., Ojamo, H., Eerikäinen, T., 2013. Constraint-based genome-scale metabolic modeling of *Clostridium acetobutylicum* behavior in an immobilized column. Bioresour. Technol. 142, 603–610.

Wang, F., Kashket, S., Kashket, E.R., 2005. Maintenance of ΔpH by a butanol-tolerant mutant of *Clostridium beijerinckii*. Microbiology 151, 607–613.

Wang, Z., Gerstein, M., Snyder, M., 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet. 10, 57–63.

Wang, Y., Li, X., Mao, Y., Blaschek, H.P., 2011. Single-nucleotide resolution analysis of the transcriptome structure of *Clostridium beijerinckii* NCIMB 8052 using RNA-Seq. BMC Genomics 12, 479.

Wang, Y., Li, X., Mao, Y., Blaschek, H.P., 2012. Genome-wide dynamic transcriptional profiling in *Clostridium beijerinckii* NCIMB 8052 using single-nucleotide resolution RNA-Seq. BMC Genomics 13, 102.

Wang, Q., Venkataramanan, K.P., Huang, H., Papoutsakis, E.T., Wu, C.H., 2013. Transcription factors and genetic circuits orchestrating the complex, multilayered response of *Clostridium acetobutylicum* to butanol and butyrate stress. BMC Syst. Biol. 7, 120.

Wang, Y., Li, X., Blaschek, H.P., 2013a. Effect of supplementary butyrate on butanol production and the metabolic switch in *Clostridium beijerinckii* NCIMB 8052: genome-wide transcriptional analysis with RNA-Seq. Biotechnol. Biofuels 6, 138.

Wang, Y., Tao, F., Tang, H., Xu, P., 2013b. Genome sequence of *Clostridium diolis* strain DSM 15410, a promising natural producer of 1,3-propanediol. *Genome Announc.* 1 (e00542-13-e00542-13).

Wang, Y., Zhang, Z.T., Seo, S.O., Choi, K.J., Lu, T., Jin, Y.S., Blaschek, H.P., 2015. Markerless chromosomal gene deletion in *Clostridium beijerinckii* using CRISPR/Cas9 system. J. Biotechnol. 200, 1–5.

Wang, Y.-F., Tian, J., Ji, Z.-H., Song, M.-Y., Li, H., 2016. Intracellular metabolic changes of *Clostridium acetobutylicum* and promotion to butanol tolerance during biobutanol fermentation. Int. J. Biochem. Cell Biol. 78, 297–306.

Wang, S., Dong, S., Wang, P., Tao, Y., Wang, Y., 2017. Genome editing in *Clostridium saccharoperbutylacetonicum* N1-4 with the CRISPR-Cas9 system. *Appl. Environ. Microbiol.* 83, e00233–17.

Webster, J.R., Reid, S.J., Jones, D.T., Woods, D.R., 1981. Purification and characterization of autolysin from *Clostridium acetobutylicum*. Appl. Environ. Microbiol. 41, 371–374.

Wietzke, M., Bahl, H., 2012. The redox-sensing protein Rex, a transcriptional regulator of solventogenesis in *Clostridium acetobutylicum*. Appl. Microbiol. Biotechnol. 96, 749–761.

Winkler, J., Kao, K.C., 2011. Transcriptional analysis of *Lactobacillus brevis* to n-butanol and ferulic acid stress responses. PLoS One 6, e21438.

Wu, Y., Xue, C., Chen, L.-J., Wan, H.-H., Bai, F.-W., 2015. Transcriptional analysis of micronutrient zinc-associated response for enhanced carbohydrate utilization and earlier solventogenesis in *Clostridium acetobutylicum*. *Sci. Rep.* 5, 16598.

Wu, P., Wang, G., Wang, G., Borresen, B.T., Liu, H., Zhang, J., 2016a. Butanol production under microaerobic conditions with a symbiotic system of *Clostridium acetobutylicum* and *Bacillus cereus*. Microb. Cell Factories 15, 8.

Wu, Y., Xue, C., Chen, L., Yuan, W., Bai, F., 2016b. Synergistic effect of calcium and zinc on glucose/xylose utilization and butanol tolerance of *Clostridium acetobutylicum*. *FEMS Microbiol. Lett.* 363, fnw023.

Wuster, A., Babu, M.M., 2008. Conservation and evolutionary dynamics of the agr cell-to-cell communication system across firmicutes. J. Bacteriol. 190, 743–746.

Xiao, Y., Francke, C., Abee, T., Wells-Bennik, M.H., 2011. Clostridial spore germination versus bacilli: genome mining and current insights. *Food Microbiol.* 28, 266–274.

Xiaoqiang, J., Liu, C., Song, H., Ding, M., Du, J., Ma, Q., Yuan, Y., 2016. Design, analysis and application of synthetic microbial consortia. *Synth. Syst. Biotechnol.* 1, 109–117.

Xu, M., Zhao, J., Yu, L., Tang, I.-C., Xue, C., Yang, S.T., 2015. Engineering *Clostridium acetobutylicum* with a histidine kinase knockout for enhanced n-butanol tolerance and production. Appl. Microbiol. Biotechnol. 99, 1011–1022.

Xue, C., Zhao, J., Lu, C., Yang, S.-T., Bai, F., Tang, I.-C., 2012. High-titer n-butanol production by *Clostridium acetobutylicum* JB200 in fed-batch fermentation with intermittent gas stripping. *Biotechnol. Bioeng.* 109, 2746–2756.

Xue, C., Zhao, J., Chen, L., Yang, S.-T., Bai, F., 2017. Recent advances and state-of-the-art strategies in strain and process engineering for biobutanol production by *Clostridium acetobutylicum*. Biotechnol. Adv. 35, 310–322.

Yang, S.T., Zhao, J. 2013. Adaptive engineering of *Clostridium* for increased butanol production. United States Patent, US 8,450,093 B1.

Yang, H., Haddad, H., Tomas, C., Alsaker, K., Papoutsakis, E.T., 2003. A segmental nearest neighbor normalization and gene identification method gives superior results for DNA-array analysis. *Proc. Natl. Acad. Sci.* 100, 1122–1127.

Yang, Y., Hoogewind, A., Moon, Y.H., Day, D., 2016. Production of butanol and isopropanol with an immobilized *Clostridium*. *Bioprocess Biosyst. Eng.* 39, 421–428.

Yang, Y., Zhang, L., Huang, H., Yang, C., Yang, S., Gu, Y., Jiang, W., 2017. A flexible binding site architecture provides new insights into CcpA global regulation in gram-positive bacteria. MBio 8. http://dx.doi.org/10.1128/mBio.02004-16.

Yen, J.Y., Nazem-Bokaee, H., Freedman, B.G., Athamneh, A.I.M., Senger, R.S., 2013. Deriving metabolic engineering strategies from genome-scale modeling with flux ratio constraints. Biotechnol. J. 8, 581–594.

Yin, J.Q., Zhao, R.C., 2007. Identifying expression of new small RNAs by microarrays. Methods 43, 123–130.

Ying, X., Cao, Y., Wu, J., Liu, Q., Cha, L., Li, W., 2011. sTarPicker: a method for efficient

prediction of bacterial sRNA targets based on a two-step model for hybridization. PLoS One 6, e22705.
Yoo, M., Bestel-Corre, G., Croux, C., Riviere, A., Meynial-Salles, I., Soucaille, P., 2015. A quantitative system-scale characterization of the metabolism of *Clostridium acetobutylicum*. MBio 6, e01808–15.
Yoo, M., Croux, C., Meynial-Salles, I., Soucaille, P., 2016. Elucidation of the roles of *adhE1* and *adhE2* in the primary metabolism of *Clostridium acetobutylicum* by combining in-frame gene deletion and a quantitative system-scale approach. Biotechnol. Biofuels 9, 92.
Youngleson, J.S., Jones, W.A., Jones, D.T., Woods, D.R., 1989. Molecular analysis and nucleotide sequence of the *adh1* gene encoding an NADPH-dependent butanol dehydrogenase in the Gram-positive anaerobe *Clostridium acetobutylicum*. Gene 78, 355–364.
Yu, M., Zhang, Y., Tang, I.-C., Yang, S.-T., 2011. Metabolic engineering of *Clostridium tyrobutyricum* for *n*-butanol production. Metab. Eng. 13, 373–382.
Yuan, Y., Bi, C., Nicolaou, S.A., Zingaro, K.A., Ralston, M., Papoutsakis, E.T., 2014. Overexpression of the *Lactobacillus planarum* peptidoglycan biosynthesis murA2 gene increases the tolerance of *Escherichia coli* to alcohols and enhances ethanol production. Appl. Microbiol. Biotechnol. 98, 8399–8411.
Zhang, Y.-X., Perry, K., Vinci, V.A., Powell, K., Stemmer, W.P.C., de Cardayré, S.B., 2002. Genome shuffling leads to rapid phenotypic improvement in bacteria. Nature 415, 644–646.
Zhang, L., Nie, X., Ravcheev, D.A., Rodionov, D.A., Sheng, J., Gu, Y., Yang, C., Jiang, W., Yang, C., 2014. Redox-responsive repressor Rex modulates alcohol production and oxidative stress tolerance in *Clostridium acetobutylicum*. *J. Bacteriol.* 196, 3949–3963.
Zhang, Y., Jiao, S., Lv, J., Du, R., Yan, X., Wan, C., Zhang, R., Han, B., 2017. Sigma factor regulated cellular response in a non-solvent producing *Clostridium beijerinckii* degenerated strain: a comparative transcriptome analysis. *Front. Microbiol.* 8. http://dx.doi.org/10.3389/fmicb.2017.00023.
Zhao, Y., Hindoeff, L.A., Chuang, A., Monroe-Augustus, M., Lyristis, M., Harrison, M.L., Rudolph, F.B., Bennett, G.N., 2003. Expression of a cloned cyclopropane fatty acid synthase gene reduces solvent formation in *Clostridium acetobutylicum* ATCC 824. Appl. Environ. Microbiol. 69, 2831–2841.
Zhao, X., Condruz, S., Chen, J., Jolicoeur, M., 2016. A quantitative metabolomics study of high sodium reponse in *Clostridium acetobutylicum* ATCC 824 acetone-butanol-ethanol (ABE) fermentation. *Sci. Rep.* 6, 28307.
Zhuang, W., Yang, J., Wu, J., Liu, D., Zhou, J., Chen, Y., Ying, H., 2016. Extracellular polymer substances and the heterogeneity of *Clostridium acetobutylicum* biofilm induced tolerance to acetic acid and butanol. *RSC Adv.* 6, 33695.
Zingaro, K.A., Papoutsakis, E.T., 2012. Toward a semisynthetic stress response system to engineer microbial solvent tolerance. MBio 3, e00308–12.
Zingaro, K.A., Papoutsakis, E.T., 2013. GroESL overexpression imparts *Escherichia coli* tolerance to i-, n-, and 2-butanol, 1,2,4-butanetriol and ethanol with complex and unpredictable pattern. *Metab. Eng.* 15, 196–205.
Zingaro, K.A., Nicolaou, S.A., Papoutsakis, E.T., 2013. Dissecting the assays to assess microbial tolerance to toxic chemicals in bioprocessing. *Trends Biotechnol.* 31, 643–653.

5.6 Article VI

SEDLAR, Karel, KOSCOVA, Pavlina, VASYLKIVSKA, Maryna, BRANSKA, Barbora, KOLEK, Jan, KUPKOVA, Kristyna, PATAKOVA, Petra and PROVAZNIK, Ivo. Transcription profiling of butanol producer Clostridium beijerinckii NRRL B-598 using RNA-Seq. *BMC Genomics*. 2018. 19(1), p. 415. (2018 IF = 3.501, Q2 in BIOTECHNOLOGY & APPLIED MICROBIOLOGY)

DOI: https://doi.org/10.1186/s12864-018-4805-8

Sedlar *et al. BMC Genomics* (2018) 19:415
https://doi.org/10.1186/s12864-018-4805-8

BMC Genomics

RESEARCH ARTICLE Open Access

Transcription profiling of butanol producer Clostridium beijerinckii NRRL B-598 using RNA-Seq

Karel Sedlar[1*], Pavlina Koscova[1], Maryna Vasylkivska[2], Barbora Branska[2], Jan Kolek[2,3], Kristyna Kupkova[1,4], Petra Patakova[2] and Ivo Provaznik[1]

Abstract

Background: Thinning supplies of natural resources increase attention to sustainable microbial production of bio-based fuels. The strain *Clostridium beijerinckii* NRRL B-598 is a relatively well-described butanol producer regarding its genotype and phenotype under various conditions. However, a link between these two levels, lying in the description of the gene regulation mechanisms, is missing for this strain, due to the lack of transcriptomic data.

Results: In this paper, we present a transcription profile of the strain over the whole fermentation using an RNA-Seq dataset covering six time-points with the current highest dynamic range among solventogenic clostridia. We investigated the accuracy of the genome sequence and particular genome elements, including pseudogenes and prophages. While some pseudogenes were highly expressed, all three identified prophages remained silent. Furthermore, we identified major changes in the transcriptional activity of genes using differential expression analysis between adjacent time-points. We identified functional groups of these significantly regulated genes and together with fermentation and cultivation kinetics captured using liquid chromatography and flow cytometry, we identified basic changes in the metabolism of the strain during fermentation. Interestingly, *C. beijerinckii* NRRL B-598 demonstrated different behavior in comparison with the closely related strain *C. beijerinckii* NCIMB 8052 in the latter phases of cultivation.

Conclusions: We provided a complex analysis of the *C. beijerinckii* NRRL B-598 fermentation profile using several technologies, including RNA-Seq. We described the changes in the global metabolism of the strain and confirmed the uniqueness of its behavior. The whole experiment demonstrated a good reproducibility. Therefore, we will be able to repeat the experiment under selected conditions in order to investigate particular metabolic changes and signaling pathways suitable for following targeted engineering.

Keywords: *Clostridium beijerinckii* NRRL B-598, RNA-Seq transcriptome, ABE fermentation

Background

While a less costly petroleum refinery still represents the main source of fuels and chemicals, limited natural resources and nature protection have increased attention to sustainable production of bio-based products. These trends make biorefinery the future lucrative producer of renewable fuels and chemicals. Especially, the microbial production of solvents such as acetone, butanol, and ethanol (ABE) is currently of great interest [1]. Solventogenic *Clostridia* are widely studied for their ability to produce biofuels from biomass in ABE fermentation [2]. Unfortunately, different genera or even strains of these rod-shaped, gram-positive anaerobes show substantial differences in phenotypic traits, i.e. the ability to utilize different substrates and to produce different substances. Thus, the findings acquired using model organisms such as *C. acetobutylicum* ATCC 824 [3], *C. pasteurianum* DSM 525 [4], or *C. beijerinckii* NCIMB 8052 [5] cannot be applied in general. Fortunately, thanks to a massive reduction in sequencing costs, a wide range of complete

* Correspondence: sedlar@feec.vutbr.cz
[1]Department of Biomedical Engineering, Brno University of Technology, Technicka 12, 616 00 Brno, Czechia
Full list of author information is available at the end of the article

or at least draft genomes of solventogenic *Clostridia* are now available. These include various strains of *C. acetobutylicum*, *C. aurantibutyricum*, *C. beijerinckii*, *C. diolis*, *C. felsineum*, *C. pasteurianum*, *C. puniceum*, *C. roseum*, *C. saccharobutylicum*, and *C. saccharoperbutylacetonicum* [6]. *C. beijerinckii* strains, utilizing a wider range of substrates for solvent production seem to be the most robust, i.e. able to endure a wide range of environmental conditions, among these [7].

However, the knowledge of the genomic sequence itself does not provide any information regarding the gene regulation, which is crucial to improvements of the strains for industrial application. The study of gene expression is therefore irreplaceable in genome engineering. Current whole transcriptome sequencing technology, referred to as RNA-Seq, allows the study of transcription on a genome-wide scale with an unlimited dynamic range, compared to the older microarrays, which only enabled researchers to track preselected genes [8]. In this paper, we present transcriptome dynamics during the cultivation of the promising butanol producer, *C. beijerinckii* NRRL B-598 [9] (formerly misidentified as *C. pasteurianum* NRRL B-598 [10]) as a result of RNA-Seq profiling. Until now, only the transcription of six selected genes involved in sporulation and solvent production was studied for this strain using RT-qPCR, yet the study supported the theory that solventogenesis is not regulated in the same way in all solventogenic clostridia [11]. Here, we further investigate the specifics of the strain *C. beijerinckii* NRRL B-598. The obtained transcriptome data includes the whole life cycle of the strain and therefore covers changes in metabolism, i.e. acidogenesis, solventogenesis and their transition state. Together with the sporulation cycle and other significant events such as changing motility and adaptation to acid/solvent stress, the whole fermentation process is reflected in this dataset. Flow cytometry, combined with fluorescent staining [12], has enabled insights into population heterogeneity and HPLC analysis of metabolites/substrate; plus, growth curve data has allowed us to better interpret the biological meaning. Moreover, the RNA-Seq technology has allowed us to study not only the temporal transcription of any gene but also to explore the accuracy of the current genome annotation. Compared to the transcription profiling of the strain *C. beijerinckii* NCIMB 8052, we reached a dynamic range that was approximately 10 times higher. To increase the robustness and validity of the experiment, each of the time-points was represented by three biological replicates rather, than verification using qPCR [13].

Results

Cultivation and fermentation kinetics

The fermentation profile of *C. beijerinckii* NRRL B-598 showed a typical two-stage course of metabolites formation with acid production in the first period followed by solvents formation (see Fig. 1a). Six time-points (T1–T6) were selected for RNA-Seq analysis to cover all metabolic stages within a period of 23 h. The latter stages were not analyzed due to a high percentage of dead and lysing cells (Fig. 1b) causing an insufficient quality of RNA samples for RNA-Seq. Individual sampling points were selected based on the fermentation pattern, which was monitored on-line as changes in a pH course (Fig. 1c). The first sample was collected after an approximate five-fold increase in optical cell density (Fig. 1d) while a sharp decrease in pH occurred, so only acidogenic, non-sporulating and mostly motile cells were expected to be present in the sample. The second time-point was proposed to cover a transient physiological state between acidogenesis and solventogenesis, which was indicated by a pH breakpoint and corresponded to the highest concentration of acids in media along with the onset of solvent formation. No cell-thickening or pre-spore formation was observed at this stage. The third sample set was withdrawn during the period of the most progressive rise in pH, suggesting a high rate of reutilization of the acids, together with solvent formation. Granulose accumulation and early phases of sporulation were observed at this stage (see Additional file 1). The second pH breakpoint was covered by the fourth sample, where the rise in pH ceased and pH again started to decline, indicating a change in metabolism. However, there was no apparent increase in the production of acids in the fermentation data. The remaining two samples were taken at the regular time-intervals, in order to cover all stages of ABE fermentation as well as the sporulation cycle. Overall culture fitness and spore formation was monitored by flow cytometry (FC) and the combined staining of cell culture by membrane disruption and enzyme activity indicators: propidium iodide (PI) and carboxyfluorescein diacetate (CFDA), respectively. A relatively high amount of double-stained cells was present in the culture at all stages. A previous study by Kolek et al. [12] considered these double-stained cells as an active population consisting of cell doublets and sporulating cells; therefore, only PI-positive cells were counted as dead cells. The staining pattern of the Clostridium culture at different time-points revealed dynamic changes in proportion of active cells within the first 13 h, with a detectable drop at the period with the lowest pH (the sixth hour), thus supporting the presumption that cells are highly-stressed by the presence of organic acids together with a low pH (when values slightly below pH 5 were reached). After the 13th hour, viability gradually decreased and during the 23rd hour the first mature spores, released from mother cells, were observed. The FC data provided a better insight into viability changes compared to sole OD measurements, according to which the culture kept on growing steadily until the

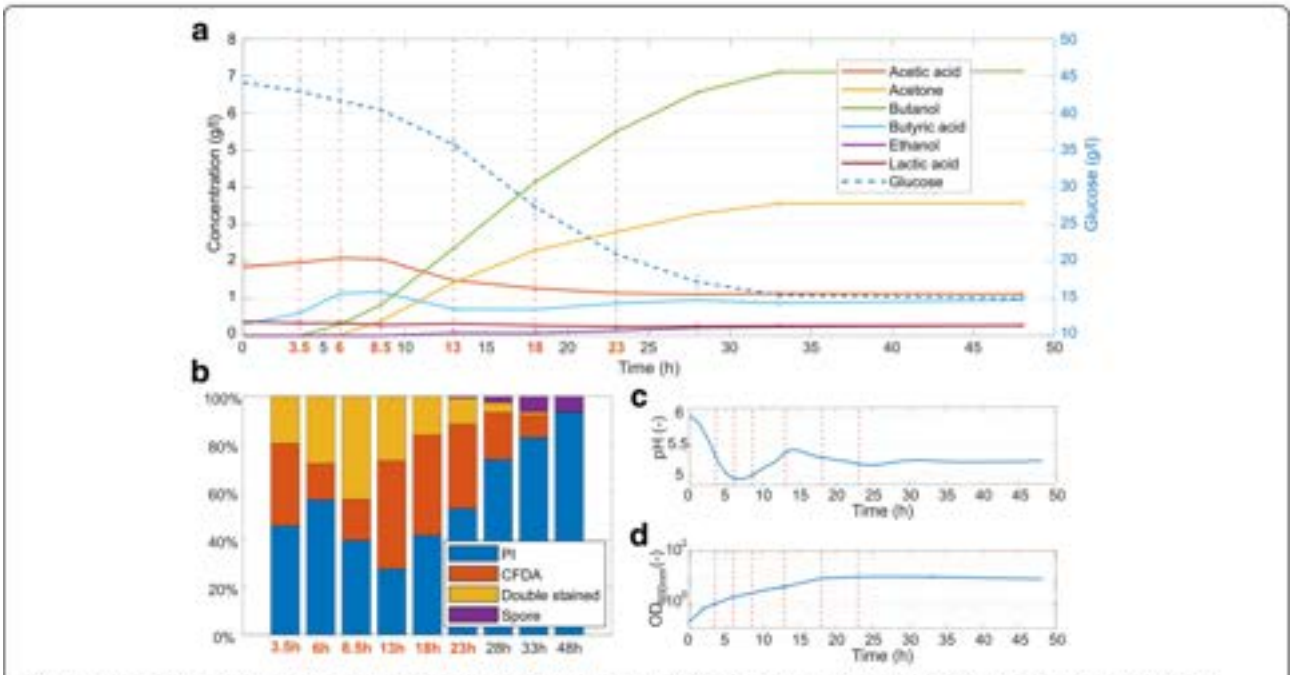


Fig. 1 Cultivation and fermentation characteristics of *Clostridium beijerinckii* NRRL B-598. (**a**) The concentration of glucose, solvents and acids during ABE fermentation. (**b**) Flow cytometry – the distribution of cells within the population according to their fluorescence pattern for combined staining using PI and CFDA. (**c**) pH curve for respective cultivation. (**d**) Cell growth measured as optical density at 600 nm. Values represent the mean of the biological replicates and error bars represent the standard deviations. Time-points (T1–T6) for samples subjected to RNA expression analysis are indicated by red vertical dotted lines and/or by red text labels

18th hour. The only noticeable changes in the OD measurements are the two slowdowns during the acidogenesis/solventogenesis transient states. The FC data clearly shows that culture viability had already started to decline at around the 13th hour, which corresponds to the apparent decrease in the number of regulated genes from that time.

A proportion of viable cells determined by FC was used to calculate the specific glucose consumption rate relating only to the active portion of clostridium culture (see Table 1). The amount of glucose consumed per time and biomass unit could help to elucidate the differences in expressions of glycolysis-related genes. The highest number of 5.16 g of utilized glucose per gram of active biomass and hour was reached at the very beginning. Surprisingly, after a decrease in the acid/solvent switch, the glucose consumption increased again and accompanied the T3–T4 transition state with the highest number of regulated genes.

Table 1 Specific rate of glucose utilization between time-points chosen for RNA-seq analysis

Samples	Time interval (h)	Specific glucose consumption rate ($g.g^{-1}_{a}.h^{-1}$)
T1-T2	3.5–6.0	5.16
T2-T3	6.0–8.5	2.20
T3-T4	8.5–13.0	2.71
T4-T5	13.0–18.0	2.50
T5-T6	18.0–23.0	1.59

[a]Values were calculated for the concentration of viable cells

Mapping statistics

The whole dataset covered three series of six samples (six time-points), in which each series represented an independent biological replicate (A, B, and C). Although series A consisted of reads that were 50 bp long and series B and C consisted of reads that were 75 bp long, the whole series could be processed in the same way. The quality assessment after the first preprocessing steps (demultiplexing, quality trimming, and adapter trimming) confirmed an overall high-quality of sequences (average Phred score Q ≈ 35) and no adapter content. The only following sequence-filtering step was the removal of the remaining residual rRNA contamination, even after the rRNA depletion. The rRNA depletion was performed prior to the library construction and the non-captured rRNAs were apparent from the high GC content in some reads. The amount of non-rRNA reads ranged from 7.3 to 20.5 million (see Fig. 2a). Subsequently, we mapped the cleansed reads to the *C. beijerinckii* NRRL B-598 genome. Most reads mapped to the genome

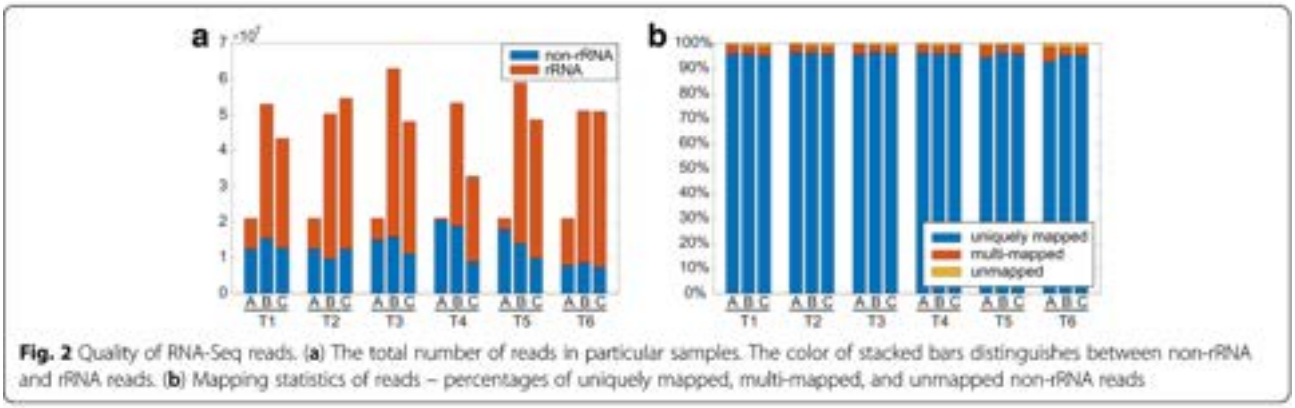


Fig. 2 Quality of RNA-Seq reads. (**a**) The total number of reads in particular samples. The color of stacked bars distinguishes between non-rRNA and rRNA reads. (**b**) Mapping statistics of reads – percentages of uniquely mapped, multi-mapped, and unmapped non-rRNA reads

unambiguously, regardless of their different length in replicates A and B, C (see Fig. 2b). Nevertheless, in order to cover the expression of duplicated genes that were present in the *C. beijerinckii* NRRL B-598 genome, the reads mapping to multiple loci were also included in the gene expression analysis (see Table 2). However, the contribution of such reads was down-weighted in the expression analysis depending on the number of times they mapped to the genome, so the sum of the total number of reads stayed intact.

The reads mapping to more genomic objects were also weighted. Such a phenomenon is caused by overlapping genes. In the current RefSeq genome (NZ_CP011966.2), 285 out of the 5230 genes predicted by NCBI PGAP [14] overlapped by at least one codon and another 66 neighboring genes had no space between them. Although none of the 198 pseudogenes overlapped with another pseudogene, 18 pseudogenes overlapped with genes directly and another 73 pseudogenes were at a distance from genes that could be covered by a single read. These reasons caused single read mapping onto two genomic objects. At the same time, the transcriptome assembly contained fewer transcripts compared to the number of genomic elements with detectable transcription (precisely 4837 transcripts vs. 5418 genomic elements) because the overlapping and nearby genes, e.g. those in the same operon, were covered by a single transcript. Due to this fact, transcripts could not have been used to resolve overlapping genes. On the other hand, their mapping to the genome helped to confirm or disprove transcriptional activity of pseudogenes and prophages.

Pseudogenes

Due to the high number of pseudogenes with detectable expression, we decided to further investigate their coverage by RNA-Seq reads. Only a single pseudogene remained completely silent when ambiguously mapping reads were used, while 184 pseudogenes had RPKM > 1 (Reads Per Kilobase per Milion mapped reads) in all six time-points. Using only uniquely mapped reads, eight pseudogenes remained completely silent and 178 were transcribed in every time-point. Although the number of transcribed pseudogenes remained almost the same across the six time-points, levels of their expression seemed to rise over time. While pseudogenes formed approximately 2.8% of *C. beijerinckii* NRRL B-598 genome, only 0.47% of all reads in T1 mapped to pseudogenes. However, this number continuously rose over the time according to the linear model $\%mapped = 0{,}1115 \cdot time - 0.0629$ (with the regression value 0.9575), resulting in 2.83% of reads to be mapped onto pseudogenes in T6.

To further analyze the activity of pseudogenes, we decided to evaluate the coverage of pseudogenes through the use of transcripts assembled from all the reads in our dataset. The accuracy of mapping transcripts to the genome is higher thanks to their length (1057 bp on average). The results are summarized in Table 3.

There are 24 pseudogenes that were not covered by any transcript. These were probably completely silent (see Additional file 2). The second group consisted of 78 pseudogenes that were not covered in their whole length. In most cases, there were only short overlaps with transcripts

Table 2 Transcriptional activity of genes and pseudogenes

Sample	T1 (3.5 h)	T2 (6 h)	T3 (8.5 h)	T4 (13 h)	T5 (18 h)	T6 (23 h)	Total
No. of genes with RPKM>1[a]	5055 (4981)	5101 (5026)	5162 (5100)	5197 (5139)	5198 (5133)	5193 (5128)	5219 (5158)
No. of pseudogenes with RPKM>1[a]	188 (179)	186 (179)	190 (184)	196 (190)	195 (188)	194 (187)	197 (190)
Max. expression (RPKM)	$4.0 \cdot 10^4$	$3.4 \cdot 10^4$	$3.4 \cdot 10^4$	$3.4 \cdot 10^4$	$3.4 \cdot 10^4$	$4.0 \cdot 10^4$	$4.0 \cdot 10^4$

[a]Values in brackets apply to uniquely mapped reads only

Table 3 Coverage of pseudogenes by transcripts

	Not covered	Partly covered	Fully covered, overlapped transcripts	Fully covered, single transcript
Frameshifted	7	45	10	33
Missing start and/or stop	15	23	3	37
Internal stop	2	6	0	8
Combined issues	0	4	3	2
Total	24	78	16	80

of active genes neighboring these pseudogenes. In some cases, only part of a transcript was mapped to a pseudogene sequence, suggesting that these are silenced duplications of an active gene. Although genes in the third group were fully covered, this coverage consisted of two or more overlapping transcripts. Therefore, the transcription in both groups (partly covered and fully covered by overlapping transcripts) was highly questionable. On the contrary, pseudogenes within the fourth group were fully covered by unique transcripts. This group consisted of pseudogenes that were transcribed and active genes that were possibly misidentified as pseudogenes due to errors in the genome assembly. In comparison with their transcripts, 23 out of 80 pseudogenes (see Additional file 3) in this group were missing one nucleotide in homopolymers. This could have been caused by previous sequencing errors, as Roche 454 in combination with PacBio were used for the genome assembly. Nevertheless, insertion of these nucleotides was not detected in all reads mapping to these positions; the figure ranged from 60% to almost 100%.

Transcription profiles and reproducibility

Only 11 genes were not transcribed at any of the six sampling points. Moreover, seven out of those 11 genes were related to 16S rRNA and these reads were filtered before mapping. Therefore, only four genes (X276_RS15615, X276_RS24570, X276_RS24585, X276_RS26445) demonstrated no transcripts. On the other hand, 5024 genes out of all 5219 transcribed genes (RPKM> 1) had detectable transcription at all time-points. Nevertheless, it is difficult to decide whether the expression of genes with low RPKM values has biological meaning, due to a high biological noise. Analysis using assembled transcripts is complicated, because most transcripts cover more than one gene and transcripts overlap. Transcription on a genome-wide scale (see Additional file 4) shows a novel pattern. While the transcriptional profiles from the first three time-points (T1, T2, and T3) correspond to the transcription of the *C. beijerinckii* NCIMB 8052 genome [5], the latter profiles do not.

Reproducibility of the experiment was verified using three biological replicates and by checking the expression of six selected genes whose transcription profiles were observed during a previous study by Kolek et al. [11] (see Fig. 3a). The samples were visualized using the t-Distributed Stochastic Neighbor Embedding (t-SNE) [15] dimensionality reduction method on the normalized expression data. This final 2D representation showed that replicates (A, B, and C) were similar to each other at particular sampling times (T1–T6), while replicates sequenced using Illumina HiSeq (A) were slightly more distant to samples from Illumina NextSeq (B and C), see Fig. 3b. Overall, samples were divided into two clusters. While one cluster contained samples corresponding to the initial phase of fermentation (up to 8.5th hour), the other cluster consisted of samples from the later fermentation phase (from 13th up to 23rd hour).

Differential expression

We explored differential expression of all genes and pseudogenes with detectable transcription among adjacent time-points, in order to analyze changes in the transcription of particular genes over the whole fermentation process (see Fig. 4). In total, transcription of 2260 annotated genomic objects, forming more than 41.5% of all protein-coding elements, was regulated during the fermentation process when the criterion of adjusted p-value < 0.05 (Benjamini-Hochberg correction) was applied. While 474 genes were regulated more than once, only 31 of them were regulated more than three times. The single gene X276_RS14155 (PTS maltose transporter subunit IIBC) was regulated four times. The majority of differentially expressed genes were covered by at least 100 reads after the normalization of expression data (see Additional file 5). In total 3168 genes had no statistically significant regulations among adjacent time-points and formed potential housekeeping genes. The complete results of the differential expression analysis, including log2fold changes and adjusted p-values, are available in Additional file 6.

A major change was detected between the third and the fourth time-point when 1582 genes were regulated. While 835 out of these genes were up-regulated, 714 were up-regulated only between these two time-points (see Fig. 4b). Similarly, 666 out of the 747 down-regulated genes were down-regulated uniquely between T3 and T4 (see Fig. 4c). However, some of the uniquely up-regulated genes were down-regulated between another couple of time points and some of the uniquely down-regulated genes were up-regulated during another

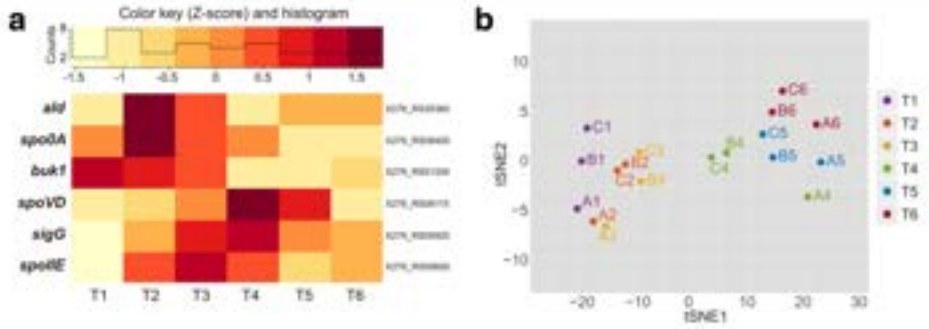


Fig. 3 Analysis of the transcriptome reproducibility. (**a**) Transcription profiles of six selected genes visualized on the heatmap using a Z-score related to an average expression of each gene. (**b**) 2D representation of the normalized expression data after dimensionality reduction by t-SNE to compare the samples collected at the six time-points (T1–T6) coded by different colors. Each point represented a sample with a text label based on the biological replicate (A, B, and C) and the time-point from which it originated (T1–T6)

transition. Therefore, the total number of uniquely regulated genes between the T3 and T4 time-points was 1174. Every pair of adjacent time-points had uniquely regulated genes except for the last T5–T6 transition, when regulation of only six already regulated genes was detected. Nevertheless, previously up-regulated genes X276_RS05345 (hypothetical protein) and X276_RS24350 (butyrate kinase) were down-regulated between these later time-points. Both up-regulated genes during this transition, X276_RS08605 (tryptophan synthase subunit beta) and X276_RS18605 (DUF4179 domain-containing protein), also had detectable growth in transcription between previous time-points and were covered by more than 1000 and 2000 reads, respectively.

Transcription of phage DNA

We searched the *C. beijerinckii* NRRL B-598 genome for phage sequences and found three prophages (see Table 4). While two of these regions were relatively short and phages were incomplete, the other phage was intact and consisted of 35 genes coding known phage proteins and six hypothetical protein-coding regions.

The expression within the first phage region corresponding to an incomplete phage was low (averaging RPKM = 47) with only two genes differentially expressed during T3–T4 change. Six genes were carried by a positive and four by a negative strand. Only four genes were fully covered by transcripts mapping to the region. The transcription within the third phage region covering the

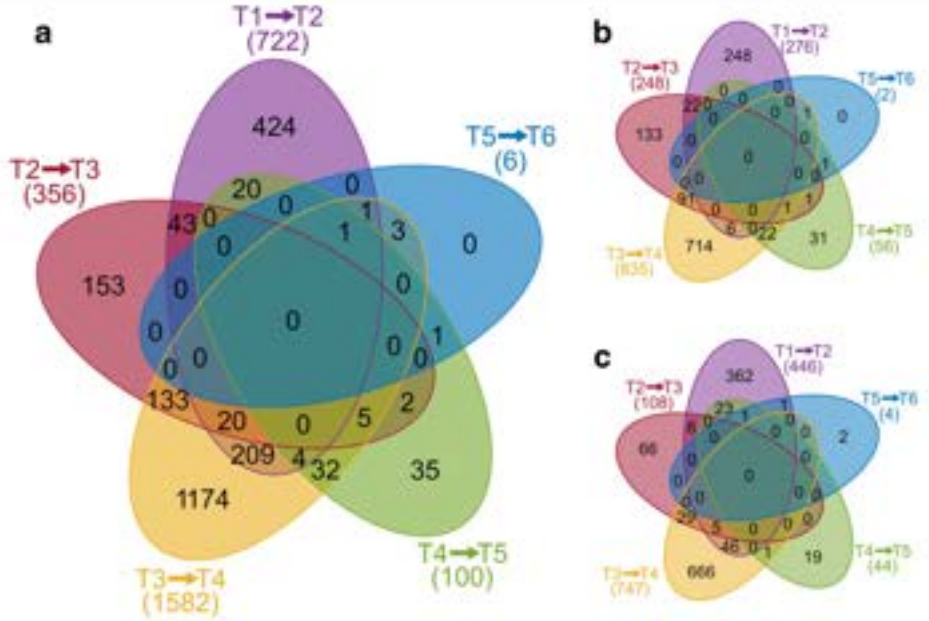


Fig. 4 Differential expression analysis. Venn diagrams showing the number of (**a**) all-regulated, (**b**) up-regulated, and (**c**) down-regulated genes between adjacent time-points

Table 4 Phage DNA within the *C. beijerinckii* NRRL B-598 genome

Region	Position	Length (bp)	Status	Total no. of proteins	No. of phage proteins
1	996,985..1006473	9488	incomplete	10	8
2	2,920,342..2960012	39,670	intact	41	35
3	4,005,361..4018720	13,357	incomplete	17	15

other incomplete phage was more active with average RPKM = 86, but none of the genes were differentially expressed during the fermentation. All genes were carried by a negative strand and 14 out of the 17 genes were covered by a single transcript, including one pseudogene (X276_RS17860) with a missing stop codon. The only region containing intact prophage consisted of 38 genes and three pseudogenes with a missing stop codon, carried by a positive strand. The whole region began with a pseudogene and had low transcription (averaging RPKM = 21). Although six genes had statistically significant differential expressions between T3 and T4, only short transcripts mapped to the region and only partly covered the genes. Thus, the phage remained silent.

Discussion

The fermentation data presented in Fig. 1 comply with standard results usually achieved by using the same TYA cultivation medium [11, 12]. Deeper insight into the population is enabled by combination of double fluorescent staining and flow cytometry. Value of flow cytometry had already been confirmed for *C. acetobutylicum* [16, 17]. Cytometric data enabled the calculation of a specific rate of glucose consumption related to metabolically active cells in the population during different time periods of the cultivation, together with information about the overall culture condition.

The high proportion of reads that mapped to the genome in particular samples unambiguously, suggested a good quality of RNA-Seq data and successful alignment even for shorter 50 bp reads in replicates A. Although we presumed that utilization of longer 75 bp reads in replicates B and C could reach even higher percentage of unique mapping, the proportion remained similar (see Fig. 2b). Nevertheless, the number of genes with detectable transcription slightly differed when reads mapping to multiple loci were used. Although high sequencing depth and rRNA depletion brought a noise to RNA-Seq [18], in our case, this bias was caused by duplicated genes rather than being a sequencing issue [19]. To prevent omitting transcription of duplicated genes and pseudogenes, we decided to include multi-mapping reads into the analysis. The majority of reads mapped to the genome without any mismatches and support an overall high quality of the genome assembly. Nevertheless, 23 indels were detected in regions of frameshifted pseudogenes.

Although pseudogenes, in bacteria defined as 'genes silenced by one or more deleterious mutations' [20], could still be transcribed [21], their number in *C. beijerinckii* NRRL B-598 was rather high. For example, the reference sequence for the closely related strain *C. beijerinckii* NCIMB 8052 [13] (NC_009617.1) contained only 112 pseudogenes predicted by NCBI PGAP. While the number of pseudogenes with an incomplete coding region or those containing internal stop was comparable for both strains, the number of pseudogenes with frameshift was almost twice as high in *C. beijerinckii* NRRL B-598 genome. Although the high number of frameshifted genes could indicate an extraordinary number of frameshifted duplicates of genes, all 23 indels were detected in homopolymers. Therefore, such pseudogenes could also be misannotated genes due to pyrosequencing errors [22] that were not filtered out using PacBio RSII sequencing used for the complete genome assembly [9]. Nevertheless, 50 bp and 75 bp long reads were too short to distinguish between a frameshifted duplicate and an assembly error as no indels were present in 100% of reads mapping to ambiguous positions. Eventually, the activity of some pseudogenes was supported in differential expression analysis, by high log2foldchange, excessing a value of three.

The transcriptome of *C. beijerinckii* NRRL B-598 had never been studied before so no correlation to the older dataset could be carried out. However, the transcription of the six selected genes under the same cultivation conditions was monitored using qRT-PCR in study of *C. beijerinckii* NRRL B-598 and its mutant strain overexpressing sporulation initiation factor *spo0A* [11]. In the mentioned study by Kolek et al. [11], an increase in expression was observed in mid-cultivation for *spoIIE* and *sigG* and in the second part of cultivation for *spoVD*. This corresponded to the results of this study (see Fig. 3a). Moreover, the expression profiles of the remaining genes also showed the same pattern. Butyrate kinase (*buk*, X276_RS1200) transcription was maximal at the beginning of the cultivation, decreased in time, and rose slightly at the end of cultivation. The expression of *ald* and *spo0A* increased in the first third of cultivation and for *ald* also at the end of cultivation. Moreover, the reproducibility of the experiment was supported by utilization of three biological replicates and their high similarity in the sampling points visualized using tSNE in Fig. 3. The tSNE coordinates were obtained by comparing distances among

samples in the original high-dimensional space, i.e. distances from the normalized expression profiles to the distances of the samples in the reduced space, i.e. the visualized points. The position of the samples in the 2D space was then optimized until the samples with similar expression profiles were placed close to each other and samples with very different expression profiles were at a further distance from each other. Two main clusters, distinguishing samples from the first and the second half of the experiment, were present. While the similarity of the replicates from the first cluster was supported mainly by the first coordinate tSNE1, the similarity in the other cluster was supported by the second coordinated tSNE2.

Wang et al. [13] observed similar clustering of RNA-Seq samples of *C. beijerinckii* NCIMB 8052, in which the first cluster was represented by samples from exponential and transition phases and the other by samples from a stationary phase. On the other hand, transcription profiles of *C. beijerinckii* NCIMB 8052 [5] and *C. beijerinckii* NRRL B-598 (see Additional file 4) on the genome-wide scale were different, especially in the later phase of cultivation. This could have been caused by structural reorganizations in the genomes of both strains or by differences in gene regulatory mechanisms. Due to the high similarity of both genomes (see Additional file 7), the latter seemed more relevant. The explanation for differences in transcription profiles of *C. beijerinckii* NRRL B-598 and *C. beijerinckii* NCIMB 8052 in the later phases could lie in the different phenotypic behavior of both strains at this stage. Although strain NCIMB 8052 ceased growing together with the start of solventogenesis [5, 13], strain NRRL B-598 continued growing until approximately half way through the solventogenic phase (see Fig. 1d). Another apparent difference was an increased number of mature spores formed by the NCIMB 8052 strain under similar cultivation conditions [12]. The genome of *C. beijerinckii* NRRL B-598 contained two housekeeping regions with stable high level of transcription activity that were not present in *C. beijerinckii* NCIMB 8052 genome. This high activity was caused by genes transcribing into cell wall binding proteins, in the first region by the gene X276_RS24890 with average RPKM $2.4{\cdot}10^4$, while in the second region by the gene X276_RS25120 with average RPKM $1.8{\cdot}10^4$. The most noticeable change in the transcription on the genome wide scale was captured between T3 and T4 time-points when the highest number of differentially expressed genes was detected. Increased activity was visible especially within the region spanning the position from 176,588 to 208,581 containing 45 genes whose average expression in RPKM rose from $1.9{\cdot}10^3$ to $3.0{\cdot}10^3$. Thirty-seven out of those genes code proteins belonged to the Clusters of Orthologous Groups of proteins (COG) functional group J associated with translation.

The massive change in the gene expression, which can be spotted in Fig. 4, was surprisingly not associated with the acidogenesis/solventogenesis switch that occurred earlier, mainly between the T2 and T3 time-points, neither with the sporulation initiation. Regarding the COG assignment of 45 abovementioned genes to group J (translation), it might be possible that at least a part of these genes corresponded with spore coat formation genes. Clostridial sporulation typically lasts 8–12 h and therefore the T4 time-point might have coincided with stage IV or V of a sporulation cycle in which formation of spore coat proteins occurred [23]. In addition to the coat proteins, a need for specific protein complexes involved in spore structures assemblies could be responsible for the increased protein formation demand.

Further transition between T4 and T5 could also show an entry to the irreversible phase of sporulation, in which two independent gene regulations were established in the mother cell and pre-spore and sporulation must be completed. Overall culture attenuation after T4 is apparent from both a decrease of specific glucose consumption (Table 1) and from cytometric data that confirmed the gradual increase in the proportion of inactive cells. An opposite phenomenon was observed between T3 and T4. An increase in the specific rate of glucose consumption, corresponding to highly regulated genes coding for COG functional group C (energy production and conversion) (see Additional file 8), was detected together with an apparently improved viability.

Even though the massive change between T3 and T4 was obvious, searching within COG categories (see Additional file 8) does not provide unambiguous clarification for this phenomenon. Mostly the same categories of regulated genes could be found between adjacent time-points within the first 13 h of cultivation with both down- and up-regulated representatives. After the 13th hour COG D and COG L related to cell cycle control and replication respectively were not differentially expressed which was fully consistent with the decrease in cell growth and declining culture viability supporting a hypothesis of the switch of a highly proliferating culture into a new strategy, securing genus preservation via ensuring a complete sporulation process. Simultaneously COG F for nucleotide metabolism transport are up-regulated within the first two compared time sets and down-regulated in the latter two. These findings were comparable to the transcriptional profile of *C. acetobutylicum* [24] unlike the category J which was in *C. acetobutylicum* down-regulated in the stationary phase. The same applied to the motility related genes (COG N) that were in our study more down-regulated even within the first measured interval and up-regulated in latter stages between T4 and T5. This might seem confusing as solventogenic clostridia are known to be motile within

the exponential and acidogenic stage [25] and after the switch to solventogenesis, motility is generally lost. *C. beijerinckii* NRRL B-598 possessed such a change in motility as well but the first sample point T1 was already characterized by highly motile cells and therefore a decrease in related genes expression copied the phenotypic profile. On the other hand, an increase in the latter stages is probably the result of culture phenotype desynchronization when all the cell types are again present, including motile cells. The predominant upregulation of COG O (post translational modification, protein turnover, chaperone function) between later stages might relate to cell stress response to increasing solvent concentrations [26].

Furthermore, some cells within the whole population might have undergone a massive change in energy metabolism and solvent production, which is associated with the switch of different genes in the period of transition between T3 and T4 time-points. The solvent formation and acidogenesis/solventogenesis switch are usually explained as a stress response induced by accumulation of acids in the cultivation medium and pH decrease. Low pH could cause depletion of ATP pool in cells because of active transport of H^+ ions across cell membrane. To prevent this event and to ensure population survival, some cells initiated sporulation, while other cells began converting acids into solvents. However, the whole population situation was no longer critical at time-point T4 and the lower concentration of acids in the medium might have induced another metabolic change, this time associated with the direct formation of butanol/acetone from glucose. As this pathway generated only a half of ATP in comparison with acidogenesis, its overall rate was probably higher. However, a significant advantage of the reduced risk of low pH outweighed this discomfort. Moreover, this hypothesis was supported by metabolites formations, glucose consumption, and pH profile (see Fig. 1a, c) and by an increase in specific glucose consumption. More than 20 years ago, Dürre et al. [27] envisaged for *C. acetobutylicum* that different genes are probably involved in early and late solventogenesis. Population heterogeneity reflected by FC and fluorescent staining (Fig. 1b) supports the hypothesis that not all cells in the population exhibit the same phenotype to cope with changing unfavorable living conditions. The population might rather choose the bet hedging strategy [28] to enable at least some cells from the population to survive.

Many bacterial genomes contain prophages or at least their remnants. Although they may represent large fraction of the strain-specific DNA sequences [29], the strain *C. beijerinckii* NRRL B-598 contained only three prophage regions while only one was complete. This could be the reason for a high genome sequence similarity with the strain *C. beijerinckii* NCIMB 8052 as the prophages are responsible for genome rearrangements and inversions [30]. Even though the complete prophage contained six differentially expressed genes between T3 and T4, their average transcription was very low suggesting false positive detection. Due to the absence of transcripts mapping to the prophage regions, all these three regions seemed to be silent. During industrial cultivations in the South Africa [31], there were several events mapped in which bacteriophages caused total collapse or reduction of solvents production due to lytic or lysogenic cycles, respectively. Therefore, the detected prophages deserve further experimental investigation.

Conclusions

Although the strain *C. beijerinckii* NRRL B-598 is a promising butanol producer, we lack a precise description of mechanisms within its fermentation metabolism, which prevent us from further modifications of the strain for industrial applications. Moreover, these mechanisms seems to be unique and different from other clostridia, including a closely related strain *C. beijerinckii* NCIMB 8052. In this study, we provided a complex analysis of its fermentation profile using HLPC, FC, and RNA-Seq technologies. Six time-points were selected to study its transcription profile, while the whole experiment was repeated in order to get three biological replicates (A, B, and C) for each time-point. This allowed us to verify the reproducibility of the experiment and to gather the RNA-Seq dataset with the currently highest dynamic range available among solventogenic clostridia. We analyzed the latest RefSeq annotation of the genome and confirmed its high accuracy. Nevertheless, through the analysis of single nucleotide variants, several putative missing nucleotides were found within the regions of frameshifted pseudogenes. Transcription regulations identified by differential expression analysis of adjacent time-points showed the greatest changes between T3 and T4 time-points. Surprisingly, this change was not directly connected to the acidogenic/solventogenic change, nor the sporulation initiation but rather to a massive change in the energy metabolism and solvent production in a part of cell population as we discuss based on auxiliary HLPC and FC data.

Furthermore, we discovered three prophage regions within the genome, which demonstrated low or no transcription activity. Nevertheless, these regions are important for further experimental investigation. The experimental design and the gathered data proved good reproducibility, therefore, repeating the experiment under different conditions will also allow us to explore gene regulatory mechanisms and signaling pathways within the strain.

Methods

Bacterial culture and fermentation experiment

The strain *C. beijerinckii* NRRL B-598 was maintained in a form of spore suspension. TYA broth, prepared according to Kolek et al. (2017) [11], containing: 50 g/l glucose, 6 g/l tryptone (Sigma Aldrich), 2 g/l yeast extract (Merck), 3 g/l ammonium acetate, 0.5 g/l KH_2PO_4, 0.3 g/l $MgSO_4{\cdot}7H_2O$, and 0.01 g/l $FeSO_4$, was used for the fermentation experiment. Multiforce 1 l bioreactors (Infors HT) with 630 ml TYA broth and agitation at 200 rpm were used for batch cultivation of the strain at 37 °C. Oxygen was removed from bioreactors by bubbling with N_2 prior to fermentation. pH was adjusted to 6.3 by 10% NaOH and all bioreactors were inoculated with 70 ml of inoculum that was cultured previously in an anaerobic chamber overnight (Concept 400; Ruskinn Technology) under an anaerobic atmosphere (90% N_2, 10% H_2). The whole experiment was repeated during different weeks to obtain three biological replicates.

Samples were taken at specific times and processed for cell concentration determination, HPLC analysis, microscopy, flow cytometry, and RNA isolation. Samples for RNA isolation were taken at 3.5, 6, 8.5, 13, 18, and 23 h of cultivation.

Culture growth and HPLC analysis

Cell concentration was determined by the optical density (OD) measurement at 600 nm with Spectrophotometer (Varian Cary 50 UV-VIS spectrophotometer, Varian) against TYA broth. For calculations of a specific glucose consumption rate, dry weight of biomass (CDW) was used. CDW was determined after drying biomass until constant weight at 105 °C. The equation was following:

$$q_p = \frac{c_{i+1} - c_i}{\overline{CDW_{i,i+1}} * \overline{X_{i,i+1}} * (t_{i+1} - t_i)}$$

where q_p is a specific substrate consumption rate related to a number of viable cells ($g.g^{-1}.h^{-1}$), c is concentration of glucose (g/L), CDW is cell dry weight (g/L), x is a proportion of viable cells in population and t is time (h). Symbols i and $i + 1$ indicate two adjacent sampling time points.

Concentrations of glucose and fermentation products (lactic acid, acetic acid, butyric acid, ethanol, acetone, and butanol) were measured by HPLC with refractive index detection (Agilent Series 1200 HPLC; Agilent) in microfiltered samples of culture broths. An IEX H+ polymer column (Watrex) was used for the separation. Conditions of analysis were as follows: isocratic elution, 5 mM H_2SO_4 as a mobile phase with flow rate of 0.5 ml min^{-1}, column temperature 60 °C, injection sample volume 20 µl. The chromatograms were processed by ChemStation for LC systems software using a set of standard samples with known concentrations to elaborate calibration curves.

Microscopy, fluorescent staining, and flow cytometry

Phase contrast microscopy (Olympus BX51; Olympus) with × 400 and × 1000 magnifications was used to determine the morphological status of cells. Population viability and heterogeneity was evaluated using flow cytometry (BD Accuri C6) in combination with fluorescent staining. A combination of propidium iodide PI (Sigma Aldrich) and carboxyfluorescein diacetate CFDA (Sigma Aldrich) was employed for the differentiation of active and damaged cells and detection of spores according to Kolek et al. (2016) [12].

RNA isolation and sequencing

Cell samples for isolation of total RNA were collected from 3 ml of culture broth (OD_{600} 0.9–1.0) by centrifugation at 10000 rpm for two minutes, washed with RNase free water and cell pellets were immediately stored at − 70 °C. RNA from the cell pellet was isolated using High Pure RNA Isolation Kit (Roche). Isolated total RNA was stored frozen at − 70 °C. The total RNA concentration was determined on DS-11 FX+ Spectrophotometer (DeNovix). Quality and integrity of the samples were assessed using the Agilent RNA 6000 Nano Kit (Agilent) with the Agilent 2100 Bioanalyzer (Agilent). RNA integrity number was measured using 2100 Bioanalyzer Expert software.

Frozen total RNA samples were thawed on ice and an aliquot of each sample containing 10 µg of RNA was taken for 16S and 23S ribosomal RNAs removal using The MICROB*Express*™ Bacterial mRNA Enrichment Kit (Ambion). Efficiency of ribosomal RNA depletion and concentration of RNA samples were checked on the Agilent 2100 Bioanalyzer (Agilent) with the Agilent RNA 6000 Nano Kit (Agilent). Library construction and sequencing of samples from the first replicate on Illumina HiSeq 4000, single-end, 50 bp, was performed by BGI Europe A/S (Copenhagen, Denmark). Library construction and sequencing of samples from two remaining replicates were performed by CEITEC Genomics core facility (Brno, Czechia) on Illumina NextSeq, single-end, 75 bp.

Bioinformatics analysis

The quality assessment after steps of the RNA-Seq reads processing was done using FastQC in combination with MultiQC to summarize the reports across all samples [32]. Reads representing 16S and 23S rRNA regions were filtered out using SortMeRNA [33] with SILVA database of known bacterial 16S and 23S rRNA genes [34] to simplify the following mapping of reads. Clean reads were mapped to the reference genome of *C. beijerinckii* NRRL B-598 (NZ_CP011966.2) using STAR [35]. Resulting SAM (Sequence Read Alignment/Map) files were

indexed and transformed into more compact BAM (Binary Read Alignment/Map) format using SAMtools [36].

Transcripts were assembled de novo from a whole dataset of 18 samples using Trinity v2.4.0 [37]. Transcripts were mapped to *C. beijerinckii* NRRL B-598 reference genome (NZ_CP011966.2) with BLAST+ v2.7.1 [38]. Mapped reads and transcripts were visualized as a graph of sequence read coverage across the genome and further explored in Integrative Genomics Viewer (IGV) v 2.4.3 [39] to capture variable regions, including identification of putative missing nucleotides in pseudogene region in the current genome assembly. On the other hand, genome-wide coverage plots were reconstructed with SAMtools using sorted reads and visualized as circular representations of genome with DNAplotter [40] integrated in Artemis [41]. Dotplot for visual comparison of *C. beijerinckii* NRRL B-598 and *C. beijerinckii* NCIMB 8052 genomes was produced in YASS genomic similarity search tool [42]. Phage regions in the *C. beijerinckii* NRRL B-598 genome were predicted with PHASTER [43] and PhiSpy [44]. In PhiSpy both available clostridial references (*C. perfringens* and *C. tetani*) were used.

A count table was reconstructed using the R/Bioconductor featureCounts function included in the Rsubread package [45] and RPKM were computed using the R/Bioconductor edgeR package [46]. Differential analysis was performed on a raw count table with R/Bioconductor DESeq2 package [47]. Data was normalized using a built-in DESeq2 function. This normalization used negative binomial distribution and handles both differences in library sizes and differences in library composition. DESeq2 identified genes that were differentially expressed in a time-dependent manner. Dimensionality reduction and visualization of normalized samples was produced with R Rtsne package using Barnes-Hut t-SNE implementation [48] in combination with ggplot2 R package [49]. Venn diagrams and heatmaps representing transcription of selected genes using Z score were generated with R packages VennDiagram [50] and gplots, respectively. Time series and bar plots were generated with Matlab 2017b.

Additional files

Additional file 1: Snapshots from microscopic observation during cultivation. (PDF 628 kb)

Additional file 2: Silent pseudogenes. (PDF 195 kb)

Additional file 3: Putative active genes misidentified as pseudogenes due to assembly errors. (PDF 210 kb)

Additional file 4: Circular plots showing average coverage of the genome by RNA-Seq reads in all six time points. The outermost and the second outermost circles represent positions of genes on the forward (red) and reverse (blue) strands respectively. The third circle (green) stands for pseudogenes. The yellow peak and shading area represents transcription greater than the average and violet lower than average. Floating window of 10,000 bp with step of 200 bp was used to render the shading area. (PDF 701 kb)

Additional file 5: Differential analysis of adjacent time points using MA plots. MA plots showing statistically differentially expressed genes in color. Color coding respect the color coding used in Venn diagrams in Fig. 4. (PDF 315 kb)

Additional file 6: Differential expression analysis. Complete results of differential expression analysis using DESeq2. (XLSX 2287 kb)

Additional file 7: Dotplot of *C. beijerinckii* NRRL B-598 and *C. beijerinckii* NCIMB 8052 genome. Dotplots showing that no major rearrangement between the two strains are present. (PDF 315 kb)

Additional file 8: COG functional categories of differential expressed genes. Barplots showing the number of COG categories associated with differentially expressed genes between adjacent time points. (PDF 226 kb)

Acknowledgements
Computational resources were partially provided by the CESNET LM2015042 and the CERIT Scientific Cloud LM2015085, under the programme "Projects of Large Research, Development, and Innovations Infrastructures". We acknowledge the CF Genomics of CEITEC supported by the NCMG research infrastructure (LM2015091 funded by MEYS CR) for their support with obtaining scientific data presented in this paper.

Funding
This work has been supported by grant project GACR 17-005515.

Availability of data and materials
The genome assembly referred in this paper is version CP011966.2 using NCBI RefSeq annotation NZ_CP011966.2. The RNA-Seq sequencing data have been deposited in the NCBI Sequence Read Archive (SRA) under the accession number SRP033480.

Authors' contributions
KS, JK, PP, and IP designed the study. MV and BB performed the experiments. KS, PK, and KK analyzed the data. KS, PP, and BB wrote the manuscript with the input from all authors. All authors discussed the results, read and approved the final manuscript.

Ethics approval and consent to participate
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Publisher's Note

Author details
[1]Department of Biomedical Engineering, Brno University of Technology, Technicka 12, 616 00 Brno, Czechia. [2]Department of Biotechnology, University of Chemistry and Technology Prague, Technicka 5, 166 28 Prague, Czechia. [3]Institute of Aquaculture and Protection of Waters, University of South Bohemia in České Budějovice, Na Sádkách 1780, 370 05 České Budějovice, Czechia. [4]Department of Biochemistry and Molecular Genetics, University of Virginia Health System, Charlottesville, VA 22908, USA.

Received: 20 February 2018 Accepted: 18 May 2018
Published online: 30 May 2018

References

1. Kujawska A, Kujawski J, Bryjak M, Kujawski W. ABE fermentation products recovery methods - A review. Renew. Sustain. Energy Rev. [Internet]. Pergamon; 2015 [cited 2017 Nov 23];48:648–661. Available from: http://www.sciencedirect.com/science/article/pii/S1364032115002981.
2. Patakova P, Linhova M, Rychtera M, Paulova L, Melzoch K. Novel and neglected issues of acetone-butanol-ethanol (ABE) fermentation by clostridia: Clostridium metabolic diversity, tools for process mapping and

continuous fermentation systems. Biotechnol Adv. [Internet]. Elsevier; 2013 [cited 2017 Nov 23];31:58–67. Available from: http://www.sciencedirect.com/science/article/pii/S0734975012000122.

3. Nölling J, Breton G, Omelchenko M V, Makarova KS, Zeng Q, Gibson G, et al. Genome sequence and comparative analysis of the solvent-producing bacterium Clostridium acetobutylicum. J Bacteriol [Internet]. American Society for Microbiology; 2001 [cited 2017 Nov 23];183:4823–4838. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11466286.
4. Poehlein A, Grosse-Honebrink A, Zhang Y, Minton NP, Daniel R. Complete genome sequence of the nitrogen-fixing and solvent-producing Clostridium pasteurianum DSM 525. Genome Announc. [Internet]. American Society for Microbiology; 2015 [cited 2017 Nov 23];3:e01591-e01514. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25700415.
5. Wang Y, Li X, Mao Y, Blaschek HP. Single-nucleotide resolution analysis of the transcriptome structure of Clostridium beijerinckii NCIMB 8052 using RNA-Seq. BMC Genomics [Internet]. BioMed Central; 2011 [cited 2017 Nov 22];12:479. Available from: http://bmcgenomics.biomedcentral.com/articles/10.1186/1471-2164-12-479.
6. Poehlein A, Solano JDM, Flitsch SK, Krabben P, Winzer K, Reid SJ, et al. Microbial solvent formation revisited by comparative genome analysis. Biotechnol Biofuels [Internet]. BioMed Central; 2017 [cited 2017 Nov 23];10:58. Available from: http://biotechnologyforbiofuels.biomedcentral.com/articles/10.1186/s13068-017-0742-z.
7. Ezeji T, Blaschek HP. Fermentation of dried distillers' grains and solubles (DDGS) hydrolysates to solvents and value-added products by solventogenic clostridia. Bioresour. Technol. [Internet]. Elsevier; 2008 [cited 2017 Dec 8];99:5232–5242. Available from: http://www.sciencedirect.com/science/article/pii/S0960852407007778.
8. Wang Z, Gerstein M, Snyder M. RNA-Seq: A revolutionary tool for transcriptomics. Nat. Rev. Genet. [Internet]. Nature Publishing Group; 2009 [cited 2017 Nov 23];10:57–63. Available from: http://www.nature.com/doifinder/10.1038/nrg2484.
9. Sedlar K, Kolek J, Skutkova H, Branska B, Provaznik I, Patakova P. Complete genome sequence of Clostridium pasteurianum NRRL B-598, a non-type strain producing butanol. J Biotechnol. 2015;214:113–4. Available from: http://www.sciencedirect.com/science/article/pii/S0168165615301279.
10. Sedlar K, Kolek J, Provazník I, Patáková P. Reclassification of non-type strain Clostridium pasteurianum NRRL B-598 as Clostridium beijerinckii NRRL B-598. J Biotechnol [Internet]. 2017 [cited 2017 Mar 3];244:1–3. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0168165617300135.
11. Kolek J, Diallo M, Vasylkivska M, Branska B, Sedlar K, López-Contreras AM, et al. Comparison of expression of key sporulation, solventogenic and acetogenic genes in C. Beijerinckii NRRL B-598 and its mutant strain overexpressing spo0A. Appl. Microbiol. Biotechnol. [Internet]. Springer Berlin Heidelberg; 2017 [cited 2017 Dec 8];101:8279–8291. Available from: http://link.springer.com/10.1007/s00253-017-8555-3.
12. Kolek J, Branska B, Drahokoupil M, Patakova P, Melzoch K. Evaluation of viability, metabolic activity and spore quantity in clostridial cultures during ABE fermentation. Sauer M, editor. FEMS Microbiol Lett [Internet]. Oxford University Press; 2016 [cited 2018 Jan 3];363:fnw031. Available from: https://academic.oup.com/femsle/article-lookup/doi/10.1093/femsle/fnw031.
13. Wang Y, Li X, Mao Y, Blaschek HP. Genome-wide dynamic transcriptional profiling in Clostridium beijerinckii NCIMB 8052 using single-nucleotide resolution RNA-Seq. BMC Genomics [Internet]. BioMed Central; 2012 [cited 2017 Nov 22];13:102. Available from: http://bmcgenomics.biomedcentral.com/articles/10.1186/1471-2164-13-102.
14. Angiuoli S V., Gussman A, Klimke W, Cochrane G, Field D, Garrity GM, et al. Toward an Online Repository of Standard Operating Procedures (SOPs) for (Meta)genomic Annotation. Omi. A J. Integr. Biol. [Internet]. 2008 [cited 2018 Jan 10];12:137–41. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18416670.
15. Maaten L van der, Hinton G. Visualizing Data using t-SNE. J Mach Learn Res [Internet]. 2008 [cited 2018 Jan 12];620:267–284. Available from: http://www.jmlr.org/papers/v9/vandermaaten08a.html.
16. Tracy BP, Gaida SM, Papoutsakis ET. Flow cytometry for bacteria: Enabling metabolic engineering, synthetic biology and the elucidation of complex phenotypes. Curr. Opin. Biotechnol. [Internet]. 2010 [cited 2018 Feb 9];21:85–99. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20206495.
17. Tracy BP, Gaida SM, Papoutsakis ET. Development and application of flow-cytometric techniques for analyzing and sorting endospore-forming clostridia. Appl. Environ. Microbiol. [Internet]. 2008 [cited 2018 Feb 9];74:7497–7506. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18931289.
18. Lahens NF, Kavakli IH, Zhang R, Hayer K, Black MB, Dueck H, et al. IVT-seq reveals extreme bias in RNA sequencing. Genome Biol. [Internet]. BioMed Central; 2014 [cited 2018 Jan 31];15:R86. Available from: http://genomebiology.biomedcentral.com/articles/10.1186/gb-2014-15-6-r86.
19. Zytnicki M, Mmquant: how to count multi-mapping reads? BMC Bioinformatics [Internet]. BioMed Central; 2017 [cited 2018 Jan 31];18:411. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28915787.
20. Goodhead I, Darby AC. Taking the pseudo out of pseudogenes. Curr. Opin. Microbiol. [Internet]. Elsevier Current Trends; 2015 [cited 2018 Jan 10];23:102–109. Available from: http://www.sciencedirect.com/science/article/pii/S1369527414001799.
21. Zheng D, Frankish A, Baertsch R, Kapranov P, Reymond A, Siew WC, et al. Pseudogenes in the ENCODE regions: Consensus annotation, analysis of transcription, and evolution. Genome Res. [Internet]. Cold Spring Harbor Laboratory Press; 2007 [cited 2018 Jan 10];17:839–851. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17568002.
22. Balzer S, Malde K, Lanzen A, Sharma A, Jonassen I. Characteristics of 454 pyrosequencing data-enabling realistic simulation with flowsim. Bioinformatics [Internet]. Oxford University Press; 2011 [cited 2018 Jan 31]. p. i420–i425. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20823302.
23. Al-Hinai MA, Jones SW, Papoutsakis ET. The Clostridium sporulation programs: diversity and preservation of endospore differentiation. Microbiol Mol Biol Rev [Internet]. American Society for Microbiology (ASM); 2015 [cited 2018 Feb 9];79:19–37. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25631287.
24. Alsaker K V, Papoutsakis ET. Transcriptional program of early sporulation and stationary-phase events in Clostridium acetobutylicum. J Bacteriol [Internet]. American Society for Microbiology; 2005 [cited 2018 Feb 14];187:7103–7118. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16199581.
25. Jones DT, Woods DR. Acetone-butanol fermentation revisited. Microbiol Rev [Internet]. American Society for Microbiology (ASM); 1986 [cited 2018 Feb 20];50:484–524. Available from: http://www.ncbi.nlm.nih.gov/pubmed/3540574.
26. Patakova P, Kolek J, Sedlar K, Koscova P, Branska B, Kupkova K, et al. Comparative analysis of high butanol tolerance and production in clostridia. Biotechnol. Adv. [Internet]. Elsevier; 2017 [cited 2018 Feb 14]; Available from: https://www.sciencedirect.com/science/article/pii/S0734975017301568.
27. Dürre P, Fischer RJ, Kuhn A, Lorenz K, Schreiber W, Stürzenhofecker B, et al. Solventogenic enzymes of Clostridium acetobutylicum: catalytic properties, genetic organization, and transcriptional regulation. FEMS Microbiol Rev [Internet]. 1995 [cited 2018 Feb 9];17:251–262. Available from: http://www.ncbi.nlm.nih.gov/pubmed/7576767.
28. Beaumont HJE, Gallie J, Kost C, Ferguson GC, Rainey PB. Experimental evolution of bet hedging. Nature [Internet]. 2009 [cited 2018 Feb 9];462:90–93. Available from: http://pubman.mpdl.mpg.de/pubman/item/escidoc:2259285/component/escidoc:2259283/Beaumont_et_al_2009.pdf.
29. Brussow H, Canchaya C, Hardt W-D. Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. Microbiol Mol Biol Rev [Internet] American Society for Microbiology (ASM); 2004 [cited 2018 Feb 1];68:560–602. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15353570.
30. Ochman H, Lawrence JG, Groisman EA. Lateral gene transfer and the nature of bacterial innovation. Nature [Internet]. Nature Publishing Group; 2000 [cited 2018 Feb 1];405:299–304. Available from: http://www.nature.com/articles/35012500.
31. Jones DT, Shirley M, Wu X, Keis S. Bacteriophage infections in the industrial acetone butanol (AB) fermentation process. J Mol Microbiol Biotechnol [Internet]. 2000 [cited 2018 Feb 9];2:21–26. Available from: https://www.caister.com/jmmb/v/v2/v2n1/03.pdf.
32. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: Summarize analysis results for multiple tools and samples in a single report. Bioinformatics [Internet]. Oxford University Press; 2016 [cited 2017 Aug 22];32:3047–3048. Available from: https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btw354.
33. Kopylova E, Noé L, Touzet H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics [Internet]. 2012 [cited 2017 Sep 14];28:3211–3217. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23071270.

34. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. Nucleic Acids Res. [Internet]. Oxford University Press; 2013 [cited 2018 Jan 4];41:D590–D596. Available from: http://academic.oup.com/nar/article/41/D1/D590/1069277/The-SILVA-ribosomal-RNA-gene-database-project.
35. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: Ultrafast universal RNA-seq aligner. Bioinformatics [Internet]. Oxford University Press; 2013 [cited 2017 Jul 26];29:15–21. Available from: https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/bts635.
36. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics [Internet]. Oxford University Press; 2009 [cited 2017 Jul 26];25:2078–9. Available from: https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btp352.
37. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. [Internet]. NIH Public Access; 2011 [cited 2018 Jan 5];29:644–652. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21572440.
38. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics [Internet]. 2009 [cited 2018 Jan 17];10:421. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20003500.
39. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat. Biotechnol. [Internet]. 2011 [cited 2018 Jan 17];29:24–26. Available from: http://www.nature.com/doifinder/10.1038/nbt.1754.
40. Carver T, Thomson N, Bleasby A, Berriman M, Parkhill J. DNAPlotter: Circular and linear interactive genome visualization. Bioinformatics [Internet]. 2009 [cited 2018 Jan 17];25:119–120. Available from: https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btn578.
41. Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, et al. Artemis: sequence visualization and annotation. Bioinformatics [Internet]. 2000 [cited 2018 Jan 29];16:944–945. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11120685.
42. Noé L, Kucherov G. YASS: Enhancing the sensitivity of DNA similarity Search Nucleic Acids Res [Internet]. Oxford University Press; 2005 [cited 2018 Jan 17];33:W540–W543. Available from: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gki478.
43. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, et al. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res. [Internet]. Oxford University Press; 2016 [cited 2018 Jan 17];44:W16–W21. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27141966.
44. Akhter S, Aziz RK, Edwards RA. PhiSpy: A novel algorithm for finding prophages in bacterial genomes that combines similarity-and composition-based strategies. Nucleic Acids Res. [Internet]. Oxford University Press; 2012 [cited 2018 Jan 17];40:e126. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22584627.
45. Liao Y, Smyth GK, Shi W. FeatureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics [Internet]. 2014 [cited 2018 Jan 29];30:923–930. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24227677.
46. Robinson MD, McCarthy DJ, Smyth GK. edgeR: A Bioconductor package for differential expression analysis of digital Gene Expr data. Bioinformatics [Internet]. Oxford University Press; 2010 [cited 2018 Jan 5];26:139–140. Available from: https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btp616.
47. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. [Internet]. BioMed Central; 2014 [cited 2018 Jan 5];15:550. Available from: http://genomebiology.biomedcentral.com/articles/10.1186/s13059-014-0550-8.
48. van der Maaten L. Accelerating t-SNE using Tree-Based Algorithms J Mach Learn Res [Internet]. 2014 [cited 2018 Jan 17];15:1–21. Available from: http://jmlr.org/papers/volume15/vandermaaten14a/vandermaaten14a.pdf.
49. Wickham H. ggplot2 Elegant Graphics for Data Analysis [Internet]. Media. Springer; 2009 [cited 2018 Jan 29]. Available from: http://had.co.nz/ggplot2/book
50. Chen H, Boutros PC. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. BMC Bioinformatics [Internet]. BioMed Central; 2011 [cited 2018 Jan 29];12:35. Available from: http://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-12-35.

5.7 Article VII

SEDLAR, Karel, KOLEK, Jan, GRUBER, Markus, JURECKOVA, Katerina, BRANSKA, Barbora, CSABA, Gergely, VASYLKIVSKA, Maryna, ZIMMER, Ralf, PATAKOVA, Petra and PROVAZNIK, Ivo. A transcriptional response of Clostridium beijerinckii NRRL B-598 to a butanol shock. *Biotechnology for Biofuels*. 2019. 12(1), p. 243. (2019 IF = 4.815, Q1 in BIOTECHNOLOGY & APPLIED MICROBIOLOGY)

DOI: https://doi.org/10.1186/s13068-019-1584-7

Article VII: A transcriptional response of *Clostridium beijerinckii* NRRL B-598 to a butanol shock

Sedlar et al. Biotechnol Biofuels (2019) 12:243
https://doi.org/10.1186/s13068-019-1584-7

Biotechnology for Biofuels

RESEARCH Open Access

A transcriptional response of *Clostridium beijerinckii* NRRL B-598 to a butanol shock

Karel Sedlar[1*], Jan Kolek[2], Markus Gruber[3], Katerina Jureckova[1], Barbora Branska[2], Gergely Csaba[3], Maryna Vasylkivska[2], Ralf Zimmer[3], Petra Patakova[2] and Ivo Provaznik[1]

Abstract

Background: One of the main obstacles preventing solventogenic clostridia from achieving higher yields in biofuel production is the toxicity of produced solvents. Unfortunately, regulatory mechanisms responsible for the shock response are poorly described on the transcriptomic level. Although the strain *Clostridium beijerinckii* NRRL B-598, a promising butanol producer, has been studied under different conditions in the past, its transcriptional response to a shock caused by butanol in the cultivation medium remains unknown.

Results: In this paper, we present a transcriptional response of the strain during a butanol challenge, caused by the addition of butanol to the cultivation medium at the very end of the acidogenic phase, using RNA-Seq. We resequenced and reassembled the genome sequence of the strain and prepared novel genome and gene ontology annotation to provide the most accurate results. When compared to samples under standard cultivation conditions, samples gathered during butanol shock represented a well-distinguished group. Using reference samples gathered directly before the addition of butanol, we identified genes that were differentially expressed in butanol challenge samples. We determined clusters of 293 down-regulated and 301 up-regulated genes whose expression was affected by the cultivation conditions. Enriched term "RNA binding" among down-regulated genes corresponded to the down-turn of translation and the cluster contained a group of small acid-soluble spore proteins. This explained phenotype of the culture that had not sporulated. On the other hand, up-regulated genes were characterized by the term "protein binding" which corresponded to activation of heat-shock proteins that were identified within this cluster.

Conclusions: We provided an overall transcriptional response of the strain *C. beijerinckii* NRRL B-598 to butanol shock, supplemented by auxiliary technologies, including high-pressure liquid chromatography and flow cytometry, to capture the corresponding phenotypic response. We identified genes whose regulation was affected by the addition of butanol to the cultivation medium and inferred related molecular functions that were significantly influenced. Additionally, using high-quality genome assembly and custom-made gene ontology annotation, we demonstrated that this settled terminology, widely used for the analysis of model organisms, could also be applied to non-model organisms and for research in the field of biofuels.

Keywords: ABE fermentation, Butanol shock, *Clostridium beijerinckii* NRRL B-598, RNA-Seq transcriptome

*Correspondence: sedlar@feec.vutbr.cz
[1] Department of Biomedical Engineering, Brno University of Technology, Technicka 12, 616 00 Brno, Czech Republic
Full list of author information is available at the end of the article

Background

Solventogenic bacteria from the *Clostridium* genus are used for their ability to produce solvents in acetone–butanol–ethanol (ABE) fermentation [1]. Although it has been more than 100 years, since the first industrial ABE fermentation process was launched, for a long time, bacterial production was replaced by cheaper chemical production from oil [2]. Due to the increasing interest in nature conservation and the fluctuating price of oil, bacterial production of bio-butanol can currently compete with synthetic production [3]. While clostridia represent a large group of organisms with various properties, among the solventogenic representatives three species, *C. acetobutylicum*, *C. beijerinckii*, and *C. pasteurianum*, are primarily of interest in butanol production [4]. This is coupled with the development of molecular tools for manipulation with these species in the last 2 decades, for example ClosTron technology and the modular shuttle plasmids system, transposon-based mutagenesis, counter-selection markers, or CRISPR-Cas-based gene editing [5]. Unfortunately, particular species or even strains can be so different that a tool designed for one strain is not easily applicable to even closely related strains. An example can be found in the strain *C. beijerinckii* NRRL B-598 [6], formerly misidentified as *C. pasteurianum* [7], presented in this study. The strain contains specific restriction–modification (R-M) systems, preventing the use of previously proposed protocols for electrotransformation, conjugation, and sonoporation [8]. Thus, knowledge gathered using the most widely described strains *C. acetobutylicum* ATCC 824 [9], *C. beijerinckii* NCIMB 8052 [10], and *C. pasteurianum* DSM 525 [11] needs to be supplemented by studies of other strains to understand the processes at the molecular level. Even a single-nucleotide variant (SNV) can be responsible for various phenotypic traits [12].

Although various genomes of solventogenic clostridia are studied and compared [13], the genomic sequence itself provides only the theoretical capabilities of an organism and transcriptomic studies are needed to reveal the active parts of a genome. Currently, there are only a few high-quality transcriptomes, which allow full analysis of gene expression and possible post-transcriptional regulation in ABE solventogenic clostridia [4]. For the butanol producing species mentioned above, these mainly include a comprehensive RNome study of *C. acetobutylicum* [14], the transcriptome of *C. beijerinckii* NCIMB 8052 under standard cultivation and with the addition of butyrate into the cultivation medium [15, 16], and our previous transcriptomic studies of *C. beijerinckii* NRRL B-598 under standard cultivation conditions [17, 18]. Therefore, few studies are insufficient to deepen an understanding of butanol production, as solventogenesis is not regulated in the same way, in all solventogenic clostridia and even the same strain can demonstrate different behavior when different cultivation conditions are established [19]. To enhance the knowledge base regarding the behavior of solventogenic clostridia, in this paper, we describe a transcriptional response of *C. beijerinckii* NRRL B-598 to butanol shock caused by the addition of butanol in a concentration of 4.5 g/L to the cultivation medium at the very end of the acidogenic phase. While the transcriptional response to the butanol shock has been mapped for *C. acetobutylicum* [20, 21], it has never been performed for *C. beijerinckii*. Butanol is considered one of the most significant stressors during ABE fermentation [2]; therefore, the butanol challenge experiment was evaluated thoroughly to reveal statistically relevant changes in gene expression. Additionally, we improved the genome assembly by sequencing genomic DNA as our previous study revealed possible misassemblies [18] and reannotated this novel assembly. To summarize the stress response, we utilized gene ontology (GO) enrichment analysis. While this kind of analysis simplifies comparison of responses between various species or strains and can be of great advantage, it is not commonly used for non-model organisms due to lack of comprehensive resources of GO annotation. We scanned various databases and constructed our own high-quality GO annotation. This novel approach can be easily used for other non-model organisms using standard languages for statistical computing. The population heterogeneity was characterized using flow cytometry (FC) coupled with fluorescent staining and, simultaneously, population dynamics and metabolite formation were thoroughly monitored.

Results

Cultivation and fermentation kinetics

The goal of the cultivation experiment was to obtain transcriptomic data describing both immediate and later responses towards a non-lethal butanol shock, performed in the phase of transition between the late acidogenic phase and early start of the solventogenesis. Butanol was added directly after sample collection at time 6 h (T_b0). The selected final concentration of added butanol was approximately 0.5% v/v, which was verified previously as unambiguously stressing, but not a lethal concentration for *C. beijerinckii* NRRL B-598 culture [22]. Based on the high-pressure liquid chromatography (HPLC) analyses, there was a small, detectable concentration of butanol produced already before the butanol was added; the exact final concentration of butanol at time 6.5 h (T_b1) was 4.5 g/L (4.42 g/L and 4.58 g/L in the two replicates) (see Fig. 1a). The shock did not stop the butanol production and the next increase in butanol concentration was

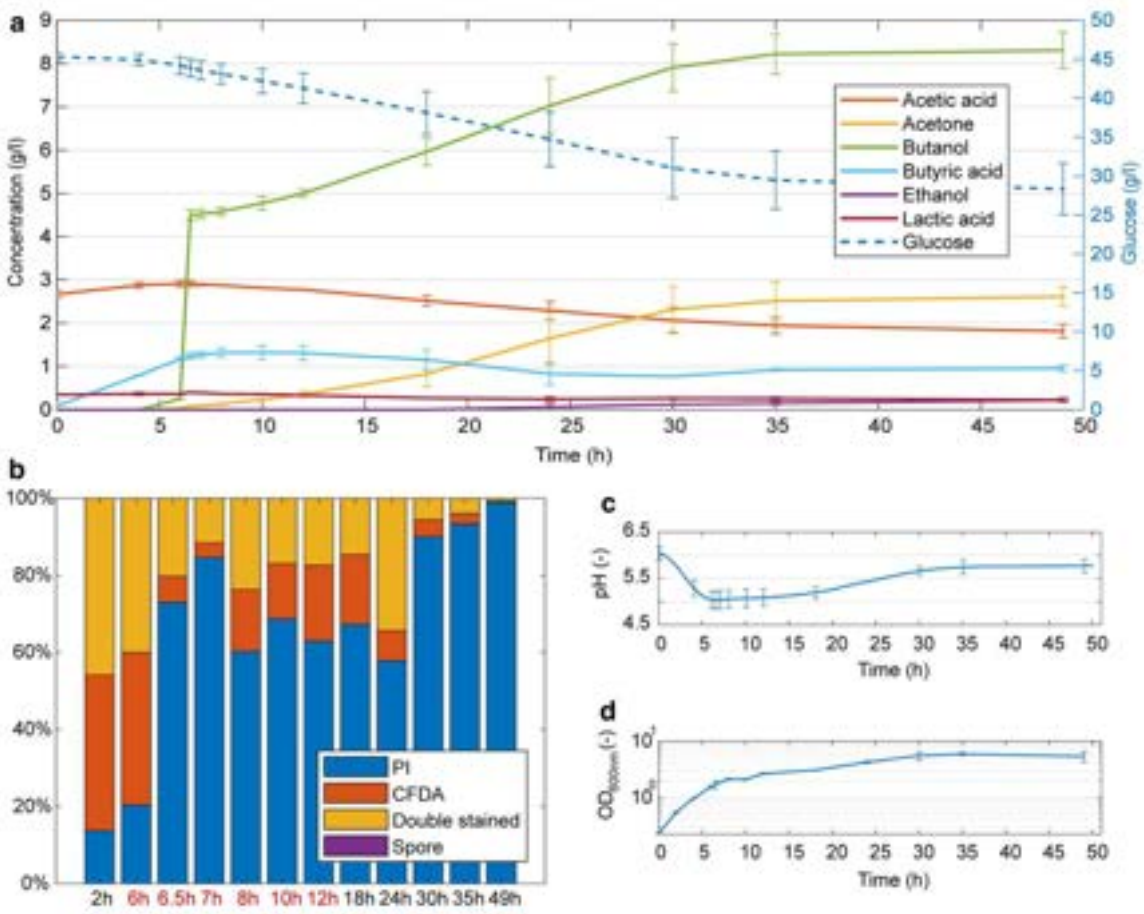


Fig. 1 Cultivation and fermentation characteristics of *Clostridium beijerinckii* NRRL B-598 during butanol shock. **a** The concentration of glucose, solvents, and acids during ABE fermentation measured using HPLC. **b** Flow cytometry—the distribution of cells within the population according to their fluorescence pattern for combined staining using PI and CFDA. **c** pH curve for the cultivation. **d** Cell growth measured as optical density at 600 nm. Values represent the mean of the biological replicates and error bars represent the standard deviations. Time-points (T_b0–T_b5) for samples subjected to RNA expression analysis are indicated by red text labels

evident immediately in the sample collected at time 7 h (T_b2). Its production continued until the cultivation was stopped. The final butanol titer was approximately 8.3 g/L (8.0 g/L and 8.6 g/L in the two replicates).

The concentration of all monitored acids (acetic, butyric, and lactic) started to decrease slightly in the culture after the addition of butanol and only the titer of butyric acid started to increase again at time 30 h (see Fig. 1a). Acetone production started around time 6 h and its concentration increased to an approximate time of 35 h. The measured ethanol concentrations were very low at all times as it is typical for this strain also during standard culture conditions [23] (see Fig. 1a and Additional file 1). The butanol shock slowed glucose consumption, compared to standard ABE fermentation. At the end of the cultivation, a relatively high amount of substrate (ca. 30 g/L) remained unused.

After the shock, the cell growth was retarded for approximately the next 4 h, as can be seen in the optical density (OD) analysis (see Fig. 1d). This corresponds well with an increased number of propidium iodide (PI) stained, i.e., non-active, cells identified by FC (see Fig. 1b). After time-point 10 h (T_b4), restored growth of the culture was evident. In the case of the pH course, the culture lacked the traditional rapid increase of pH after the onset of solventogenesis, the so-called metabolic shift (see Fig. 1c and Additional file 1).

The culture produced no spores as determined by a flow cytometry analysis (see Fig. 1b) as well as by light microscopy (Additional file 2). The cells were rod shaped with rather longer chains at the final stages of the experiment. The largest fraction of live cells, carboxyfluorescein diacetate (CFDA) stained, were observed at the beginning of the cultivation prior to the butanol shock at times

2 h and 6 h. Immediately after the addition of butanol, an inhibiting effect was observed. At time 6.5 h as well as 7 h, a number of CFDA stained cells (reflecting those cells with highly active esterases) dropped dramatically and a corresponding increase in cells with damaged cell membrane function, PI stained, was observed. At time 8 h, cell viability was partly restored (39.8% of cells) and the fraction of active cells remained more or less constant up to at least time 24 h. Metabolically active cells were still clearly detectable at time 35 h, but nearly no living cells were found in the last sample (49 h).

Genome assembly improvement and GO annotation

We used paired-end reads from DNA sequencing for refinement of the previous genome assembly. After adapter and quality trimming, 4 million 150 bp paired-end reads of an overall high quality (average Phred score $Q \approx 35$) were mapped to the previous CP011966.2 assembly and used for the construction of the augmented assembly, currently available in GenBank under accession number CP011966.3. The novel assembly is 114 bp longer than the previous one (6,186,993 bp vs. 6,186,879 bp). The differences were almost exclusively single-nucleotide changes, except for a single-dinucleotide deletion, and can be divided into three groups: (i) substitutions, (ii) insertions, and (iii) deletions (see Additional file 3). (i) Substitutions affect seven positions, of which four are located in protein-coding regions and the remaining three are in pseudogene regions according to the novel annotation. (ii) Deletions affect seven positions: a single deletion is located in protein-coding region, five in a pseudogene, and the remaining one in an intergenic region. (iii) The largest group is formed of 122 insertions: 86 in protein-coding regions, 31 in intergenic regions, and 5 in pseudogenes. This group is responsible for the majority of changes in the annotation, as in the previous assembly: 75 of these positions were located in pseudogenes, 35 in intergenic regions, 11 in protein-coding regions, and the remaining insertion affected a position where a protein-coding region and a pseudogene overlapped.

The novel assembly was reannotated and the annotation was compared to the previous one (see Table 1). The total number of annotated elements in the augmented assembly is slightly higher, while the number of pseudogenes is reduced. This reduction is caused by a number of insertions mentioned above, resulting in a substantial reduction (100 to 42) of frameshifts detected in pseudogenes. Nevertheless, the changes are not simply caused by the addition of novel loci and the reannotation of pseudogenes as genes (see Additional file 4). In total, 58 loci of the previous assembly were completely discarded from the annotation. The main part, 36 loci, was previously

Table 1 Comparison of genome annotations

	CP011966.2	CP011966.3
Protein-coding genes	5084	5128
RNAs	149	148
Pseudogenes	199	166
Total number of elements	5432	5442

labeled as protein-coding genes, 21 as pseudogenes, and a single locus as non-coding RNA. On the contrary, 68 new loci were introduced in the genome, most of them (44) as pseudogenes and 24 as protein-coding genes. The remaining 96 modifications in the annotation are due to changes of biotypes. While 76 pseudogenes were reannotated as protein-coding genes, 20 protein-coding genes are now labeled as pseudogenes.

We paid a special attention to the improvement of the GO annotation of the novel assembly. We searched for GO terms assigned to the *C. beijerinckii* NRRL B-598 genome and found 22,013 terms assigned to 3917 distinct genomic elements. Some of these terms were duplicated, since there were four different sources of annotation: UniProt [24], InterPro [25], Gene Ontology Consortium (GOC) [26], and RNAcentral [27]. After the removal of duplications, 16,271 uniquely assigned terms remained in the annotation. The remaining genomic elements, without any assigned GO term, were subjected to sequence-based annotation in InterPro and GO databases. To find relevant homologies, protein BLAST [28] searches against the whole bacterial domain were used. After filtering out duplications and obsolete terms, 1702 distinct GO terms were assigned to 4455 genomic elements in 18,020 unique assignments. The resulting annotation was summarized in a map file (see Additional file 5) that can be used for GO enrichment analysis in the R/Bioconductor package topGO [29]. We also added a brief overview of the GO annotation by assigning levels (their longest distance from the root) to assigned terms (see Additional file 6). The most common term is GO:0016021 "integral component of membrane", from the cellular component (CC) category, assigned to 1251 genes. The most abundant terms from the biological process (BP) and molecular function (MF) categories are GO:0055114 "oxidation–reduction process" with 430 genes and GO:0016740 "transferase activity" with 610 genes, respectively. Nevertheless, these values are extreme and a median value of the times of a GO term assignment is two.

RNA-Seq transcriptome

Our RNA-Seq data set of *C. beijerinckii* NRRL B-598 response to a butanol shock covers six time-points (T_b0–T_b5) by two independent biological replicates, labeled as

F and G (as we continue to label our RNA-Seq samples of the strain in alphabetical order, A–E were assigned to standard ABE fermentation in our previous studies [17, 18]). The whole data set contains almost 450 million 75 bp single-end reads. Despite the rRNA depletion performed prior to the library construction, reads corresponding to rRNA were detected and removed prior to the mapping in silico. The amount of remaining non-rRNA reads ranged from 1.4 to 5.3 million per sample (see Additional file 7). Although the quality assessment after the first preprocessing steps (demultiplexing, quality trimming, and adapter trimming) confirmed an overall high-quality of sequences (average Phred score $Q \approx 35$), in some samples, almost 20% of reads could not have been mapped unambiguously (see Additional file 7). Reads mapping to the genome more than ten times were discarded and counted as unmapped. To cover the expression of duplicated genes, the reads mapping to the genome up to ten times were included in the gene expression analysis (see Table 2). However, the contribution of such reads was down-weighted in the expression analysis, depending on the number of times they mapped to the genome, so the sum of the number of counted reads remained the same. Similarly, reads mapping to more than one genomic object were also down-weighted. In the current assembly, there are 311 overlapping loci. The majority of them are formed by 294 pairs of overlapping protein-coding genes, the additional 16 genes overlap with pseudogenes, and the remaining single case corresponds to two overlapping pseudogenes. In total, 33 protein-coding genes and four pseudogenes demonstrated no transcripts (RPKM < 1) at any of the six sampling points.

Reproducibility of the experiment was supported by the utilization of two biological replicates and by the comparison of replicates to the previously gathered data sets. An overview of the data set produced by the t-Distributed Stochastic Neighbor Embedding (t-SNE) [30] dimensionality reduction method applied to the normalized expression data suggested a partitioning of the samples into three separate clusters (see Fig. 2a). The first was formed by samples obtained directly before butanol addition to the cultivation medium. Samples from the following three time-points formed the second cluster and samples from the remaining two time-points formed the third cluster. Differences between samples before and after butanol addition are particularly visible

Table 2 Transcriptional activity of genes and pseudogenes

Sample	T_b0 (6 h)	T_b1 (6.5 h)	T_b2 (7 h)	T_b3 (8 h)	T_b4 (10 h)	T_b5 (12 h)	Total
No. of genes with RPKM > 1[a]	4942 (4891)	4943 (4888)	4967 (4907)	4972 (4918)	5003 (4951)	5003 (4968)	5095 (5054)
No. of pseudogenes with RPKM > 1[a]	112 (141)	147 (142)	146 (143)	147 (144)	152 (148)	147 (142)	162 (160)
Max. expression (RPKM)	4.5×10^4	8.2×10^4	6.3×10^4	7.8×10^4	7.7×10^4	8.0×10^4	8.2×10^4

[a] Values in brackets apply to uniquely mapped reads only

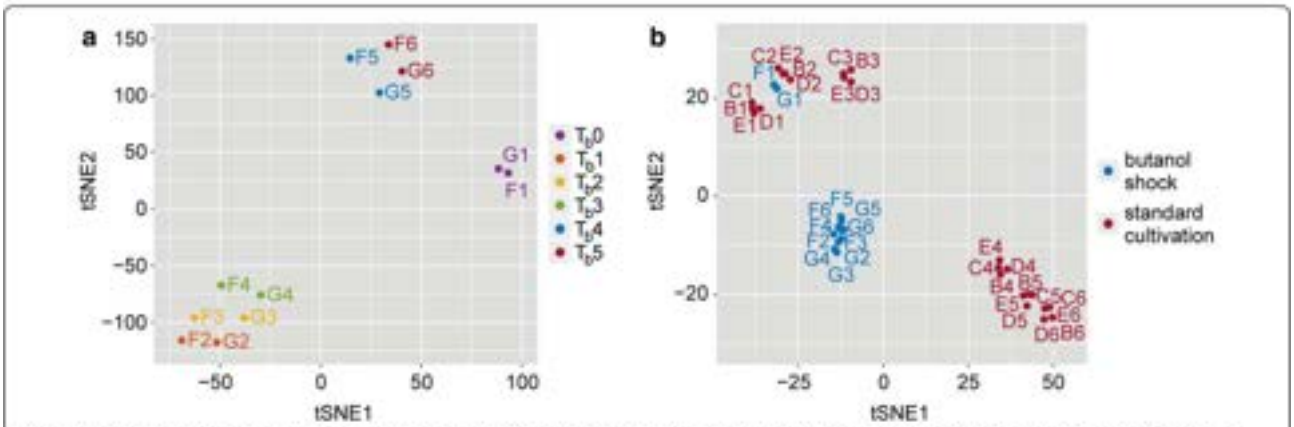


Fig. 2 Overall comparison of RNA-Seq samples. 2D representation of the normalized expression data after dimensionality reduction by t-SNE. **a** Comparison of the samples collected at the six time-points (T_b0–T_b5) coded by different colors. Each point represents a sample with a text label indicating the biological replicate (F, G) and the time-point from which it originated (T_b0–T_b5). **b** Comparison of the samples collected during butanol shock cultivation (red) and the samples from our previous studies [17, 18] during standard cultivation (blue). Again, points represent samples with a text labels indicating biological replicates (B, C, D, and E for standard cultivation and F and G for butanol shock). Samples F1 and G1 collected before butanol addition at time-point $T_b0 = 6$ h correspond to samples B2, C2, D2, and E2 collected at $T2 = 6$ h during standard cultivation

in comparison to previously gathered samples during standard cultivation [17, 18] (see Fig. 2b). While samples before butanol addition cluster to the corresponding samples from standard cultivation, samples after butanol addition form a separate cluster. To perform the comparison, we mapped samples from the previous studies to the novel genome assembly CP011966.3.

Differential expression

To further analyze particular samples, we performed differential expression analysis of adjacent time-points and showed the results as respective Venn diagrams (see Fig. 3 and Additional file 8). In accordance with the previous dimensionality reduction, the main regulation was detected directly after butanol addition (between T_b0 and T_b1), when 1443 loci were regulated (adjusted p value < 0.05, Benjamini–Hochberg correction) and the second highest regulation between T_b3 and T_b4, when 300 loci were differentially expressed. In total, 1499 protein-coding genes were regulated at least once between adjacent time-points, 303 of these more than once. The remaining 3629 protein-coding genes had no statistically significant regulations among adjacent time-points. Only 14 out of 166 pseudogenes were regulated, 13 were regulated once, and a single pseudogene was regulated twice. Only a single non-coding RNA gene X276_26885 was regulated once, directly after butanol addition. The complete results of the differential expression analysis among adjacent time-points, including log2FoldChanges and adjusted p values, are available in Additional file 9.

We explored differentially expressed genes at particular time-points against the reference time-point T_b0, prior to the butanol addition, to find gene expression changes elicited by butanol addition. There were 2037 genomic loci with at least one statistically significant differential expression (adjusted p value < 0.05, Benjamini–Hochberg correction). Based on their log2FoldChanges in all five comparisons, genes were distributed into three clusters. Although all selected loci had at least one significant change in expression, loci within the first cluster of 1443 elements demonstrated zero log2FoldChanges on average. Genes within the second (293 elements) and the third cluster (301 elements) are significantly down-regulated and up-regulated, respectively (see Fig. 4). While the first cluster also captures noise and contains loci of various biotypes, including four rRNA genes, the second cluster of down-regulated elements is formed exclusively by protein-coding genes. The third cluster of up-regulated elements is formed mainly by protein-coding genes, but it also contains nine pseudogenes, a single non-coding RNA gene, and a tRNA gene.

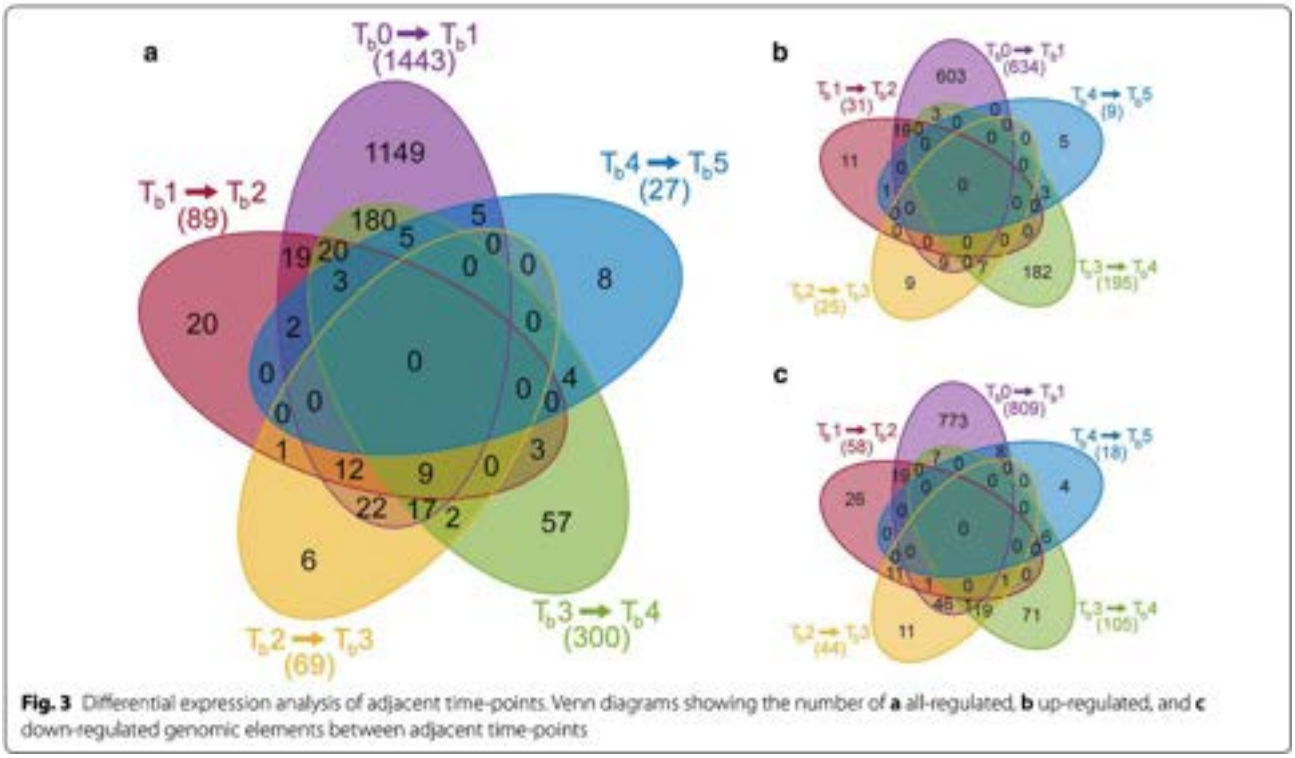


Fig. 3 Differential expression analysis of adjacent time-points. Venn diagrams showing the number of **a** all-regulated, **b** up-regulated, and **c** down-regulated genomic elements between adjacent time-points

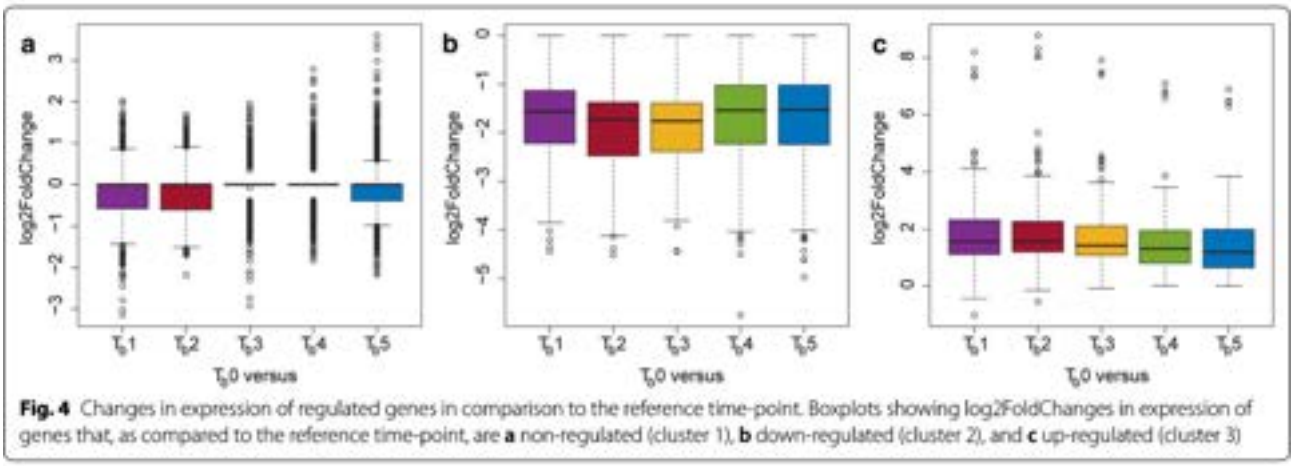


Fig. 4 Changes in expression of regulated genes in comparison to the reference time-point. Boxplots showing log2FoldChanges in expression of genes that, as compared to the reference time-point, are **a** non-regulated (cluster 1), **b** down-regulated (cluster 2), and **c** up-regulated (cluster 3)

Gene ontology enrichment

To explore and describe the functional response to the butanol shock, we performed MF GO enrichment analysis in all three clusters of genes using all 2037 regulated genomic loci as the gene universe. MF GO terms significantly enriched (p value < 0.05, Fisher's exact test) in cluster 1 were especially terms related to "iron ion binding", "methyltransferase", "nuclease activity", "helicase activity", and others (see Table 3). Among the genes annotated with the term "iron ion binding" are genes for ferredoxin, acyl-CoA-dehydrogenase, genes involved in Fe–S proteins biosynthesis, pyruvate:ferredoxin (flavodoxin) oxidoreductase, and many more genes which are indispensable or house-keeping (see Additional file 10).

In cluster 2 (down-regulated), we can recognize as main recurring terms "dsDNA binding", "RNA/rRNA binding", and several terms which are connected to transports like "ATPase activity", "amine transmembrane transporter activity", "organic acid transmembrane transporter", or "anion/organic anion transmembrane transporter" (see Table 4). Under term "ATPase activity", we can distinguish many ABC transporters with various functions. Reflecting growth attenuation, down-regulation of distinctive group of genes involved in proteosynthesis like ribosome components (see Fig. 5 and Additional file 11) can be found in terms referring to "structural constituent of ribosome", "structural molecule activity", and "RNA/rRNA binding". Aborted preparation for sporulation is connected with down-regulation of group of genes coding small acid-soluble spore proteins (see Fig. 5 and Additional file 11), which can be found associated with term "dsDNA binding."

Up-regulated genes in cluster 3 are significantly enriched in terms like "transcriptional regulation", "protein binding", or "ATP binding" (see Table 5). GO term "secondary active transport" is also significantly enriched. The third cluster contains genes coding molecular chaperones like DnaKJ, GroESL, HptG, and several other heat-shock proteins (HSPs), which can be found associated with the term "protein/ATP binding" (see Fig. 5 and Additional file 12). A large group of genes coding putative TetR/AcrR regulation factors are also part of cluster

Table 3 GO enrichment results in cluster 1

GO.ID	Term	Annotated	Significant	Expected	classicFisher
GO:0004518	Nuclease activity	19	18	13.42	0.012
GO:0004386	Helicase activity	18	17	12.71	0.016
GO:0016741	Transferase activity, transferring one-carbon groups	55	46	38.85	0.019
GO:0043169	Cation binding	211	162	149.03	0.019
GO:0046872	Metal ion binding	207	159	146.21	0.020
GO:0010181	FMN binding	17	16	12.01	0.021
GO:0004519	Endonuclease activity	11	11	7.77	0.021
GO:0005506	Iron ion binding	22	20	15.54	0.023
GO:0008168	Methyltransferase activity	45	38	31.78	0.024
GO:0046914	Transition metal ion binding	74	59	52.27	0.048

Table 4 GO enrichment results in cluster 2

GO.ID	Term	Annotated	Significant	Expected	classicFisher
GO:0003735	Structural constituent of ribosome	54	22	7.38	4.70E−07
GO:0005198	Structural molecule activity	56	22	7.65	9.90E−07
GO:0019843	rRNA binding	37	17	5.06	1.40E−06
GO:1901682	Sulfur compound transmembrane transporter activity	7	6	0.96	3.80E−05
GO:0015116	Sulfate transmembrane transporter activity	5	5	0.68	4.60E−05
GO:0015419	ATPase-coupled sulfate transmembrane transporter activity	5	5	0.68	4.60E−05
GO:0031177	Phosphopantetheine binding	6	5	0.82	0.00024
GO:0072341	Modified amino acid binding	6	5	0.82	0.00024
GO:0008509	Anion transmembrane transporter activity	23	10	3.14	0.00041
GO:0003690	Double-stranded DNA binding	10	6	1.37	0.00079
GO:0043225	ATPase-coupled inorganic anion transmembrane transporter activity	8	5	1.09	0.00179
GO:0003723	RNA binding	78	20	10.66	0.00269
GO:0022857	Transmembrane transporter activity	141	31	19.27	0.00298
GO:0005215	Transporter activity	148	32	20.22	0.00338
GO:0005275	Amine transmembrane transporter activity	6	4	0.82	0.00406
GO:0015424	Amino acid-transporting ATPase activity	6	4	0.82	0.00406
GO:0031263	Amine-transporting ATPase activity	6	4	0.82	0.00406
GO:0033283	Organic acid-transporting ATPase activity	6	4	0.82	0.00406
GO:0033284	Carboxylic acid-transporting ATPase activity	6	4	0.82	0.00406
GO:0015318	Inorganic molecular entity transmembrane transporter activity	55	15	7.52	0.00493
GO:0015103	Inorganic molecular entity transmembrane transporter activity	10	5	1.37	0.00639
GO:0015171	Amino acid transmembrane transporter activity	10	5	1.37	0.00639
GO:0016765	Transferase activity, transferring alkyl or aryl (other than methyl) groups	10	5	1.37	0.00639
GO:0033218	Amide binding	10	5	1.37	0.00639
GO:0004794	L-Threonine ammonia-lyase activity	4	3	0.55	0.00905
GO:0015075	Ion transmembrane transporter activity	60	15	8.2	0.0117
GO:0042626	ATPase activity, coupled to transmembrane movement of substances	39	11	5.33	0.01196
GO:0043492	ATPase activity, coupled to movement of substances	39	11	5.33	0.01196
GO:0005342	Organic acid transmembrane transporter activity	12	5	1.64	0.01594
GO:0046943	Carboxylic acid transmembrane transporter activity	12	5	1.64	0.01594
GO:0004124	Cysteine synthase activity	2	2	0.27	0.01859
GO:0004421	Hydroxymethylglutaryl-CoA synthase activity	2	2	0.27	0.01859
GO:0004779	Sulfate adenylyltransferase activity	2	2	0.27	0.01859
GO:0004781	Sulfate adenylyltransferase (ATP) activity	2	2	0.27	0.01859
GO:0015087	Cobalt ion transmembrane transporter activity	2	2	0.27	0.01859
GO:0016887	ATPase activity	92	20	12.57	0.01898
GO:0008514	Organic anion transmembrane transporter activity	13	5	1.78	0.02308
GO:0015399	Primary active transmembrane transporter activity	44	11	6.01	0.0294
GO:0015405	P-P-bond-hydrolysis-driven transmembrane transporter activity	44	11	6.01	0.0294
GO:0019842	Vitamin binding	39	10	5.33	0.03149
GO:0008982	Protein-N(PI)-phosphohistidine-sugar phosphotransferase activity	14	5	1.91	0.03201
GO:0015144	Carbohydrate transmembrane transporter activity	14	5	1.91	0.03201
GO:0016841	Ammonia-lyase activity	6	3	0.82	0.03666
GO:0022804	Active transmembrane transporter activity	76	16	10.38	0.04519

3 and term "DNA binding" (see Fig. 5 and Additional file 12); ctsR, hrcA, or putative sigma factors can also be found in the same group.

Discussion

Although the previous version of the genome CP011966.2 was reconstructed using a combination of next

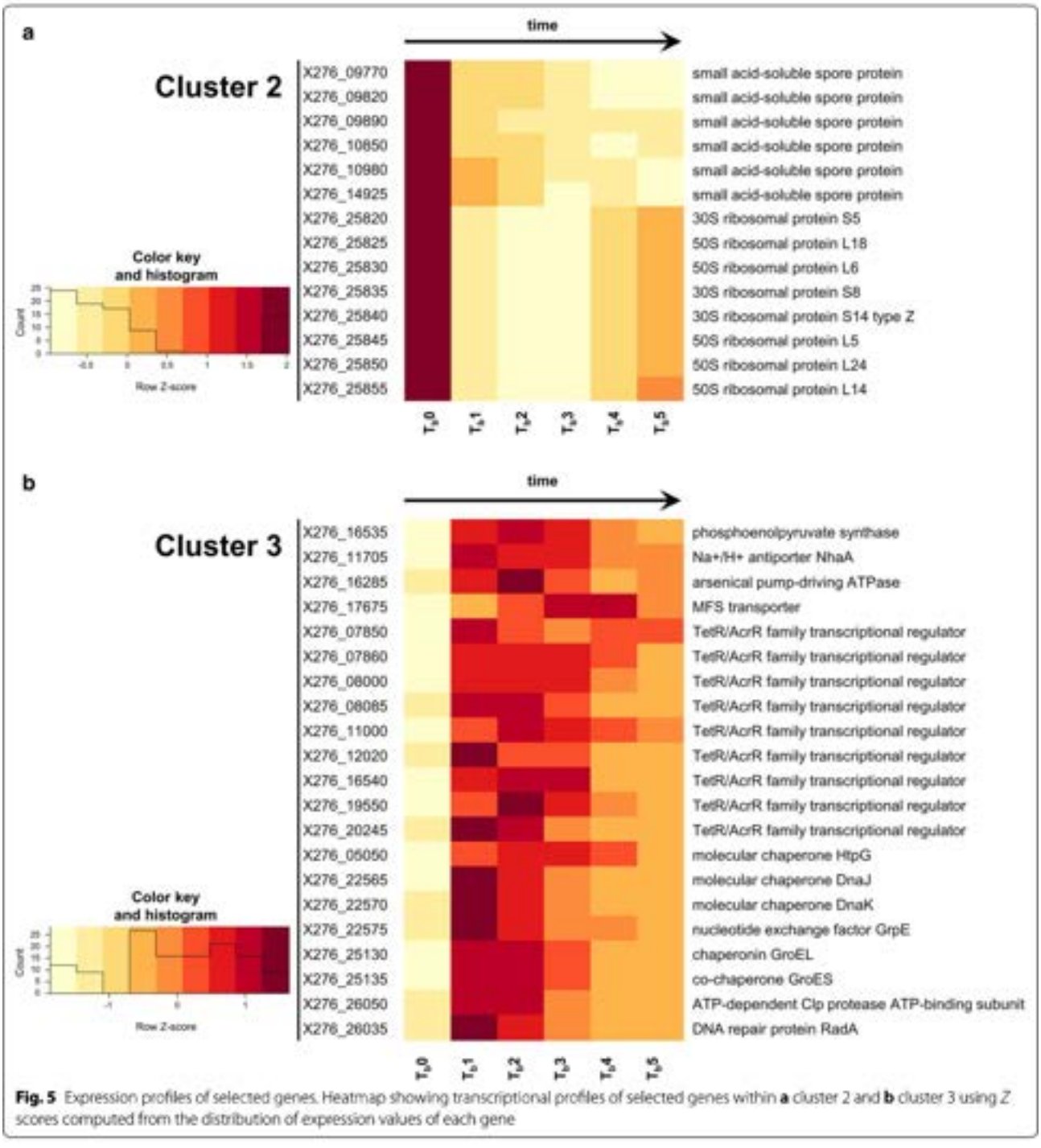


Fig. 5 Expression profiles of selected genes. Heatmap showing transcriptional profiles of selected genes within **a** cluster 2 and **b** cluster 3 using *Z* scores computed from the distribution of expression values of each gene

generation sequencing and third-generation sequencing, the assembly suffered from the inability of Roche 454 pyrosequencing to adjust low-quality PacBio RSII sequencing, especially in homopolymeric regions of the genome [31]. This was apparent from our previous transcriptomic study of the strain, where Illumina sequencing revealed possible indels in coding regions [18]. Therefore, we decided to employ additional DNA sequencing, since even an SNV can be responsible for substantial phenotypic differences in solventogenic clostridia [12, 32]. A number of insertions and deletions introduced in the novel version of the genome CP011966.3 (see Additional file 3) confirmed errors in the homopolymeric regions and led to the substantial reduction of frameshifts in

Table 5 GO enrichment results in cluster 3

GO.ID	Term	Annotated	Significant	Expected	classicFisher
GO:0003677	DNA binding	221	57	34.7	1.70E−05
GO:0004803	Transposase activity	8	6	1.26	3.00E−04
GO:0005515	Protein binding	33	13	5.18	0.00077
GO:0008519	Ammonium transmembrane transporter activity	5	4	0.79	0.00261
GO:0051082	Unfolded protein binding	5	4	0.79	0.00261
GO:0050567	Glutaminyl-tRNA synthase (glutamine-hydrolyzing) activity	3	3	0.47	0.00383
GO:0003700	DNA-binding transcription factor activity	76	20	11.93	0.01011
GO:0140110	Transcription regulator activity	77	20	12.09	0.01176
GO:0005315	Inorganic phosphate transmembrane transporter activity	4	3	0.63	0.01354
GO:0030554	Adenyl nucleotide binding	233	48	36.59	0.01813
GO:0000150	Recombinase activity	2	2	0.31	0.02457
GO:0004139	Deoxyribose-phosphate aldolase activity	2	2	0.31	0.02457
GO:0008880	Glucuronate isomerase activity	2	2	0.31	0.02457
GO:0005488	Binding	802	140	125.94	0.02519
GO:0005524	ATP binding	232	47	36.43	0.02616
GO:0032559	Adenyl ribonucleotide binding	232	47	36.43	0.02616
GO:0097159	Organic cyclic compound binding	625	112	98.15	0.02678
GO:1901363	Heterocyclic compound binding	625	112	98.15	0.02678
GO:0046983	Protein dimerization activity	5	3	0.79	0.02993
GO:0003676	Nucleic acid binding	316	61	49.62	0.03063
GO:0016879	Ligase activity, forming carbon–nitrogen bonds	30	9	4.71	0.03468
GO:0140097	Catalytic activity, acting on DNA	36	10	5.65	0.04418
GO:0015291	Secondary active transmembrane transporter activity	22	7	3.45	0.0446

detected open reading frames and to the overall reduction in a number of genomic elements annotated as pseudogenes. Moreover, all 12 insertions and three non-synonymous substitutions in protein-coding sequences resulted in proteins more similar to other proteins produced by bacteria from the *Clostridium* genus. The annotation of the augmented genome sequence introduced several changes (see Additional file 4). A number of elements coding hypothetical proteins were reduced as 48 of these elements were discarded from the genome and only 26 were newly introduced. An additional 14 hypothetical proteins were identified by changes in pseudogenes. Twenty-two of the twenty-three pseudogenes that were selected as putative active genes in our previous study by Sedlar et al. [18] were automatically reannotated as protein-coding genes due to the changes in the augmented assembly. Thus, the current version of the genome confirmed our previous findings.

Even though BLAST-based GO annotation tends to capture all true assignments, its overall precision is hampered by a number of false positive assignments [33]. We reduced possible misannotations by merging BLAST-based annotation with InterPro annotation, which has higher precision, yet lower recall, in Blast2GO suite [34]. Our manually curated annotation shows a distribution of GO term levels very similar to the annotation reconstructed from database searches only (see Additional file 6) and the median value of the times of a GO term assignment is the same. Although purely computationally inferred GO annotations are sufficient for many analyses [35], we consider our curation steps to be a quality improvement. While dimensionality reduction of butanol shock data suggested division of time-points into three clusters (see Fig. 2a), differences between clusters formed by T_b1–T_b3 and T_b4–T_b5 time-points are not so evident when the whole data set is compared to the RNA-Seq data set from a standard cultivation (see Fig. 2b). The visible difference between samples from the first time-point T_b0 to those at the remaining time-points was supported by differential expression analysis, when the number of regulated genes was the highest (see Fig. 3a). The second highest number of differentially expressed genes was recorded between T_b3 and T_b4 time-points, and confirmed the difference between T_b1–T_b3 and T_b4–T_b5 clusters. While the difference between T_b0 and T_b1–T_b3 can be accredited to a defense reaction to butanol shock, an increased number of regulated genes between T_b3 and T_b4 are connected to the restored growth of population. Even though it was reported that viability of *C. beijerinckii* NRRL B-598 was not altered

when a butanol challenge of approximately 5 g/L was added prior to inoculation [36], the addition of butanol at a late acidogenic stage induced a loss of vital function in a significantly high number of cells. This, together with abandoned sporulation, is probably the reasons that T_b4 and T_b5 samples did not cluster with the respective stage from standard cultivation, even though no negative regulation or any visible interference between butanol addition and production was observed. This correlates with results obtained for *C. acetobutylicum* [20, 21], where butanol addition up-regulated its synthesis.

The final butanol titer at the end of cultivation was approximately 8.3 g/L including added butanol, which means that the final concentration of produced butanol was roughly 4 g/L. This indicates that, in butanol challenge cultivation, butanol probably reached the maximally tolerated titer for metabolic activity of the cells, such that further butanol production has been inhibited. A similar maximal concentration was also reached using *C. beijerinckii* NRRL B-598 during the same butanol shock, but with an initial glucose concentration 20 g/L [22].

To summarize the response to a butanol shock, we used our novel GO annotation (Additional file 5) to perform a GO enrichment analysis. Pairwise comparison of the samples measured before butanol addition with samples after butanol addition allowed us to focus on the subset of genes that were differentially expressed because of butanol addition. While the total number of differentially expressed genes was relatively high (2037), log-2FoldChange-based clustering revealed further division of these genes into three clusters. The first and the largest cluster of 1443 genes demonstrated high variance of values and a lot of outliers, but almost zero median value. Therefore, we consider these genes as non-regulated due to the butanol shock. Statistically significant differential expressions in this cluster are like due to noise, biological as well as technical. First, the cell cycle within the culture is unsynchronized, and thus, regulations of genes that were not caused by the butanol shock can be captured. Second, there is technical noise remaining in the data. Although the data were carefully filtered, contaminations always remain. This is apparent, for example, from four regulated rRNA genes within the first cluster caused by remaining rRNA reads. While the number of reads mapping to rRNA loci is very low, similarly low changes in their abundance between different samples can be incorrectly identified as differential expression. The truly down- and up-regulated genes due to the butanol shock can be found in cluster 2 and cluster 3, respectively. Both clusters contain around 300 genes (293 and 301, respectively), which are only small fractions of the total number of genes in the genome of *C. beijerinckii* NRRL B-598 suitable for proper GO enrichment analysis during the butanol shock.

Although cluster 1 contained genes that were likely not regulated by the butanol shock, we decided to perform a GO enrichment analysis to summarize these genes. The cluster was formed by a mixture of genes with various functions, which resulted in only ten significantly enriched GO terms at the significance level $\alpha = 0.05$. Moreover, no p value of Fisher's exact test was lower than 0.01. Further inspection of genes associated with enriched GO terms revealed that some of these genes are probably indispensable, house-keeping (see Additional file 10), or coding enzymes necessary for DNA maintenance (e.g., DNA polymerase, primase, helicase, topoisomerase, or methyltransferase).

GO enrichment analysis in clusters of down-regulated (cluster 2) and up-regulated (cluster 3) genes revealed similar physiological response as described by Alsaker et al. [21], where global response was expressed as representation of differentially expressed genes in different clusters of orthologous genes (COG) categories. Among others, GO terms like "structural constituent of ribosome" (GO:0003735), "structural molecule activity" (GO:0005198), and "RNA/rRNA binding" (GO:0003723/GO:0019843) were enriched in cluster 2, which is in accordance with the significant down-regulation in COG category J (translation) for *C. acetobutylicum* [21]. Enrichment of these terms is caused by a group of genes that are assigned a couple of GO terms, even all of these four GO terms. These terms are close neighbors in the GO graph, which hints at the possibility of further slimming the GO annotation for solventogenic clostridia in the future. The highest percentage of up-regulated genes after butanol addition to *C. acetobutylicum* culture was found in COG category O (post-translational modification, protein turnover, and chaperones) [21]. Similarly, up-regulated HSPs in our study can be found associated with the GO term "protein/ATP binding" (GO:0005515/GO:0005524) in the GO enrichment analysis of cluster 3. HSPs are able to help with protein folding to native conformation, dsDNA stabilization, or can induce next changes in expression in the role of stress transcription factors [37]. Expression of HSPs during butanol production or butanol shock has been previously described in many works [2, 38–40] and several HSPs are the most probably involved in butanol stress reaction *C. beijerinckii* NRRL B-598, as well [17]. During standard cultivation, it was shown that production of class I HSPs, including DnaKJ and GroESL, were particularly regulated by pH stress and acid production, while genes coding alternative sigma-factor SigI, related theoretically to class II HSPs expression, were regulated in accordance with highest butanol titer. Similarly, genes for class

III HSPs and uncategorized HSP HptG were also highly expressed when butanol started to be produced in higher concentrations [17]. Strong up-regulation of *dnaK, dnaJ, groES, groEL, grpE, radA*, or *hptG* was also evident after butanol addition during butanol challenge cultivation (see Fig. 5). This fully supports the premise and already published results obtained for *C. acetobutylicum* [20, 21] that HSPs play a fundamental role in overcoming butanol stress. Although some GO terms may appear generic, their connection to butanol tolerance is meaningful. For example, term "DNA-binding transcription" factor activity (GO:0003700) was also found to be enriched during *n*-butanol challenge in *Escherichia coli* [41].

It is evident from FC analysis and microscopy that culture did not produce any matured spores, prespores, or even thick, so-called "clostridial" cells accumulating granulose during cultivations with butanol addition (see Fig. 1b and Additional file 2). This is, as expected, in contrast to standard cultivation experiments under the same cultivation conditions (see Additional file 1) [17] and also does not correlate with the response of *C. acetobutylicum* to butanol shock [20, 21], where sporulation remained unaffected. Moreover, sporulation suppression and, at the same time, intact solventogenesis can be considered another evidence for independent regulation of sporulation and solventogenesis in *C. beijerinckii* NRRL B-598, which fully correlates with already published results [17, 19, 36]. The fact that sporulation was not induced could have been caused by relatively small final density of cells in comparison with standard cultivation (see Additional file 1). An Agr-based quorum sensing system can be responsible for the initiation of granulose formation and subsequent sporulation in solventogenic clostridia, as postulated previously [42]. The differences in butanol elicited stress response in *C. beijerinckii* NRRL B-598, and *C. acetobutylicum* ATCC 824 might result in different organization of Agr quorum sensing genes in both genomes and no found homologies in the respective genes in both strains [17, 43]. Thus, quorum sensing could be a reason why sporulation was not started and, therefore, several genes related to spore formation were found in cluster 2. Apparent down-regulation was detected for small, acid-soluble proteins (SASPs), small proteins coating DNA in matured spores with putative peroxidase activity, which play a fundamental role in DNA protection [44, 45]. Observed expression of SASPs is in contrast with standard expression of SASPs in *C. perfringens* where SASPs are expressed after the start of sporulation [46] and are expressed under regulation of *sigG* and *sigF* in *C. acetobutylicum* [47]. On the other hand, Wetzel et al. [47] assert that SASPs can bind DNA in vitro which implies that SASPs could potentially protect DNA against nucleases, not only in matured spores.

Conclusions

Mechanisms preventing solventogenic clostridia from producing a higher titer of biofuels are widely studied yet remain unclarified. There are several reasons for this. First, solventogenic clostridia are non-model organisms whose genome sequences started to be explored only recently. Although genomes of more and more strains are being sequenced and assembled, only a few of them are robustly assembled using various sequencing techniques to fix assembly errors caused by specific biases or errors. Since even single-nucleotide changes in genomic sequences are responsible for various phenotypic traits, comparison of different strains may be difficult. Second, there is a lack of further exploration of different strains under various cultivation conditions. Moreover, a unified annotation summarizing behavior of various strains or a selected strain under different conditions is missing. Here, we overcame these obstacles by resequencing the genome of *C. beijerinckii* NRRL B-598 to produce the high-quality assembly with unified GO annotation and by exploring the transcriptional processes during butanol challenge cultivation using RNA-Seq and auxiliary HPLC and FC techniques.

The main change in transcriptional regulation was captured directly after butanol addition. When compared to the samples from a standard cultivation, samples from a butanol challenge forms a distinguished group. Still, they can be further divided into two groups. The first group is formed by samples obtained within 2 h after butanol addition and can be assigned to a defense reaction to the butanol shock. The second group captures samples where growth of population was restored; still expression of genes is different from the standard cultivation samples. To summarize the transcriptional response connected to the butanol shock, we selected only genes that are differentially expressed in a majority of pairwise comparisons of samples gathered during butanol challenge to samples gathered before butanol addition. We utilized our custom-made GO annotation to characterize the clusters of up- and down-regulated genes. This allowed us to describe the response to the butanol shock in detail using a well-defined terminology. Moreover, this analysis has been compared to a somewhat coarser analysis of the response of *C. acetobutylicum* to a butanol shock using clusters of orthologous genes. The butanol response in both species resulted in up-regulation of heat-shock protein genes and did not intervene with solventogenesis. On the other hand, there was a significant difference in sporulation. While sporulation and also granulose formation were suppressed in *C. beijerinckii* NRRL B-598, these life cycle events remained unaffected in *C. acetobutylicum* which may serve as further indirect evidence for uncoupling sporulation and solventogenesis regulation in

C. beijerinckii NRRL B-598. We believe that the proposed novel high-quality assembly and annotation will be very useful for the future exploration of the strain and will inspire others to start using this well-defined terminology when describing transcriptional responses of solventogenic clostridia.

Methods

Bacterial culture and fermentation experiment

Culture of the strain *C. beijerinckii* NRRL B-598 was obtained from NRRL (ARS) collection of microorganisms and was maintained as a spore suspension in 4 °C in distilled water. For all manipulation, TYA broth [19] containing 20 g/L or 50 g/L of glucose was used. The bacterial strain was cultivated in parallel Multifors 1 L bioreactors (INFORS HT, Bottmingen, Switzerland). Preparation process of the culture inoculum and initial cultivation parameters were chosen the same as in Patakova et al. [17]. At the beginning of cultivation, pH of the culture was adjusted to 6.3 by NaOH solution addition and pH was monitored, but not controlled during the following cultivation.

Directly after collection of samples at time 6 h of cultivation, butanol shock was performed by addition of pure, HPLC-grade butanol (Sigma-Aldrich, Praha, Czechia) to final concentration approximately 0.5% v/v. Control sampling prior to and after addition were conducted for specification of precise added butanol concentration. Butanol was added to the bioreactor under strictly sterile and anaerobic conditions.

Culture growth and HPLC analysis

Optical density measurement at 600 nm was used for culture growth monitoring. Samples were processed by the procedure as published previously by Patakova et al. [17]. Substrate consumption and metabolite production were detected and quantified using HPLC with refractive index detection (Agilent Series 1200 HPLC, Agilent, Santa Clara, CA, USA). Sample preparation and analysis were performed identically to Patakova et al. [17].

Microscopy, fluorescent staining, and flow cytometry

Cell morphology was determined in the native culture using phase contrast microscopy (BX51, Olympus, Tokio, Japan) using 400× and 1000× magnification. Cell culture viability and the amount of endospores were determined using flow cytometry (BD Accuri C6, Accuri Cytometers Inc., Ann Arbor, MI, USA) combined with PI (Sigma-Aldrich) and CFDA (Sigma-Aldrich) fluorescent staining using protocol published in Branska et al. [36].

DNA extraction and sequencing

DNeasy UltraClean Microbial Kit (Qiagen, Hilden, Germany) was used for genomic DNA extraction. DNA was extracted from an exponentially growing culture; the quality of isolated genomic DNA was controlled using a nanodrop machine (DeNovix, Wilmington, DE, USA). Library construction and sequencing of the sample was performed by CEITEC Genomics core facility (Brno, Czechia) on Illumina NextSeq, pair-end, 150 bp.

RNA extraction and sequencing

High Pure RNA Isolation Kit (Roche, Basel, Switzerland) was used for total RNA isolation from samples. The MICROB*Express*™ Bacterial mRNA Enrichment Kit (Ambion, Austin, TX, USA) was used for ribosomal RNAs' depletion from total RNA samples. All RNA samples were stored at −70 °C without next defrosting to prevent freeze–thaw damage. For control of quality of extracted total RNA, depleted mRNA, and to prevent DNA contaminations, an Agilent 2100 bioanalyzer with the RNA 6000 Nano Kit (Agilent, Santa Clara, CA, USA) in combination with routine spectrophotometric control on nanodrop machine (DeNovix, Wilmington, DE, USA) was used. Library construction and sequencing of samples were performed by CEITEC Genomics core facility (Brno, Czechia) on Illumina NextSeq, single-end, 75 bp.

Bioinformatics analysis

The quality assessment of sequencing data (DNA and RNA) after all processing steps was done using FastQC in combination with MultiQC to summarize the reports across all samples [48]. Adapter and quality trimming was performed using Trimmomatic [49]. For the genome reassembly, reads from DNA sequencing were mapped to the previous genome sequence CP011966.2 with BWA [50]. The new assembly was constructed with Pilon [51]. Our improved assembly was used as a reference for the second mapping of reads and the second round of assembly polishing with Pilon. The resulting assembly was uploaded to GenBank as CP011966.3 version of the *C. beijerinckii* NRRL B-598 genome. RNA-Seq reads were cleansed of reads corresponding to 16S and 23S rRNA using SortMeRNA [52] and the SILVA database [53] of known bacterial 16S and 23S rRNA genes to simplify the following mapping task that was performed with STAR [54]. Resulting SAM (Sequence Read Alignment/Map) files were indexed and transformed into more compact BAM (Binary Read Alignment/Map) format using SAMtools [55].

The R/Bioconductor featureCounts function included in the Rsubread package [56] was used to compute count tables. Differential analysis was performed on

raw count tables with the R/Bioconductor DESeq2 package [57] using DESeq2 built-in normalization. For the analysis of adjacent time-points presented in Venn diagrams, all samples were normalized at once. For separate analysis of particular time-points against the reference time-point, only compared samples were used for normalization. Visual comparison of samples was performed via t-SNE dimensionality reduction of a count table after regularized log transformation using the Rtsne [58] and ggplot2 [59] R packages. Venn diagrams and heatmaps representing transcription of selected genes using Z scores were generated with R packages VennDiagram [60] and gplots, respectively. Time series and bar plots were generated with Matlab 2017b and gplots.

The GO annotation map file was compiled from annotations obtained with QuickGO [61] and Blast2GO [62] with custom-made R/Bioconductor scripts using functions from the genomeIntervals, Biostrings, and topGO packages [29]. Basic statistics of the GO annotation were computed using the dnet and igraph R packages [63, 64]. GO enrichment analysis was performed using the topGO package [29].

Supplementary information

Supplementary information accompanies this paper at https://doi.org/10.1186/s13068-019-1584-7.

Additional file 1. Comparison of cultivation and fermentation characteristics of *Clostridium beijerinckii* NRRL B-598 during standard cultivation and butanol shock.

Additional file 2. *Clostridium beijerinckii* NRRL B-598 microphotograph.

Additional file 3. Differences between assemblies.

Additional file 4. Differences between genome annotations.

Additional file 5. *Clostridium beijerinckii* NRRL B-598 Gene Ontology annotation.

Additional file 6. A brief overview of the *C. beijerinckii* NRRL B-598 GO annotation.

Additional file 7. Quality of RNA-Seq reads and mapping.

Additional file 8. Differential expression analysis of adjacent time-points using MA plots.

Additional file 9. Complete differential expression analysis of adjacent time-points.

Additional file 10. Genes under enriched GO terms in cluster 1.

Additional file 11. Genes under enriched GO terms in cluster 2.

Additional file 12. Genes under enriched GO terms in cluster 3.

Abbreviations

ABE: acetone–butanol–ethanol; BP: biological process; CC: cellular component; CFDA: carboxyfluorescein diacetate; COG: clusters of orthologous genes; FC: flow cytometry; GO: gene ontology; HPLC: high-pressure liquid chromatography; MF: molecular function; OD: optical density; PI: propidium iodide; R-M: restriction-modification; SASPs: small, acid-soluble proteins; SNV: single-nucleotide variant.

Acknowledgements

This article is based upon support of the international mobility project MeMoV, No. CZ.02.2.69/0.0/0.0/16_027/0008371O funded by the Ministry of Education Youth and Sports, Czech Republic. Computational resources were partially provided by the CESNET LM2015042 and the CERIT Scientific Cloud LM2015085, under the program "Projects of Large Research, Development, and Innovations Infrastructures" and the LMU München. We acknowledge the CF Genomics of CEITEC supported by the NCMG research infrastructure (LM2015091 funded by MEYS CR) for their support with obtaining scientific data presented in this paper.

Authors' contributions

KS, BB, GC, RZ, PP, and IP designed the study. MV and BB performed the experiments. KS, MG, and KJ analyzed the data. KS, JK, BB, and PP wrote the manuscript with the input from all authors. All authors discussed the results. All authors read and approved the final manuscript.

Funding

This work has been supported by grant project GACR 17-00551S.

Availability of data and materials

The genome assembly referred in this paper is the version CP011966.3. The genome sequencing and RNA-Seq data have been deposited in the NCBI Sequence Read Archive (SRA) under the accession number SRP033480 (SRX6419026 for F replicates, SRX6419027 for G replicates, and SRX6419139 for genome resequencing, respectively).

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

[1] Department of Biomedical Engineering, Brno University of Technology, Technicka 12, 616 00 Brno, Czech Republic. [2] Department of Biotechnology, University of Chemistry and Technology Prague, Technicka 5, 166 28 Prague, Czech Republic. [3] Institut für Informatik, Ludwig-Maximilians-Universität München, Amalienstraße 17, 80333 Munich, Germany.

Received: 10 July 2019 **Accepted:** 4 October 2019
Published online: 13 October 2019

References

1. Kujawska A, Kujawski J, Bryjak M, Kujawski W. ABE fermentation products recovery methods—a review. Renew Sustain Energy Rev. 2015;48:648–61.
2. Patakova P, Kolek J, Sedlar K, Koscova P, Branska B, Kupkova K, et al. Comparative analysis of high butanol tolerance and production in clostridia. Biotechnol Adv. 2018;36:721–38.
3. Green EM. Fermentative production of butanol-the industrial perspective. Curr Opin Biotechnol. 2011;22:337–43.
4. Charubin K, Bennett RK, Fast AG, Papoutsakis ET. Engineering Clostridium organisms as microbial cell-factories: challenges & opportunities. Metab Eng. 2018;50:173–91.
5. Joseph RC, Kim NM, Sandoval NR. Recent developments of the synthetic biology toolkit for Clostridium. Front Microbiol. 2018;9:154. https://doi.org/10.3389/fmicb.2018.00154/full.
6. Sedlar K, Kolek J, Skutkova H, Branska B, Provaznik I, Patakova P. Complete genome sequence of *Clostridium pasteurianum* NRRL B-598, a non-type strain producing butanol. J Biotechnol. 2015;214:113–4.
7. Sedlar K, Kolek J, Provaznik I, Patakova P. Reclassification of non-type strain *Clostridium pasteurianum* NRRL B-598 as *Clostridium beijerinckii* NRRL B-598. J Biotechnol. 2017;244:1–3.
8. Kolek J, Sedlar K, Provaznik I, Patakova P. Dam and Dcm methylations prevent gene transfer into *Clostridium pasteurianum* NRRL B-598:

development of methods for electrotransformation, conjugation, and sonoporation. Biotechnol Biofuels. 2016;9:14.
9. Nolling J, Breton G, Omelchenko MV, Makarova KS, Zeng Q, Gibson R, et al. Genome sequence and comparative analysis of the solvent-producing bacterium *Clostridium acetobutylicum*. J Bacteriol. 2001;183:4823–38.
10. Wang Y, Li X, Mao Y, Blaschek HP. Single-nucleotide resolution analysis of the transcriptome structure of *Clostridium beijerinckii* NCIMB 8052 using RNA-Seq. BMC Genomics. 2011;12:479.
11. Poehlein A, Grosse-Honebrink A, Zhang Y, Minton NP, Daniel R. Complete genome sequence of the nitrogen-fixing and solvent-producing *Clostridium pasteurianum* DSM 525. Genome Announc. 2015;3:e01591-14.
12. Sandoval-Espinola WJ, Makwana ST, Chinn MS, Thon MR, Andrea Azcárate-Peril M, Bruno-Bárcena JM. Comparative phenotypic analysis and genome sequence of *Clostridium beijerinckii* SA-1, an offspring of NCIMB 8052. Microbiology. 2013;159:2558–70. https://doi.org/10.1099/mic.0.069534-0.
13. Poehlein A, Solano JDM, Flitsch SK, Krabben P, Winzer K, Reid SJ, et al. Microbial solvent formation revisited by comparative genome analysis. Biotechnol Biofuels. 2017;10:58. https://doi.org/10.1186/s13068-017-0742-z.
14. Venkataramanan KP, Jones SW, McCormick KP, Kunjeti SG, Ralston MT, Meyers BC, et al. The Clostridium small RNome that responds to stress: the paradigm and importance of toxic metabolite stress in *C. acetobutylicum*. BMC Genomics. 2013;14:849. https://doi.org/10.1186/1471-2164-14-849.
15. Wang Y, Li X, Blaschek HP. Effects of supplementary butyrate on butanol production and the metabolic switch in *Clostridium beijerinckii* NCIMB 8052: genome-wide transcriptional analysis with RNA-Seq. Biotechnol Biofuels. 2013;6:138. https://doi.org/10.1186/1754-6834-6-138.
16. Wang Y, Li X, Mao Y, Blaschek HP. Genome-wide dynamic transcriptional profiling in *Clostridium beijerinckii* NCIMB 8052 using single-nucleotide resolution RNA-Seq. BMC Genomics. 2012;13:102. https://doi.org/10.1186/1471-2164-13-102.
17. Patakova P, Branska B, Sedlar K, Vasylkivska M, Jureckova K, Kolek J, et al. Acidogenesis, solventogenesis, metabolic stress response and life cycle changes in *Clostridium beijerinckii* NRRL B-598 at the transcriptomic level. Sci Rep. 2019;9:1371.
18. Sedlar K, Koscova P, Vasylkivska M, Branska B, Kolek J, Kupkova K, et al. Transcription profiling of butanol producer *Clostridium beijerinckii* NRRL B-598 using RNA-Seq. BMC Genomics. 2018;19:415. https://doi.org/10.1186/s12864-018-4805-8.
19. Kolek J, Diallo M, Vasylkivska M, Branska B, Sedlar K, López-Contreras AM, et al. Comparison of expression of key sporulation, solventogenic and acetogenic genes in *C. beijerinckii* NRRL B-598 and its mutant strain overexpressing spo0A. Appl Microbiol Biotechnol. 2017;101:8279–91.
20. Tomas CA, Beamish J, Papoutsakis ET. Transcriptional analysis of butanol stress and tolerance in *Clostridium acetobutylicum*. J Bacteriol. 2004;186:2006–18.
21. Alsaker KV, Paredes C, Papoutsakis ET. Metabolite stress and tolerance in the production of biofuels and chemicals: gene-expression-based systems analysis of butanol, butyrate, and acetate stresses in the anaerobe *Clostridium acetobutylicum*. Biotechnol Bioeng. 2010;105:1131–47.
22. Kolek J, Patakova P, Melzoch K, Sigler K, Rezanka T. Changes in membrane plasmalogens of *Clostridium pasteurianum* during butanol fermentation as determined by lipidomic analysis. PLoS ONE. 2015;10:e0122058.
23. Lipovsky J, Patakova P, Paulova L, Pokorny T, Rychtera M, Melzoch K. Butanol production by *Clostridium pasteurianum* NRRL B-598 in continuous culture compared to batch and fed-batch systems. Fuel Process Technol. 2016;144:139–44.
24. Bateman A, Martin M-J, Orchard S, Magrane M, Alpi E, Bely B, et al. UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res. 2018;47:D506–15.
25. Mitchell AL, Attwood TK, Babbitt PC, Blum M, Bork P, Bridge A, et al. InterPro in 2019: improving coverage, classification and access to protein sequence annotations. Nucleic Acids Res. 2018;47:D351–60.
26. Carbon S, Douglass E, Dunn N, Good B, Harris NL, Lewis SE, et al. The gene ontology resource: 20 years and still GOing strong. Nucleic Acids Res. 2018;47:D330–8.
27. Sweeney BA, Petrov AI, Burkov B, Finn RD, Bateman A, Szymanski M, et al. RNAcentral: a hub of information for non-coding RNA sequences. Nucleic Acids Res. 2018;47:D221–9.
28. Altschul S, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997;25:3389–402. https://doi.org/10.1093/nar/25.17.3389.
29. Alexa A, Rahnenfuhrer J. topGO: enrichment analysis for gene ontology. R package version 2.30.1. 2016.
30. Van Der Maaten L, Hinton G. Visualizing Data using t-SNE. J Mach Learn Res. 2008;620:267–84.
31. Au KF, Underwood JG, Lee L, Wong WH. Improving PacBio long read accuracy by short read alignment. PLoS ONE. 2012;7:e46679. https://doi.org/10.1371/journal.pone.0046679.
32. Xu M, Zhao J, Yu L, Tang I-C, Xue C, Yang S-T. Engineering *Clostridium acetobutylicum* with a histidine kinase knockout for enhanced *n*-butanol tolerance and production. Appl Microbiol Biotechnol. 2015;99:1011–22. https://doi.org/10.1007/s00253-014-6249-7.
33. Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, von Mering C, et al. Fast genome-wide functional annotation through orthology assignment by eggNOG-Mapper. Mol Biol Evol. 2017;34:2115–22.
34. Gotz S, Garcia-Gomez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, et al. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res. 2008;36:3420–35. https://doi.org/10.1093/nar/gkn176.
35. Skunca N, Altenhoff A, Dessimoz C. Quality of computationally inferred gene ontology annotations. PLoS Comput Biol. 2012;8:e1002533. https://doi.org/10.1371/journal.pcbi.1002533.
36. Branska B, Pechacova Z, Kolek J, Vasylkivska M, Patakova P. Flow cytometry analysis of *Clostridium beijerinckii* NRRL B-598 populations exhibiting different phenotypes induced by changes in cultivation conditions. Biotechnol Biofuels. 2018;11:99. https://doi.org/10.1186/s13068-018-1096-x.
37. Whitley D, Goldberg SP, Jordan WD. Heat shock proteins: a review of the molecular chaperones. J Vasc Surg. 1999;29:748–51.
38. Mann MS, Dragovic Z, Schirmacher G, Lütke-Eversloh T. Over-expression of stress protein-encoding genes helps *Clostridium acetobutylicum* to rapidly adapt to butanol stress. Biotechnol Lett. 2012;34:1643–9. https://doi.org/10.1007/s10529-012-0951-2.
39. Liao Z, Zhang Y, Luo S, Suo Y, Zhang S, Wang J. Improving cellular robustness and butanol titers of *Clostridium acetobutylicum* ATCC824 by introducing heat shock proteins from an extremophilic bacterium. J Biotechnol. 2017;252:1–10.
40. Tomas CA, Welker NE, Papoutsakis ET. Overexpression of groESL in *Clostridium acetobutylicum* results in increased solvent production and tolerance, prolonged metabolism, and changes in the cell's transcriptional program. Appl Environ Microbiol. 2003;69:4951–65.
41. Reyes LH, Almario MP, Kao KC. Genomic library screens for genes involved in n-butanol tolerance in *Escherichia coli*. PLoS ONE. 2011;6:e17678.
42. Xue C, Zhao J, Chen L, Yang ST, Bai F. Recent advances and state-of-the-art strategies in strain and process engineering for biobutanol production by *Clostridium acetobutylicum*. Biotechnol Adv. 2017;35:310–22.
43. Steiner E, Scott J, Minton NP, Winzer K. An agr quorum sensing system that regulates granulose formation and sporulation in *Clostridium acetobutylicum*. Appl Environ Microbiol. 2012;78:1113–22.
44. Setlow B, Setlow P. Binding of small, acid-soluble spore proteins to DNA plays a significant role in the resistance of *Bacillus subtilis* spores to hydrogen peroxide. Appl Environ Microbiol. 1993;59:3418–23.
45. Setlow P. I will survive: DNA protection in bacterial spores. Trends Microbiol. 2007;15:172–80.
46. Raju D, Waters M, Setlow P, Sarker MR. Investigating the role of small, acid-soluble spore proteins (SASPs) in the resistance of *Clostridium perfringens* spores to heat. BMC Microbiol. 2006;6:50.
47. Wetzel D, Fischer R-J. Small acid-soluble spore proteins of *Clostridium acetobutylicum* are able to protect DNA in vitro and are specifically cleaved by germination protease GPR and spore protease YyaC. Microbiology. 2015;161:2098–109.
48. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics. 2016;32:3047–8. https://doi.org/10.1093/bioinformatics/btw354.
49. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114–20.
50. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics. 2009;25:1754–60. https://doi.org/10.1093/bioinformatics/btp324.

51. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS ONE. 2014;9:e112963. https://doi.org/10.1371/journal.pone.0112963.
52. Kopylova E, Noé L, Touzet H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics. 2012;28:3211–7.
53. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 2013;41:D590–6.
54. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29:15–21. https://doi.org/10.1093/bioinformatics/bts635.
55. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. Bioinformatics. 2009;25:2078–9. https://doi.org/10.1093/bioinformatics/btp352.
56. Liao Y, Smyth GK, Shi W. FeatureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30:923–30.
57. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:550. https://doi.org/10.1186/s13059-014-0550-8.
58. Van Der Maaten L. Accelerating t-SNE using tree-based algorithms. J Mach Learn Res. 2014;15:1–21.
59. Wickham H. ggplot2 elegant graphics for data analysis. Cham: Springer; 2009.
60. Chen H, Boutros PC. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. BMC Bioinform. 2011;12:35. https://doi.org/10.1186/1471-2105-12-35.
61. Binns D, Dimmer E, Huntley R, Barrell D, O'Donovan C, Apweiler R. QuickGO: a web-based tool for gene ontology searching. Bioinformatics. 2009;25:3045–6. https://doi.org/10.1093/bioinformatics/btp536.
62. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics. 2005;21:3674–6. https://doi.org/10.1093/bioinformatics/bti610.
63. Fang H, Gough J. The 'dnet' approach promotes emerging research on cancer patient survival. Genome Med. 2014;6:64. https://doi.org/10.1186/s13073-014-0064-8.
64. Csárdi G, Nepusz T. The igraph software package for complex network research. InterJournal Complex Syst. 2006;1695:1–9.

Publisher's Note

5.8 Article VIII

VASYLKIVSKA, Maryna, BRANSKA, Barbora, **SEDLAR, Karel**, JURECKOVA, Katerina, PROVAZNIK, Ivo and PATAKOVA, Petra. Phenotypic and Genomic Analysis of Clostridium beijerinckii NRRL B-598 Mutants With Increased Butanol Tolerance. *Frontiers in Bioengineering and Biotechnology*. 2020. 8, p. 1307. (2020 IF = 5.890, Q1 in MULTIDISCIPLINARY SCIENCES)

DOI: https://doi.org/10.3389/fbioe.2020.598392

frontiers in Bioengineering and Biotechnology

ORIGINAL RESEARCH
published: 05 November 2020
doi: 10.3389/fbioe.2020.598392

Phenotypic and Genomic Analysis of *Clostridium beijerinckii* NRRL B-598 Mutants With Increased Butanol Tolerance

Maryna Vasylkivska[1], Barbora Branska[1], Karel Sedlar[2], Katerina Jureckova[2], Ivo Provaznik[2] and Petra Patakova[1]*

[1] Department of Biotechnology, University of Chemistry and Technology, Prague, Prague, Czechia, [2] Department of Biomedical Engineering, Faculty of Electrical Engineering and Communication, Brno University of Technology, Brno, Czechia

OPEN ACCESS

Edited by:
Fu-Li Li,
Qingdao Institute of Bioenergy and Bioprocess Technology (CAS), China

Reviewed by:
Yi Wang,
Auburn University, United States
Noppol - Leksawasdi,
Chiang Mai University, Thailand
Thaddeus Chukwuemeka Ezeji,
The Ohio State University, United States

***Correspondence:**
Maryna Vasylkivska
vasylkim@vscht.cz

Specialty section:
This article was submitted to Bioprocess Engineering, a section of the journal Frontiers in Bioengineering and Biotechnology

Received: 24 August 2020
Accepted: 20 October 2020
Published: 05 November 2020

Citation:
Vasylkivska M, Branska B, Sedlar K, Jureckova K, Provaznik I and Patakova P (2020) Phenotypic and Genomic Analysis of Clostridium beijerinckii NRRL B-598 Mutants With Increased Butanol Tolerance. Front. Bioeng. Biotechnol. 8:598392. doi: 10.3389/fbioe.2020.598392

N-Butanol, a valuable solvent and potential fuel extender, can be produced via acetone-butanol-ethanol (ABE) fermentation. One of the main drawbacks of ABE fermentation is the high toxicity of butanol to producing cells, leading to cell membrane disruption, low culture viability and, consequently, low produced concentrations of butanol. The goal of this study was to obtain mutant strains of *Clostridium beijerinckii* NRRL B-598 with improved butanol tolerance using random chemical mutagenesis, describe changes in their phenotypes compared to the wild-type strain and reveal changes in the genome that explain improved tolerance or other phenotypic changes. Nine mutant strains with stable improved features were obtained by three different approaches and, for two of them, ethidium bromide (EB), a known substrate of efux pumps, was used for either selection or as a mutagenic agent. It is the rst utilization of this approach for the development of butanol-tolerant mutants of solventogenic clostridia, for which generally there is a lack of knowledge about butanol efux or efux mechanisms and their regulation. Mutant strains exhibited increase in butanol tolerance from 36% up to 127% and the greatest improvement was achieved for the strains for which EB was used as a mutagenic agent. Additionally, increased tolerance to other substrates of efux pumps, EB and ethanol, was observed in all mutants and higher antibiotic tolerance in some of the strains. The complete genomes of mutant strains were sequenced and revealed that improved butanol tolerance can be attributed to mutations in genes encoding typical stress responses (chemotaxis, autolysis or changes in cell membrane structure), but, also, to mutations in genes X276_07980 and X276_24400, encoding efux pump regulators. The latter observation conrms the importance of efux in butanol stress response of the strain and offers new targets for rational strain engineering.

Keywords: butanol tolerance, random chemical mutagenesis, solventogenic Clostridium species, genome sequence, butanol efux

INTRODUCTION

Butanol can be produced from renewable feedstocks of dierent kinds, including agricultural waste materials, by acetone-butanol-ethanol (ABE) fermentation using solventogenic clostridia (Lee et al., 2008). The most famous species of the group of solventogenic clostridia are *Clostridium acetobutylicum* and *Clostridium beijerinckii*, both of which share a common process bottleneck low tolerance to butanol.

Butanol is a toxic metabolite that tends to incorporate into the cell membrane, increases membrane uidity and may disrupt membrane functions (Bowles and Ellefson, 1985; Lepage et al., 1987; Patakova et al., 2018; Vasylkivska and Patakova, 2020). In *C. acetobutylicum*, at inhibitory levels, eects of butanol on cell membrane results in lower ATP generation by the cell, a malfunction in nutrient uptake and an inability of the cell to maintain its internal pH (Bowles and Ellefson, 1985). In Gram-negative bacteria, butanol damages the inner and outer membranes and can also result in a change in cell shape (Fletcher et al., 2016). Such damage, at the cellular level, results in low culture viability and decreased growth rate and, as a consequence, low nal butanol concentrations achieved after fermentation (Ingram, 1986).

Dierent methods have been used to obtain butanol-tolerant strains of solventogenic clostridia, including targeted modications, e.g., overexpression of genes encoding heat-shock proteins or modication of fatty acids synthesis (Tomas et al., 2003; Zhao et al., 2003; Mann et al., 2012; Liao et al., 2017), serial transfer and adaptation (Lin and Blaschek, 1983; Baer et al., 1987; Soucaille et al., 1987; Xue et al., 2012; Liu et al., 2013; Yang and Zhao, 2013) or random mutagenesis (Hermann et al., 1985; Matta-el-Ammouri et al., 1986; Annous and Blaschek, 1991; Jain et al., 1994; Mao et al., 2010; Kong et al., 2016; Tanaka et al., 2017). Use of these methods resulted in improved survival in the presence of butanol at a concentration of 1012 g/L for the wild-type strain (WTS) to 1618 g/L for mutant strains (Vasylkivska and Patakova, 2020), sometimes up to 23 g/L, as in the case of mutant strain *C. beijerinckii* BA101 (Qureshi and Blaschek, 2001). Surprisingly, random and targeted mutagenesis have apparently produced very similar improvements in butanol tolerance, although this probably reects our incomplete understanding of butanol tolerance mechanisms and their regulation (Patakova et al., 2018). In most cases, obtained mutant strains in addition to higher tolerance exhibited higher butanol production, usually an improvement from approximately 912 g/L for WTS to 13 16 g/L for mutant strains, sometimes up to 1921 g/L for selected mutants (Vasylkivska and Patakova, 2020). However, some authors have reported increased tolerance, but not increased butanol production for obtained mutants, both for random (Baer et al., 1987; Gallardo et al., 2017; M[illegible] de G[illegible]rando et al., 2018) and targeted (Zhao et al., 2003; Alsaker et al., 2004; Mann et al., 2012; Jones et al., 2016) mutagenesis. Thus, although it is commonly accepted that increased butanol tolerance leads to higher production, such data suggest strain-specic dependence between butanol tolerance and production or even the absence of any direct connection.

Eux is one of the innate mechanisms of stress response in bacteria and it is based on active transport of the substances from the cell. This mechanism is mostly associated with antibiotic resistance, however, it has been shown that it also takes part in the butanol stress response of *Pseudomonas putida* and *Escherichia coli* (Fisher et al., 2014; Bui et al., 2015; Basler et al., 2018; Zhang et al., 2018). It was suggested that eux can be very eective when cells are dealing with high solvent concentrations (Segura et al., 2012), and enhancement of eux pump activity could possibly shift metabolic ux, resulting in higher production (Mukhopadhyay, 2015).

Eux, particularly butanol eux, is rarely studied in solventogenic clostridia (Vasylkivska and Patakova, 2020). Up-regulation of ATP-binding cassette transporters (Schwarz et al., 2012) and putative eux pump regulators (Sedlar et al., 2019) observed under cultivation with butanol stress are the only evidence of innate butanol eux in the group. Recently, a butanol eux pump from *P. putida* S12 was expressed in *Clostridium saccharoperbutylacetonicum*, resulting in improved butanol tolerance (Jiménez-Bonilla et al., 2020) and demonstrating that butanol eux studies in solventogenic *Clostridium* have potential for the development of butanol-tolerant production strains. As no native butanol eux pumps have yet been reported for solventogenic clostridia, targeted engineering cannot be used for the study. Nevertheless, it was shown that eux enhancement can be achieved even by a point mutation in the eux pump gene sequence (Bohnert et al., 2007) or in the eux pump promoter or regulator. To generate such mutations, ethyl methanesulfonate (EMS) or ethidium bromide (EB) can be used. EMS can alkylate guanine bases in DNA, resulting in unidirectional random transition mutations between GC and AT base pairs. Such a type of mutation can lead to an amino acid change and even loss of protein function. EMS was previously successfully used by Jain et al. (1994) to obtain a high butanol producing stable asporogenic mutant, *C. beijerinckii* ATCC 55025, formerly *C. acetobutylicum* ATCC 55025 (Jain et al., 1994). EB is a known substrate of dierent eux pumps in both Gram-positive (Patel et al., 2010) and Gram-negative bacteria (Paixao et al., 2009) and also a chemical mutagen (Ohta et al., 2001). Use of EB may enhance eux in the strain, resulting in improved butanol tolerance. Therefore, for this study, we have chosen three approaches to develop butanol-tolerant mutants of solventogenic *C. beijerinckii* NRRL B-598:

(1) Random chemical mutagenesis using EMS with selection on butanol (EMS + butanol mutants).
(2) Random chemical mutagenesis using EMS with selection on EB (EMS + EB mutants).
(3) Random chemical mutagenesis using EB as a mutagenic agent, where strains were selected directly on agar plates containing EB (EB mutants).

Phenotypic behavior of selected mutant strains was observed by testing tolerance to dierent substances such as butanol, EB, ethanol and antibiotics, and metabolite production. The complete genomes of mutant strains that exhibited improved butanol tolerance were sequenced to understand their changes in

phenotype and to contribute to knowledge about mutations that lead to increased butanol tolerance. To the best of our knowledge, the genomic sequence of only a few butanol-tolerant strains obtained by random mutagenesis are available.

MATERIALS AND METHODS

Bacterial Strain

Clostridium beijerinckii NRRL B-598 (WTS), former *Clostridium pasteurianum* NRRL B-598 (Sedlar et al., 2017), obtained from the Agricultural Research Service Culture Collection (1815 N. University Street, Peoria, IL 61604) was used in this study. The culture was maintained in the form of a spore suspension [containing about $2.2 \cdot 10^8$ spores/ml, determined by method described by Branska et al. (2018)] in sterile distilled water at 4°C. For each cultivation experiment, 450 µl of spore preserve was used for 100 ml of cultivation medium.

Mutagenesis and Strain Selection

Mutagenesis

Mutant strains described in this work were obtained using three dierent approaches of random chemical mutagenesis (**Figure 1**). Firstly, ethyl methanesulfonate (EMS) was used as a mutagenic agent in combination with selection in butanol (EMS + butanol mutants), secondly, EMS was used for mutagenesis but selection was carried out in ethidium bromide (EB) (EMS + EB mutants) and, nally, EB was used as a mutagenic agent and strains were directly selected on agar plates containing EB without exposition to EMS (EB mutants).

For the rst two approaches, when EMS was used as a mutagenic agent, a spore suspension of *C. beijerinckii* NRRL B-598 was heated to 80°C for 2 min, vortexed and transferred to Erlenmeyer asks with TYA medium (Vasylkivska et al., 2019) containing 20 g/L of glucose (analytical reagent grade, PENTA, Chrudim, Czechia). The strain was cultivated in a Concept 400 anaerobic chamber (Ruskinn Technology) under a stable N_2 atmosphere at 37°C (**Figure 1**, step 1). After 24 h of cultivation, 2 ml × 2 ml of cell suspension were transferred into sterile Eppendorf tubes, centrifuged and 0.5 ml of fresh sterile TYA medium were added to the cell pellet. Eppendorf tubes were vortexed and 20 µl/ml of EMS (pure, Sigma-Aldrich) were added (**Figure 1**, step 2). The exposition time was 40 min, during which Eppendorf tubes were placed in the anaerobic chamber. The cell suspension with mutagen was further centrifuged and washed twice with sterile physiological solution. The washed cell suspension was transferred to Petri dishes (250 µl of suspension on each plate) containing HPLC-grade butanol (Sigma-Aldrich) (**Figure 1**, step A3) or EB (for molecular biology, 10 mg/mL in H_2O, Sigma-Aldrich) (**Figure 1**, step B3) for selection. Tolerance of WTS to butanol and EB was tested prior to mutagenesis procedure (details are described in section Tolerance Testing) and twice as high a concentration of butanol (for EMS + butanol mutants) or twice as high a concentration of EB (for EMS + EB mutants) than WTS was able to tolerate were used for mutants selection on Petri dishes (**Figure 1**, step 3). Inoculated petri dishes were cultivated for 48 h at 37°C in the anaerobic chamber. Colonies obtained were transferred to test tubes containing 10 ml of TYA medium with 20 g/L of glucose, and cultivated overnight in the anaerobic chamber (**Figure 1**, step 4). The cell suspension was cryopreserved in 30% (v/v) glycerol (analytical reagent grade, PENTA, Chrudim, Czechia) solution below −80°C.

EB mutant strains, prepared by direct cultivation on agar plates containing EB without exposition to EMS, were obtained as follows: the spore suspension of *C. beijerinckii* NRRL B-598 was heated to 80°C for 2 min (**Figure 1**, step C1), vortexed and 50 µl of the spore suspension were transferred onto Petri dishes containing either 0.5 mg/L or 2 mg/L of EB (**Figure 1**, step C2). Agar plates were cultivated in the anaerobic chamber for 24 h, chosen colonies (**Figure 1**, step C3) were cultivated in TYA medium and the cell culture was cryopreserved (**Figure 1**, step 4).

Mutant strains obtained by all methods were tested for butanol tolerance (details are described in section Tolerance Testing) (**Figure 1**, step 5). Selected mutant strains with improved butanol tolerance compared to the wild-type strain (WTS) were cultivated in TYA medium (EB mutants) or on TYA agar plates (EMS + butanol and EMS + EB mutants) until sporulation was observed (Olympus BX51 microscope); spore preserves were prepared (**Figure 1**, step 6). Inocula were prepared from the spore preserves and mutant strains were tested one more time for butanol tolerance (**Figure 1**, step 7). After this step, the selected strains were further used in the experiments described in this article.

Tolerance Testing

Inocula of the wild-type strain and of the mutant strains (except for the rst butanol tolerance testing shown as step 5 in **Figure 1**) were prepared from spore suspensions, and inocula of the mutant strains for the rst butanol tolerance testing were prepared from the cryopreserves. Prior to inoculation, the spore suspensions of strains were heated to 80°C for 2 min and vortexed (heat-shock); cryopreserves were thawed at room temperature and vortexed. TYA medium containing 20 g/L of glucose was used for the preparation of inocula and inoculated test tubes were transferred to the anaerobic chamber and cultivated at 37°C overnight.

Tolerance to various substances was tested in microtiter plates containing 120 µl of TYA medium and 10 µl of cell culture. Medium for the experiment additionally contained 0.02 g/L of acid base indicator, bromocresol purple (suitable for indicator, dye content 90%, Sigma-Aldrich), and each substance in dierent concentrations. Eight substances were tested: two metabolites [butanol in the range of 0 to 30 g/L and ethanol (analytical reagent grade, PENTA, Chrudim, Czechia) in the range of 0 to 65 g/L], eux pump inducer ethidium bromide EB in the range of 06 mg/L and ve antibiotics [chloramphenicol (≥98% (HPLC), Sigma-Aldrich] in the range of 0 to 150 mg/L, tetracycline [98.0102.0% (HPLC), Sigma-Aldrich] in the range of 030 mg/L, streptomycin in a form of streptomycin sulfate [>95.0% (TN), Tokyo Chemical Industry] in the range of 0 to 35 mg/L, ampicillin in a form of sodium salt (pure Ph. EUr., AppliChem) in the range of 0100 mg/L and erythromycin [(for microbiological assay, Sigma-Aldrich) in the range of 0100 mg/L]. Inoculated microtiter plates were cultivated in the anaerobic chamber at 37°C for 24 h. Results of the

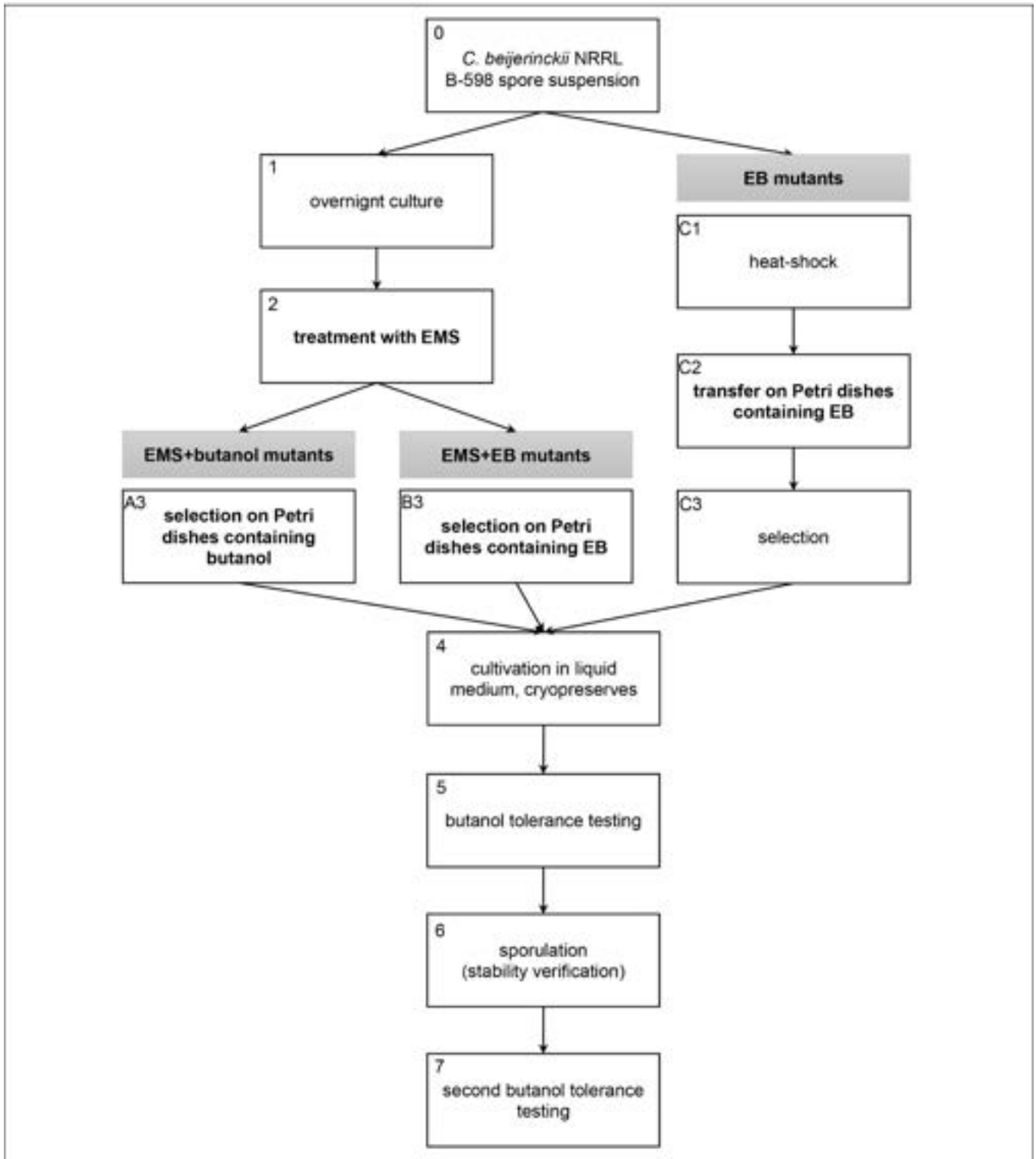


FIGURE 1 | Scheme of three random chemical mutagenesis approaches that were used to obtain mutant strains of *C. beijerinckii* NRRL B-598 with increased butanol tolerance. EMS, ethyl methanesulfonate; EB, ethidium bromide.

testing were evaluated visually as a change in color of the medium, from purple to yellow due to acid production and medium pH shift indicating growth of the strain. For each strain and each substrate, butanol tolerance was tested in at least three repetitions and TYA medium without additions was used as a control.

Cultivation Experiments

Cultivation in Erlenmeyer Flasks

For the determination of glucose consumption rate and metabolite production, WTS and mutant strains were cultivated in triplicates in non-shaken Erlenmeyer asks. TYA medium containing 40 g/L of glucose was used for both inoculum preparation and cultivation experiment itself. The inoculum was prepared from the spore preserves after heat-shock (as described in section Tolerance Testing) and cultivated in an anaerobic chamber under a stable N_2 atmosphere at 37°C overnight. For the cultivation experiment, asks were inoculated with 10% (v/v) cell culture and cultivated in the anaerobic chamber for 72 h. At the end of cultivation, samples were taken for pH measurement and subsequent HPLC analysis (details of the analysis are described in section Analytical Methods).

Cultivation in triplicates in non-shaken Erlenmeyer asks with pH control was performed the same way as described above, but TYA medium after sterilization prior to inoculation was supplemented with $CaCO_3$ so that concentration 10 g/L of $CaCO_3$ was achieved.

Cultivation in Bioreactors

According to the result of experiment described in Section Cultivation in Erlenmeyer Flasks, one selected mutant strain was chosen for batch cultivation alongside with WTS in triplicates in parallel Multiforce 1L bioreactors (Infors HT).

TYA medium for the inoculum was prepared with glucose concentrations of 20 g/L and for cultivation in bioreactors with concentrations of 40 g/L. The inoculum was prepared from the spore preserves as described in Section Cultivation in Erlenmeyer Flasks.. Prior to inoculation of bioreactors, oxygen was exchanged with N_2 and the pH of the medium was adjusted to 6.4. Bioreactors were inoculated with 10% (v/v) cell culture. Cultivation temperature was 37°C and agitation speed was set to 3.3 s^{-1} throughout the cultivation, the pH was not controlled. During the cultivation, samples were taken for OD measurements and subsequent HPLC analysis (details of the analysis are described in section Analytical Methods).

Analytical Methods

Culture growth was measured as the optical density OD of the culture broth at 600 nm (Varian Cary 50 UV-Vis spectrophotometer, Agilent) against TYA medium as a blank (Vasylkivska et al., 2019).

The concentrations of lactic acid (retention time t_R 6.9 min), acetic acid (t_R 8.1 min), ethanol (t_R 11.7 min), acetone (t_R 12.3 min), butyric acid (t_R 13.5 min), butanol (t_R 24.9 min), and glucose (t_R 4.8 min) were measured using HPLC with refractive index detection (Agilent Series 1200 HPLC). A column with stationary phase of Polymer IEX H from 8 µm (Watrex) was used for the separation. Samples of culture broth were centrifuged and the supernatants were microltered. The sample injection volume was 20 µl, the column temperature was 60°C, and 5 mM H_2SO_4 was used as the mobile phase with a ow-rate of 1 ml/min. The concentration of substances was determined from calibration curves (Sedlar et al., 2018).

DNA Isolation and Genome Sequencing

DNA was isolated from exponentially growing cultures prepared by inoculation of TYA medium with heat-shocked spores (as described above) using commercially available isolation kit DNeasy UltraClean Microbial Kit (Qiagen, Hilden, Germany), following recommended instructions. Library construction and sequencing of the sample was performed by CEITEC Genomics core facility (Brno, Czechia) on Illumina NextSeq 500, pair-end, 150 bp.

Bioinformatics Analysis

The quality assessment of particular steps of data processing were done using FastQC in combination with MultiQC to summarize the reports across all samples (Ewels et al., 2016). Adapter and quality trimming and ltering of singletons was performed with Trimmomatic v0.36 (Bolger et al., 2014). Remaining high quality paired reads were mapped to the genome sequence of the wild type strain with BWA-mem v0.7.15 (Li and Durbin, 2009). The latest genome assembly of the *C. beijerinckii* NRRL B-598 CP011966.3 (Sedlar et al., 2019) was used as a reference. Resulting SAM (Sequence Read Alignment/Map) les were indexed and transformed into more compact BAM (Binary Read Alignment/Map) format using SAMtools v1.9 (Li et al., 2009). Sequences in BAM les were cleaned, sorted, deduplicated and les were indexed with Picard Tools v2.21.6. Single nucleotide variants in genome sequences between wild type strain and particular mutant strains were called with GATK v4.1.4.1 (McKenna et al., 2010). Detected variants were further ltered in order to reduce the number of false positives. For this purpose, WTS resequencing data gathered within the same sequencing run, were used. All variants that were called simultaneously in the WTS and mutant were ltered out. Moreover, we used coverage of false detections in the WTS to set a threshold for ltering, and only variants covered by more than 25.9% of an average coverage of a strain were used. This threshold corresponded to the highest coverage of falsely detected mutations in the WTS. Moreover, not all variants were detected in the whole population. We again set a threshold using analysis of the WTS. Only variants that were called in at least 30.8% of the population were counted. Furthermore, we detected structural variations to the genome sequences, including copy number variations (CNV) with Pilon v 1.22 (Walker et al., 2014). All these analyses were performed with R/Bioconductor using functions from the genomeIntervals v1.42 (Gagneur et al., 2020), Biostrings v2.54 (Pagès et al., 2020), vcfR v1.11 (Knaus and Grünwald, 2017), and ggplot2 (Wickham, 2009) packages.

RESULTS

Mutant Strains, Tolerance Testing to Butanol, Ethanol and Tolerance to Known Efux Pump Substrates

As a rst step, WTS *C. beijerinckii* NRRL B-598 was tested for its tolerance to butanol, ethanol and EB. Testing revealed

that the strain was able to grow in medium containing no more than 11 ± 1 g/L of butanol, 38.5 ± 2.1 g/L of ethanol or 1.75 ± 0.40 mg/L of EB. EMS + butanol and EMS + EB mutants were, therefore, picked (**Figure 1**, step 3) from agar plates containing approximately twice as high a concentration of butanol or EB, which were 20 g/L and 4 g/L, respectively, prior to both butanol tolerance testing in microtiter plates (**Figure 1, steps 5** and 7).

Number of strains being reduced through screening process is shown in **Table 1**. For mutagenesis using EMS as mutagenic agent, total of 45 colonies were selected during step A3 (**Figure 1**) and 45 during B3 (**Figure 1**). The rst butanol tolerance testing (**Figure 1**, step 5) revealed that 32 strains exhibited increased butanol tolerance compared to WTS. To ensure that increased butanol tolerance was a result of mutation and not adaptation, these 32 strains were cultivated on TYA agar plates until sporulation was observed (**Figure 1**, step 6) and then spores were germinated and the strains were once again tested for butanol tolerance (**Figure 1**, step 7). After the second round of butanol tolerance testing, six strains exhibiting increased butanol tolerance were selected for further experiments: EMS + butanol mutants B33 and B44 and EMS + EB mutants E15, E28, E32, and E33 (**Table 1**).

Using another approach, mutagenesis of WTS was carried out on agar plates containing EB with no exposition to EMS (EB mutants, **Figure 1** steps C1C3). 24 colonies (**Table 1**) were picked during C3 step (**Figure 1**). After the subsequent butanol tolerance test in microtiter plates (**Figure 1**, step 5), three EB mutants strains exhibited an increase in tolerance, strains A, B, and C. These strains were also cultivated until sporulation was observed (**Figure 1**, step 6) and then the spores were germinated and the strains were again tested for butanol tolerance (**Figure 1**, step 7). This conrmed that the acquired increase in tolerance was a stable phenotype.

All of mutant strains showed a signicant increase in butanol and ethanol tolerance compared to WTS ($p < 0.05$, two-sample *t*-test) and were able to grow in a medium containing up to 25.0 g/L of butanol or up to 55.0 g/L of ethanol (**Figure 2**).

As the goal of the study was to increase the eux capacity of mutant strains, the tolerance of mutant strains and the WTS to dierent substrates of eux pumps, i.e., EB and antibiotics, was also tested. Mutants acquired increased EB (up to 5 mg/L) tolerance, see **Figure 3**.

Wild-type strain was tolerant to tetracycline, chloramphenicol and streptomycin at concentrations of 3.0 ± 0.0, 19.0 ± 1.4, and 20.0 ± 0.0 µg/ml, respectively, but was sensitive to ampicillin and erythromycin. The sensitivity toward ampicillin and erythromycin remained unchanged in all mutant strains, but tolerance to tetracycline and chloramphenicol was modied and, in some cases, increased. Surprisingly, tolerance to streptomycin decreased in all cases and strains B33 (EMS + butanol mutant), E15, E32, and E33 (EMS + EB mutants) exhibited complete growth inhibition in the presence of this antibiotic, see **Figure 4**.

Metabolites Production by WTS and Mutant Strains

Production of solvents and acids was tested in asks containing TYA medium. Mutant strains A and C exhibited similar fermentation proles as WTS and produced butanol in similar concentrations under standard cultivation conditions (**Supplementary Table 1**). Other mutant strains, B (EB mutant), B33, B44 (EMS + butanol mutants), E15, E28, E32, and E33 (EMS + EB mutants), probably developed the so-called acid crash phenotype and their respective fermentation outputs achieved higher concentrations of butyric, acetic or lactic acids compared to WTS.

Inspired by the study of the mutant strain *C. beijerinckii* BA105, which at rst was seem to only exhibit the acid crash phenotype but it was later revealed that signicantly higher butanol production can be achieved by the strain under pH regulation (Seo et al., 2017), we tested TYA medium with the addition of 10 g/L $CaCO_3$ for partial pH control (**Table 2**). WTS and mutant strains A and C (EB mutants) produced 1116% higher concentrations of butanol and consumed the total amount of glucose present in the medium. Furthermore, mutant strain B (EB mutant) was able to produce butanol at a concentration of 4.7 ± 0.5 g/L, which is signicantly ($p < 0.05$, two-sample *t*-test), almost 12 times, higher than that achieved in the medium without supplementation. However, EMS + butanol and EMS + EB stayed in the acidogenic phase of fermentation and the only dierence compared to previous experiments was a signicantly ($p < 0.05$, two-sample *t*-test) higher concentration of butyric acid at the end of fermentation.

The behavior of the mutant strain with the overall highest butanol production, EB mutant C, was compared with WTS

TABLE 1 | Number of mutant strains being reduced through screening process[†].

Mutagenesis method	Colonies picked after mutagenesis*	Strains exhibiting increased butanol tolerance (after 1st butanol testing)**	Strains exhibiting increased butanol tolerance (after 2d butanol testing)***	Strains selected for further experiments
EB	24	3	3	A, B and C
EMS + butanol	45	13	2	B33 and B44
EMS + EB	45	19	4	E15, E28, E32 and E33

[†]**Steps A3, B3, and C3 from **Figure 1**; **step 5 from **Figure 1**; ***step 7 from **Figure 1**.*
EB strains obtained by random chemical mutagenesis using EB as a mutagenic agent, when strains were selected directly on agar plates containing EB; EMS + butanol strains obtained by random chemical mutagenesis using EMS with selection on butanol; EMS + EB strains obtained by random chemical mutagenesis using EMS with selection on EB.

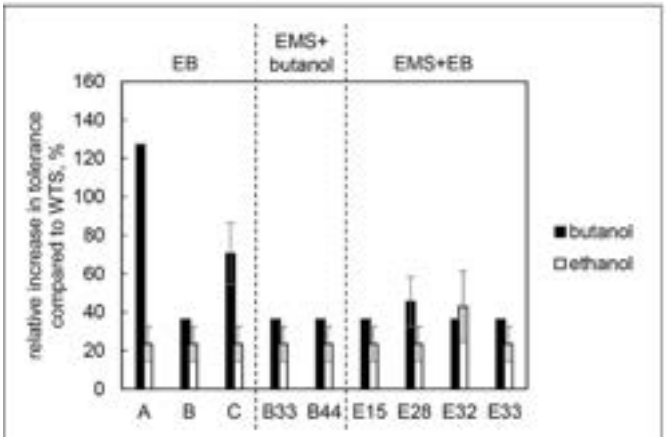


FIGURE 2 | Relative increase in butanol and ethanol tolerance of mutant strains compared to *C. beijerinckii* NRRL B-598 (WTS). Tolerance levels are expressed as the maximum concentration of butanol or ethanol in TYA medium at which growth of the strains was still detectable. The original tolerance of WTS to butanol and ethanol was 11 ± 1 g/L and 38.5 ± 2.1 g/L, respectively. EB- strains obtained by random chemical mutagenesis using EB as a mutagenic agent, when strains were selected directly on the agar plates containing EB; EMS + butanol- strains obtained by random chemical mutagenesis using EMS with selection on butanol; EMS + EB- strains obtained by random chemical mutagenesis using EMS with selection on EB.

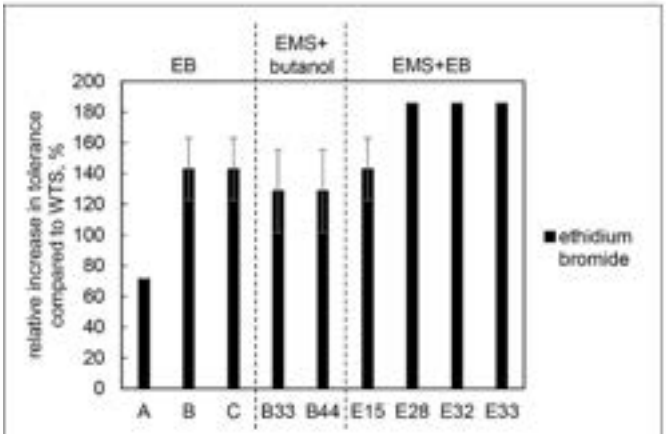


FIGURE 3 | Relative increase in tolerance to EB of mutant strains compared to *C. beijerinckii* NRRL B-598 (WTS). Tolerance levels are expressed as the maximum concentration of substances in TYA medium at which growth of the strains was still detectable. Original tolerance of WTS to EB was 1.75 ± 0.4 mg/L. EB- strains obtained by random chemical mutagenesis using EB as a mutagenic agent, when strains were selected directly on agar plates containing EB; EMS + butanol - strains obtained by random chemical mutagenesis using EMS with selection on butanol; EMS + EB- strains obtained by random chemical mutagenesis using EMS with selection on EB.

in parallel batch bioreactor fermentation (**Supplementary Figure 1**). At rst sight, the fermentation proles, growth curves and glucose consumption curves looked very similar, however, a thorough analysis of fermentation data revealed dierences in the fermentation dynamics. The pH and growth curves show that the mutant strain C switched to solventogenesis earlier than the WTS and grew faster during the exponential phase. This was also conrmed by calculation of the specic growth rate (μ) for the exponential phase of growth (rst 5 h of cultivation as no lag phase was observed), which was 0.32 ± 0.03 h^{-1} for the WTS and 0.48 ± 0.01 h^{-1} for the mutant strain C. Comparison of fermentation parameters for the rst 24 h of cultivation and total fermentation time (48 h) is given in **Supplementary Table 2**. Although the strains reached the same butanol yield (**Supplementary Table 2**), productivity for mutant C was somewhat higher when calculated for the rst 24 h of fermentation, which is in accordance with its higher glucose consumption rate during this period (**Supplementary Table 2**).

Genomic Analysis of Mutant Strains

To identify the causes of phenotypic changes in mutant strains, their genomes were sequenced and compared with the WTS genome. After quality trimming and deduplication, 2.44.5 million high quality (average Phred score $Q \approx 35$) paired sequences mapped to the reference genome suggesting coverage from 115 × to 221 × per mutant strain (see **Supplementary Table 3**). Sequencing of mutant strains revealed that most of the mutations were single-nucleotide polymorphisms (SNP). In unltered data, we detected from 21 to 51 SNPs or short indels per mutant strain and 18 false positive detections in the WTS (see **Supplementary Files 4**). Nevertheless, after ltering (see section Materials and Methods), the number of SNPs was reduced to the range from one to 17 per mutant strain (see **Table 3**). In total, 21 non-synonymous mutations were captured, including a mutation disrupting an open reading frame. SNPs in six genes were observed in at least two mutant strains (**Figure 4**). SNPs in X276_13415 encoding a putative S-layer family protein was detected in ve mutant strains, SNPs in X276_14460 encoding cytochrome b5 was observed in four strains. EB mutants strain A, B and C included SNPs in the same genes, X276_22865 and X276_03000 encoding a carbohydrate ABC transporter permease and an AAA family ATPase, respectively (**Figure 4** and **Table 3**), and no mutations were observed in these gene when EMS was used as the mutagen.

In addition to SNPs, we detected several longer genome changes, including copy number variations (CNVs) (**Tables 4, 5**). No copy number variations were detected for mutant strain E32. CNVs were detected for similar positions in strains A, B. C, B44, E28, and E33 (**Table 5**).

DISCUSSION

Using three dierent approaches of random chemical mutagenesis, we were able to obtain nine mutant strains of *C. beijerinckii* NRRL B-598 exhibiting from 36 to 127% increases in butanol tolerance (**Figure 2**). Similar increases in tolerance have been described for other solventogenic clostridia; in fact, the most well-studied and best-performing mutant strain, *C. beijerinckii* BA101, was likewise obtained by random chemical mutagenesis (Annous and Blaschek, 1991). *C. beijerinckii* BA101, an ospring of *C. beijerinckii* NCIMB 8052, exhibited around a 110% increase in butanol tolerance (Qureshi and Blaschek, 2001). Similarly, the asporogenic mutant *C. beijerinckii* ATCC

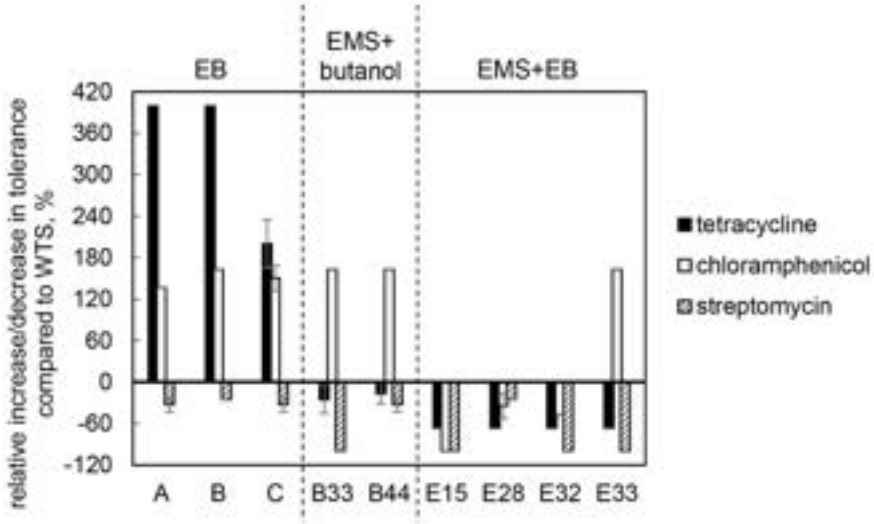


FIGURE 4 | Relative increase/decrease in tolerance to antibiotics tetracycline, chloramphenicol and streptomycin in mutant strains compared to *C. beijerinckii* NRRL B-598 (WTS). Tolerance levels are expressed as the maximum concentration of substances in TYA medium at which growth of the strains was still detectable. Tolerance of WTS was 3.0 ± 0.0, 19.0 ± 1.4, and 20.0 ± 0.0 μg/ml of tetracycline, chloramphenicol and streptomycin, respectively. EB strains obtained by random chemical mutagenesis using EB as a mutagenic agent, when strains were selected directly on agar plates containing EB; EMS + butanol strains obtained by random chemical mutagenesis using EMS with selection on butanol; EMS + EB strains obtained by random chemical mutagenesis using EMS with the selection on EB.

TABLE 2 | Concentrations of glucose, acids and solvents and pH reached by *C. beijerinckii* NRRL B-598 (WTS) and its mutant strains in TYA medium containing 10 g/L of $CaCO_3$ after 72 h cultivation[1].

Mutagenesis method	Strain	Consumed glucose, g/L	Lactic acid, g/L	Acetic acid, g/L	Ethanol, g/L	Acetone, g/L	Butyric acid, g/L	Butanol, g/L	Final pH
	WTS	38.4 ± 0.0	0.3 ± 0.0	3.0 ± 0.2	0.4 ± 0.0	1.3 ± 0.1	1.8 ± 0.1	**8.1 ± 0.2**	6.2 ± 0.0
EB	A	38.4 ± 0.0	0.2 ± 0.0	2.5 ± 0.1	0.3 ± 0.0	1.6 ± 0.1	1.7 ± 0.1	**8.1 ± 0.2**	6.4 ± 0.1
EB	B	34.4 ± 1.6	0.2 ± 0.0	3.3 ± 0.5	0.3 ± 0.1	0.1 ± 0.1	9.3 ± 0.7	**4.7 ± 0.5**	5.7 ± 0.0
EB	C	38.4 ± 0.0	0.2 ± 0.0	2.6 ± 0.1	0.4 ± 0.1	1.6 ± 0.1	1.7 ± 0.2	**8.1 ± 0.1**	6.4 ± 0.0
EMS + butanol	B33	26.1 ± 2.0	0.1 ± 0.1	1.5 ± 0.3	0.0 ± 0.0	0.0 ± 0.0	14.2 ± 0.8	**0.0 ± 0.0**	5.6 ± 0.0
EMS + butanol	B44	22.9 ± 0.5	0.1 ± 0.1	1.2 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	12.7 ± 0.3	**0.0 ± 0.1**	5.7 ± 0.0
EMS + EB	E15	26.5 ± 1.1	0.2 ± 0.0	1.0 ± 0.2	0.0 ± 0.0	0.0 ± 0.0	15.2 ± 0.6	**0.0 ± 0.0**	5.7 ± 0.0
EMS + EB	E28	22.4 ± 0.6	0.1 ± 0.1	1.3 ± 0.1	0.0 ± 0.0	0.0 ± 0.0	12.8 ± 0.2	**0.1 ± 0.0**	5.7 ± 0.0
EMS + EB	E32	24.5 ± 1.0	0.2 ± 0.1	1.9 ± 0.2	0.0 ± 0.0	0.0 ± 0.0	13.4 ± 0.4	**0.0 ± 0.0**	5.7 ± 0.0
EMS + EB	E33	26.7 ± 0.8	0.2 ± 0.0	1.7 ± 0.1	0.0 ± 0.0	0.0 ± 0.0	15.0 ± 0.4	**0.0 ± 0.0**	5.7 ± 0.0

[1] *EB, strains obtained by random chemical mutagenesis using EB as a mutagenic agent, when strains were selected directly on agar plates containing EB; EMS + butanol, strains obtained by random chemical mutagenesis using EMS with selection on butanol; EMS + EB, strains obtained by random chemical mutagenesis using EMS with selection on EB. Bold values are the most important outcomes of the experiment.*

55025, an ospring of *C. acetobutylicum* ATCC 4259, was able to tolerate about 11.4 g/L of butanol while its WTS only 5 g/L (Jain et al., 1994). Interestingly, two of our nine mutant strains of *C. beijerinckii* NRRL B-598 with the highest increase in butanol tolerance were obtained by mutagenesis on agar plates containing EB (EB mutants) (**Figure 2**), a method that has not previously been used for this purpose in solventogenic *Clostridium*.

Despite increased butanol tolerance, no mutant strain exhibited an increase in butanol production compared with WTS (**Supplementary Table 1** and **Table 2**), which is a divergence from well-known mutant strains obtained in a similar way. For example, *C. beijerinckii* BA101 was able to produce around 20 g/L of butanol during batch cultivation (an increase of over 100%) (Annous and Blaschek, 1991; Chen and Blaschek, 1999) and *C. beijerinckii* ATCC 55025 displayed an increase in butanol concentration between 22 and 38% (Jain et al., 1994). However, it has previously been shown in multiple studies, both for random (Baer et al., 1987; Gallardo et al., 2017; M[illegible] de G[illegible]rando et al., 2018) and targeted (Zhao et al., 2003; Alsaker et al., 2004; Mann et al., 2012; Jones et al., 2016) mutagenesis, that increased butanol tolerance does not always result in improved butanol production.

EMS + butanol and EMS + EB mutant strains exhibited the acid crash phenotype under standard cultivation conditions and under pH regulation via $CaCO_3$ supplementation (**Supplementary Table 1** and **Table 2**), producing high concentrations of butyric acid. It was shown for the WTS

TABLE 3 | List of single-nucleotide polymorphisms that occurred in mutant strains of *C. beijerinckii* NRRL B-598 exhibiting high butanol tolerance[1].

	Position	Ref	Alt	Locus	Product	Start	End	Strand	Feature	Aa_ref	Aa_alt
Mutant strain A*											
1	863781	T	G	X276_22865	Carbohydrate ABC transporter permease	863698	864525	+		I	M
2	1661343	C	T	X276_19395	Peptidase S8 and S53 subtilisin kexin sedolisin	1661151	1662989	+		L	F
3	2775912	T	C	X276_14460	Cytochrome b5	2775189	2775996	+		T	T
4	3007463	C	A	X276_13415	S-layer family protein	3005247	3009269	+		V	V
5	3008654	T	G	X276_13415	S-layer family protein	3005247	3009269	+		T	T
6	5443138	C	T	X276_03000	AAA family ATPase	5441786	5444533	–		E	K
Mutant strain B*											
1	863781	T	G	X276_22865	Carbohydrate ABC transporter permease	863698	864525	+		I	M
2	1190011	G	T						Non-coding		
3	2052101	G	T	X276_17580	Peptidase S8	2051180	2052916	+		G	C
4	3394324	C	T	X276_11885	16S ribosomal RNA	3393103	3394616	–		T	T
5	4087693	C	A						Non-coding		
6	5173544	A	AT	X276_04045	Chemotaxis protein CheC	5173215	5173817	–			
7	5443138	C	T	X276_03000	AAA family ATPase	5441786	5444533	–		E	K
Mutant strain C*											
1	863781	T	G	X276_22865	Carbohydrate ABC transporter permease	863698	864525	+		I	M
2	968296	C	T	X276_22430	Peptidase S8	967488	969206	+		H	Y
3	2775238	G	T	X276_14460	Cytochrome b5	2775189	2775996	+		E	*
4	2919727	T	A	X276_13825	Hypothetical protein	2919158	2919820	+		S	S
5	3004110	C	T	X276_13420	Collagen-like protein	3003304	3004570	+	Pseudogene	G	G
6	3025567	C	A	X276_13335	Collagen-like protein	3025298	3027013	+		G	G
7	4087693	C	A						Non-coding		
8	5443138	C	T	X276_03000	AAA family ATPase	5441786	5444533	–		E	K
9	5649791	C	A						Non-coding		
Mutant strain B33**											
1	81253	G	A	X276_26470	Prolipoprotein diacylglyceryl transferase	81069	81839	+		G	E
2	1115679	G	C						Non-coding		
3	1457462	G	A						Non-coding		
4	1913236	G	A	X276_18195	ABC transporter substrate-binding protein	1912234	1913397	+		D	N
5	3567763	C	T	X276_11100	Methyl-accepting chemotaxis protein	3566625	3568346	–		S	N
6	4087693	C	A						Non-coding		
7	4257246	C	T	X276_07910	DNA polymerase III subunit epsilon	4256489	4257421	–		S	N
8	4560040	C	T	X276_06635	Pyruvate, phosphate dikinase	4560023	4562566	–		V	I
9	5150062	T	A	X276_27350	Hypothetical protein	5146861	5150820	–		T	T
10	5150524	A	T	X276_27350	Hypothetical protein	5146861	5150820	–		S	S
11	5150526	A	T	X276_27350	Hypothetical protein	5146861	5150820	–		S	T
12	5150530	C	T	X276_27350	Hypothetical protein	5146861	5150820	–		V	V
13	5150532	C	T	X276_27350	Hypothetical protein	5146861	5150820	–		V	M
14	5150533	A	T	X276_27350	Hypothetical protein	5146861	5150820	–		I	I
15	5150538	C	T	X276_27350	Hypothetical protein	5146861	5150820	–		V	I
16	5150545	T	C	X276_27350	Hypothetical protein	5146861	5150820	–		G	G
17	5150568	T	C	X276_27350	Hypothetical protein	5146861	5150820	–		I	V

(Continued)

TABLE 3 | Continued

	Position	Ref	Alt	Locus	Product	Start	End	Strand	Feature	Aa_ref	Aa_alt
Mutant strain B44**											
1	3006134	C	A	X276_13415	S-layer family protein	3005247	3009269	+		T	T
2	3007658	C	A	X276_13415	S-layer family protein	3005247	3009269	+		G	G
3	4087693	C	A						Non-coding		
Mutant strain E15***											
1	654069	C	T	X276_23670	Nitrogenase iron protein	653549	654403	+		A	V
2	685966	C	T	X276_23545	tRNA 2-thiocytidine biosynthesis protein TtcA	685570	686436	+		P	S
3	724396	GT	G						Non-coding		
4	1809500	CA	C	X276_18715	YggS family pyridoxal phosphate-dependent enzyme	1809001	1809681	+			
5	2376845	G	A	X276_16220	MFS transporter	2376293	2377702	–		Y	Y
6	2453563	G	A	X276_15915	Sigma-54-dependent Fis family transcriptional regulator	2451932	2453896	+		E	E
7	2468045	G	A	X276_15855	HlyC/CorC family transporter	2467109	2468398	+		A	T
8	2482557	C	T	X276_15775	Bifunctional 4-hydroxy-2-oxoglutarate aldolase/2-dehydro-3-deoxy-phosphogluconate aldolase	2482286	2482915	+		A	V
9	3008051	C	A	X276_13415	S-layer family protein	3005247	3009269	+		T	T
10	3141441	C	CA	X276_12855	DUF4179 domain-containing protein	3141433	3142683	+			
11	4087693	C	A						Non-coding		
12	4243433	C	CT	X276_07980	MarR family transcriptional regulator	4243239	4243676	–			
Mutant strain E28***											
1	491837	G	A	X276_24400	MerR family transcriptional regulator	491399	492223	+		E	K
2	851273	T	G	X276_22910	16S ribosomal RNA	851024	852535	+		*	G
3	851321	T	G	X276_22910	16S ribosomal RNA	851024	852535	+		*	G
4	2775807	C	A	X276_14460	Cytochrome b5	2775169	2775996	+		G	G
5	2775912	T	C	X276_14460	Cytochrome b5	2775169	2775996	+		T	T
6	3004250	C	A	X276_13420	Collagen-like protein	3003304	3004570	+	Pseudogene	P	Q
7	3007799	C	A	X276_13415	S-layer family protein	3005247	3009269	+		T	T
8	3008654	T	G	X276_13415	S-layer family protein	3005247	3009269	+		T	T
Mutant strain E32***											
1	2207690	C	CT	X276_16910	IS110 family transposase	2206548	2207848	–	Pseudogene		
2	2775912	T	C	X276_14460	Cytochrome b5	2775169	2775996	+		T	T
3	3007397	C	A	X276_13415	S-layer family protein	3005247	3009269	+		G	G
4	3876973	C	A	X276_09805	Glycoside hydrolase	3876787	3877821	–		M	I
5	4087693	C	A						Non-coding		
6	4243379	C	CAT	X276_07980	MarR family transcriptional regulator	4243239	4243676	–			
7	4960470	G	A	X276_04930	Chemotaxis protein	4958594	4963015	–		T	I
8	4982628	C	T	X276_04840	Response regulator	4982393	4982745	–	Pseudogene	I	I
9	5011570	C	T	X276_04700	Non-ribosomal peptide synthase	5010682	5018304	–		W	*
10	5053669	C	T	X276_04620	Hypothetical protein	5053325	5055736	–		D	N
11	5083112	C	T	X276_04450	YigZ family protein	5082972	5083619	–		V	I

(Continued)

TABLE 3 | Continued

	Position	Ref	Alt	Locus	Product	Start	End	Strand	Feature	Aa_ref	Aa_alt
12	5084395	C	T	X276_04445	PLP-dependent aminotransferase family protein	5083996	5085438	–		M	I
13	5275399	G	A	X276_03630	Carboxynorspermidine decarboxylase	5274563	5275702	–		H	Y
Mutant strain E33* **											
1	2663675	G	A	X276_14895	PFL family protein	2663567	2664922	+		G	R

[1]*Ref, DNA base in the wild-type strain; Alt, DNA base in mutant strain; Start, position of the gene in the genome, start; End, position of the gene in the genome, end; Aa_ref, encoded amino acid in the wild-type strain; Aa_alt, encoded amino acid in mutant strain. *Strains obtained by random chemical mutagenesis using EB as a mutagenic agent, when strains were selected directly on agar plates containing EB; **strains obtained by random chemical mutagenesis using EMS with selection on butanol; ***strains obtained by random chemical mutagenesis using EMS with selection on EB.*

TABLE 4 | List of mutations that caused longer changes in the genomes of mutant strains of *C. beijerinckii* NRRL B-598 exhibiting high butanol tolerance[1].

Strain	Mutation position	Size (bp)	Note
B33*	5311613-5311767	155	Intergenic
E15**	2982416-2982635	220	Part of the gene X276_13505 encoding DNA mismatch repair protein MutS
	5311613-5311767	155	Intergenic
E33**	4243635-4244393	759	Part of the gene X276_07980 encoding MarR family transcriptional regulator

[1]*Strains obtained by random chemical mutagenesis using EMS with the selection on butanol; **strains obtained by the random chemical mutagenesis using EMS with selection on ethidium bromide.*

that it is possible to cultivate the strain at larger scale under constant pH regulation to produce butyric acid as the main fermentation product (Drahokoupil and Patáčková, 2020). Therefore, these mutant strains can be further studied as alternative butyric acid producers.

Along with higher butanol tolerance, mutant strains of *C. beijerinckii* NRRL B-598 exhibited improved tolerance to ethanol and EB (**Figures 2**, **3**). Tolerance testing revealed that mutant strains obtained by dierent approaches behaved dierently. While tolerance to ethanol increased similarly in all mutant strains (**Figure 2**), EMS + EB mutants exhibited higher tolerance to EB than other strains (**Figure 3**). In the case of antibiotics, EB mutants generally exhibited higher tolerance (**Figure 4**). On the other hand, antibiotic tolerance of EMS + EB strains was lower than the WTS, except for strain E33 when tested for chloramphenicol (**Figure 4**). These dierences can be probably explained by dierent mutations that occurred in the strains, so to reveal the dierences at the gene level, genomes of all mutant strains were sequenced and variant callings were analyzed.

Variant calling in bacteria is a neglected topic and mainly, attention is paid to eukaryotes. Nevertheless, one of the rst benchmarking studies dealing with bacterial variant calling by Bush et al. (2020) showed that a combination of BWA for mapping and GATK HaplotypeCaller brought the best results for closely related strains. As this was the case in our study, we used these tools to capture SNPs (**Supplementary Files 4**). Moreover, we took advantage of our data containing resequencing of the WTS that was recently used to update its genome assembly (CP011966.3) (Sedlar et al., 2019). We used false positive detections of SNPs in the WTS sequencing data to infer our own ltering rules (see section Materials and Methods) (**Table 3**). While this approach is good for SNP detection, it is unable to call structural variants. Thus, we utilized Pilon to detect longer changes in genomic sequences (**Tables 4**, **5**). The combination of various approaches for detection of short and longer changes is quite common for eukaryotes (Long et al., 2013). Utilization of Pilon is advantageous in our case as the reference sequence was constructed using this approach, and thus, we were again able to remove false positives by ltering variants that were falsely called in the WTS data. SNPs in individual genes of the mutants are discussed further in the text, however, it is dicult to discuss CNVs because the topic has been scarcely studied in bacteria. CNVs are usually studied in comparative analyses (Vetrovský and Baldrian, 2013; Greenblum et al., 2015) and the approach cannot be applied in our case.

Only few genome sequences of butanol-tolerant mutant strains obtained by random mutagenesis are currently available [*C. beijerinckii* SA-1 (Sandoval-Espinola et al., 2013), *C. pasteurianum* M150B (Sandoval et al., 2015), *C. acetobutylicum* ATCC 55025 and *C. acetobutylicum* JB200 (Xu et al., 2017)]. Genomic sequences of *C. acetobutylicum* ATCC 55025 and *C. acetobutylicum* JB200, as well as *C. pasteurianum* M150B, included multiple mutations, for example, 143 SNPs and 67s SNP for *C. acetobutylicum* ATCC 55025 and *C. pasteurianum* M150B, respectively. On the other hand, similar to the case of *C. beijerinckii* SA-1, for which 10 genetic polymorphisms were conrmed, including eight SNPs, genomes of our mutant strains of *C. beijerinckii* NRRL B-598 included one to 17 SNPs (2151 prior to ltering), one longer change in strains B33, E15, and E33 and one or more copy number variations

TABLE 5 | List of copy number variations in the genomes of mutant strains of *C. beijerinckii* NRRL B-598 exhibiting high butanol tolerance[1].

Strain	Position	Size	Locus	Putative product
A*	341278349235	7.96 kb	X276_25195	DNA-3-methyladenine glycosylase 2 family protein
			X276_25190	Site-specic integrase
			X276_25185	XRE family transcriptional regulator
			X276_25180	Hypothetical protein
			X276_25175	Transcription factor
			X276_25170	Hypothetical protein
			X276_25165	Replication protein
			X276_25160	Hypothetical protein
			X276_25155	Hypothetical protein
			X276_25150	Hypothetical protein
			X276_25145	Hypothetical protein
B*	341243349259	8.02 kb	X276_25195	DNA-3-methyladenine glycosylase 2 family protein
			X276_25190	Site-specic integrase
			X276_25185	XRE family transcriptional regulator
			X276_25180	Hypothetical protein
			X276_25175	Transcription factor
			X276_25170	Hypothetical protein
			X276_25165	Replication protein
			X276_25160	Hypothetical protein
			X276_25155	Hypothetical protein
			X276_25150	Hypothetical protein
			X276_25145	Hypothetical protein
C*	341297349217	7.92 kb	X276_25195	DNA-3-methyladenine glycosylase 2 family protein
			X276_25190	Site-specic integrase
			X276_25185	XRE family transcriptional regulator
			X276_25180	Hypothetical protein
			X276_25175	Transcription factor
			X276_25170	Hypothetical protein
			X276_25165	Replication protein
			X276_25160	Hypothetical protein
			X276_25155	Hypothetical protein
			X276_25150	Hypothetical protein
			X276_25145	Hypothetical protein
B33**	167230167409	180 bp	X276_26025	PIN/TRAM domain-containing protein
	499439499573	135 bp	Only intergenic	
	575045575230	186 bp	Only intergenic	
	728816728957	142 bp	X276_23400	Hypothetical protein
	60427056042895	191 bp	X276_00695	Methyl-accepting chemotaxis protein
	61378716138337	467 bp	X276_00305	23S rRNA [pseudouridine(1915)-N(3)]-methyltransferase RlmH
B44**	341237349270	8.03 kb	X276_25195	DNA-3-methyladenine glycosylase 2 family protein
			X276_25190	Site-specic integrase
			X276_25185	XRE family transcriptional regulator
			X276_25180	Hypothetical protein
			X276_25175	Transcription factor
			X276_25170	Hypothetical protein
			X276_25165	Replication protein
			X276_25160	Hypothetical protein
			X276_25155	Hypothetical protein
			X276_25150	Hypothetical protein
			X276_25145	Hypothetical protein
E15***	49015166	266 bp	X276_26795	DNA topoisomerase (ATP-hydrolyzing) subunit B

(Continued)

TABLE 5 | Continued

Strain	Position	Size	Locus	Putative product
E28***	341278349256	7.98 kb	X276_25195	DNA-3-methyladenine glycosylase 2 family protein
			X276_25190	Site-specic integrase
			X276_25185	XRE family transcriptional regulator
			X276_25180	Hypothetical protein
			X276_25175	Transcription factor
			X276_25170	Hypothetical protein
			X276_25165	Replication protein
			X276_25160	Hypothetical protein
			X276_25155	Hypothetical protein
			X276_25150	Hypothetical protein
			X276_25145	Hypothetical protein
E33***	341393349134	7.74 kb	X276_25190	Site-specic integrase
			X276_25185	XRE family transcriptional regulator
			X276_25180	Hypothetical protein
			X276_25175	Transcription factor
			X276_25170	Hypothetical protein
			X276_25165	Replication protein
			X276_25160	Hypothetical protein
			X276_25155	Hypothetical protein
			X276_25150	Hypothetical protein
			X276_25145	Hypothetical protein

[1] *"Strains obtained by random chemical mutagenesis using EB as a mutagenic agent, when strains were selected directly on agar plates containing EB; **strains obtained by random chemical mutagenesis using EMS with selection on butanol; ***strains obtained by random chemical mutagenesis using EMS with selection on EB.*

in all of the strains, except E32 (**Tables 3 5**). Some of the mutations revealed in *C. beijerinckii* NRRL B-598 mutant strains with increased butanol tolerance can be connected to tolerance mechanisms, while others may play roles in metabolic enhancement rather than tolerance improvement. Many of the mutations were membrane-related, which correlates with the fact that normal cellular responses of the strains to the solvent was mainly at the membrane level (Patakova et al., 2018). A similar result was observed in *E. coli* when a genomic library enrichment strategy was used under butanol challenge (Reyes et al., 2011). Examples of membrane-related mutations in *C. beijerinckii* NRRL B-598 mutant strains are mutations in genes encoding peptidases, transporters or cytochrome b5 (**Table 3**). Interestingly, mutations in genes encoding peptidase S8 were observed not only in our EB mutants A (X276_19395), B (X276_17580), and C (X276_22430) (**Table 3**), but also in a mutant strain of *C. beijerinckii* SA-1, the ospring of *C. beijerinckii* NCIMB 8052 (Sandoval-Espinola et al., 2013). Additionally, it seems that EB mutants A, B and C not only exhibited dierent phenotypes from other mutants for which EMS was used as a mutagenic agent, but were also dierent in terms of mutations and formed a distinguishable cluster, as shown in **Figure 5**.

We analyzed changes in the genomes of mutant strains to determine whether some of the mutations occurring in EMS + butanol and EMS + EB strains could lead to the acid-crash phenotype. In the case of strain E15, the mutation occurred in gene X276_15915 encoding the sigma-54-dependent Fis family transcriptional regulator (**Table 3**). This may be causing the acid crash phenotype. Sigma-54 regulates sugar consumption and carbon metabolism in *C. beijerinckii* (Hocq et al., 2019). Thus, a higher glucose uptake rate and acid production rate, the reasons for the acid crash phenotype (Maddox et al., 2000), may occur due to mutations in regulatory genes such as sigma-54. A similar hypothesis has already been proposed for the degenerate strain *C. beijerinckii* DG-8052 (Lv et al., 2016). Unltered data contained a mutation that might be responsible for the phenotype; it is situated in gene X276_15350 encoding NADP-dependent glyceraldehyde-3-phosphate dehydrogenase in mutant strain E28 (**Supplementary Files 4**). The gene product plays a role in the formation of NADPH, which is necessary for the biosynthetic processes and modulation of redox potential (Liu et al., 2015). The importance of the gene for metabolite production was shown when expression of NADP-dependent glyceraldehyde-3-phosphate dehydrogenase of *C. acetobutylicum* in *E. coli* resulted in improved productivity of lycopene and ε-caprolactone (Martínez et al., 2008). As in the case of solventogenic strain *C. saccharoperbutylacetonicum*, where defects in genes encoding enzymes responsible for NADH formation resulted in strain degeneration (Hayashida and Yoshino, 1990), mutations in genes encoding NADPH formation in mutant strain E28 could aect the cells in a similar way. Nevertheless, this mutation (in gene X276_15350) was later ltered (**Table 3**) and, therefore, acid crash in the E28 strain probably happened due to some other changes. The reason for the acid crash phenotype in other mutant strains was not clear and needs further investigation.

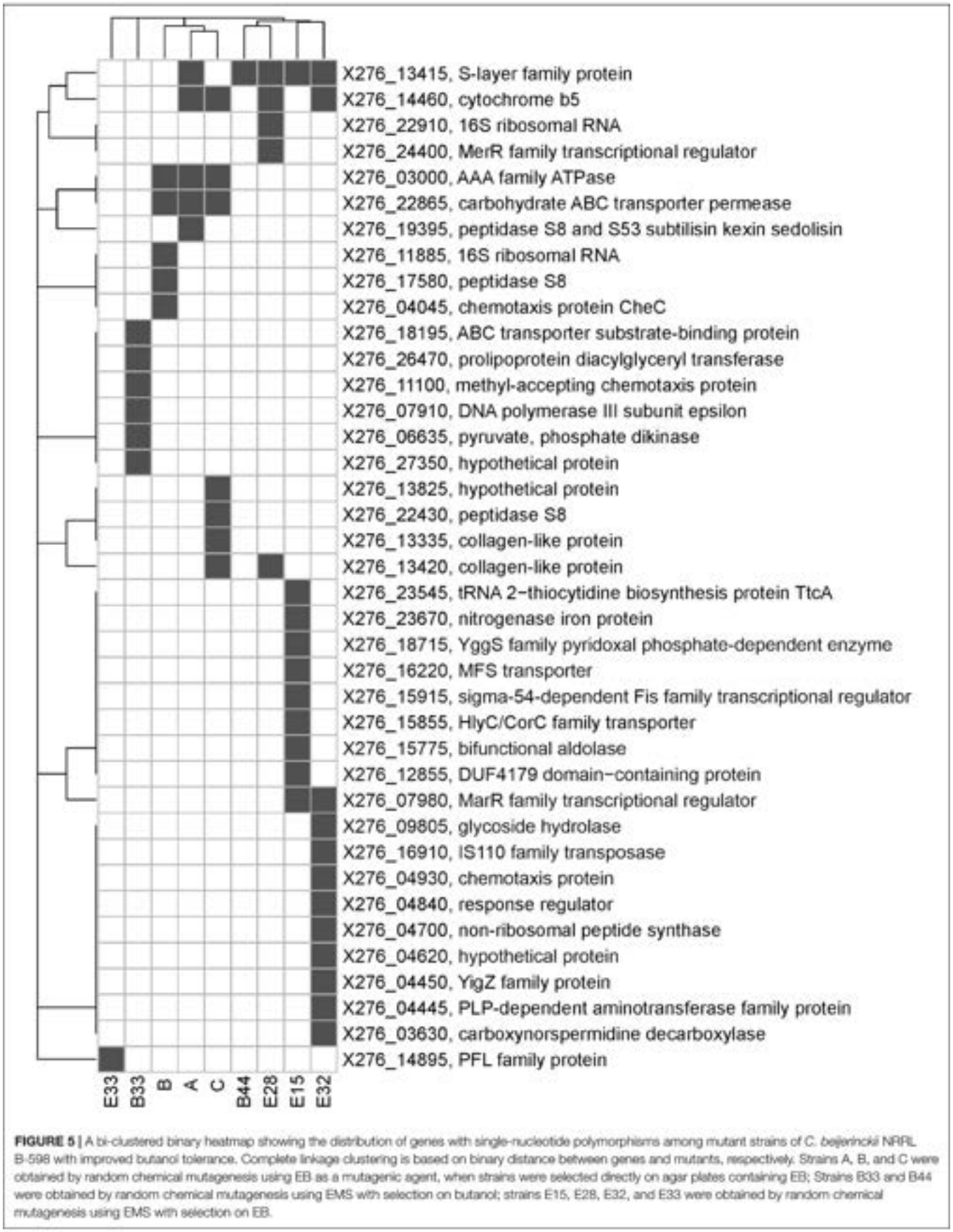


FIGURE 5 | A bi-clustered binary heatmap showing the distribution of genes with single-nucleotide polymorphisms among mutant strains of *C. beijerinckii* NRRL B-598 with improved butanol tolerance. Complete linkage clustering is based on binary distance between genes and mutants, respectively. Strains A, B, and C were obtained by random chemical mutagenesis using EB as a mutagenic agent, when strains were selected directly on agar plates containing EB; Strains B33 and B44 were obtained by random chemical mutagenesis using EMS with selection on butanol; strains E15, E28, E32, and E33 were obtained by random chemical mutagenesis using EMS with selection on EB.

The faster growth and somewhat higher butanol productivity of mutant C, observed during the rst 24 h of cultivation (**Supplementary Table 2**), can probably be attributed to mutations in genes encoding a carbohydrate ABC transporter permease (X276_22865) and an AAA family ATPase (X276_03000), in which mutations were also detected for the other EB mutants, A and B (**Figure 5**). Use of non-PTS mechanisms of glucose uptake during the solventogenic phase, such as ABC transporters, resulted in more complete glucose utilization and increased butanol production in *C. beijerinckii* BA101 (Lee et al., 2005). A BLASTp (Gish and States, 1993) analysis of AAA family ATPase X276_03000 showed that it contained motifs similar to the PTS operon transcription anti-terminator in *Clostridioides dicile* 630, therefore it probably plays a part in PTS regulation. It was reported that the genome of the butanol-overproducing mutant *C. beijerinckii* SA-1 included mutations in PTS genes (Sandoval-Espinola et al., 2013) and that the mutant strain *C. beijerinckii* BA101 had a partially defective PTS (Lee and Blaschek, 2001; Lee et al., 2005). Therefore, a connection between PTS and regulation of solvent production, contributing to butanol overproduction, was hypothesized (Sandoval-Espinola et al., 2013).

Mutations in genes connected to butanol tolerance can also be identied in mutant strains of *C. beijerinckii* NRRL B-598. For example, in genes encoding chemotaxis proteins in strains B (X276_04045), B33 (X276_11100 and X276_00695), and E32 (X276_04930) (**Tables 3, 5**). Chemotaxis is one the important mechanisms of adaptation to environmental stress (Zhao et al., 2007), in our case, the presence of produced butanol in the medium. It was shown that butanol acts as a repellent for *C. acetobutylicum*, meaning that it induces negative chemotaxis in the strain (Gutierrez and Maddox, 1987). For strain E32, mutations were observed in the gene X276_09805 encoding glycoside hydrolase, which belongs to family 25 (**Table 3**). This family includes enzymes with lysozyme activity and ones connected to autolysin production, for example, lyc gene (CA_C0554) of *C. acetobutylicum* ATCC 824. It was reported that enzymes that take part in autolysis and cell wall recycling in solventogenic clostridia are also connected to butanol tolerance (Patakova et al., 2018). In mutant strain B33, mutations were detected in X276_26470 encoding prolipoprotein diacylglyceryl transferase (**Table 3**), which catalyzes attachments of lipoproteins to cell membrane. Lipoproteins take part in multiple processes in the cell, including membrane transport, modications of the cell wall, and antibiotic tolerance (Nielsen and Lampen, 1982; Chimalapati et al., 2012); therefore, mutations in prolipoprotein diacylglyceryl transferase can also be connected to butanol tolerance mechanisms.

One of most interesting mutations connected to improved butanol tolerance was revealed in EMS + EB mutants E15, E32, and E33 in the gene X276_07980 encoding the MarR family transcriptional regulator (**Tables 3, 4**). Similarly to our result, mutations in genes encoding the MarR family transcriptional regulator occurred for the hyper butanol-tolerant and -producing strain, *C. acetobutylicum* JB200 and the authors hypothesized that this mutation contributed to enhanced butanol tolerance (Xu et al., 2017). Interestingly, MarR is a regulator that contributes to antibiotic resistance, as well as resistance to organic solvents and oxidative stress agents, by modulating the eux pump and porin expression (Sharma et al., 2017). Thus, other regulators may also contribute to enhanced butanol tolerance. For example, mutations in gene X276_24400 encoding the MerR family transcriptional regulator in mutant strain E28 (**Table 3**). A BLASTp (Gish and States, 1993) analysis showed that the MerR gene shares similarity with the BmrR regulator from *Bacillus subtilis*, which controls the Bmr eux pump for removal of antibiotics, dyes and disinfectants (Ahmed et al., 1994). Therefore, mutations in eux pump regulators were found in all EMS + EB mutants.

These results suggest that the increased or decreased tolerance of mutant strains to antibiotics, EB or solvents (butanol and ethanol) (**Figures 2 4**) may also be due to altered activities of eux pumps controlled by regulators. Transcriptional analysis of *C. beijerinckii* NRRL B-598 revealed upregulation of the TetR/AcrR family regulators putatively involved in eux pump gene transcription after butanol addition to the medium (Sedlar et al., 2019), suggesting that eux pumps might be involved in overcoming butanol stress in the strain. In the *C. beijerinckii* NRRL B-598 genome, 55 genes were identied as putatively encoding eux pumps (Jureckova et al., 2018). Some of these pumps may be able to remove antibiotics, EB or solvents from a cell. For example, genes encoding Gram-positive eux pumps capable of transporting tested antibiotics have been described for *Streptococcus pneumoniae*, *B. subtilis*, *Enterococcus faecalis*, *Lactococcus lactis*, and *Corynebacterium glutamicum* (Schindler and Kaatz, 2016). An eux pump responsible for increased ethanol tolerance was reported for *Saccharomyces cerevisiae* BY4741 (Yang et al., 2013). Furthermore, native eux pumps able to actively remove butanol were recently discovered in *Pseudomonas putida* and *E. coli* (Basler et al., 2018; Zhang et al., 2018), and heterologous expression of such a butanol eux pump from *P. putida* in butanol-producing *C. saccharoperbutylacetonicum* led to increased butanol tolerance of the strain (Jiménez-Bonilla et al., 2020).

CONCLUSION

Our study shows that random chemical mutagenesis using EB can be successfully used for the generation of butanol-tolerant mutant strains in solventogenic *Clostridium*. While mutagenesis with EMS as a mutagenic agent, which is more commonly used for such a purpose, resulted in mutants exhibiting the acid-crash phenotype, use of EB alone did not disrupt solvent production. Moreover, use of EB for both mutagenesis and selection resulted in increased tolerance to several dierent substrates of eux pumps.

We speculate that the acid crash phenotype in mutant strain E15 was acquired due to mutations in the gene encoding the sigma-54-dependent Fis family transcriptional regulator. Further investigations are needed to reveal the reasons for the phenotype in other EMS + butanol and EMS + EB strains.

Higher butanol tolerance of the strains may be connected to mutations in genes connected to the stress response, for example,

glycoside hydrolase or prolipoprotein diacylglyceryl transferase. However, the most prominent change in tolerance to substrates of eux pumps, including butanol, can be explained by mutations in genes encoding eux pump regulators, which were found in EMS + EB mutants. These regulators can be further studied in research connected to butanol tolerance mechanisms using targeted mutagenesis.

DATA AVAILABILITY STATEMENT

The genome sequencing data have been deposited in the NCBI Sequence Read Archive (SRA) under the project accession number PRJNA229510 (https://www.ncbi.nlm.nih.gov/sra/?term=PRJNA229510). WTS data are available under accession number SRX6419139 and mutant strains data under accession number within the range from SRX8614691 to SRX8614699.

AUTHOR CONTRIBUTIONS

MV, KS, IP, and PP designed the study. MV and BB performed the experiments and analyzed the data. KS and KJ performed the bioinformatics analyses. MV wrote the original draft, KS, BB, and PP contributed to manuscript writing. All authors read and approved the nal manuscript.

FUNDING

This work was supported by GACR (Grantova Agentura Ceske Republiky), Czechia, project number 17-00551S.

ACKNOWLEDGMENTS

Computational resources were partially provided by the CESNET LM2015042 and the CERIT Scientic Cloud LM2015085, under the programme Projects of Large Research, Development, and Innovations Infrastructures. We acknowledge the CF Genomics of CEITEC supported by the NCMG research infrastructure (LM2015091 funded by MEYS CR) for their support in obtaining the scientic data presented in this paper.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fbioe.2020.598392/full#supplementary-material

REFERENCES

Ahmed, M., Borsch, C. M., Taylor, S. S., Vázquez-Laslop, N., and Neyfakh, A. A. (1994). A protein that activates expression of a multidrug eux transporter upon binding the transporter substrates. *J. Biol. Chem.* 269, 2850628513.

Alsaker, K. V., Spitzer, T. R., and Papoutsakis, E. T. (2004). Transcriptional analysis of *spo0A* overexpression in *Clostridium acetobutylicum* and its eect on the cells response to butanol stress. *J. Bacteriol.* 186, 19591971. doi: 10.1128/JB.186.7.1959-1971.2004

Annous, B. A., and Blaschek, H. P. (1991). Isolation and characterization of *Clostridium acetobutylicum* mutants with enhanced amylolytic activity. *Appl. Environ. Microbiol.* 57, 25442548.

Baer, S. H., Blaschek, H. P., and Smith, T. L. (1987). Eect of butanol challenge and temperature on lipid composition and membrane uidity of butanol-tolerant *Clostridium acetobutylicum*. *Appl. Environ. Microbiol.* 53, 28542861.

Basler, G., Thompson, M., Tullman-Ercek, D., and Keasling, J. (2018). A *Pseudomonas putida* eux pump acts on short-chain alcohols. *Biotechnol. Biofuels* 11:136. doi: 10.1186/s13068-018-1133-9

Bohnert, J. A., Schuster, S., Hähnrich, E., Trittler, R., and Kern, W. V. (2007). Altered spectrum of multidrug resistance associated with a single point mutation in the *Escherichia coli* RND-type MDR eux pump YhiV (MdtF). *J. Antimicrob. Chemother.* 59, 12161222. doi: 10.1093/jac/dkl426

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a exible trimmer for Illumina sequence data. *Bioinformatics* 30, 21142120. doi: 10.1093/bioinformatics/btu170

Bowles, L. K., and Ellefson, W. L. (1985). Eects of butanol on *Clostridium acetobutylicum*. *Appl. Environ. Microbiol.* 50, 11651170.

Branska, B., Pechacova, Z., Kolek, J., Vasylkivska, M., and Patakova, P. (2018). Flow cytometry analysis of *Clostridium beijerinckii* NRRL B-598 populations exhibiting dierent phenotypes induced by changes in cultivation conditions. *Biotechnol. Biofuels* 11:99. doi: 10.1186/s13068-018-1096-x

Bui, L. M., Lee, J. Y., Geraldi, A., Rahman, Z., Lee, J. H., and Kim, S. C. (2015). Improved n-butanol tolerance in *Escherichia coli* by controlling membrane related functions. *J. Biotechnol.* 204, 3344. doi: 10.1016/j.jbiotec.2015.03.025

Bush, S. J., Foster, D., Eyre, D. W., Clark, E. L., De Maio, N., Shaw, L. P., et al. (2020). Genomic diversity aects the accuracy of bacterial single-nucleotide polymorphism-calling pipelines. *Gigascience* 9:giaa007. doi: 10.1093/gigascience/giaa007

Chen, C. K., and Blaschek, H. P. (1999). Acetate enhances solvent production and prevents degeneration in *Clostridium beijerinckii* BA101. *Appl. Microbiol. Biotechnol.* 52, 170173. doi: 10.1007/s002530051504

Chimalapati, S., Cohen, J. M., Camberlein, E., MacDonald, N., Durmort, C., Vernet, T., et al. (2012). Eects of deletion of the *Streptococcus pneumoniae* lipoprotein diacylglyceryl transferase gene Lgt on ABC transporter function and on growth in vivo. *PLoS One* 7:e41393. doi: 10.1371/journal.pone.0041393

Drahokoupil, M., and Patáková, P. (2020). Production of butyric acid at constant pH by a solventogenic strain of *Clostridium beijerinckii*. *Czech J. Food Sci.* 38, 185191. doi: 10.17221/95/2020-CJFS

Ewels, P., Magnusson, M., Lundin, S., and Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 32, 30473048. doi: 10.1093/bioinformatics/btw354

Fisher, M. A., Boyarskiy, S., Yamada, M. R., Kong, N., Bauer, S., and Tullman-Ercek, D. (2014). Enhancing tolerance to short-chain alcohols by engineering the *Escherichia coli* AcrB eux pump to secrete the non-native substrate n-butanol. *ACS Synth. Biol.* 3, 3040. doi: 10.1021/sb400065q

Fletcher, E., Pilizota, T., Davies, P. R., McVey, A., and French, C. E. (2016). Characterization of the eects of n-butanol on the cell envelope of *E. coli*. *Appl. Microbiol. Biotechnol.* 100, 96539659. doi: 10.1007/s00253-016-7771-6

Gagneur, J., Toedling, J., Bourgon, R., and Delhomme, N. (2020). *genomeIntervals: Operations on Genomic Intervals. R package version* 1.44.2. Available online at: https://bioconductor.org/packages/release/bioc/html/genomeIntervals.html (accessed June 13, 2020).

Gallardo, R., Alves, M., and Rodrigues, L. R. (2017). Inuence of nutritional and operational parameters on the production of butanol or 1,3-propanediol from glycerol by a mutant *Clostridium pasteurianum*. *N. Biotechnol.* 34, 5967. doi: 10.1016/j.nbt.2016.03.002

Gish, W., and States, D. J. (1993). Identication of protein coding regions by database similarity search. *Nat. Genet.* 3, 266272. doi: 10.1038/ng0393-266

Greenblum, S., Carr, R., and Borenstein, E. (2015). Extensive strain-level copy-number variation across human gut microbiome species. *Cell* 160, 583594. doi: 10.1016/j.cell.2014.12.038

Gutierrez, N. A., and Maddox, I. S. (1987). Role of chemotaxis in solvent production by *Clostridium acetobutylicum*. *Appl. Environ. Microbiol.* 53, 1924 1927.

Hayashida, S., and Yoshino, S. (1990). Degeneration of solventogenic *Clostridium* caused by a defect in NADH generation. *Agric. Biol. Chem.* 54, 427435. doi: 10.1271/bbb1961.54.427

Hermann, M., Fayolle, F., Marchal, R., Podvin, L., Sebald, M., and Vandecasteele, J. P. (1985). Isolation and characterization of butanol-resistant mutants of *Clostridium acetobutylicum*. *Appl. Environ. Microbiol.* 50, 12381243.

Hocq, R., Bouilloux-Lafont, M., Lopes Ferreira, N., and Wasels, F. (2019). σ 54 (σ L) plays a central role in carbon metabolism in the industrially relevant *Clostridium beijerinckii*. *Sci. Rep.* 9:7228. doi: 10.1038/s41598-019-43822-2

Ingram, L. O. (1986). Microbial tolerance to alcohols: role of the cell membrane. *Trends Biotechnol.* 4, 4044. doi: 10.1016/0167-7799(86)90152-6

Jain, M. K., Beacom, D., and Datta, R. (1994). Mutant strain of *C. acetobutylicum and process for making butanol*. *Patent* 12:242. doi: 10.1016/0734-9750(94) 90895-8

Jiménez-Bonilla, P., Zhang, J., Wang, Y., Blersch, D., De-Bashan, L.-E., Guo, L., et al. (2020). Enhancing the tolerance of *Clostridium saccharoperbutylacetonicum* to lignocellulosic-biomass-derived inhibitors for ecient biobutanol production by overexpressing eux pumps genes from *Pseudomonas putida*. *Bioresour. Technol* 312:123532. doi: 10.1016/j.biortech.2020.123532

Jones, A. J., Venkataramanan, K. P., and Papoutsakis, T. (2016). Overexpression of two stress-responsive, small, non-coding RNAs, 6S and tmRNA, imparts butanol tolerance in *Clostridium acetobutylicum*. *FEMS Microbiol. Lett.* 363:fnw063. doi: 10.1093/femsle/fnw063

Jureckova, K., Koscova, P., Sedlar, K., Kolek, J., Patakova, P., and Provazník, I. (2018). *In silico* prediction of genes coding eux pumps in *Clostridium beijerinckii* NRRL B-598, in *Proceedings of the 6th International Conference on Chemical Technology*, eds M. Vesely, Z. Hrdlicka, J. Hanika, and J. Lubojacky (Czech Republic: Czech Society of Industrial Chemistry), 8690.

Knaus, B. J., and Grünwald, N. J. (2017). vcfr: a package to manipulate and visualize variant call format data in R. *Mol. Ecol. Resour.* 17, 4453. doi: 10.1111/1755-0998.12549

Kong, X., He, A., Zhao, J., Wu, H., Ma, J., Wei, C., et al. (2016). Ecient acetone-butanol-ethanol (ABE) production by a butanol-tolerant mutant of *Clostridium beijerinckii* in a fermentation-pervaporation coupled process. *Biochem. Eng. J.* 105, 9096. doi: 10.1016/j.bej.2015.09.013

Lee, J., and Blaschek, H. P. (2001). Glucose uptake in *Clostridium beijerinckii* NCIMB 8052 and the solvent-hyperproducing mutant BA101. *Appl. Environ. Microbiol.* 67, 50255031.

Lee, J., Mitchell, W. J., Tangney, M., and Blaschek, H. P. (2005). Evidence for the presence of an alternative glucose transport system in *Clostridium beijerinckii* NCIMB 8052 and the solvent-hyperproducing mutant BA101. *Appl. Environ. Microbiol.* 71, 33843387. doi: 10.1128/AEM.71.6.3384-3387.2005

Lee, S. Y., Park, J. H., Jang, S. H., Nielsen, L. K., Kim, J., and Jung, K. S. (2008). Fermentative butanol production by clostridia. *Biotechnol. Bioeng.* 101, 209228. doi: 10.1002/bit.22003

Lepage, C., Fayolle, F., Hermann, M., and Vandecasteele, J. P. (1987). Changes in membrane lipid composition of *Clostridium acetobutylicum* during acetone-butanol fermentation: eects of solvents, growth temperature and pH. *Microbiology* 133, 103110. doi: 10.1099/00221287-133-1-103

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 17541760. doi: 10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078 2079. doi: 10.1093/bioinformatics/btp352

Liao, Z., Zhang, Y., Luo, S., Suo, Y., Zhang, S., and Wang, J. (2017). Improving cellular robustness and butanol titers of *Clostridium acetobutylicum* ATCC824 by introducing heat shock proteins from an extremophilic bacterium. *J. Biotechnol.* 252, 110. doi: 10.1016/j.jbiotec.2017.04.031

Lin, Y. L., and Blaschek, H. P. (1983). Butanol production by a butanol-tolerant strain of *Clostridium acetobutylicum* in extruded corn broth. *Appl. Environ. Microbiol.* 45, 966973. doi: 10.1128/aem.45.3.966-973.1983

Liu, J., Qi, H., Wang, C., and Wen, J. (2015). Model-driven intracellular redox status modulation for increasing isobutanol production in *Escherichia coli*. *Biotechnol. Biofuels* 8:108.

Liu, X. B., Gu, Q. Y., and Yu, X. B. (2013). Repetitive domestication to enhance butanol tolerance and production in *Clostridium acetobutylicum* through articial simulation of bio-evolution. *Bioresour. Technol.* 130, 638643. doi: 10.1016/j.biortech.2012.12.121

Long, Q., Rabanal, F. A., Meng, D., Huber, C. D., Farlow, A., Platzer, A., et al. (2013). Massive genomic variation and strong selection in *Arabidopsis thaliana* lines from Sweden. *Nat. Genet.* 45, 884890. doi: 10.1038/ng.2678

Lv, J., Jiao, S., Du, R., Zhang, R., Zhang, Y., and Han, B. (2016). Proteomic analysis to elucidate degeneration of *Clostridium beijerinckii* NCIMB 8052 and role of Ca2+ in strain recovery from degeneration. *J. Ind. Microbiol. Biotechnol.* 43, 741750. doi: 10.1007/s10295-016-1754-6

Maddox, I. S., Steiner, E., Hirsch, S., Wessner, S., Gutierrez, N. A., Gapes, J. R., et al. (2000). The cause of acid crash and acidogenic fermentations during the batch acetone-butanol-ethanol (ABE-) fermentation process. *J. Mol. Microbiol. Biotechnol.* 2, 95100.

Mann, M. S., Dragovic, Z., Schirrmacher, G., and Lutke-Eversloh, T. (2012). Over-expression of stress protein-encoding genes helps *Clostridium acetobutylicum* to rapidly adapt to butanol stress. *Biotechnol. Lett.* 34, 16431649. doi: 10.1007/s10529-012-0951-2

Mao, S., Luo, Y., Zhang, T., Li, J., Bao, G., Zhu, Y., et al. (2010). Proteome reference map and comparative proteomic analysis between a wild type *Clostridium acetobutylicum* DSM 1731 and its mutant with enhanced butanol tolerance and butanol yield. *J. Proteome Res.* 9, 30463061. doi: 10.1021/pr9012078

Martínez, I., Zhu, J., Lin, H., Bennett, G. N., and San, K. Y. (2008). Replacing *Escherichia coli* NAD-dependent glyceraldehyde 3-phosphate dehydrogenase (GAPDH) with a NADP-dependent enzyme from *Clostridium acetobutylicum* facilitates NADPH dependent pathways. *Metab. Eng.* 10, 352359. doi: 10.1016/j.ymben.2008.09.001

Méhé de Gérando, H., Wasels, F., Bisson, A., Clement, B., Bidard, F., Jourdier, E., et al. (2018). Genome and transcriptome of the natural isopropanol producer *Clostridium beijerinckii* DSM6423. *BMC Genom.* 19:242. doi: 10.1186/s12864-018-4636-7

Matta-el-Ammouri, G., Janati-Idrissi, R., Rambourg, J.-M., Petitdemange, H., and Gay, R. (1986). Acetone butanol fermentation by a *Clostridium acetobutylicum* mutant with high solvent productivity. *Biomass* 10, 109119. doi: 10.1016/0144-4565(86)90059-4

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010). The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 12971303. doi: 10.1101/gr.107524.110

Mukhopadhyay, A. (2015). Tolerance engineering in bacteria for the production of advanced biofuels and chemicals. *Trends Microbiol.* 23, 498508. doi: 10.1016/j.tim.2015.04.008

Nielsen, J. B. K., and Lampen, J. O. (1982). Membrane-bound penicillinases in gram-positive bacteria. *J. Biol. Chem.* 257, 44904495.

Ohta, T., Tokishita, S. I., and Yamagata, H. (2001). Ethidium bromide and SYBR Green I enhance the genotoxicity of UV-irradiation and chemical mutagens in *E. coli*. *Mutat. Res. Genet. Toxicol. Environ. Mutagen.* 492, 9197. doi: 10.1016/S1383-5718(01)00155-3

Pagès, H., Aboyoun, P., Gentleman, R., and DebRoy, S. (2020). *Biostrings: Ecient Manipulation of Biological Strings. R package version* 2.56.0. Available online at: https://bioconductor.org/packages/release/bioc/html/Biostrings.html (accessed June 13, 2020).

Paixao, L., Rodrigues, L., Couto, I., Martins, M., Fernandes, P., de Carvalho, C. C. C. R., et al. (2009). Fluorometric determination of ethidium bromide eux kinetics in *Escherichia coli*. *J. Biol. Eng.* 3:18. doi: 10.1186/1754-1611-3-18

Patakova, P., Kolek, J., Sedlar, K., Koscova, P., Branska, B., Kupkova, K., et al. (2018). Comparative analysis of high butanol tolerance and production in clostridia. *Biotechnol. Adv.* 36, 721738. doi: 10.1016/j.biotechadv.2017.12.004

Patel, D., Kosmidis, C., Seo, S. M., and Kaatz, G. W. (2010). Ethidium bromide MIC screening for enhanced eux pump gene expression or eux activity

in *Staphylococcus aureus*. *Antimicrob. Agents Chemother.* 54, 50705073. doi: 10.1128/AAC.01058-10

Qureshi, N., and Blaschek, H. P. (2001). Recent advances in ABE fermentation: Hyper-butanol producing *Clostridium beijerinckii* BA101. *J. Ind. Microbiol. Biotechnol.* 27, 287291. doi: 10.1038/sj.jim.7000114

Reyes, L. H., Almario, M. P., and Kao, K. C. (2011). Genomic library screens for genes involved in n-butanol tolerance in *Escherichia coli*. *PLoS One* 6:e17678. doi: 10.1371/journal.pone.0017678

Sandoval, N. R., Venkataramanan, K. P., Groth, T. S., and Papoutsakis, E. T. (2015). Whole-genome sequence of an evolved *Clostridium pasteurianum* strain reveals Spo0A deciency responsible for increased butanol production and superior growth. *Biotechnol. Biofuels* 8:227. doi: 10.1186/s13068-015-0408-7

Sandoval-Espinola, W. J., Makwana, S. T., Chinn, M. S., Thon, M. R., Andrea Azcárrate-Peril, M., and Bruno-Bárcena, J. M. (2013). Comparative phenotypic analysis and genome sequence of *Clostridium beijerinckii* SA-1, an ospring of NCIMB 8052. *Microbiology* 159, 25582570. doi: 10.1099/mic.0.069534-0

Schindler, B. D., and Kaatz, G. W. (2016). Multidrug eux pumps of Gram-positive bacteria. *Drug Resist. Updat.* 27, 113. doi: 10.1016/j.drup.2016.04.003

Schwarz, K. M., Kuit, W., Grimmler, C., Ehrenreich, A., and Kengen, S. W. M. (2012). A transcriptional study of acidogenic chemostat cells of *Clostridium acetobutylicum* - Cellular behavior in adaptation to n-butanol. *J. Biotechnol.* 161, 366377. doi: 10.1016/j.jbiotec.2012.03.018

Sedlar, K., Kolek, J., Gruber, M., Jureckova, K., Branska, B., Csaba, G., et al. (2019). A transcriptional response of *Clostridium beijerinckii* NRRL B-598 to a butanol shock. *Biotechnol. Biofuels* 12:243. doi: 10.1186/s13068-019-1584-7

Sedlar, K., Kolek, J., Provazník, I., and Patáková, P. (2017). Reclassication of non-type strain *Clostridium pasteurianum* NRRL B-598 as *Clostridium beijerinckii* NRRL B-598. *J. Biotechnol.* 244, 13. doi: 10.1016/j.jbiotec.2017.01.003

Sedlar, K., Koscova, P., Vasylkivska, M., Branska, B., Kolek, J., Kupkova, K., et al. (2018). Transcription proling of butanol producer *Clostridium beijerinckii* NRRL B-598 using RNA-Seq. *BMC Genom.* 19:415. doi: 10.1186/s12864-018-4805-8

Segura, A., Molina, L., Fillet, S., Krell, T., Bernal, P., Muñoz-Rojas, J., et al. (2012). Solvent tolerance in Gram-negative bacteria. *Curr. Opin. Biotechnol.* 23, 415421. doi: 10.1016/j.copbio.2011.11.015

Seo, S.-O., Janssen, H., Magis, A., Wang, Y., Lu, T., Price, N. D., et al. (2017). Genomic, transcriptional, and phenotypic analysis of the glucose derepressed *Clostridium beijerinckii* mutant exhibiting acid crash phenotype. *Biotechnol. J.* 12, 1700182. doi: 10.1002/biot.201700182

Sharma, P., Haycocks, J. R. J., Middlemiss, A. D., Kettles, R. A., Sellars, L. E., Ricci, V., et al. (2017). The multiple antibiotic resistance operon of enteric bacteria controls DNA repair and outer membrane integrity. *Nat. Commun.* 8:1444. doi: 10.1038/s41467-017-01405-7

Soucaille, P., Joli, G., Izard, A., and Goma, G. (1987). Butanol tolerance and autobacteriocin production by *Clostridium acetobutylicum*. *Curr. Microbiol.* 14, 295299. doi: 10.1007/BF01568139

Tanaka, Y., Kasahara, K., Hirose, Y., Morimoto, Y., Izawa, M., and Ochi, K. (2017). Enhancement of butanol production by sequential introduction of mutations conferring butanol tolerance and streptomycin resistance. *J. Biosci. Bioeng.* 124, 400407. doi: 10.1016/j.jbiosc.2017.05.003

Tomas, C. A., Welker, N. E., and Papoutsakis, E. T. (2003). Overexpression of *groESL* in *Clostridium acetobutylicum* results in increased solvent production and tolerance, prolonged metabolism, and changes in the cells transcriptional program. *Appl. Environ. Microbiol.* 69, 49514965. doi: 10.1128/AEM.69.8.4951-4965.2003

Vasylkivska, M., Jureckova, K., Branska, B., Sedlar, K., Kolek, J., Provazník, I., et al. (2019). Transcriptional analysis of amino acid, metal ion, vitamin and carbohydrate uptake in butanol-producing *Clostridium beijerinckii* NRRL B-598. *PLoS One* 14:e0224560. doi: 10.1371/journal.pone.0224560

Vasylkivska, M., and Patakova, P. (2020). Role of eux in enhancing butanol tolerance of bacteria. *J. Biotechnol.* 320, 1727. doi: 10.1016/j.jbiotec.2020.06.008

Vetrovský, T., and Baldrian, P. (2013). The variability of the 16S rRNA gene in bacterial genomes and its consequences for bacterial community analyses. *PLoS One* 8:e57923. doi: 10.1371/journal.pone.0057923

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., et al. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963. doi: 10.1371/journal.pone.0112963

Wickham, H. (2009). *ggplot2*. New York, NY: Springer. doi: 10.1007/978-0-387-98141-3

Xu, M., Zhao, J., Yu, L., and Yang, S. T. (2017). Comparative genomic analysis of *Clostridium acetobutylicum* for understanding the mutations contributing to enhanced butanol tolerance and production. *J. Biotechnol.* 263, 3644. doi: 10.1016/j.jbiotec.2017.10.010

Xue, C., Zhao, J., Lu, C., Yang, S. T., Bai, F., and Tang, I. C. (2012). High-titer n-butanol production by *Clostridium acetobutylicum* JB200 in fed-batch fermentation with intermittent gas stripping. *Biotechnol. Bioeng.* 109, 2746 2756. doi: 10.1002/bit.24563

Yang, K. M., Woo, J. M., Lee, S. M., and Park, J. B. (2013). Improving ethanol tolerance of *Saccharomyces cerevisiae* by overexpressing an ATP-binding cassette eux pump. *Chem. Eng. Sci.* 103, 7478. doi: 10.1016/j.ces.2012.09.015

Yang, S.-T., and Zhao, J. (2013). *Adaptive engineering of Clostridium for increased butanol production*. U.S. Patent No US8450093B1. Washington, DC: U.S. Patent and Trademark Oce.

Zhang, Y., Dong, R., Zhang, M., and Gao, H. (2018). Native eux pumps of *Escherichia coli* responsible for short and medium chain alcohol. *Biochem. Eng. J.* 133, 149156. doi: 10.1016/j.bej.2018.02.009

Zhao, K., Liu, M., and Burgess, R. R. (2007). Adaptation in bacterial agellar and motility systems: from regulon members to foraging-like behavior in *E. coli*. *Nucleic Acids Res.* 35, 44414452. doi: 10.1093/nar/gkm456

Zhao, Y., Hindor, L. A., Chuang, A., Monroe-Augustus, M., Lyristis, M., Harrison, M. L., et al. (2003). Expression of a cloned cyclopropane fatty acid synthase gene reduces solvent formation in *Clostridium acetobutylicum* ATCC 824. *Appl. Environ. Microbiol.* 69, 28312841. doi: 10.1128/AEM.69.5.2831-2841.2003

Conict of Interest: The authors declare that the research was conducted in the absence of any commercial or nancial relationships that could be construed as a potential conict of interest.

5.9 Article IX

SEDLAR, Karel, VASYLKIVSKA, Maryna, MUSILOVA, Jana, BRANSKA, Barbora, PROVAZNIK, Ivo and PATAKOVA, Petra. Phenotypic and genomic analysis of isopropanol and 1,3-propanediol producer Clostridium diolis DSM 15410. *Genomics*. 2021. 113(1), p. 1109–1119. (published online in 2020, 2020 IF = 5.736, Q1 in BIOTECHNOLOGY & APPLIED MICROBIOLOGY)

DOI: https://doi.org/10.1016/j.ygeno.2020.11.007

Article IX: Phenotypic and genomic analysis of isopropanol and 1,3-propanediol producer *Clostridium diolis*

Genomics 113 (2021) 1109–1119

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Original Article

Phenotypic and genomic analysis of isopropanol and 1,3-propanediol producer *Clostridium diolis* DSM 15410

Karel Sedlar [a,*], Maryna Vasylkivska [b], Jana Musilova [a], Barbora Branska [b], Ivo Provaznik [a], Petra Patakova [b]

[a] Department of Biomedical Engineering, Faculty of Electrical Engineering and Communication, Brno University of Technology, Technicka 12, Brno, Czech Republic
[b] Department of Biotechnology, University of Chemistry and Technology Prague, Technicka 5, 166 28 Prague, Czech Republic

ARTICLE INFO

Keywords:
Clostridium diolis
Complete genome
Solventogenic
DSM
1,3-propanediol

ABSTRACT

Clostridium diolis DSM 15410 is a type strain of solventogenic clostridium capable of conducting isopropanol-butanol-ethanol fermentation. By studying its growth on different carbohydrates, we verified its ability to utilize glycerol and produce 1,3-propanediol and discovered its ability to produced isopropanol. Complete genome sequencing showed that its genome is a single circular chromosome and belongs to the cluster I (sensu scricto) of the genus *Clostridium*. By cultivation analysis we highlighted its specific behavior in comparison to two selected closely related strains. Despite the fact that several CRISPR loci were found, 16 putative prophages showed the ability to receive foreign DNA. Thus, the strain has the necessary features for future engineering of its 1,3-propanediol biosynthetic pathway and for the possible industrial utilization in the production of biofuels.

1. Introduction

The production of bio-based chemicals, such as solvents or acids, and bio-hydrogen from renewable waste feedstock, for example, food and horticultural waste or byproducts of meat-processing industry [1–4], is the cornerstone of sustainable circular economies [5]. In this scenario, microbial production plays an important role, however, at an industrial level, the wider utilization of microbial cell factories depends on the ability to select suitable strains and determine their potential for metabolic engineering [6]. While genome editing tools, including promising clustered regularly interspaced short palindromic repeats (CRISPR) technologies [7], exist, their utilization on non-model organisms can be challenging and requires detailed knowledge of previously engineered strains [8]. Although the basic phenotype features of many bacterial strains are known, their potential industrial capacity remains hidden in unknown genotypes as complete genome sequences are available for only selected species. Moreover, phenotypic heterogeneity may not reflect genetic differences or similarities [9].

An important group of industrially usable bacteria can be found in solventogenic clostridia, typically rod-shaped and spore-forming anaerobes, whose phylogenetic affinities are still unsettled as they are continuously being reclassified or reidentified [10–12]. This applies also to *C. diolis* species as it was reclassified as heterotypic synonym of *C. beijerinckii* earlier this year [13]. The most widely studied species within this group is *Clostridium acetobutylicum*, whose genome was sequenced in 2001 for the first time [14]. By that time, various genomes of solventogenic clostridia were analyzed and assembled into high quality complete genome sequences, including other promising solvent producers such as *C. beijerinckii* [15,16], *C. pasteurianum* [17], *C. butylicum* [18], and many others. Unfortunately, the diversity of solventogenic clostridia makes the knowledge gathered for a particular species or strain hardly applicable for another, even closely related, species or strain [19]. Therefore, analyses of particular strains on a genome-wide scale are needed as one of the key parameters for the engineering of clostridia as microbial cell factories lies in well-annotated genomes [20]. Despite advances in sequencing and genome assembly, a complete genome sequence of the type strain *Clostridium diolis* DSM 15410 remained unassembled until now.

C. diolis DSM 15410 (formerly known as *C. butyricum* DSM 5431) is a type strain with ability to produce, except common metabolites for solventogenic bacteria from genus *Clostridium*, also 1,3-propanediol [21]. It utilizes a wide range of substrates, including glycerol [22]. Although the genome of the strain was sequenced in the past, no high-quality genome assembly was achieved. On the contrary, its first draft genome by Wang et al. [23] was marked as contaminated in the GenBank database. In this paper, we present the first complete genome

* Corresponding author.
E-mail address: sedlar@feec.vutbr.cz (K. Sedlar).

https://doi.org/10.1016/j.ygeno.2020.11.007
Received 7 August 2020; Received in revised form 29 October 2020; Accepted 4 November 2020
Available online 7 November 2020

sequence of *Clostridium diolis* DSM 15410. We annotated the genome, predicted the operon structure and searched for prophage DNA, restriction-modification (R-M) systems, and CRISPR arrays. Moreover, we performed a phylogenomic analysis and compared selected genes of the central metabolism to genes in other strains of *C. beijerinckii*/*diolis* species and performed series of comparative cultivation experiments using strains *C. beijerinckii* NCIMB 8052 and *C. beijerinckii* NRRL B-598.

2. Materials and methods

2.1. Bacterial strain and cultivation condition

2.1.1. Bacterial strains

Culture of the strains *C. diolis* DSM 15410, *C. beijerinckii* NCIMB 8052, and *C. beijerinckii* NRRL B-598 were obtained from the German Collection of Microorganisms and Cell Cultures (DSMZ), National Collection of Industrial, Food and Marine Bacteria (NCIMB), and Agricultural Research Service Culture Collection (ARS/NRRL), respectively. The cultures were maintained as spore preserves in sterile distilled water at 4 °C.

2.1.2. Ability to grow on different substrates

The ability of *C. diolis* DSM 15410 to grow on different substrates was tested in modified Reinforced clostridial medium (RCM) medium containing 10 g/L tryptone (Sigma-Aldrich), 3 g/L yeast extract (Merck), 10 g/L meat extract (Merck), 3 g/L sodium acetate, 5 g/L sodium chloride and one of each tested carbohydrate (arabinose, cellobiose, fructose, glucose, glycerol, lactose, maltose, mannose, and xylose) in concentration 20 g/L. RCM medium without carbohydrate supplementation was used as a control and the test was done in triplicates for each carbohydrate. Inoculum was prepared as an overnight culture from pre-heated (2 min, 80 °C) spore preserves and cultivated in modified RCM medium containing 20 g/L glucose in an anaerobic chamber at 37 °C. Each test tube containing 10 ml of medium with carbohydrate was inoculated with 0.5 ml of overnight culture and cultivated for 24 h in an anaerobic chamber at 37 °C. After cultivation, an optical density (OD) at 600 nm (Varian Cary 50 UV-VIS spectrophotometer, Varian) was measured in each sample against the respective medium without inoculation as blank.

2.1.3. Bioreactor cultivation

Multifors 1 L bioreactors (Infors HT) and modified RCM medium containing 50 g/L of glucose were used for the batch cultivation of *C. diolis* DSM 15410 (triplicate). Prior to the inoculation, the pH of the medium was adjusted to 6.3 and oxygen from the bioreactor was removed by N_2 bubbling. The inoculum was prepared the same way as described above. The amount of added inoculum in working volume (700 ml) was 10%. After inoculation, the pH was measured, but not controlled, and samples were taken for the consequent OD measurement and HPLC analysis of produced metabolites and consumed glucose. Samples were analyzed on HPLC (Agilent Series 1200 HPLC; Agilent) with refractive index detection and an IEX H+ polymer column (Watrex). The parameters of the HPLC analysis were as follows: an injection sample volume of 20 µl, 5 mM H_2SO_4 as a mobile phase, a flow rate of 0.5 ml/min, and a column temperature of 60 °C.

2.1.4. 1,3-Propanediol production

For an analysis of 1,3-propanediol production, *C. diolis* DSM 15410was cultivated in a modified RCM medium and a medium optimized for the production of propanediol in *C. diolis* [24], both of which contained 50 g/L of glycerol. The production of 1,3-propanediol and other metabolites was analyzed using HPLC as described above.

A phase contrast microscopy (Olympus BX51; Olympus) was used to determine the morphological state of the cells. The HPLC data, OD at 600 nm, and pH course were processed and plotted using Matlab 2019a.

2.1.5. Comparison of cell morphology, carbohydrate consumption and metabolites production

Cell morphology, carbohydrate consumption and metabolites production of *C. diolis* DSM 15410, *C. beijerinckii* NCIMB 8052 and *C. beijerinckii* NRRL B-598 were analyzed in four media:

1) Tryptone-yeast extract-acetate medium (TYA) with glucose as a substrate containing 40 g/L glucose, 2 g/L yeast extract (Merck), 6 g/L tryptone (Sigma-Aldrich,), 0.5 g/L KH_2PO_4, 3 g/L ammonium acetate, 0.3 g/L $MgSO_4.7H_2O$ and 0.01 g/L $FeSO_4$);
2) RCM medium (as described above) containing 40 g/L glucose;
3) RCM medium (as described above) containing 50 g/L glycerol;
4) Medium optimized for the production of propanediol in *C. diolis* [24] containing 50 g/L glycerol.

Inocula were prepared in TYA medium from heat-shocked spore preserves cultivated overnight in anaerobic chamber at 37 °C. For the analysis, strains were cultivated for 72 h. Samples for HPLC analysis and microscopy control (Olympus BX51 microscope) were taken at the end of cultivation.

2.2. DNA extraction and sequencing

DNA was extracted from an exponentially growing culture of *C. diolis* DSM 15410 using a High Pure PCR Template Preparation Kit (Roche). The extraction was followed by an RNAase treatment by RNase A (Sigma-Aldrich) and subsequently purified by applying the High Pure PCR Product Purification Kit (Roche). The whole genome sequencing was performed using a combination of Illumina Next-Seq and PacBio RSII platforms. For the Illumina sequencing, the library construction and sequencing of the sample was performed by CEITEC Genomics core facility (Brno, Czechia) on Illumina NextSeq, pair-end (PE), 150 bp. For the PacBio sequencing, library construction and sequencing of the sample was performed by SeqMe (Dobris, Czechia) on PacBio RS II, 2× SMRT cell. The presence of plasmid(s) was tested using a GeneJET Plasmid Miniprep Kit (Thermo Scientific) and was performed, by different researchers, in two triplicates – no plasmid was obtained.

2.3. Genome assembly

The initial quality assessment of raw reads was done using a combination of FastQC v0.11.5 and MultiQC v1.7 [25]. The adapter and quality trimming was performed using Trimmomatic v1.36 [26]. Trimmed Illumina PE reads were combined with PacBio continuous long reads (CLR) into an initial draft genome assembly with SPAdes v3.11.1 [27]. High-quality circular consensus sequencing (CCS) reads were generated using SMRTlink v7.0.1.66975 and used for closing the gaps in the initial draft assembly with GMcloser v1.6.2 [28]. The resulting one-contig assembly was polished using all reads in a two-step procedure. The first step of polishing was performed with the SMRTlink Resequencing protocol by mapping all subreads and generating a polished contig. The second step was done by mapping Illumina PE reads to the polished contig with BWA v0.7.17 [29] and using Pilon v1.23 [30] for its second round of polishing. Files with mapped reads were handled with SAMtools v1.7 [31] and, during particular steps, mapping quality and quality of assemblies were controlled using Qualimap v2.2.1 [32]. The polished assembly was examined for circularity based on the presence of overlapping sequences at both ends of the contig. The overlap was found using MUMmer v3.23 [33] and the duplicated sequence was manually trimmed from one end of the contig. Finally, the replication origin (oriC) was predicted using Ori-finder [34] and the whole sequence was rearranged according to its position, so the *DnaA* gene is the first gene in the complete genome assembly.

2.4. Genome annotation and analysis

A genome annotation was added by the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) [35]. An operon prediction was completed using Operon-mapper [36] and the results were manually added to the genome record. The functional annotation of the protein coding genes was performed by assigning clusters of orthologous group (COG) categories from the eggNOG database with eggNOG-mapper [37]. The circular genome map was prepared with DNAplotter [38] integrated in Artemis [39]. Prophage DNA was searched with PhiSpy 3.7.8 [40]. Methylation motifs were inferred using base modification and motif analysis protocol in SMRTlink v7.0.1.66975. Further processing and analysis of the methylated motifs was completed by the internal tools of the REBASE database [41]. The annotated genome sequence was further analyzed for presence of CRISPR loci using CRISPRDetect tool [42]. A phylogenomic tree was produced with PhyloPhlAn 3.0 [43] using genes conserved across the bacterial domain.

3. Results and discussion

3.1. The general characterization of the strain

C. diolis DSM 15410 is a mesophilic heterofermentive rod-shaped spore-forming bacterium, see Fig. 1, that was able to grow in a medium containing a wide range of carbohydrates; the lowest increase in OD was observed in an arabinose medium and the highest increase was in a cellobiose and maltose medium, see Table 1. Additionally, mixed sugar utilization experiment, performed by Xin et al. (2016), revealed that the strain is able to simultaneously co-utilize glycerol with glucose, xylose or arabinose [22].Thus, the strain demonstrates metabolic flexibility and future perspective of its commercial fermentation on waste substrates.

During batch cultivation in a modified RCM medium with glucose, the strain exhibited a two-phase fermentation profile typical for other solventogenic *Clostridia* – acidogenesis, until approximately the 7th hour of cultivation and, after that, solventogenesis, see Fig. 2a, b. Interestingly, the strain produced very low concentrations of acetone and produced isopropanol, which, to the best of our knowledge, was not previously reported for the strain despite being mentioned in over 100 articles. This was probably due to the focus on 1,3-propanediol production from glycerol in most of the publication rather than on ABE/IBE production from glucose. Except for the metabolites shown in Fig. 2b, the strain produced trace amounts of acetone and ethanol. A decreased concentration of acids after the 8th hour of cultivation, see Fig. 2b, was due to their partial reutilization, which is one of the protective mechanisms cells use against low pH [44].

Table 1
Ability of *C. diolis* DSM 15410 to grow on different carbohydrates.

Carbohydrate	OD 600 nm after 24 h cultivation*
Arabinose	0.53 ± 0.04
Cellobiose	4.51 ± 1.03
Fructose	2.66 ± 0.96
Glucose	3.14 ± 0.04
Glycerol	1.73 ± 0.17
Lactose	2.10 ± 0.51
Maltose	3.99 ± 0.60
Mannose	3.00 ± 0.75
Xylose	2.86 ± 0.67
Control (medium without carbohydrate addition)	0.58 ± 0.05

* Average and standard deviation from three experiments.

Similar fermentation profiles and also comparable isopropanol concentrations were found in *C. beijerinckii* IBE-performing strains, such as *C. beijerinckii* DSM 6423 [45,46]. IBE fermentation has a certain advantages over ABE fermentations especially if the goal is to use produced solvents as biofuel [47,48]. Production of isopropanol instead of acetone is preferred because acetone is highly corrosive to engine and has low energy density [48]. On the other hand, isopropanol can be used as an octane booster for production of high-octane gasolines and is cheaper than other tested additives [49].

When cultivated in a medium with glycerol, *C. diolis* DSM 15410 produced 1,3- propanediol, however, it only reached a low concentration (0.3–0.4 g/L) when cultivated in an RCM medium. Therefore, we tested production in a medium optimized for 1,3-propanediol production [24], resulting in the propanediol concentration increasing considerably to 5.48 ± 0.32 g/L. The glycerol consumption was 14.50 ± 0.51 g/L, and the butanol concentration was 0.70 ± 0.06 g/L. The concentration of 1,3-propanediol obtained in this study was higher than described by Kaur et al. (2012) [24], however, this might be caused by another experimental design. The production of 1,3-propanediol exclusively from glycerol but not glucose was expected, as it was previously shown by a ^{13}C labeling experiment that the metabolic trait for 1,3-propanediol starts from glycerol in the strain [22].

Just a low number of strains can conduct IBE fermentation and strains that are able to produce both isopropanol and 1,3-propanediol are exception among solvent-producing microorganisms. Search of optimal cultivation conditions to effectively produce both solvents can be one of the important future prospective of *C. diolis* DSM 15410 study.

3.2. The characteristics of the C. diolis DSM 15410 genome

The length of the final genome assembly was 5,940,808 bp and the

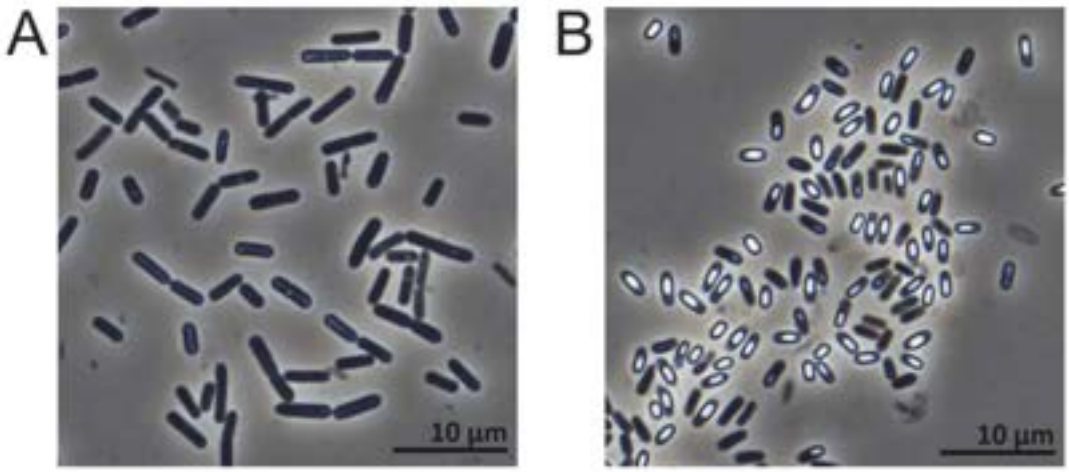


Fig. 1. Vegetative cells (A) and spores (B) of *C. diolis* DSM 15410 obtained in RCM medium after 24 and 48 h of cultivation, respectively.

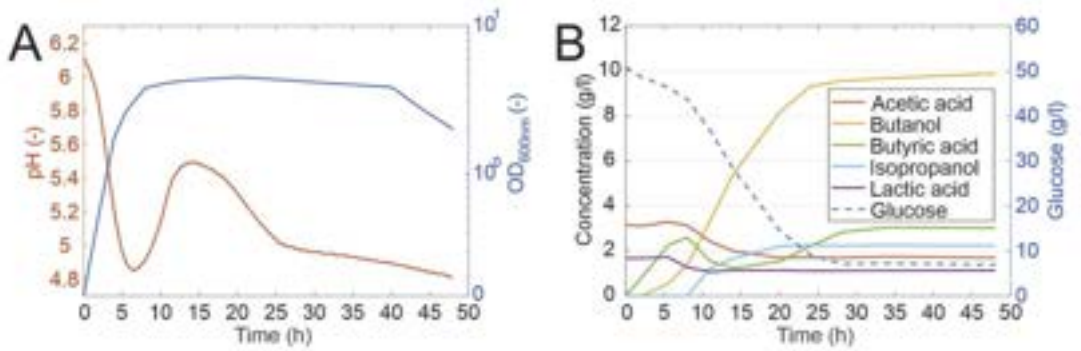


Fig. 2. The fermentation characteristics of *C. diolis* DSM 15410 cultivated in a bioreactor in an RCM medium containing glucose. (A) The growth curve represented by OD_{600nm} and pH course. (B) The glucose consumption and production of metabolites. The standard error of the measurements did not exceed 5%.

sequence has been deposited at DDBJ/EMBL/GenBank under accession No. **CP043998**. Coverage of the assembly after the filtering steps reached almost 700× and the assembly was reconstructed with the contribution of more than 4.4 million paired Illumina reads (83% of all Illumina reads) and more than 1.38 million PacBio reads (84% of all PacBio reads). The remaining reads consisted of unpaired Illumina reads and chimeric PacBio reads that were detected during quality trimming and combining Illumina and PacBio reads into the first draft assembly. PacBio chimeras are not standard PCR-generated chimeras, but are generated during library preparation [50]. They cannot be removed by standard filtering tools, e.g. UCHIME [51], but can be filtered out using ccs reads and by mapping Illumina reads on PacBio reads during hybrid assembly [52,53]. We used both of these approaches (see Materials and Methods). Although the number of remaining reads was very high, they did not form any kind of longer contigs. Therefore, the complete genome of the strain *Clostridium diolis* DSM 15410 is formed by one circular chromosome, whose circularity was proved by searching for overlaps during genome assembly. Moreover, the absence of plasmid DNA was supported by a negative result during the trials for plasmid DNA isolation. Genetic information stored exclusively on the chromosoene is one of the prerequisites for the future engineering of the strain as chromosomal pathway integration is preferred over plasmid-based expression due to its higher genetic stability [6].

The GC content of the sequence was calculated as 29.8% which was according to our presumptions as solventogenic clostridia form a low GC content group of gram-positive bacteria. The genome contains 5244 annotated open reading frames (ORFs) divided into 3224 operons, see Table 2. The majority of ORFs consisted of protein coding genes, but 147 pseudogenes were also found. The sequences of 108 pseudogenes were found incomplete, 36 were frameshifted, 25 contained internals stop, and 19 suffered from multiple problems. The positions of particular features within the genome is shown in Fig. 3. Protein coding genes and pseudogenes were also assigned COG categories. Unfortunately, 596 CDSs were not assigned any COG and additional 990 CDSs were assigned group S with an unknown function. Nevertheless, remaining 3516 CDSs (out of all 5102 protein coding genes and pseudogenes) were divided into the remaining GOG categories, see supplementary Table S1.

Table 2
Genome features of *Clostridium diolis* DSM 15410.

Feature	Chromosome
Length (bp)	5,940,808
GC content (%)	29.8
Total number of ORFs	5244
Total number of operons	3224
Protein coding genes	4955
Pseudogenes	147
rRNA genes (5S, 16S, 23S)	15, 14, 14
tRNA	93
ncRNA	6

Additionally, we decided to test annotated ORFs for prophage genetic code. They may represent a large fraction of the strain-specific DNA sequences as they serve as anchors for genomic rearrangements [54]. We detected 16 prophages in total using PhiSpy, see supplementary Table S2. The cumulative length of prophages was 479,165 bp forming slightly more than 8% of the genome. Such information might be useful for future explorations of the genome rearrangements among *C. beijerinckii* and *C. diolis* strains. Unfortunately, the total number of prophages is questionable as we were not able to reproduce the result of detection using additional online tools. PHASTER [55] predicted only two phages from which only one matched PhiSpy prediction. On the contrary, Prophage Hunter [56] predicted 22 prophages. Thus, future experimental work is needed for the analysis of viral DNA hidden in the *C. diolis* DSM 15410 genome.

3.3. *Restriction-modification systems and CRISPR arrays*

Future engineering of the strain might be limited by restriction-modification (R-M) systems that bacteria use to protect their own DNA. Such a limitation was already described for closely related species *C. beijerinckii* [57] (please note that the referenced strain *C. pasteurianum* NRRL B-598 was reidentified as *C. beijerinckii* NRRL B-598 [12]). We used PacBio sequencing data to study methylations on a genome wide scale. We detected 1217 m6A and 2023 m4C methylated positions and additional 40,787 modified bases in total. Roughly half of the detected methylations (622 m6A and 610 m5C) and 6773 modified bases were used to infer nine methylation motifs, see supplementary Table S3. The data was deposited in the REBASE PacBio database and processed with an internal rebase tool to match the methylated motifs with R-M systems. Seven motifs were found to be unique, while two has already been described before. Unfortunately, none of the detected motifs seemed to be genuine, rather they were the results of miscalls for the m5C motif. Although one 5mC type II methyltransferase was found in the *C. diolis* DSM 15410 genome (M.Cdi15410CORF24710P), its recognition site was not matched with detected motifs as PacBio gives unreliable results for m5C motifs. Besides from type II methylase, a type IV methyl-directed restriction enzyme (Cdi15410CORF14235P) was found, see supplementary Table S4. Activity of R-M system in *C. diolis* DSM 15410 was recently proved by Li et al. [58] as foreign DNA could not be transformed into the cell without pre-methylation. Surprisingly, utilization of its own methyltransferase (M.Cdi15410CORF24710P) did not lead to successful transformation. This result suggests that type IV (Cdi15410CORF14235P) restriction enzyme or additional hitherto unknown R-M system is active in *C. diolis* DSM 15410. Unfortunately, PacBio data are insufficient to resolve this problem, because unknown system must recognize cytosine residues, and needs to be supplemented with bisulfite sequencing in the future to capture active R-M system. Nevertheless, Li et al. [58] meanwhile proved that *C. diolis* DSM 15410 can be transformed using DNA pre-methylated by methylatransferases

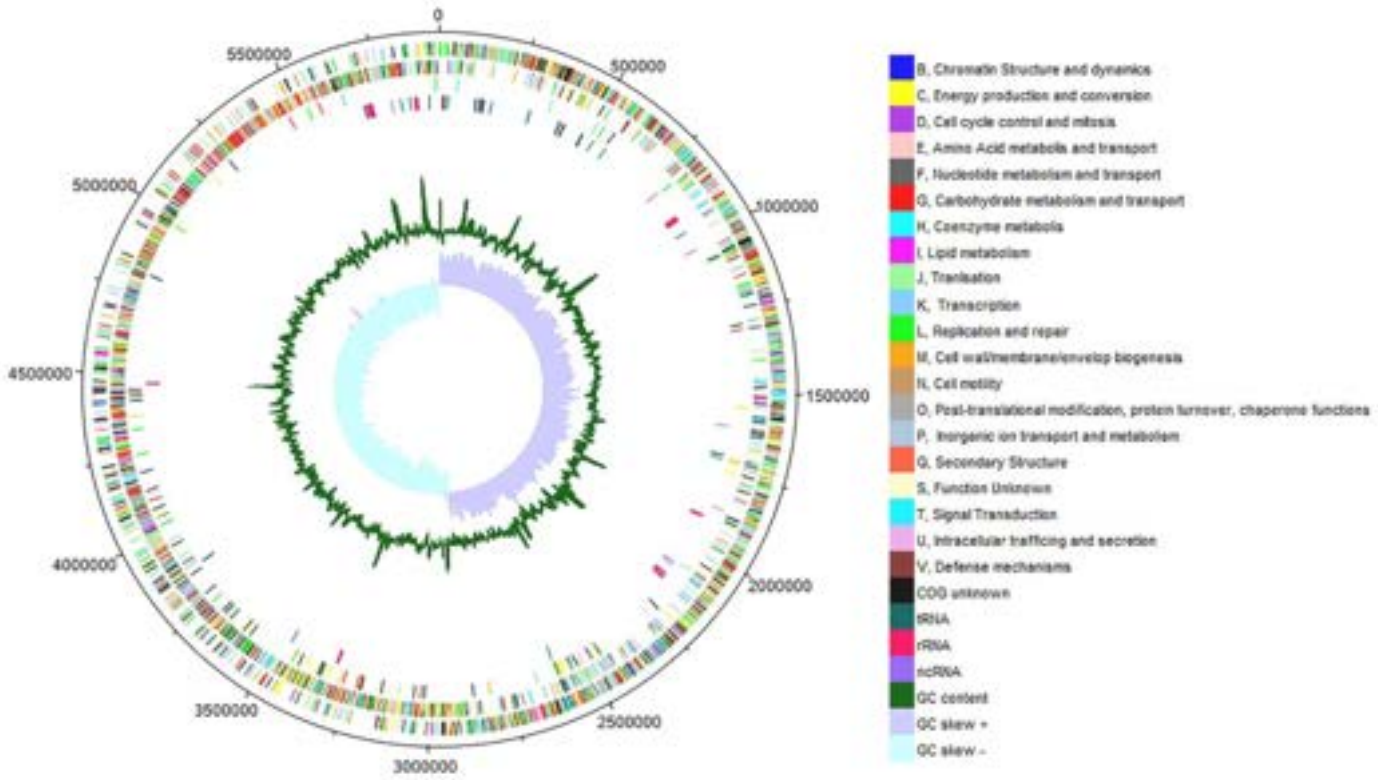


Fig. 3. A circular chromosome map of the *C. diolis* DSM 15410 genome. The outermost and second outermost circles represent CDSs on the forward and reverse strands, respectively. The third circle represents pseudogenes and the colors represent the COG functional classification. The fourth circle represents rRNA genes, while the colors distinguish between tRNA, rRNA, and ncRNA. The inner shaded area represents (from outside in) GC content and GC skew plotted using a 10-kb window with step of 200 bp.

(M. Cce743I and M. Cce743II) from *C. cellulovorans* DSM 743B.

Additional information useful for the future engineering of the strain can be found in CRISPR arrays. A CRISPR-associated system (Cas) forms a kind of bacterial immune system that provides protection from foreign genetic material, including plasmids [59]. On the other hand, CRISPR-Cas9 systems can be also used for genome editing. Although the technique found its utilization mainly for the genome editing of eukaryotes, it has already been applied to closely related (see below) strain of *C. beijerinckii* [60]. We found four CRISPR arrays in the *C. diolis* DSM 15410 genome, see supplementary Table S5. The sizes of arrays ranged from 149 bp to 844 bp and two to 11 spacer units. Except for the shortest one, the remaining three arrays had *cas* or *cas* like genes in their neighborhoods, see supplementary Table S6. Unfortunately, none of the *cas* genes coded the Cas9 protein used for genome editing. Nevertheless,

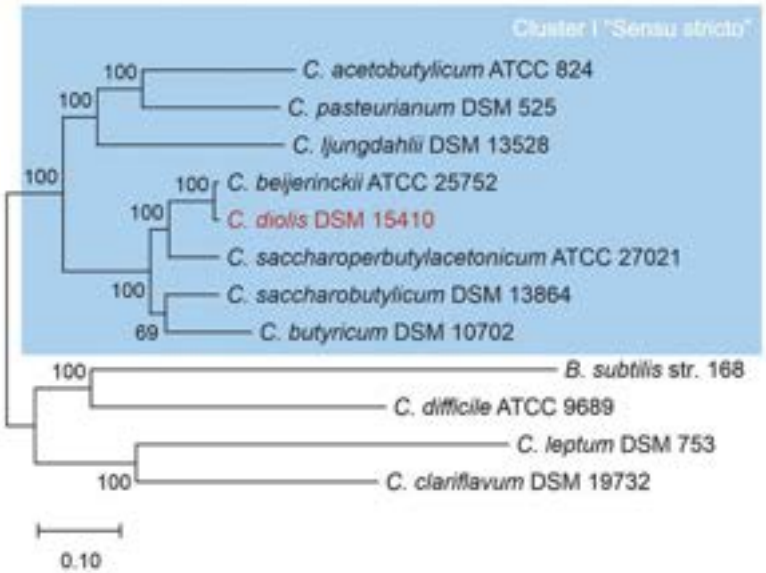


Fig. 4. The phylogenetic placement of *C. diolis* DSM 15410. The blue rectangle contains species from the cluster I "Sensu stricto" of the genus *Clostridium*. The tree was constructed using Phylophlan 3.0, using its internal database of circa 400 genes conserved across bacterial domain. The values represent the boostrap support based on 100 replicates. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

this does not prevent the CRISPR-Cas9 beeing utilized for *C. diolis* DSM 15410 genome editing as the *Streptococcus pyogenes* CRISPR-Cas9 was used for *C. beijerinckii* NCIMB 8052 genome editing.

3.4. Phylogeny

Although the evolution and taxonomy of clostridia was updated last year in a comprehensive phylogenomic study, the *C. diolis* species was omitted due to the missing high quality genome assembly [61]. As Kobayashi et al. [13] meanwhile proposed a reclassification of *C. diolis* as heterotypic synonym for *C. beijerinckii*, meaning they are the same species, it is evident that *C. diolis* DSM 15410 belongs to the cluster I (Sensu stricto) representing the "true" *Clostridium* genus, with *C. butyricum* being the type species. As the reclassification was based on incomplete genomes, we decided to verify its taxonomic placement and selected several solventogenic clostridial species from the cluster I, several "not true" clostridia from other clusters, and *Bacillus subtilis* to perform a phylogenomic analysis, see Fig. 4. Our results showed that *C. diolis* DSM 15410 truly belongs to the subcluster 7 of the cluster I, as defined by Cruz-Morales [61], with *C. beijerinckii*, *C. saccharoperbutylacetonicum*, *C. saccharobutylicum*, and *C. butyricum* being the most closely related species. Thus, we confirmed results of the study by Kobayashi et al. [13] that was unlike our study carried out using only draft genomes.

Digital DNA to DNA hybridization (dDDH) analysis using type strain genome server (TYGS) [62] and newly assembled complete genome confirmed that type strains *C. diolis* DSM 15410 and *C. beijerinckii* DSM 791 reached value of 85.2% confirming they are the same species. Nevertheless, using known complete genome sequences of *C. beijerinckii* strains, we found out that the values is higher (88.8%) for the strain *C. beijerinckii* NRRL B-598, suggesting this strains and *C. diolis* DSM 15410 might be closely related.

We performed comparative cultivation experiment of *C. diolis* DSM 15410 with two closely-related *C. beijerinckii* strains, *C. beijerinckii* NCIMB 8052 and *C. beijerinckii* NRRL B-598 (former *C. pasteurianum* NRRL B-598 [12]), see Table 3, which revealed some differences between them and *C. diolis* DSM 15410. Firstly, *C. diolis* DSM 15410 exhibited better glucose consumption and solvent production in RCM medium with glucose rather that TYA for *C. beijerinckii* strains. At the same time, a much lower frequency of random "acid crash" events, when solvents production was not initiated or was suppressed, was observed on TYA medium without pH control, in comparison to both *C. beijerinckii* NRRL B-598 and NCIMB 8052 strains. Secondly, ability to utilize glycerol and produce 1,3-propanediol and isopropanol in glucose medium was again confirmed for *C. diolis* DSM 15410, which was not observed for the two tested *C. beijerinckii* strains. Nevertheless, these abilities were observed for other *C. beijerinckii* strains, such as, 1,3-propanediol production by type strain *C. beijerinckii* DSM 791 [63] or isopropanol production by *C. beijerinckii* DSM 6423 or *C. beijerinckii* BGS1 [48,64]. Observable growth of two *C. beijerinckii* strains was surprisingly detected in RCM medium containing glycerol. However, this can be attributed to the composition of the medium that is rich and complex and contains various alternative sources of carbon and energy. The slight decrease of glycerol concentration shown in Table 3 might be thus attributed to other than utilization processes, for example, the adhesion of glycerol to cell walls or cell debris.

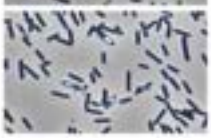
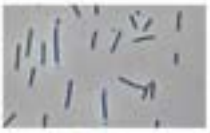
3.5. Selected genes in central metabolism

Homologues of all identified genes playing a key role in the central metabolism of *C. beijerinckii* NRRL B-598 [65] and *C. beijerinckii* NCIMB 8052 [15,66] were found in the *C. diolis* DSM 15410 genome, see Fig. 5. Acidogenic enzymes phosphate acetyltransferase and acetate kinase encoded by *pta* (F3K33_06270) and *ack* (F3K33_06275), phosphate butyryltransferase and butyrate kinase encoded by *ptb* (F3K33_01180) and *buk* (F3K33_01185) catalyze the production of acetate and butyrate from their respective CoA precursors in the strain. These genes, except *buk*, were found in a single copy in the genome; homologous *buk* genes are F3K33_20500 and F3K33_23485. The re-assimilation of acids, observed during solventogenesis, was catalyzed by a CoA-transferase, subunits A and B of which are encoded by *ctfA* (F3K33_19590) and *ctfB* (F3K33_19595). We believe that these genes encoding CoA-transferase together, with two genes encoding solventogenic enzymes aldehyde dehydrogenase *ald* (F3K33_19585) and acetoacetate decarboxylase *adc* (F3K33_19600), form the *sol* operon in *C. diolis* DSM 15410, even though Operon-mapper labeled the *ald* gene as a separate operon. This structure of the *sol* operon falls under type II *sol* operon, which can also be found in *C. beijerinckii*, *C. puniceum*, *C. saccharobutylicum* and *C. saccharoperbutylacetonicum* [67]. Multiple homologous genes encoding butanol dehydrogenase (*bdh*) and *ald* were annotated in the genome, however, a more detailed study is needed to confirm which ones are actively expressed in the strain. The production of isopropanol from acetone was probably catalyzed by NADP-dependent isopropanol dehydrogenase encoded by *adh* (F3K33_14815), which shares high sequence similarity (95.64%) with CIBE_3470 from *C. beijerinckii* DSM 6423 [68]. Production of isopropanol from acetone is an interesting feature of the strain, which is not common among solventogenic clostridia and is mostly performed by *C. beijerinckii* strains [47]. It is possible that NADP-dependent isopropanol dehydrogenase encoded by *adh* (F3K33_14815) can be expressed in more studied clostridial strains to obtain IBE producers. This was already performed for some *C. acetobutylicum* strains, however, expressed gene was cloned from *C. beijerinckii* DSM 6423 (= *C. beijerinckii* NRRL B-593) [69–71].

The biosynthetic pathway of 1,3-propanediol in *C. diolis* DSM 15410 is probably catalyzed by coenzyme B12-independent glycerol dehydratase, glycerol dehydratase activator and 1,3-propanediol dehydrogenase, see Fig. 5, encoded by *dhaB1* (F3K33_19825), *dhaB2* (F3K33_19820), and *dhaT* (F3K33_19810), respectively. This pathway structure is similar to the 1,3-propanediol pathway of *Clostridium butyricum* VPI1718 [72], however, because multiple homologues encoding glycerol dehydratase can be found in the genome of *C. diolis* DSM 15410, other genes may also take part in biosynthesis.

The biosynthetic pathway of 1,3-propanediol can be a perspective target for future research and strain engineering as it allows 1,3-propanediol production from glycerol and close to none acetone production, the desirable state when cultivation is done to produce biofuels. Recently, the first protocol for genetic manipulation of the strain was tested [58] demonstrating enhancement of butanol or 1,3-propanediol productions and butyl acetate formation as a proof of concept. Only few published studies describing expression of 1,3-propanediol pathway in solventogenic clostridia exist. In one of the studies, 1,3-propanediol pathway from *C. butyricum* was expressed in *C. acetobutylicum*, and when engineered strain was cultivated as fed-batch culture, a higher concentration and productivity of 1,3-propanediol than that of the natural producer was observed [73]. Expression of the pathway is more often studied in *E. coli*, where genes from the *Klebsiella pneumoniae* are usually expressed [74,75]. Nevertheless, also genes from *C. butyricum* were successfully used [76]. Because majority of natural 1,3-propanediol producers, including *K. pneumoniae* and *C. pasteurianum*, carry coenzyme B12-dependent glycerol dehydratase, production process requires supplementation of a high-cost vitamin B12. However, supplementation is not required for cultivation of *C. butyricum* VPI1718, bearing coenzyme B12-independent glycerol dehydratase [72]. The same applies to *C. diolis* DSM 15410 as high production of 1,3-propanediol was observed in optimized medium not supplemented with the vitamin [24]. The explanation is the same B12-independent glycerol dehydratase in the *C. diolis* DSM 15410 genome (F3K33_19825) sharing high sequence similarity of the protein product (98.60%) with the corresponding enzyme in *C. butyricum*.

Table 3
Comparison of *C. diolis* DSM 15410, *C. beijerinckii* NCIMB 8052 and *C. beijerinckii* NRRL B-598 (former *C. pasteurianum* NRRL B-598) cell morphology, carbohydrate consumption and metabolites production in different media.

Medium	Morphology	Microscopic image	Glucose consumption (g/L)	Glycerol consumption (decrease) (g/L)	1,3-Propanediol (g/L)	Isopropanol (g/L)	Total solvents (g/L)	Butyric acid (g/L)
C. diolis DSM 15410								
TYA with glucose	Massive sporulation		20.9 ± 1.1	–	0.0 ± 0.0	0.7 ± 0.0	7.7 ± 0.3	0.5 ± 0.2
RCM with glucose	Sporulation		24.5 ± 0.4	–	0.0 ± 0.0	1.0 ± 0.0	8.3 ± 0.3	0.3 ± 0.1
RCM with glycerol	Sporulation limited but few spores recorded		–	5.7 ± 0.9	0.3 ± 0.0	0.1 ± 0.0	1.4 ± 0.3	3.1 ± 0.1
Optimized medium for 1,3-propanediol production [24]	Sporulation limited but few spores recorded		–	13.0 ± 0.3	6.1 ± 0.3	0.1 ± 0.0	7.2 ± 0.4	1.5 ± 0.3
C. beijerinckii NCIMB 8052								
TYA with glucose	Long chains without sporulation		21.7 ± 4.0	–	0.0 ± 0.0	0.0 ± 0.0	6.9 ± 1.4	1.0 ± 0.0
RCM with glucose	Short chins without spores		6.7 ± 0.1	–	0.0 ± 0.0	0.0 ± 0.0	1.7 ± 0.1	1.8 ± 0.1
RCM with glycerol	Various shape, random spores		–	2.6 ± 0.2	0.0 ± 0.0	0.0 ± 0.0	0.2 ± 0.0	3.4 ± 0.1
Optimized medium for 1,3-propanediol production [24]	Limited growth, short chains		–	0.4 ± 0.8	0.0 ± 0.0	0.0 ± 0.0	0.1 ± 0.0	0.4 ± 0.1
C. beijerinckii NRRL B-598 (former *C. pasteurianum* NRRL B-598)								
TYA with glucose	Long chains, no spores		18.7 ± 3.7	–	0.0 ± 0.0	0.0 ± 0.0	5.0 ± 1.0	3.3 ± 0.0
RCM with glucose	Short rods, no spores		11.9 ± 0.6	–	0.0 ± 0.0	0.0 ± 0.0	1.7 ± 0.3	4.2 ± 0.1
RCM with glycerol	Low cell density, visible sporulation		–	2.0 ± 0.4	0.0 ± 0.0	0.0 ± 0.0	0.1 ± 0.1	2.9 ± 0.2
Optimized medium for 1,3-propanediol production [24]	Limited growth, various rods, cell debris		–	0.1 ± 0.1	0.0 ± 0.0	0.0 ± 0.0	0.1 ± 0.0	0.4 ± 0.0

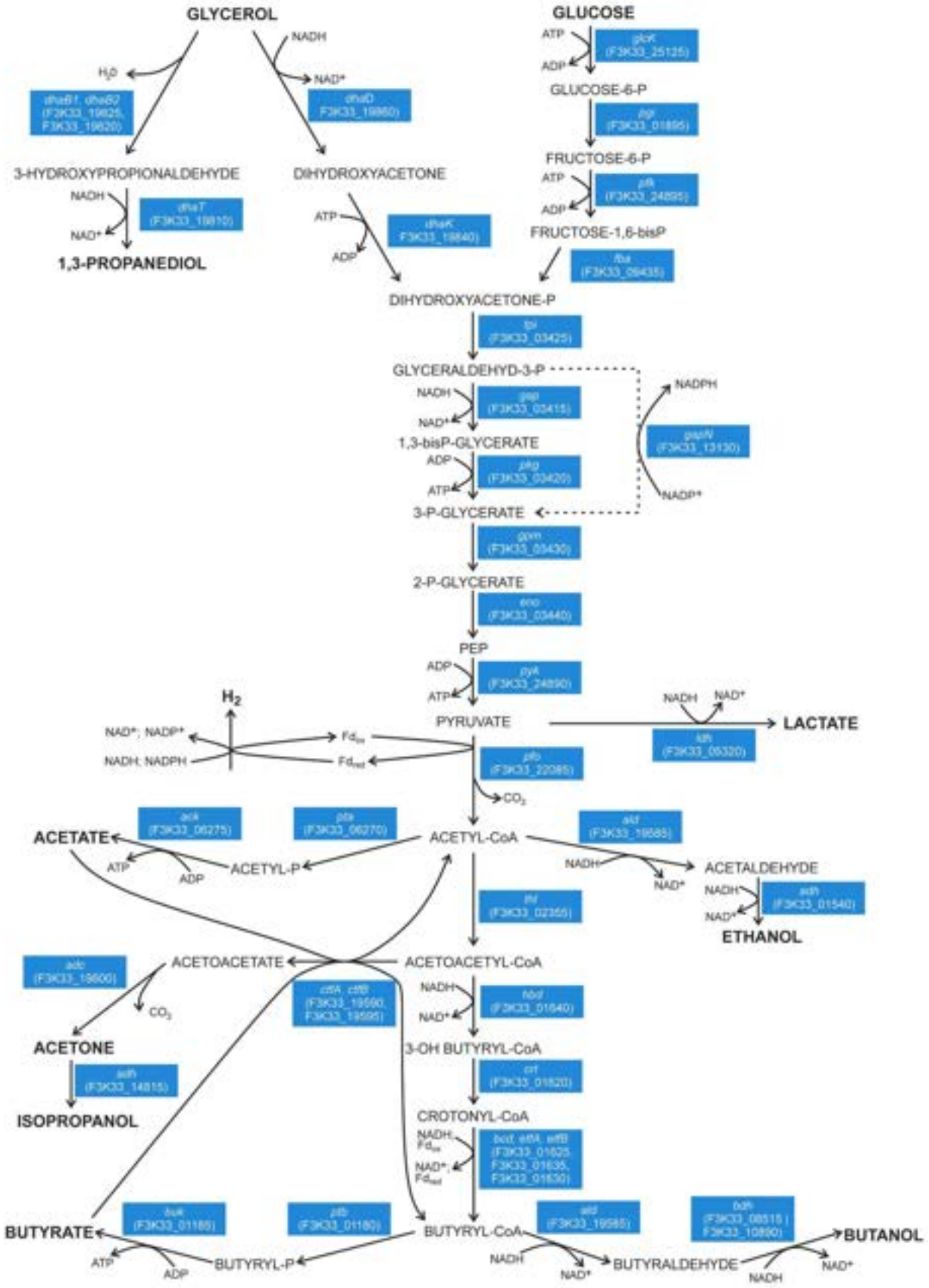


Fig. 5. Central metabolism of *Clostridium diolis* DSM 15410. Main substrates (glycerol, glucose) and main products (1,3-propanediol, H_2, lactate, acetate, acetone, ethanol, isopropanol, butyrate, and butanol) are highlighted in bold. Homologues of genes having a key role in the central metabolism of *C. beijerinckii* NRRL B-598 and *C. beijerinckii* NCIMB 8052 are indicated in blue boxes. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4. Conclusions

We assembled the first complete genome sequence of the type strain *Clostridium diolis* DSM 15410. Genomic analysis of the strain DSM 15410 showed that its genome is a single circular chromosome with a size of 5,940,808 bp that contains 5244 ORFs divided into 3224 operons. We supplemented genome data with the general characterization of the strain using microscopy, HPLC, and other techniques to demonstrate phenotype abilities of the strain for utilization in the production of biofuels. Moreover, we performed a phylogenomic analysis to confirm its placement within cluster I (Sensu stricto) of the genus *Clostridium* and to confirm that *C. beijerinckii* and *C. diolis* are the same species. On the other hand, we highlighted differences of the strain *C. diolis* DSM 15410 by comparative cultivation with two selected *C. beijerinckii* strains. Although we found four CRISPR arrays, three of them with *cas* genes, which could serve as an immune system against foreign DNA, we also found 16 putative prophages in the *C. diolis* DSM 15410 genome suggesting that future engineering of the strain with additional DNA would be possible. We analyzed and reported R-M systems that could prevent gene transfer into *C. diolis* DSM 15410 genome. Although PacBio analysis proved no significant m4C and m6A methylations, and the remaining m5C methylase lacked coupled restriction enzyme, additional exploration by bisulfite sequencing will be needed to fully describe R-M systems in *C. diolis* DSM 15410. Eventually, we analyzed genes of the central metabolism and found genes encoding isopropanol production from acetone and 1,3-propanediol biosynthetic pathway, perspective targets for future research and strain engineering.

Data availability

The genome assembly referred in this paper is the version **CP043998.1**. The whole genome sequencing data have been deposited in the NCBI Sequence Read Archive (SRA) under the project accession number **PRJNA565754**.

Funding

This work has been supported by grant project GACR 17-005515.

Acknowledgements

We acknowledge the CF Genomics of CEITEC supported by the NCMG research infrastructure (LM2015091 funded by MEYS CR) for their support with obtaining scientific data presented in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ygeno.2020.11.007.

References

[1] C. Zhang, T. Li, G. Su, J. He, Enhanced direct fermentation from food waste to butanol and hydrogen by an amylolytic *Clostridium*, Renew. Energy 153 (2020) 522–529, https://doi.org/10.1016/j.renene.2020.01.151.

[2] B. Branska, L. Fortova, M. Dvoraková, H. Liu, P. Patakova, J. Zhang, M. Melzoch, Chicken feather and wheat straw hydrolysate for direct utilization in biobutanol production, Renew. Energy 145 (2020) 1941–1948, https://doi.org/10.1016/j.renene.2019.07.094.

[3] R.K. Mahato, D. Kumar, G. Rajagopalan, Biohydrogen production from fruit waste by *Clostridium* strain BOH3, Renew. Energy 153 (2020) 1368–1377, https://doi.org/10.1016/j.renene.2020.02.092.

[4] P. Patakova, B. Branska, Z. Lin, P. Wu, H. Liu, M. Drahokoupil, Y. Zhou, L. Paulova, J. Zhang, K. Melzoch, Microbial production of butanol from food industry waste, in: Food Ind. Wastes, Elsevier, 2020, pp. 163–180, https://doi.org/10.1016/b978-0-12-817121-9.00008-5.

[5] S. Venkata Mohan, G.N. Nikhil, P. Chiranjeevi, C. Nagendranatha Reddy, M. V. Rohit, A.N. Kumar, O. Sarkar, Waste biorefinery models towards sustainable circular bioeconomy: critical review and future perspectives, Bioresour. Technol. 215 (2016) 2–12, https://doi.org/10.1016/j.biortech.2016.03.130.

[6] M. Gustavsson, S.Y. Lee, Prospects of microbial cell factories developed through systems metabolic engineering, Microb. Biotechnol. 9 (2016) 610–617, https://doi.org/10.1111/1751-7915.12385.

[7] K.R. Choi, S.Y. Lee, CRISPR technologies for bacterial systems: current achievements and future directions, Biotechnol. Adv. 34 (2016) 1180–1209, https://doi.org/10.1016/j.biotechadv.2016.08.002.

[8] I. Mougiakos, E.F. Bosma, W.M. de Vos, R. van Kranenburg, J. van der Oost, Next generation prokaryotic engineering: the CRISPR-Cas toolkit, Trends Biotechnol. 34 (2016) 575–587, https://doi.org/10.1016/J.TIBTECH.2016.02.004.

[9] M. Ackermann, A functional perspective on phenotypic heterogeneity in microorganisms, Nat. Rev. Microbiol. 13 (2015) 497–508, https://doi.org/10.1038/nrmicro3491.

[10] C.D. Moon, D.M. Pacheco, W.J. Kelly, S.C. Leahy, D. Li, J. Kopečný, G.T. Attwood, Reclassification of *Clostridium proteoclasticum* as *Butyrivibrio proteoclasticus* comb. nov., a butyrate-producing ruminal bacterium, Int. J. Syst. Evol. Microbiol. 58 (2008) 2041–2045, https://doi.org/10.1099/ijs.0.65845-0.

[11] P.A. Lawson, F.A. Rainey, Proposal to restrict the genus Clostridium prazmowski to Clostridium butyricum and related species, Int. J. Syst. Evol. Microbiol. 66 (2016) 1009–1016, https://doi.org/10.1099/ijsem.0.000824.

[12] K. Sedlar, J. Kolek, I. Provazník, P. Patakova, Reclassification of non-type strain Clostridium pasteurianum NRRL B-598 as Clostridium beijerinckii NRRL B-598, J. Biotechnol. 244 (2017) 1–3, https://doi.org/10.1016/j.jbiotec.2017.01.003.

[13] H. Kobayashi, Y. Tanizawa, M. Sakamoto, Y. Nakamura, M. Ohkuma, M. Tohno, Reclassification of Clostridium diolis Biebl and Spröer 2003 as a later heterotypic synonym of Clostridium beijerinckii Donker 1926 (approved lists 1980) emend. Keis et al. 2001, Int. J. Syst. Evol. Microbiol. 70 (2020) 2463–2466, https://doi.org/10.1099/ijsem.0.004059.

[14] J. Nölling, G. Breton, M.V. Omelchenko, K.S. Makarova, Q. Zeng, R. Gibson, H. M. Lee, J. Dubois, D. Qiu, J. Hitti, Y.I. Wolf, R.L. Tatusov, F. Sabathe, L. Doucette-Stamm, P. Soucaille, M.J. Daly, G.N. Bennett, E.V. Koonin, D.R. Smith, D.R. Smith, Genome sequence and comparative analysis of the solvent-producing bacterium *Clostridium acetobutylicum*, J. Bacteriol. 183 (2001) 4823–4838, https://doi.org/10.1128/JB.183.16.4823-4838.2001.

[15] Y. Wang, X. Li, Y. Mao, H.P. Blaschek, Single-nucleotide resolution analysis of the transcriptome structure of Clostridium beijerinckii NCIMB 8052 using RNA-Seq, BMC Genomics 12 (2011) 479, https://doi.org/10.1186/1471-2164-12-479.

[16] K. Sedlar, J. Kolek, H. Skutkova, B. Branska, I. Provazník, P. Patakova, Complete genome sequence of Clostridium pasteurianum NRRL B-598, a non-type strain producing butanol, J. Biotechnol. 214 (2015) 113–114, https://doi.org/10.1016/j.jbiotec.2015.09.022.

[17] A. Poehlein, A. Grosse-Honebrink, Y. Zhang, N.P. Minton, R. Daniel, Complete genome sequence of the nitrogen-fixing and solvent-producing Clostridium pasteurianum DSM 525, Genome Announc. 3 (2015) e01591–e01614, https://doi.org/10.1128/genomeA.01591-14.

[18] C. Li, Y. Wang, G. Xie, B. Peng, B. Zhang, W. Chen, X. Huang, H. Wu, B. Zhang, Complete genome sequence of Clostridium butyricum JKY6D1 isolated from the pit mud of a Chinese flavor liquor-making factory, J. Biotechnol. 220 (2016) 23–24, https://doi.org/10.1016/j.jbiotec.2016.01.003.

[19] P. Patakova, M. Linhova, M. Rychtera, L. Paulova, K. Melzoch, Novel and neglected issues of acetone–butanol–ethanol (ABE) fermentation by clostridia: Clostridium metabolic diversity, tools for process mapping and continuous fermentation systems, Biotechnol. Adv. 31 (2013) 58–67, https://doi.org/10.1016/J.BIOTECHADV.2012.01.010.

[20] K. Charubin, R.K. Bennett, A.G. Fast, E.T. Papoutsakis, Engineering Clostridium organisms as microbial cell-factories: challenges & opportunities, Metab. Eng. 50 (2018) 173–191, https://doi.org/10.1016/j.ymben.2018.07.012.

[21] B. Otte, E. Grunwaldt, O. Mahmoud, S. Jennewein, Genome shuffling in Clostridium diolis DSM 15410 for improved 1,3-propanediol production, Appl. Environ. Microbiol. 75 (2009) 7610–7616, https://doi.org/10.1128/AEM.01774-09.

[22] B. Xin, Y. Wang, F. Tao, L. Li, C. Ma, P. Xu, Co-utilization of glycerol and lignocellulosic hydrolysates enhances anaerobic 1,3-propanediol production by Clostridium diolis, Sci. Rep. 6 (2016) 19044, https://doi.org/10.1038/srep19044.

[23] Y. Wang, F. Tao, H. Tang, P. Xu, Genome sequence of Clostridium diolis strain DSM 15410, a promising natural producer of 1,3-propanediol, Genome Announc. 1 (2013), https://doi.org/10.1128/genomeA.00542-13.

[24] G. Kaur, A.K. Srivastava, S. Chand, Determination of kinetic parameters of 1,3-propanediol fermentation by Clostridium diolis using statistically optimized medium, Bioprocess Biosyst. Eng. 35 (2012) 1147–1156, https://doi.org/10.1007/s00449-012-0700-x.

[25] P. Ewels, M. Magnusson, S. Lundin, M. Käller, MultiQC: summarize analysis results for multiple tools and samples in a single report, Bioinformatics. 32 (2016) 3047–3048, https://doi.org/10.1093/bioinformatics/btw354.

[26] A.M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data, Bioinformatics. 30 (2014) 2114–2120, https://doi.org/10.1093/bioinformatics/btu170.

[27] A. Bankevich, S. Nurk, D. Antipov, A.A. Gurevich, M. Dvorkin, A.S. Kulikov, V. M. Lesin, S.I. Nikolenko, S. Pham, A.D. Prjibelski, A.V. Pyshkin, A.V. Sirotkin, N. Vyahhi, G. Tesler, M.A. Alekseyev, P.A. Pevzner, SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing, J. Comput. Biol. 19 (2012) 455–477, https://doi.org/10.1089/cmb.2012.0021.

[28] S. Kosugi, H. Hirakawa, S. Tabata, GMcloser: closing gaps in assemblies accurately with a likelihood-based selection of contig or long-read alignments, Bioinformatics. 31 (2015) 3733–3741, https://doi.org/10.1093/bioinformatics/btv465.

[29] H. Li, R. Durbin, Fast and accurate short read alignment with burrows-wheeler transform, Bioinformatics. 25 (2009) 1754–1760, https://doi.org/10.1093/bioinformatics/btp324.
[30] B.J. Walker, T. Abeel, T. Shea, M. Priest, A. Abouelliel, S. Sakthikumar, C. A. Cuomo, Q. Zeng, J. Wortman, S.K. Young, A.M. Earl, Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement, PLoS One 9 (2014) e112963, https://doi.org/10.1371/journal.pone.0112963.
[31] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, The sequence alignment/map format and SAMtools, Bioinformatics. 25 (2009) 2078–2079, https://doi.org/10.1093/bioinformatics/btp352.
[32] K. Okonechnikov, A. Conesa, F. García-Alcalde, Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data, Bioinformatics. 32 (2016) 292–294, https://doi.org/10.1093/bioinformatics/btv566.
[33] S. Kurtz, A. Phillippy, A.L. Delcher, M. Smoot, M. Shumway, C. Antonescu, S. L. Salzberg, Versatile and open software for comparing large genomes, Genome Biol. 5 (2004) R12, https://doi.org/10.1186/gb-2004-5-2-r12.
[34] F. Gao, C.T. Zhang, Ori-finder: a web-based system for finding oriCs in unannotated bacterial genomes, BMC Bioinformatics 9 (2008) 79, https://doi.org/10.1186/1471-2105-9-79.
[35] T. Tatusova, M. DiCuccio, A. Badretdin, V. Chetvernin, E.P. Nawrocki, L. Zaslavsky, A. Lomsadze, K.D. Pruitt, M. Borodovsky, J. Ostell, NCBI prokaryotic genome annotation pipeline, Nucleic Acids Res. 44 (2016) 6614–6624, https://doi.org/10.1093/nar/gkw569.
[36] B. Taboada, K. Estrada, R. Ciria, E. Merino, Operon-mapper: a web server for precise operon identification in bacterial and archaeal genomes, Bioinformatics. 34 (2018) 4118–4120, https://doi.org/10.1093/bioinformatics/bty496.
[37] J. Huerta-Cepas, D. Szklarczyk, D. Heller, A. Hernández-Plaza, S.K. Forslund, H. Cook, D.R. Mende, I. Letunic, T. Rattei, L.J. Jensen, C. Von Mering, P. Bork, EggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses, Nucleic Acids Res. 47 (2019) D309–D314, https://doi.org/10.1093/nar/gky1085.
[38] T. Carver, N. Thomson, A. Bleasby, M. Berriman, J. Parkhill, DNAPlotter: circular and linear interactive genome visualization, Bioinformatics. 25 (2009) 119–120, https://doi.org/10.1093/bioinformatics/btn578.
[39] K. Rutherford, J. Parkhill, J. Crook, T. Horsnell, P. Rice, M.A. Rajandream, B. Barrell, Artemis: sequence visualization and annotation, Bioinformatics. 16 (2000) 944–945, https://doi.org/10.1093/bioinformatics/16.10.944.
[40] S. Akhter, R.K. Aziz, R.A. Edwards, PhiSpy: a novel algorithm for finding prophages in bacterial genomes that combines similarity-and composition-based strategies, Nucleic Acids Res. 40 (2012) e126, https://doi.org/10.1093/nar/gks406.
[41] R.J. Roberts, T. Vincze, J. Posfai, D. Macelis, REBASE-a database for DNA restriction and modification: enzymes, genes and genomes, Nucleic Acids Res. 43 (2015) D298–D299, https://doi.org/10.1093/nar/gku1046.
[42] A. Biswas, R.H.J. Staals, S.E. Morales, P.C. Fineran, C.M. Brown, CRISPRDetect: a flexible algorithm to define CRISPR arrays, BMC Genomics 17 (2016) 356, https://doi.org/10.1186/s12864-016-2627-0.
[43] N. Segata, D. Börnigen, X.C. Morgan, C. Huttenhower, PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes, Nat. Commun. 4 (2013) 2304, https://doi.org/10.1038/ncomms3304.
[44] B. Branska, Z. Pechacova, J. Kolek, M. Vasylkivska, P. Patakova, Flow cytometry analysis of Clostridium beijerinckii NRRL B-598 populations exhibiting different phenotypes induced by changes in cultivation conditions, Biotechnol. Biofuels 11 (2018) 99, https://doi.org/10.1186/s13068-018-1096-z.
[45] S.A. Survase, G. Jurgens, A. Van Heiningen, T. Granström, Continuous production of isopropanol and butanol using Clostridium beijerinckii DSM 6423, Appl. Microbiol. Biotechnol. 91 (2011) 1305–1313, https://doi.org/10.1007/s00253-011-3322-3.
[46] H. Mâté de Gérando, F. Wasels, A. Bisson, B. Clement, F. Bidard, E. Jourdier, A. M. López-Contreras, N. Lopes Ferreira, Genome and transcriptome of the natural isopropanol producer Clostridium beijerinckii DSM6423, BMC Genomics 19 (2018) 242, https://doi.org/10.1186/s12864-018-4626-7.
[47] C.F. dos Santos Vieira, F. Maugeri Filho, R. Maciel Filho, A. Pinto Mariano, Acetone-free biobutanol production: past and recent advances in the isopropanol-butanol-ethanol (IBE) fermentation, Bioresour. Technol. 287 (2019) 121425, https://doi.org/10.1016/j.biortech.2019.121425.
[48] C. Zhang, T. Li, J. He, Characterization and genome analysis of a butanol-isopropanol-producing Clostridium beijerinckii strain BGS1, Biotechnol. Biofuels 11 (2018) 280, https://doi.org/10.1186/s13068-018-1274-x.
[49] V.G. Rassadin, O.Y. Shlygin, N.M. Likhterova, V.N. Slavin, A.V. Zharov, Problems in production of high-octane, unleaded automotive gasolines, Chem. Technol. Fuels Oils 42 (2006) 235–242, https://doi.org/10.1007/s10553-006-0064-5.
[50] E.B. Fichot, R.S. Norman, Microbial phylogenetic profiling with the Pacific biosciences sequencing platform, Microbiome. 1 (2013) 10, https://doi.org/10.1186/2049-2618-1-10.
[51] R.C. Edgar, B.J. Haas, J.C. Clemente, C. Quince, R. Knight, UCHIME improves sensitivity and speed of chimera detection, Bioinformatics. 27 (2011) 2194–2200, https://doi.org/10.1093/bioinformatics/btr381.
[52] C.-S. Chin, D.H. Alexander, P. Marks, A.A. Klammer, J. Drake, C. Heiner, A. Clum, A. Copeland, J. Huddleston, E.E. Eichler, S.W. Turner, J. Korlach, Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data, Nat. Methods 10 (2013) 563–569, https://doi.org/10.1038/nmeth.2474.
[53] S. Koren, M.C. Schatz, B.P. Walenz, J. Martin, J.T. Howard, G. Ganapathy, Z. Wang, D.A. Rasko, W.R. McCombie, E.D. Jarvis, A.M. Phillippy, Hybrid error correction and de novo assembly of single-molecule sequencing reads, Nat. Biotechnol. 30 (2012) 693–700, https://doi.org/10.1038/nbt.2280.
[54] H. Brüssow, C. Canchaya, W.-D. Hardt, Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion, Microbiol. Mol. Biol. Rev. 68 (2004) 560–602, https://doi.org/10.1128/mmbr.68.3.560-602.2004.
[55] D. Arndt, J.R. Grant, A. Marcu, T. Sajed, A. Pon, Y. Liang, D.S. Wishart, PHASTER: a better, faster version of the PHAST phage search tool, Nucleic Acids Res. 44 (2016) W16–W21, https://doi.org/10.1093/nar/gkw387.
[56] W. Song, H.X. Sun, C. Zhang, L. Cheng, Y. Peng, Z. Deng, D. Wang, Y. Wang, M. Hu, W. Liu, H. Yang, Y. Shen, J. Li, L. You, M. Xiao, Prophage hunter: an integrative hunting tool for active prophages, Nucleic Acids Res. 47 (2019) W74–W80, https://doi.org/10.1093/nar/gkz380.
[57] J. Kolek, K. Sedlar, I. Provaznik, P. Patakova, Dam and Dcm methylations prevent gene transfer into Clostridium pasteurianum NRRL B-598: development of methods for electrotransformation, conjugation, and sonoporation, Biotechnol. Biofuels 9 (2016) 1–14, https://doi.org/10.1186/s13068-016-0436-y.
[58] A. Li, Z. Wen, D. Fang, M. Lu, Y. Ma, Q. Xie, M. Jin, Developing Clostridium diolis as a biorefinery chassis by genetic manipulation, Bioresour. Technol. 305 (2020) 123066, https://doi.org/10.1016/j.biortech.2020.123066.
[59] R. Sorek, C.M. Lawrence, B. Wiedenheft, CRISPR-mediated adaptive immune Systems in Bacteria and Archaea, Annu. Rev. Biochem. 82 (2013) 237–266, https://doi.org/10.1146/annurev-biochem-072911-172315.
[60] Y. Wang, Z.T. Zhang, S.O. Seo, P. Lynn, T. Lu, Y.S. Jin, H.P. Blaschek, Bacterial genome editing with CRISPR-Cas9: deletion, integration, single nucleotide modification, and desirable "clean" mutant selection in Clostridium beijerinckii as an example, ACS Synth. Biol. 5 (2016) 721–732, https://doi.org/10.1021/acssynbio.6b00060.
[61] P. Cruz-Morales, C.A. Orellana, G. Moutafis, G. Moonen, G. Rincon, L.K. Nielsen, E. Marcellin, R. Bapteste, Revisiting the evolution and taxonomy of clostridia, a Phylogenomic update, Genome Biol. Evol. 11 (2019) 2035–2044, https://doi.org/10.1093/gbe/evz096.
[62] J.P. Meier-Kolthoff, M. Göker, TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy, Nat. Commun. 10 (2019) 1–10, https://doi.org/10.1038/s41467-019-10210-3.
[63] D. Wischral, J. Zhang, C. Cheng, M. Lin, L.M.G. De Souza, F.L.P. Pessoa, N. Pereira, S.T. Yang, Production of 1,3-propanediol by *Clostridium beijerinckii* DSM 791 from crude glycerol and corn steep liquor: process optimization and metabolic engineering, Bioresour. Technol. 212 (2016) 100–110, https://doi.org/10.1016/j.biortech.2016.04.020.
[64] E. Rochón, F. Cebreiros, M.D. Ferrari, C. Lareo, Isopropanol-butanol production from sugarcane and sugarcane-sweet sorghum juices by *Clostridium beijerinckii* DSM 6423, Biomass Bioenergy 128 (2019) 105331, https://doi.org/10.1016/j.biombioe.2019.105331.
[65] P. Patakova, B. Branska, K. Sedlar, M. Vasylkivska, K. Jureckova, J. Kolek, P. Koscova, I. Provaznik, Acidogenesis, solventogenesis, metabolic stress response and life cycle changes in Clostridium beijerinckii NRRL B-598 at the transcriptomic level, Sci. Rep. 9 (2019) 1371, https://doi.org/10.1038/s41598-018-37679-0.
[66] Y. Wang, X. Li, Y. Mao, H.P. Blaschek, Genome-wide dynamic transcriptional profiling in clostridium beijerinckii NCIMB 8052 using single-nucleotide resolution RNA-Seq, BMC Genomics 13 (2012) 102, https://doi.org/10.1186/1471-2164-13-102.
[67] A. Poehlein, J.D.M. Solano, S.K. Flitsch, P. Krabben, K. Winzer, S.J. Reid, D. T. Jones, E. Green, N.P. Minton, R. Daniel, P. Dürre, Microbial solvent formation revisited by comparative genome analysis, Biotechnol. Biofuels 10 (2017) 58, https://doi.org/10.1186/s13068-017-0742-z.
[68] M. Diallo, A.D. Simons, H. van der Wal, F. Collas, B. Houweling-Tan, S.W. M. Kengen, A.M. López-Contreras, L-Rhamnose metabolism in Clostridium beijerinckii strain DSM 6423, Appl. Environ. Microbiol. 85 (2018) e02656–e02718, https://doi.org/10.1128/AEM.02656-18.
[69] J. Lee, Y.S. Jang, S.J. Choi, J.A. Im, H. Song, J.H. Cho, D.Y. Seung, E. Terry Papoutsakis, G.N. Bennett, S.Y. Lee, Metabolic engineering of *Clostridium acetobutylicum* ATCC 824 for isopropanol-butanol-ethanol fermentation, Appl. Environ. Microbiol. 78 (2012) 1416–1423, https://doi.org/10.1128/AEM.06382-11.
[70] C. Wang, F. Xin, X. Kong, J. Zhao, W. Dong, W. Zhang, J. Ma, H. Wu, M. Jiang, Enhanced isopropanol-butanol-ethanol mixture production through manipulation of intracellular NAD(P)H level in the recombinant *Clostridium acetobutylicum* XY16, Biotechnol. Biofuels 11 (2018) 12, https://doi.org/10.1186/s13068-018-1024-0.
[71] Z. Dai, H. Dong, Y. Zhu, Y. Zhang, Y. Li, Y. Ma, Introducing a single secondary alcohol dehydrogenase into butanol-tolerant *Clostridium acetobutylicum* Rh8 switches ABE fermentation to high level IBE fermentation, Biotechnol. Biofuels 5 (2012) 44, https://doi.org/10.1186/1754-6834-5-44.
[72] C. Raynaud, P. Sarçabal, I. Meynial-Salles, C. Croux, P. Soucaille, Molecular characterization of the 1,3-propanediol (1,3-PD) operon of Clostridium butyricum, Proc. Natl. Acad. Sci. U. S. A. 100 (2003) 5010–5015, https://doi.org/10.1073/pnas.0734105100.
[73] M. González-Pajuelo, I. Meynial-Salles, F. Mendes, J.C. Andrade, I. Vasconcelos, P. Soucaille, Metabolic engineering of *Clostridium acetobutylicum* for the industrial production of 1,3-propanediol from glycerol, Metab. Eng. 7 (2005) 329–336, https://doi.org/10.1016/j.ymben.2005.06.001.

[74] I.T. Tong, H.H. Liao, D.C. Cameron, 1,3-Propanediol production by *Escherichia coli* expressing genes from the *Klebsiella pneumoniae* dha regulon, Appl. Environ. Microbiol. 57 (1991) 3541–3546.

[75] J.H. Lee, S. Lama, J.R. Kim, S.H. Park, Production of 1,3-propanediol from glucose by recombinant *Escherichia coli* BL21(DE3), Biotechnol. Bioprocess Eng. 23 (2018) 250–258, https://doi.org/10.1007/s12257-018-0017-y.

[76] X. Tang, Y. Tan, H. Zhu, K. Zhao, W. Shen, Microbial conversion of glycerol to 1,3-propanediol by an engineered strain of *Escherichia coli*, Appl. Environ. Microbiol. 75 (2009) 1628–1634, https://doi.org/10.1128/AEM.02376-08.

5.10 Article X

MUSILOVA, Jana, KOURILOVA, Xenie, BEZDICEK, Matej, LENGEROVA, Martina, OBRUCA, Stanislav, SKUTKOVA, Helena, and **SEDLAR, Karel**. First Complete Genome of the Thermophilic Polyhydroxyalkanoates-Producing Bacterium Schlegelella thermodepolymerans DSM 15344. *Genome Biology and Evolution*. 2021. 13(4), p. evab007. (2021 IF = 4.065, Q2 in EVOLUTIONARY BIOLOGY)

DOI: https://doi.org/10.1093/gbe/evab007

GBE

First Complete Genome of the Thermophilic Polyhydroxyalkanoates-Producing Bacterium *Schlegelella thermodepolymerans* DSM 15344

Jana Musilova[1], Xenie Kourilova[2], Matej Bezdicek[3], Martina Lengerova[3], Stanislav Obruca[2], Helena Skutkova[1], and Karel Sedlar [1,*]

[1]Department of Biomedical Engineering, Faculty of Electrical Engineering and Communication, Brno University of Technology, Czech Republic

[2]Department of Food Chemistry and Biotechnology, Faculty of Chemistry, Brno University of Technology, Czech Republic

[3]Department of Internal Medicine—Hematology and Oncology, University Hospital Brno, Czech Republic

*Corresponding author: E-mail: sedlar@feec.vutbr.cz.

Accepted: 6 January 2021

Abstract

Schlegelella thermodepolymerans is a moderately thermophilic bacterium capable of producing polyhydroxyalkanoates—biodegradable polymers representing an alternative to conventional plastics. Here, we present the first complete genome of the type strain *S. thermodepolymerans* DSM 15344 that was assembled by hybrid approach using both long (Oxford Nanopore) and short (Illumina) reads. The genome consists of a single 3,858,501-bp-long circular chromosome with GC content of 70.3%. Genome annotation identified 3,650 genes in total, whereas 3,598 open reading frames belonged to protein-coding genes. Functional annotation of the genome and division of genes into clusters of orthologous groups revealed a relatively high number of 1,013 genes with unknown function or unknown clusters of orthologous groups, which reflects the fact that only a little is known about thermophilic polyhydroxyalkanoates-producing bacteria on a genome level. On the other hand, 270 genes involved in energy conversion and production were detected. This group covers genes involved in catabolic processes, which suggests capability of *S. thermodepolymerans* DSM 15344 to utilize and biotechnologically convert various substrates such as lignocellulose-based saccharides, glycerol, or lipids. Based on the knowledge of its genome, it can be stated that *S. thermodepolymerans* DSM 15344 is a very interesting, metabolically versatile bacterium with great biotechnological potential.

Key words: de novo assembly, hybrid assembly, functional annotation, PHA.

Significance

The type strain *Schlegelella thermodepolymerans* DSM 15344 is a thermophilic bacterium capable of production of polyhydroxyalkanoates (PHA)—microbial polyesters representing "green" alternative to petrochemical polymers. The fact that the bacterium grows and thrives at elevated temperatures brings numerous benefits (e.g., reduced risk of contamination) which might positively influence the process of PHA production. Moreover, due to enormous metabolic flexibility, *S. thermodepolymerans* DSM 15344 is capable of PHA production from human food chain non-competing substrates such as lignocelluloses and other waste streams of the agro-food industry. Even though the strain DSM 15344 is biotechnologically promising, genomic information on this strain is scarce. Here, we describe the first complete genome of the *S. thermodepolymerans* DSM 15344 that will facilitate studies on its PHA metabolism and enable us to further investigate and improve its biotechnological potential.

Introduction

Polyhydroxyalkanoates (PHA) are polyesters of hydroxyalkanoic acids. As the PHA are produced naturally by microbial fermentation, they can be regarded as an environmental friendly alternative to petroleum-based polymers (Muhammadi et al. 2015; Sabapathy et al. 2020). Although some facts regarding PHA fermentation are known, for example, microorganisms use PHA to store unused energy and carbon into cytoplasm in a form of intracellular granules and these granules help the organism to cope with stressors (Obruca et al. 2018), additional basic knowledge is needed to establish viable industrial processes. Although production of bioplastics is considered to be the future way and inseparable part of circular economy, less than 1% of the total plastic production comes from bioplastics industry (Shogren et al. 2019).

The type strain *S. thermodepolymerans* DSM 15344 is a thermophilic, Gram-negative bacterium that was originally investigated for its ability to degrade extracellular PHA materials such as copolymers of 3-hydroxybutyrate and 3-mercaptopropionate (Elbanna et al. 2003). So far, two draft genome assemblies of the strain were published. The assembly available under the GenBank accession number GCA_002933415.1 submitted by Zhejiang Academy of Agricultural Sciences consists of 48 contigs with N50 length of 174,537 bp and the assembly GCA_003349825.1 by DOE Joint Genome Institute contains 28 scaffolds with N50 length of 324,832 bp. Although these represent relatively high-quality draft assemblies, probably due to missing high-quality complete genome assembly and functional annotation of the genome, other important features of the strain remained hidden. Only recently, its ability to produce PHA was reported together with the unique capability of xylose utilization (Kourilova et al. 2020). Optimal growth temperature of *S. thermodepolymerans* DSM 15344, 55 °C, reduces the risk of microbial contamination; therefore, the strain presents an ideal organism for utilization in the "Next Generation Industrial Biotechnology" concept in which biotechnological process is conducted under unsterile conditions (Chen and Jiang 2018). In this article, we present its first high-quality complete genome sequence, which is currently a reference sequence for *S. thermodepolymerans* species in GenBank database. We annotated the genome, predicted the operon structure, and searched for prophage DNA and CRISPR arrays.

Materials and Methods

Growth Conditions, DNA Extraction, and Sequencing

Schlegelella thermodepolymerans DSM 15344 was obtained from Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures Braunschweig, Germany. Optimal temperature for bacterial growth was 55 °C. The inoculum was developed in nutrient-rich medium Nutrient Broth (HiMedia, India) containing peptone 10.0 g/l, beef extract 10.0 g/l, and sodium chloride 5.0 g/l. After 24 h, 5% of the bacterial suspension was inoculated into a mineral salt medium composed of $Na_2HPO_4 \cdot 12\ H_2O$ (9.0 g/l), KH_2PO_4 (1.5 g/l), NH_4Cl (1.0 g/l), $MgSO_4 \cdot 7\ H_2O$ (0.2 g/l), $CaCl_2 \cdot 2\ H_2O$ (0.02 g/l), $Fe^{(III)}NH_4$citrate (0.0012 g/l), yeast extract (0.5 g/l), 1 ml/l of microelements solution (EDTA [50.0 g/l], $FeCl_3 \cdot 6\ H_2O$ [13.8 g/l], $ZnCl_2$ [0.84 g/l], $CuCl_2 \cdot 2\ H_2O$ [0.13 g/l], $CoCl_2 \cdot 6\ H_2O$ [0.1 g/l], $MnCl_2 \cdot 6\ H_2O$ [0.016 g/l], H_3BO_3 [0.1 g/l], dissolved in distilled water), and a xylose (20.0 g/l) as a carbon substrate. *Schlegelella thermodepolymerans* DSM 15344 was cultivated in mineral salt medium under the same conditions as the inoculum.

Genomic DNA was extracted using MagAttract HMW DNA Kit (Qiagene, NL). The DNA purity was checked using NanoDrop (Thermo Fisher Scientific), the concentration was measured using Qubit 2.0 Fluorometr (Thermo Fisher Scientific), and the proper length of the isolated DNA was confirmed using Agilent 4200 TapeStation (Agilent technologies). The sequencing library for Oxford Nanopore sequencing was prepared using Ligation sequencing 1D Kit (Oxford Nanopore Technologies, UK). The sequencing was performed using the R9.4.1 flowcell and the MinION platform (Oxford Nanopore Technologies). The sequencing library for short-read sequencing was prepared using KAPA HyperPlus kit and was carried out using Miseq Reagent Kit v2 (500 cycles) and Illumina MiSeq platform (Illumina).

Genome Assembly

The Nanopore reads were basecalled using Guppy v3.4.4 (https://community.nanoporetech.com, last accessed September 18, 2020), and quality was checked using MinIONQC (Lanfear et al. 2019). Subsequently, the reads were assembled with Flye v2.8.1 (https://github.com/fenderglass/Flye, last accessed October 2, 2020). Polishing was done using Racon v1.4.13 (Vaser et al. 2017) and Medaka (https://github.com/nanoporetech/medaka, last accessed October 8, 2020); auxiliary PAF files were generated using minimap2 (Li 2018). The Illumina paired-end (PE) reads were initially quality-checked using FastQC v0.11.5 and MultiQC v1.7 (Ewels et al. 2016). Trimmomatic v1.36 (Bolger et al. 2014) was subsequently used for the adapter and quality trimming. In the next step, trimmed Illumina PE reads were mapped to the Nanopore initial assembly using BWA v07.17 (Hi 2013). Finally, the obtained assembly was polished using Pilon v1.23 (Walker et al. 2014); auxiliary BAM files were obtained using SAMtools (Li et al. 2009). As the last step, the final assembly was rearranged according to the replication origin (oriC) identified using Ori-finder (Luo et al. 2019), so the *DnaA* gene is the first gene.

Genome Annotation and Analysis

Genome annotation was done through the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) (Tatusova et al. 2016).

An operon prediction was performed using Operon-mapper (Taboada et al. 2018), and the results were manually inserted into the genome record. Functional annotation of the protein-coding genes was performed by classifying them into clusters of orthologous groups (COG) from the eggNOG database using the eggNOG-mapper (Huerta-Cepas et al. 2019). Chromosomal map of the circular genome was subsequently produced with the Artemis (Rutherford et al. 2000)-integrated DNAPlotter (Carver et al. 2009). Prophage DNA was searched using Prophage Hunter (Song et al. 2019) and PHASTER (Arndt et al. 2016). Finally, the annotated genome sequence was further analyzed for presence of CRISPR loci using CRISPRDetect tool (Biswas et al. 2016).

Results and Discussion

Genome Assembly and Properties

Schlegelella thermodepolymerans DSM 15344 initial genome assembly was reconstructed from nearly 1.8 million Oxford Nanopore Technologies reads with a median read length of 4.9 kb and finalized by mapping more than 2.4 million high-quality (average Phred score Q ≈ 35) Illumina read pairs (88% of all Illumina reads) to the initial assembly. Whole process resulted into the final assembly consisting of one circular chromosome with coverage exceeding 5,500×. The genome has been deposited at the DDBJ/EMBL/GenBank under accession number CP064338.1.

The genome length is 3,858,501 bp and contains 3,650 genes in total, divided into 1,729 operons. Most of the genes are protein-coding sequences (CDSs), but 33 pseudogenes were also found, which is less than 44 and 50 pseudogenes detected in previously published draft genome sequences PSNY00000000.1 and QQAP00000000.1, respectively. The GC content reached the value of 70.28% which is more than the average for Gram-negative bacteria (Li and Du 2014). However, it met our expectations, as it corresponded to the value 70.3% of the previously published draft genomes. High GC content can be associated with the adaptation of the bacterium to high-temperature environments. Although only single copies of rRNA genes were detected in draft genomes of *S. thermodepolymerans* DSM 15344, the complete genome sequence contains 5S, 16S, and 23S rRNA genes in duplicates. Moreover, copies of 16S and 23S rRNA genes differ in three and one positions, respectively. Such information is useful for future identification of *S. thermodepolymerans* in metagenomics studies and quantification of its abundance in microbial studies based on amplicon sequencing. The overall sequence features are summarized in table 1.

Table 1
Genomic Features of *Schlegelella thermodepolymerans* DSM 15344

Feature	Chromosome
Length (bp)	3,858,501
GC content (%)	70.28
Genes	3,650
Operons	1,729
CDS	3,589
Pseudogenes	33
rRNA (5S, 16S, 23S)	2, 2, 2
tRNA	51
ncRNA	4

Functional Annotation

The protein-coding genes were classified according to COG into 22 categories. In total, 2,576 CDSs were assigned a COG category with the most prevalent groups E—amino acid metabolism and transport containing 7.80% of the total number of CDS (280 out of 3,589) and C—energy production and conversion containing 7.52% of the total number of CDS (270 out of 3,589). This suggests that *S. thermodepolymerans* has a functional apparatus capable of utilizing a wide range of substrates as reported recently (Kourilova et al. 2020). Unfortunately, 9.33% (335 genes) were not assigned any COG and 18.89% (678 genes) were assigned group S with an unknown function. In fact, such a result was expected as only a little is known about genomes of thermophilic bacteria capable of PHA synthesis so far. (For details of each group, including the number of assigned genes assigned see supplementary table S1, Supplementary Material online.)

The position of individual features in the circular genome is shown in figure 1. Each COG is marked with a different color. Moreover, RNAs are divided into tRNA, rRNA, and ncRNA categories and displayed in the fourth outermost circle.

Searching for viral DNA resulted only in inconclusively identified prophages. Although Prophage Hunter identified five putative prophages, PHASTER results consisted of a single incomplete prophage that overlapped with one candidate indentified by Prophage Hunter. None of these phages was identified as active. This is according to our expectations as phages are viruses for which temperature is a crucial factor for survivability (Nasser and Oman 1999). Optimal temperature for growth of the strain (55 °C) is too high for most phages (Farrell and Campbell 1969). Although a group of thermophilic phages also exists, they usually occur in specific environment (Jończyk et al. 2011) and were not identified in the *S. thermodepolymerans* DSM 15344 genome. Only a single 164-bp-long CRISPR array containing two spacer units was found in the *S. thermodepolymerans* DSM 15344 genome. Unfortunately, no *cas* or *cas*-like genes were found in its neighborhood. Nevertheless, this does not prevent the CRISPR-Cas9 being utilized for *S. thermodepolymerans* DSM 15344 genome editing as a foreign system could be used.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

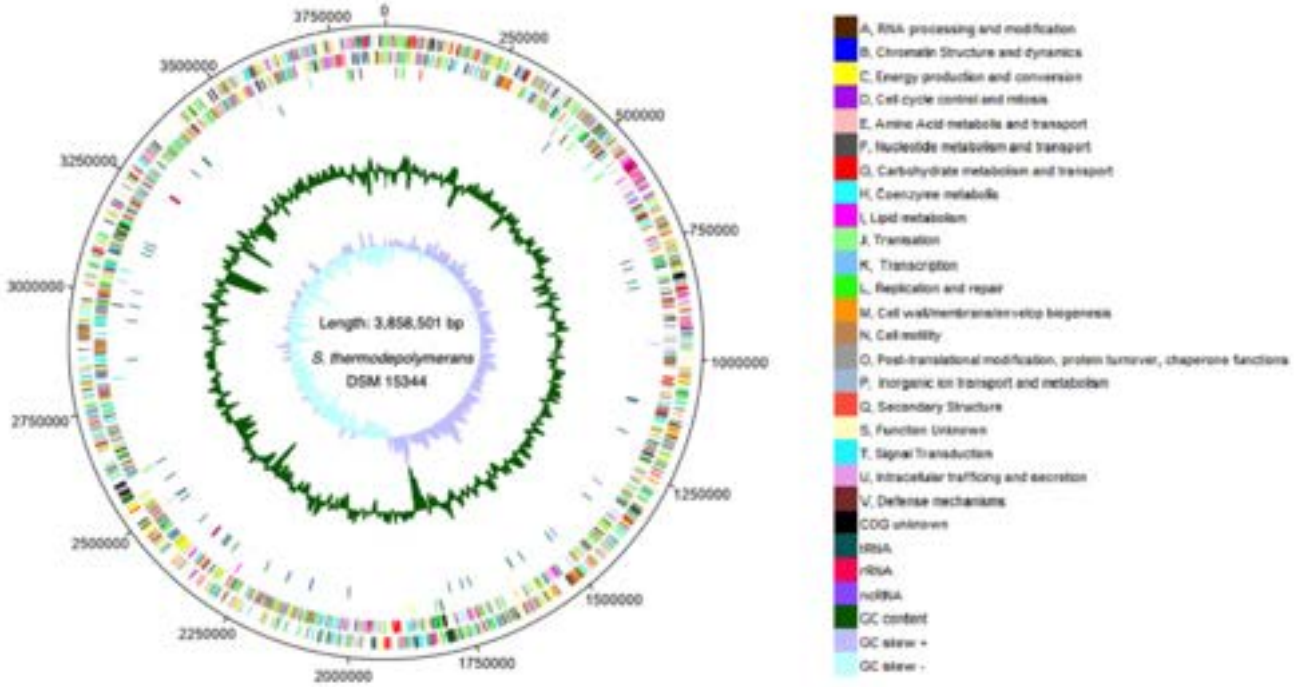


Fig. 1.—A chromosomal map of the *S. thermodepolymerans* DSM 15344 genome. The first two outermost circles represent CDSs on the forward and backward strands, respectively; the third circle represents pseudogenes. The colors represent the functional classification of COG. The fourth outermost circle represents RNA genes, distinguishing among tRNA, rRNA, and ncRNA. The inner area represents the GC content and GC skew.

Acknowledgments

This study was funded by the project GA17-01821S of the Czech Science Foundation (GACR) and supported by the Brno University of Technology intra-university junior project FCH/FEKT-J-20-6399. PhD study of Jana Musilova is supported by the Brno Ph.D. Talent Scholarship—Funded by the Brno City Municipality.

Data Availability

The whole-genome sequence has been deposited at DDBJ/ENA/GenBank under the accession number CP064338.1. The NCBI BioProject and BioSample IDs are PRJNA674774 and SAMN16675030, respectively. The raw reads have been deposited at NCBI SRA database under the accession numbers SRR13206616 (PE Illumina) and SRR13206615 (Oxford Nanopore Technologies).

Literature Cited

Arndt D, et al. 2016. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res. 44(W1):W16–W21.

Biswas A, Staals RHJ, Morales SE, Fineran PC, Brown CM. 2016. CRISPRDetect: a flexible algorithm to define CRISPR arrays. BMC Genomics 17(1):356.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30(15):2114–2120.

Carver T, Thomson N, Bleasby A, Berriman M, Parkhill J. 2009. DNAPlotter: circular and linear interactive genome visualization. Bioinformatics 25(1):119–120.

Chen GQ, Jiang XR. 2018. Next generation industrial biotechnology based on extremophilic bacteria. Curr Opin Biotechnol. 50:94–100.

Elbanna K, et al. 2003. *Schlegelella thermodepolymerans* gen. nov., sp. nov., a novel thermophilic bacterium that degrades poly(3-hydroxybutyrate-co-3-mercaptopropionate). Int J Syst Evol Microbiol. 53(4):1165–1168.

Ewels P, Magnusson M, Lundin S, Käller M. 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics 32(19):3047–3048.

Farrell J, Campbell LL. 1969. Thermophilic bacteria and bacteriophages. Adv Microb Physiol. 3:83–109.

Hi L. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv: Genomics. Available from: http://arxiv.org/abs/1303.3997.

Huerta-Cepas J, et al. 2019. EggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. 47(D1):D309–D314.

Jończyk E, Kłak M, Międzybrodzki R, Górski A. 2011. The influence of external factors on bacteriophages—review. Folia Microbiol. 56(3):191–200.

Kourilova X, et al. 2020. Production of polyhydroxyalkanoates (PHA) by a thermophilic strain of Schlegelella thermodepolymerans from xylose rich substrates. Bioresour Technol. 315:123885.

Lanfear R, Schalamun M, Kainer D, Wang W, Schwessinger B. 2019. MinIONQC: fast and simple quality control for MinION sequencing data. Bioinformatics 35(3):523–525.

Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34(18):3094–3100.

Li H, et al. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25(16):2078–2079.

Li X-Q, Du D. 2014. Variation, evolution, and correlation analysis of C+G content and genome or chromosome size in different kingdoms and phyla. PLoS One 9(2):e88339.

Luo H, Quan C-L, Peng C, Gao F. 2019. Recent development of Ori-Finder system and DoriC database for microbial replication origins. Brief Bioinform. 20(4):1114–1124.

Muhammadi S, Afzal M, Hameed S. 2015. Bacterial polyhydroxyalkanoates-eco-friendly next generation plastic: production, biocompatibility, biodegradation, physical properties and applications. Green Chem Lett Rev. 8:56–77.

Nasser AM, Oman SD. 1999. Quantitative assessment of the inactivation of pathogenic and indicator viruses in natural water sources. Water Res. 33(7):1748–1752.

Obruca S, Sedlacek P, Koller M, Kucera D, Pernicova I. 2018. Involvement of polyhydroxyalkanoates in stress resistance of microbial cells: biotechnological consequences and applications. Biotechnol Adv. 36(3):856–870.

Rutherford K, et al. 2000. Artemis: sequence visualization and annotation. Bioinformatics 16(10):944–945.

Sabapathy PC, et al. 2020. Recent developments in Polyhydroxyalkanoates (PHAs) production – a review. Bioresour Technol. 306:123132.

Shogren R, Wood D, Orts W, Glenn G. 2019. Plant-based materials and transitioning to a circular economy. Sustain Prod Consum. 19:194–215.

Song W, et al. 2019. Prophage Hunter: an integrative hunting tool for active prophages. Nucleic Acids Res. 47(W1):W74–W80.

Taboada B, Estrada K, Ciria R, Merino E. 2018. Operon-mapper: a web server for precise operon identification in bacterial and archaeal genomes. Bioinformatics 34(23):4118–4120.

Tatusova T, et al. 2016. NCBI prokaryotic genome annotation pipeline. Nucleic Acids Res. 44(14):6614–6624.

Vaser R, Sović I, Nagarajan N, Šikić M. 2017. Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res. 27(5):737–746.

Walker BJ, et al. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9(11):e112963.

Associate editor: Ochman Howard

5.11 Article XI

JURECKOVA, Katerina, RASCHMANOVA, Hana, KOLEK, Jan, VASYLKIVSKA, Maryna, BRANSKA, Barbora, PATAKOVA, Petra, PROVAZNIK, Ivo and **SEDLAR, Karel**. Identification and Validation of Reference Genes in Clostridium beijerinckii NRRL B-598 for RT-qPCR Using RNA-Seq Data. *Frontiers in Microbiology*. 2021. 12, p. 476. (2021 IF = 6.064, Q1 in MICROBIOLOGY)

DOI: https://doi.org/10.3389/fmicb.2021.640054

frontiers in Microbiology

ORIGINAL RESEARCH
published: 18 March 2021
doi: 10.3389/fmicb.2021.640054

Identication and Validation of Reference Genes in *Clostridium beijerinckii* NRRL B-598 for RT-qPCR Using RNA-Seq Data

*Katerina Jureckova[1], Hana Raschmanova[2], Jan Kolek[2], Maryna Vasylkivska[2], Barbora Branska[2], Petra Patakova[2], Ivo Provaznik[1] and Karel Sedlar[1]**

[1] Department of Biomedical Engineering, Faculty of Electrical Engineering and Communication, Brno University of Technology, Brno, Czechia, [2] Department of Biotechnology, University of Chemistry and Technology Prague, Prague, Czechia

Gene expression analysis through reverse transcription-quantitative real-time polymerase chain reaction (RT-qPCR) depends on correct data normalization by reference genes with stable expression. Although *Clostridium beijerinckii* NRRL B-598 is a promising Gram-positive bacterium for the industrial production of biobutanol, validated reference genes have not yet been reported. In this study, we selected 160 genes with stable expression based on an RNA sequencing (RNA-Seq) data analysis, and among them, seven genes (*zmp*, *rpoB1*, *rsmB*, *greA*, *rpoB2*, *topB2*, and *rimO*) were selected for experimental validation by RT-qPCR and gene ontology (GO) enrichment analysis. According to statistical analyses, *zmp* and *greA* were the most stable and suitable reference genes for RT-qPCR normalization. Furthermore, our methodology can be useful for selection of the reference genes in other strains of *C. beijerinckii* and it also suggests that the RNA-Seq data can be used for the initial selection of novel reference genes, however, their validation is required.

Keywords: HKG, housekeeping genes, non-model organisms, biofuel, Clostridium

OPEN ACCESS

Edited by:
Krishnaveni Mishra,
University of Hyderabad, India

Reviewed by:
Yi Wang,
Auburn University, Auburn, United States
Peter-D rre,
University of Ulm, Ulm, Germany

***Correspondence:**
Karel Sedlar
sedlar@vut.cz

Specialty section:
This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal Frontiers in Microbiology

Received: 10 December 2020
Accepted: 17 February 2021
Published: 18 March 2021

Citation:
Jureckova K, Raschmanova H, Kolek J, Vasylkivska M, Branska B, Patakova P, Provaznik I and Sedlar K (2021) Identication and Validation of Reference Genes in Clostridium beijerinckii NRRL B-598 for RT-qPCR Using RNA-Seq Data. Front. Microbiol. 12:640054. doi: 10.3389/fmicb.2021.640054

INTRODUCTION

Reverse transcriptase-quantitative real-time polymerase chain reaction (RT-qPCR) is the most commonly used technique to quantify gene expression due to its high sensitivity, specicity, and reproducibility (Ginzinger, 2002). Correct quantication of mRNA relies on data normalization that removes dierences in the extraction yield, reverse-transcriptase activity and eciency of PCR amplication (Vandesompele et al., 2002; Bustin et al., 2009). The most commonly used normalization method utilizes the so-called reference genes against which are gene expression data relatively quantied. Reference genes should maintain a constant mRNA expression regardless of the experimental conditions, dierent tissues, cells, or life cell phases (Derveaux et al., 2010). Thus, correct selection plays a crucial role in accurate data normalization. However, there is no universal reference gene and many studies have shown that the expression of commonly used reference genes is not always stable (Liang et al., 2014; Chapman and Waldenström, 2015; Jo et al., 2019), and reference genes should be selected individually for each organism and experimental condition (Liang et al., 2014). Furthermore, literature mining is not an appropriate approach for

their selection and cannot be used in most cases. On the other hand, recent studies use RNA sequencing (RNA-Seq) technology for the evaluation of the whole transcriptomes to nd novel candidates for reference genes (Carvalho et al., 2014; Hu et al., 2016; Pombo et al., 2017).

Thus far, suitable reference genes have been determined in dierent *Clostridium* species, such as *Clostridium dicile* (Metcalf et al., 2010), *Clostridium botulinum* (Kirk et al., 2014), and *Clostridium ljungdahlii* (Liu et al., 2013). In *Clostridium beijerinckii* NCIMB 8052, 177 putative housekeeping genes were previously identied based on transcriptomic data (Wang et al., 2011). However, thus far, no evaluation study of reference genes has been performed for this species. *C. beijerinckii* has been found to be a promising microorganism for industrial production of biobutanol, and eorts to increase butanol productivity by means of genetic and metabolic engineering of its strains have been recently reported (Agu et al., 2019; Xin et al., 2020). However, the strain engineering of *C. beijerinckii* is hindered by insucient understanding of cellular physiology and regulatory mechanisms of gene expression. Despite the recent progress in CRISPR-associated methods of genetic engineering tailored to *C. beijerinckii* (Wang et al., 2016), the identication of valid endogenous reference genes for RT-qPCR is lacking.

In the present study, we selected seven putative reference genes (*zmp*, *rpoB1*, *rsmB*, *greA*, *rpoB2*, *topB2*, and *rimO*) for *C. beijerinckii* NRRL B-598 based on the RNA-Seq data that were obtained under dierent experimental conditions and at dierent time-points. The seven candidate reference genes were described and summarized by the gene ontology (GO) enrichment analysis and further tested for expression stability by RT-qPCR experiments and evaluated by four statistical algorithms: NormFinder (Andersen et al., 2004), RefFinder (Xie et al., 2012), Coecient of variation (CV) analysis (Boda et al., 2009), and Pairwise ΔCt method (Silver et al., 2006). According to the stability rating, we identied a novel set of reference genes that can be used for the normalization of RT-qPCR data of *C. beijerinckii* NRRL B-598. As the strain *C. beijerinckii* NRRL B-598 shares high homology with other *C. beijerinckii* strains (Sedlar et al., 2017), the results will be useful for all strains of the species.

MATERIALS AND METHODS

RNA-Seq Data Pre-processing

RNA sequencing data were obtained in our previous studies in which we rst observed the transcription changes during standard ABE fermentation of *C. beijerinckii* (Sedlar et al., 2018; Patakova et al., 2019). The RNA-Seq data consisted of ve replicates and six samples/time-points per replicate. Next, we evaluated the transcriptional response of *C. beijerinckii* to butanol shock (Sedlar et al., 2019) with two replicates and six samples/time-points each. Together, the available RNA-Seq transcriptomes consisted of 42 samples collected across 12 diverse time-points and conditions (see **Table 1**).

Pre-processing of the RNA-Seq data was performed in the same manner as in our previous studies. However, the analysis was recalculated to ensure the utilization of the same versions of particular tools for all samples. The data quality assessment was conducted by FastQC (v0.11.5) and summarized reports across samples were generated by MultiQC (v1.7) (Ewels et al., 2016). Trimmomatic software (v1.36) (Bolger et al., 2014) was used for quality and adapter trimming. The RNA-Seq reads corresponding to 16S and 23S rRNA genes sequences were ltered out by SortMeRNA (v2.1) (Kopylova et al., 2012) and SILVA database (Quast et al., 2013) (v132). Cleansed reads were mapped by STAR (v 2.7.3a) (Dobin et al., 2013) to a reference genome of *C. beijerinckii* available in the GenBank database under accession CP011966.3. Mapping results in SAM (Sequence Read Alignment/Map) le format were indexed and converted to BAM (Binary Read Alignment/Map) les by SAM-tools (v1.7) (Li et al., 2009).

Mapped reads were counted by the featureCounts function from R/Bioconductor Rsubread (v2.2.6) package (Liao et al., 2014, 2019), utilizing two counting strategies: one for uniquely mapped reads and the other for multi-mapping reads. Raw count tables were further processed in R (v4.0.2), using custom-made scripts, and then used for estimation of TPM (transcript per million) values for each gene and sample. From obtained TPM values, we determined the mean value and CV [standard deviation (SD) to the mean (μ)] of TPM values for each gene.

Furthermore, we performed dierential expression analysis on raw count tables using R/Bioconductor DESeq2 (v1.28.1) package (Love et al., 2014) between all 12 time-points. A GO enrichment analysis was conducted by the R/Bioconductor topGO (v2.40.0) package (Alexa and Rahnenfurher, 2020) based on the *C. beijerinckii* GO annotation map created in our previous study (Sedlar et al., 2019).

Selection of Candidate Reference Genes

The selection of candidate reference genes was conducted by a series of ltering steps according to the results of processed RNA-Seq data: TPM values, CV of TPM values and results from dierential expression analysis. Based on the results from dierential expression analysis between all 66 time-points pairs, we counted the number of times each gene was not signicantly regulated, *p*-adjust value > 0.1 (BenjaminiHochberg correction) and ltered out all genes that did not pass the threshold of 50 insignicant changes. In the next step, we eliminated genes with a mean TPM value lower or equal to 35 TPM. Finally, we removed genes with the CV of TPM greater or equal to 30%. After each data ltering step, we compared results from both counting methods (unique and multi-mapping options), and only the genes reported by both methods were preserved for further processing.

Samples for RT-qPCR

The gene expression of the seven candidate reference genes *zmp*, *rpoB1*, *rsmB*, *greA*, *rpoB2*, *topB2*, and *rimO* were assessed in the following cultivation samples: T1, T6, T_b0, T_b2, t0, t1, and t1_CH (see **Supplementary Material 1**). The origin of the samples T1, T6, T_b0, and T_b2 is described in our previous transcriptomic studies (Sedlar et al., 2018, 2019) (see **Table 1**). Briey, samples T1 and T6 were obtained during a bioreactor batch cultivation of *C. beijerinckii* NRRL B-598 on the TYA medium at time points of 3.5 and 23 h, respectively (Sedlar et al., 2018). Samples T_b0 and

TABLE 1 | RNA-Seq samples.

Sample name	Number of replicates	Cultivation conditions	Time-point (h of cultivation)	Sample description	References
T1	5	Bioreactor cultivation, TYA medium containing glucose	3.5	Taken in the middle of acidogenic phase when OD 600 nm reached approximately 1	Sedlar et al., 2018; Patakova et al., 2019; Vasylkivska et al., 2019
T6			23	Taken during solventogenic phase, when rst mature spores were observed	
T_b0	2	Bioreactor cultivation, TYA medium containing glucose, butanol addition (at 6 h)	6	Taken directly before butanol addition, close to shift between acidogenesis and solventogenesis	Sedlar et al., 2019; Patakova et al., 2020
T_b2			7	Taken 1 h after approximately 4.5 g/L (non-lethal concentration) of butanol was added to the medium to capture short-term response to the shock	

T_b2 were obtained during bioreactor cultivation of *C. beijerinckii* NRRL B-598 on a TYA medium with added butanol at time points 0 and 1 h after butanol addition, respectively (Sedlar et al., 2019). The samples t0, t1, and t1_CH were obtained as follows: 480 ml of the TYA medium was inoculated with a spore suspension of *C. beijerinckii* NRRL B-598 and cultured overnight at 37°C in 90% N_2/10% H_2 atmosphere in an anaerobic chamber (Concept 400; Ruskinn Technology). An overnight grown culture (12 h old) was split into two parallels per 240 ml, and the antibiotic chloramphenicol (30 µg ml^{-1}) was added to one of the parallels. This point was set as time point zero (t0). After 1 h of incubation, a sample was withdrawn from both untreated (t1) and chloramphenicol-treated (t1_CH) cultures.

RNA Isolation and Reverse Transcription

Samples of the culture broth for the RNA isolation were centrifuged (13,000 *g*, 2 min, 4°C). The pellets were then washed with ice-cold distilled water and resuspended and diluted to reach optical density (measured at 600 nm) of approximately 1.0. Next, 3 ml of the suspension with the $OD_{600} \approx 1.0$ sample were centrifuged, and the cell pellet was immediately stored at −80°C for subsequent RNA isolation. Frozen samples were thawed on ice, and the total RNA was isolated using the High Pure RNA Isolation Kit (Roche Life Science, Basel, Switzerland), according to the manufacturers instructions. The quality and concentration of the RNA samples were checked on a nanodrop machine (Thermo Fisher Scientic, Waltham, MA, United States).

Reverse transcription of RNA samples was performed with the Transcriptor First Strand cDNA Synthesis Kit (Roche Life Science, Basel, Switzerland), according to the manufacturers instructions.

Quantitative Real-Time PCR

All RT-qPCR analyses were performed in QuantStudio 5 instrument (ThermoFisher Scientic, Waltham, MA, United States) using the 5 × HOT FIREPol EvaGreen qPCR Mix Plus (ROX) (Solis BioDyne, Tartu, Estonia). The reaction mix (20 µl) was prepared according to the manufacturers instructions, using each primer (see **Table 2**) in the nal concentration of 200 nM. The cycling conditions were set according to the manufacturers instructions, using a primer annealing temperature of 61°C. Primer specicity was conrmed through melt curve analysis after the cycling stage (95°C for 15 min, 61°C for 1 min, 95°C for 1 s). All RT-qPCR analyses were performed in triplicates, and the absence of contamination was conrmed by running no-template and no-RT controls.

Polymerase chain reaction eciency was determined for each sample using a 5x serial dilution of cDNA samples (5x, 25x, 125x, and 625x), ranging between 94 and 110% for all samples with the correlation coecient $R^2 > 0.99$. For each gene in each sample, the average C_q value from the dilution interval 5625x was used for further data analysis.

RT-qPCR Data Analysis

Four dierent algorithms or statistical methods were used for the evaluation of the stability of gene expression. These included

TABLE 2 | Sequences of primers used for quantitative real-time PCR.

Target gene	Forward primer	Reverse primer
rnp (K276_20795)	TGCATCAACAA GGGCTTTGAA	TATGTCATTG CTGCTGCGTC
rpoB1 (K276_25940)	ACGAAGCTAG GACCAGAGGA	TGGCGTTACC TTTCCAACCA
rsmB (K276_20790)	AGTGCACCTG GTGGTAAAACT	CCAAGCCCTG AACAAGGAAC
greA (K276_26205)	GCAGAAGCGGA CCCTATGAA	TCCGTCTGGA ACTGTTACCTC
rpoB2 (K276_25935)	AAGAGCGCC AAGAGAATGGT	ACAGCCATTT GGTCCCCATC
topB2 (K276_03960)	TATTTAGCG CAGCCCCCATT	TCTGCTTTTG CCGCATCTTC
rimO (K276_20450)	GCTGCTAAAA TGGATGGGCA	GAGCCATTTCAT AGCTTCTTCCA

TABLE 3 | Candidate reference genes of *C. beijerinckii* NRRL B-598.

Gene annotation	Abbr.	Locus tag	References
Zinc metallopeptidase	*zmp*	X276_20795	Chi et al., 2016
DNA-directed RNA polymerase subunit beta	*rpoB1*	X276_25940	Metcalf et al., 2010; Kirk et al., 2014; Wang et al., 2014; Gomes et al., 2018
16S rRNA [cytosine(967)-C(5)]-methyltransferase RsmB	*rsmB*	X276_20790	Metcalf et al., 2010; Peng et al., 2014; Liu et al., 2017; Delorenzo and Moon, 2018
Transcription elongation factor GreA	*greA*	X276_26205	Song et al., 2019
DNA-directed RNA polymerase subunit beta	*rpoB2*	X276_25935	Metcalf et al., 2010; Kirk et al., 2014; Wang et al., 2014; Gomes et al., 2018
Type IIA DNA topoisomerase subunit B	*topB2*	X276_03960	Metcalf et al., 2010; Carvalho et al., 2014
30S ribosomal protein S12 methylthiotransferase RimO	*rimO*	X276_20450	Song et al., 2019

These genes were selected based on the RNA-Seq data, their biological function, and the literature review.

NormFinder (Andersen et al., 2004), RefFinder (Xie et al., 2012), CV analysis (Boda et al., 2009) and Pairwise ΔCt method (Silver et al., 2006).

RESULTS

RNA-Seq Data Analysis

In order to identify putative candidates for reference genes, we evaluated 5,442 predicted genes of *C. beijerinckii* and applied several ltering steps to reveal genes with stable expression indierent in time and under dierent conditions (see section Materials and Methods).

First, we removed 4,370 genes based on the results from the dierential expression analysis. Next, we eliminated genes according to their mean TPM and the number of candidates dropped from 1,072 to 448 genes. Finally, we applied ltering step based on the CV of TPM and only 160 genes remained as candidates for reference genes (see **Supplementary Material 2**).

In the next step, we narrowed the list of the candidate genes by considering their biological function as well as the fact that the genes with the same function were reported to be putative reference genes and/or used as reference genes in RT-qPCR experiments in other publications (see **Table 3**).

Calculated CV values of seven selected genes based on TPM values ranging from 18.7 to 29.4%. The most stable genes were *rsmB*, *zmp*, and *rimO* (see **Table 4**).

TABLE 4 | Coefcient of variation of transcript per million (TPM) values from both count tables (uniquely mapped reads and multi-mapping reads).

Gene abbreviation	Uniquely mapped reads	Multimapping reads
	CV (%)	CV (%)
rsmB	18.7	18.8
zmp	19.3	20.4
rimO	21.9	22.2
greA	22.7	22.6
rpoB2	22.8	22.8
rpoB1	25.1	24.9
topB2	29.4	28.9

Finally, we performed GO enrichment analysis of the nal set of genes using all genomic loci as the gene universe to describe their functionality. We found 22 biological process (BP) GO terms (see **Table 5**) and 12 molecular function (MF) GO terms (see **Table 6**) that were signicantly enriched (p-value < 0.05, Fishers exact test). From BP terms, the most signicant terms related to nucleic acid metabolic process, macromolecule metabolic process or cellular nitrogen compound metabolic process. The most common term for MF GO terms was nucleic acid binding.

Analysis of Expression Stability Based on RT-qPCR Data

The expression levels, represented by the C_q values, of the seven candidate reference genes (*zmp*, *rpoB1*, *rsmB*, *greA*, *rpoB2*, *topB2*, and *rimO*) were assessed by RT-qPCR in seven cultivation samples. The samples were acquired in experiments with dierent cultivation conditions (standard, i.e., untreated culture, butanol-treated culture, and chloramphenicol-treated culture) and at dierent time points. The mean C_q values ranged from 20.2 to 33.3 for dierent dilutions and dierent genes in dierent samples (see **Supplementary Material 1**). An average C_q value was calculated for each gene in each sample across all dilutions, which was used for further analysis of expression stability. The median C_q for the dierent candidate genes ranged from 24.3 to 28.6, indicating that the expression levels of the dierent genes were not dramatically dierent (see **Figure 1**).

According to NormFinder, CV and Pairwise ΔCt analyses, the genes with the most stable expression were *zmp*, *greA*, and *rpoB2* (see **Table 7** and **Supplementary Material 1**). Using NormFinder (Andersen et al., 2004), the samples were gathered into three groups with respect to the three dierent cultivation experiments from which they originated. According to RefFinder, including multiple analytical tools (Delta CT, BestKeeper, Normnder, and Genorm), the comprehensive ranking of the genes with respect to expression stability was *zmp*, *greA*, *rsmB*, *topB2*, *rpoB2*, *rpoB1*, and *rimO*, with *zmp* and *greA* appearing as the best combination based on three out of the four tools (see **Figure 2**). Using the RefFinder tools, the samples were not gathered into groups. Taken together, *zmp* and *greA* appear to be the most suitable combination of genes that can be used for normalization of RT-qPCR experiments under tested conditions in *C. beijerinckii* NRRL B-598.

TABLE 5 | Biological process (BP) gene ontology (GO) enrichment results.

GO.ID	Term	Annotated	Signicant	Expected	Classic Fisher
GO:0090304	Nucleic acid metabolic process	710	6	1.41	0.00016
GO:0006139	Nucleobase-containing compound metabolic process	885	6	1.75	0.00062
GO:0016070	RNA metabolic process	505	5	1.00	0.00066
GO:0044260	Cellular macromolecule metabolic process	983	6	1.95	0.00116
GO:0006725	Cellular aromatic compound metabolic process	993	6	1.97	0.00123
GO:0010467	Gene expression	576	5	1.14	0.00124
GO:0046483	Heterocycle metabolic process	996	6	1.97	0.00126
GO:1901360	Organic cyclic compound metabolic process	1,014	6	2.01	0.00140
GO:0018197	Peptidyl-aspartic acid modication	1	1	0.00	0.00198
GO:0018339	Peptidyl-L-beta-methylthioaspartic acid biosynthetic process from peptidyl-aspartic acid	1	1	0.00	0.00198
GO:0034641	Cellular nitrogen compound metabolic process	1,084	6	2.15	0.00209
GO:0009451	RNA modication	44	2	0.09	0.00299
GO:0006351	Transcription, DNA-templated	390	4	0.77	0.00329
GO:0097659	Nucleic acid-templated transcription	391	4	0.78	0.00332
GO:0032774	RNA biosynthetic process	394	4	0.78	0.00342
GO:0043170	Macromolecule metabolic process	1,198	6	2.37	0.00381
GO:0018198	Peptidyl-cysteine modication	2	1	0.00	0.00396
GO:0034470	ncRNA processing	55	2	0.11	0.00464
GO:0006354	DNA-templated transcription, elongation	3	1	0.01	0.00594
GO:0032784	Regulation of DNA-templated transcription	3	1	0.01	0.00594
GO:0006396	RNA processing	63	2	0.12	0.00606
GO:0034654	Nucleobase-containing compound biosynthesis	522	4	1.03	0.00979

TABLE 6 | Molecular function (MF) gene ontology (GO) enrichment results.

GO.ID	Term	Annotated	Signicant	Expected	Classic Fisher
GO:0003899	DNA-directed 5′3 ′ RNA polymerase activity	6	2	0.01	3.6e-05
GO:0034062	5′3 ′ RNA polymerase activity	7	2	0.01	5.0e-05
GO:0097747	RNA polymerase activity	7	2	0.01	5.0e-05
GO:0140098	Catalytic activity, acting on RNA	97	3	0.16	0.00037
GO:0103039	Protein methylthiotransferase activity	1	1	0.00	0.00169
GO:0003676	Nucleic acid binding	803	5	1.36	0.00285
GO:0019899	Enzyme binding	2	1	0.00	0.00338
GO:0035596	Methylthiotransferase activity	2	1	0.00	0.00338
GO:0070063	RNA polymerase binding	2	1	0.00	0.00338
GO:0016779	Nucleotidyltransferase activity	65	2	0.11	0.00472
GO:0050497	Transferase activity, transferring alkylthio groups	3	1	0.01	0.00506
GO:0003918	DNA topoisomerase type II (double strand cut, ATP-hydrolyzing) activity	5	1	0.01	0.00842

DISCUSSION

Normalization by reference genes is required for a precise analysis of gene expression by RT-qPCR. The choice of the reference genes, however, should not solely rely on the gold standard used for the particular cell type, such as 16S rRNA in the case of bacteria (Rocha et al., 2015), but should always be systematically and experimentally validated (Bustin et al., 2009). Moreover, according to the MIQE guidelines (Bustin et al., 2009), normalization against more than one reference gene is preferred.

In the case of *C. beijerinckii* NCIMB 8,052, a list of putative reference genes based on the analysis of transcriptomic data was previously published (Wang et al., 2011), and one of the candidate genes (peptidase T, Cbei_2428) was subsequently used as a reference for RT-qPCR experiments (Wang et al., 2012). However, completely dierent set of putative reference genes was generated for closely-related *C. beijerinckii* NRRL B-598 during our analysis of RNA-Seq data of the strain. Only Cbei_1214 was matched by X276_20450 in *C. beijerinckii* NRRL B-598, and X276_14515, a homolog of Cbei_2428 used in Wang et al. (2012) study, has not found to be a suitable reference gene in our case (see **Supplementary Material 2** and section Discussion below). We investigated the gene expression stability in *C. beijerinckii* NRRL B-598 by evaluating both transcriptomic and RT-qPCR

data. Based on these, we suggest a set of appropriate reference genes regarding the given set of tested conditions (see section Materials and Methods).

Identication of candidate reference genes in *C. beijerinckii* was performed by bioinformatics analysis of our previously published RNA-Seq data (Sedlar et al., 2018, 2019; Patakova et al., 2019). All data were processed again with up-to-date software and packages and we have introduced a new approach to the selection of reference genes based on RNA-Seq data. Pre-processing of our RNA-Seq data remained the same until the counting of mapped reads. Here, we decided to create two dierent count tables with two counting options. The rst counting method is strict counting of unique reads mapped to the genomic loci and the second method also considers multi-mapping reads. Contribution of all multi-mapping reads were split equally among all genomic objects they mapped to. Therefore, the number of reads in the sample remained the same. Multi-mapping reads present a problem for downstream analyses as they reduce sensitivity (Canzar et al., 2016). Although several strategies and specialized tools were proposed to count multi-mapping reads (Roberts et al., 2013; Zytnicki, 2017), for our purposes, specicity was the principal issue. By selecting only genes where dierential expression was not detected using uniquely neither multi mapped reads, we reduced the possibility of a type I error and, therefore, improved specicity. We are aware that this approach led to a higher possibility of type II error and, thus, to lower sensitivity. However, this would be problem only for following dierential analysis aiming at nding dierentially expressed genes. In our case, the aim was to completely reduce ambiguities. Next, we decided to calculate TPM values from obtained count tables, as TPM corrects dierences in library sizes and gene lengths and enables comparisons among samples (Wagner et al., 2012).

A selection of housekeeping genes was performed by evaluation of the counts of insignicant changes in expression obtained from dierential expression analysis, mean TPM values, and the CV of TPM values. First, ltering based on the results of the dierential expression analysis, which we used to nd genes with stable expression and minimal regulation, removed genes that were regulated in more than 25% of possible time-points pairs. In the next step, we eliminated genes with low mean TPM values as we were looking for genes with moderate stable expression. Low TPM values can result from technical noise in the data even after ltration. In the last step, we focused on the ltration of genes with a gradual increase or decrease in expression, which would not be discovered by dierential expression analysis. For ltering, we used the CV of TPM values that can reect those slow changes in expression and is often used in other studies geared toward the selection of reference genes (Carvalho et al., 2014; Hu et al., 2016; Pombo et al., 2017). Only 16 out of 160 genes identied by the RNA-Seq data analysis corresponded to the previously identied 177 housekeeping genes in *C. beijerinckii* NCIMB 8,052 (Wang et al., 2011) (see **Supplementary Material 2**). While this might suggest that the reference genes identied in this study are not utilizable for other *C. beijerinckii* strains, we believe this is rather a matter of the simplied methodology used by Wang et al. (2011). In

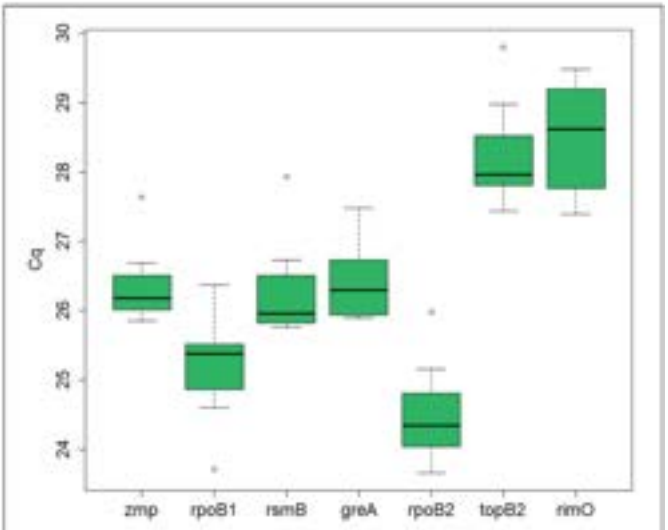


FIGURE 1 | RT-qPCR Cq values for the candidate reference genes in *C. beijerinckii* NRRL B-598. Depicted is the median Cq value calculated for each candidate reference gene across all samples. The boxes indicate the 25th and 75th percentiles and the error bars indicate the maximum and minimum values gained across the samples for each gene.

their study, the identication of HKGs was only a supplementary result and calculating CV from RPKM values appeared to be inappropriate (Wagner et al., 2012). Moreover, identication of HKG based solely on the RNA-Seq data is less specic, see section Discussion below.

Final selection of the seven candidate reference genes (see **Table 3**) for RT-qPCR experiments was summarized by a GO enrichment analysis. GO enrichment analysis of biological processes revealed 22 enriched terms at signicance level $\alpha = 0.05$ of the Fishers exact test (see **Table 5**) and most of the enriched terms related to nucleic acid metabolic process (GO:0090304) and RNA metabolic process (GO:0016700) terms, which correspond with terms identied in other bacterial strains (Rocha et al., 2015). MF GO enrichment analysis revealed ten enriched terms and the most common term nucleic acid binding (GO:0003676) corresponds with identied biological processes (see **Table 6**).

Based on the RT-qPCR experiments, *zmp* (zinc metallopeptidase), *greA* (transcription elongation factor GreA), and *rsmB* (16S rRNA (cytosine(967)-C(5))-methyltransferase RsmB) were the most stably expressed genes (see **Table 7** and **Figure 2**). In pathogenic *Clostridium* species, the zinc metalloproteases act like neurotoxins (Breidenbach and Brunger, 2004) or are involved in the degradation of extracellular substrates (Cafardi et al., 2013). The transcription factor GreA is evolutionarily conserved and widely distributed in prokaryotes (Marr and Roberts, 2000). The 16S rRNA-methyltransferases ensure methylation of 16S rRNA and in Gram-negative bacteria are involved in resistance to aminoglycosides (Lioy et al., 2014). In Gram-positive bacteria, 16S rRNA-methyltransferases was required for resistance to tetracycline antibiotics in the case of *Streptococcus pneumoniae* (Lupien et al., 2015), and to oxidative

TABLE 7 | Expression stability of candidate reference genes in *C. beijerinckii* NRRL B-598 assessed by different analytical tools.

Ranking	NormFinder (Andersen et al., 2004)		Coefcient of variation analysis (Boda et al., 2009)		Pairwise Δ Ct method (Silver et al., 2006)	
	Gene	Stability value	Gene	CV*1 (%)	Gene	Average SD*2
1	*zmp*	0.129	*zmp*	30.9	*rpoB2*	0.455
2	*greA*	0.283	*greA*	38.8	*zmp*	0.460
3	*rpoB2*	0.357	*topB2*	43.1	*topB2*	0.481
4	*rsmB*	0.379	*rpoB2*	46.3	*rpoB1*	0.488
5	*topB2*	0.403	*rsmB*	48.6	*rsmB*	0.650
6	*rpoB1*	0.495	*rpoB1*	63.7	*greA*	0.846
7	*rimO*	0.630	*rimO*	70.3	*rimO*	0.880
Best combination	*zmp + greA*		*zmp + greA*		*rpoB2 + zmp*	

*1 *coefcient of variation and* * 2 *standard deviation.*

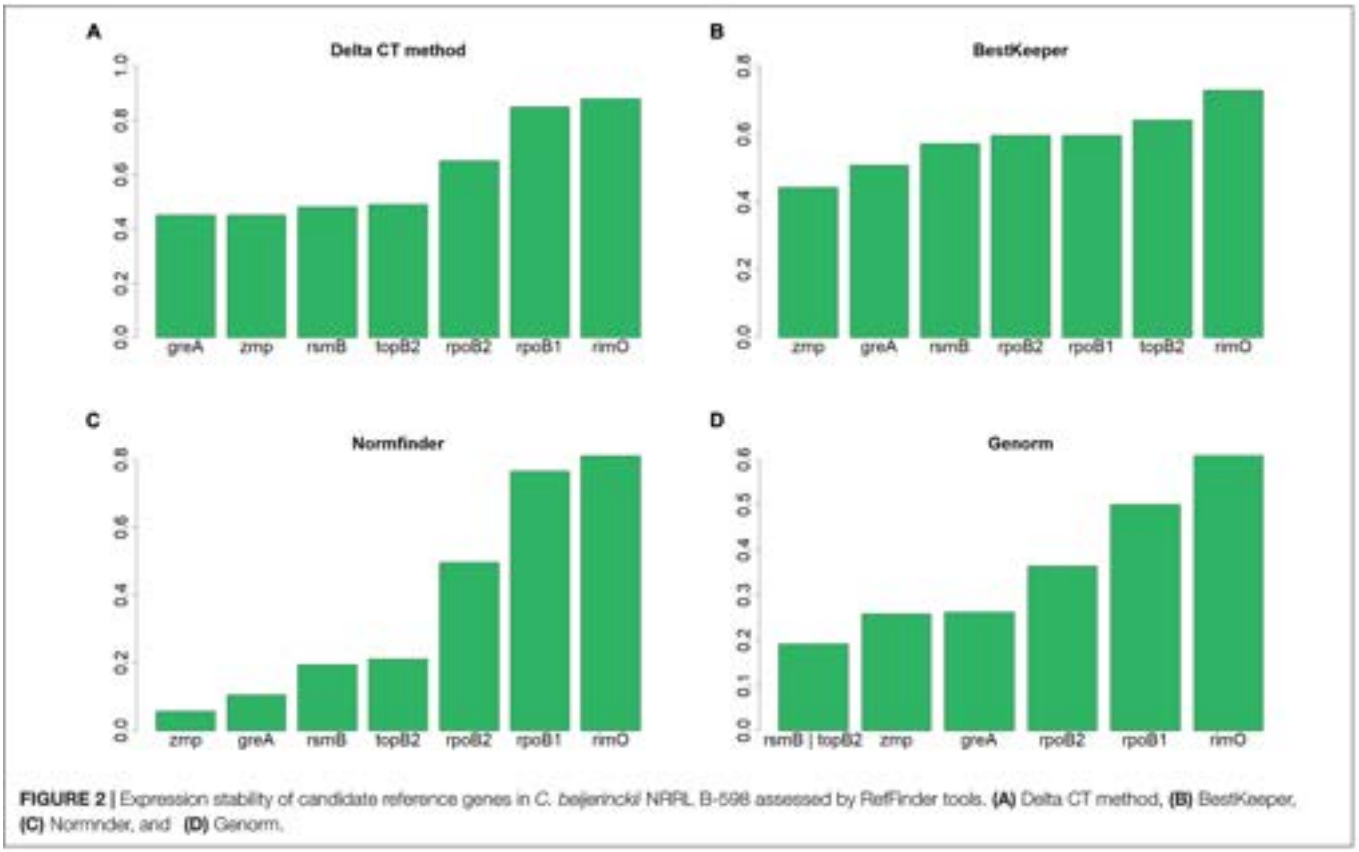


FIGURE 2 | Expression stability of candidate reference genes in *C. beijerinckii* NRRL B-598 assessed by RefFinder tools. **(A)** Delta CT method, **(B)** BestKeeper, **(C)** Normnder, and **(D)** Genorm.

stress in the case of *Staphylococcus aureus* (Kyuma et al., 2015). The *zmp* gene gured amongst the best candidate genes in *Vigna angularis* (Chi et al., 2016), the transcription elongation factor gene in *Vernicia fordii* Hemsl. (Han et al., 2012), and the 16S rRNA methyltransferase gene in *Rhodococcus opacus* (Delorenzo and Moon, 2018). Generally, the three most commonly tested and validated bacterial reference genes are 16S ribosomal RNA, DNA gyrase A, and recombinase A (Rocha et al., 2015). In our case, the genes of 16S rRNA and recombinase were not within the list of the 160 candidate genes after the bioinformatics analysis of the transcriptomic data (see **Supplementary Material 2**). This was expected for the 16S rRNA gene as wet-lab ribodepletion was performed before sequencing and the remaining contamination was removed *in silico* with SortMeRNA. The DNA topoisomerase (gyrase) subunit was within the list, though did not rank amongst the genes with the most stable expression, according to the RT-qPCR results. In *C. beijerinckii* NCIMB 8,052, gene encoding peptidase T was chosen as a reference gene (Wang et al., 2012). In *C. dicile* , the genes of 16S rRNA, adenylate kinase, and 30S ribosomal protein S10 displayed the most stable expression within eight tested candidate genes (Metcalf et al., 2010). Nevertheless, the ranking diered for three dierent *C. dicile* strains tested (Metcalf et al., 2010), conrming the need for a careful selection of reference genes for each species (Bustin et al., 2009). The 16S rRNA gene was also chosen as a reference in *C. botulinum* (Kirk et al., 2014). The most suitable genes for

normalization of gene expression in *C. ljungdahlii* were genes of gyrase subunit A, transcriptional termination factor, and formate-tetrahydrofolate ligase (Liu et al., 2013). Within the set of the tested candidate genes, the bacterial standards of 16S rRNA and recombinase A genes (Rocha et al., 2015) displayed the least stable expression in *C. ljungdahlii* (Liu et al., 2013).

The need for the validation of genes selected using only RNA-Seq data is supported by the comparison of the coecients of variation calculated from C_q (see **Table 7**) and the TPM values (see **Table 4**). Although genes *zmp* and *greA* maintain a high rating in both statistics, suggesting that rankings can in some cases correlate, results for *rsmB* and *rimO* genes tell otherwise. The *rsmB* gene was the most stable according to the RNA-Seq data, yet RT-qPCR experiments show that the other three respectively, four genes were more suitable as candidate genes (see **Table 7**). In the case of the *rimO* gene, which was the third most stable gene according to the TPM values, it dropped to the last place after experimental validation. Moreover, the average CV of C_q values was more than two times higher than the average CV of TPM values (2.14 × for uniquely mapped and 2.13 × for multi mapped reads) (see **Tables 4**, **7**), indicating that the RNA-Seq provides only less specic results selecting a wider range of genes on a given CV threshold value. These results suggest that the selection of reference genes cannot be performed by RNA-Seq data analysis alone, yet it can be used for the compilation of candidate genes list. However, validation by RT-qPCR experiments is always needed.

CONCLUSION

We identied and validated a novel set of reference genes of *C. beijerinckii* NRRL B-598. We selected 160 candidate reference genes based on analysis of all currently available RNA-Seq data for the strain covering several dierent experimental conditions. Selection of seven genes (*zmp*, *rpoB1*, *rsmB*, *greA*, *rpoB2*, *topB2*, and *rimO*) was summarized by GO enrichment analysis and further validated by RT-qPCR assays and statistical testing by four statistical algorithms (NormFinder, RefFinder, CV analysis, and Pairwise ΔCt method). The analysis ranked seven genes by their expression stability, presenting zinc metallopeptidase (*zmp*) and transcription elongation factor GreA (*greA*) as an appropriate set of reference genes regarding tested set of conditions. Our results can be helpful for selection of reference genes in other *C. beijerinckii* strains, and our methodology suggests that RNA-Seq data can be used for identication of novel reference genes, but their validation by RT-qPCR experiments is always needed.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: all computational analyses using sequencing data in this work are done against the reference sequence deposited at DDBJ/ENA/GenBank under the accession number CP011966.3. RNA-Seq data used in this study are available from NCBI SRA database under the accession numbers SRR6375604, SRR10556738SRR10556761 and SRR10569082SRR10569093.

AUTHOR CONTRIBUTIONS

KJ: conceptualization, methodology, formal analysis, data curation, visualization, and writing-original draft preparation. HS: conceptualization, methodology, validation, formal analysis, visualization, and writing-original draft preparation. JK: writing-original draft preparation. MV: validation and investigation. BB: conceptualization and investigation. PP and IP: supervision, resources, and funding. KS: conceptualization, methodology, formal analysis, and writing-original draft preparation. All authors contributed to the article and approved the submitted version.

FUNDING

This work has been supported by grant project GACR 17-00551S.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.640054/full#supplementary-material

REFERENCES

Agu, C. V., Ujor, V., and Ezeji, T. C. (2019). Metabolic engineering of *Clostridium beijerinckii* to improve glycerol metabolism and furfural tolerance. *Biotechnol. Biofuels* 12:50. doi: 10.1186/s13068-019-1388-9

Alexa, A., and Rahnenfurher, J. (2020). *topGO: Enrichment Analysis for Gene Ontology.*

Andersen, C. L., Jensen, J. L., and Ørntoft, T. F. (2004). Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Res.* 64, 52455250. doi: 10.1158/0008-5472.CAN-04-0496

Boda, E., Pini, A., Hoxha, E., Parolisi, R., and Tempia, F. (2009). Selection of reference genes for quantitative real-time RT-PCR studies in mouse brain. *J. Mol. Neurosci.* 37, 238253. doi: 10.1007/s12031-008-9128-9

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a exible trimmer for Illumina sequence data. *Bioinformatics* 30, 21142120. doi: 10.1093/bioinformatics/btu170

Breidenbach, M. A., and Brunger, A. T. (2004). Substrate recognition strategy for botulinum neurotoxin serotype A. *Nature* 432, 925929. doi: 10.1038/nature03123

Bustin, S. A., Benes, V., Garson, J. A., Hellemans, J., Huggett, J., Kubista, M., et al. (2009). The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin. Chem.* 55, 611622. doi: 10.1373/clinchem.2008.112797

Cafardi, V., Biagini, M., Martinelli, M., Leuzzi, R., Rubino, J. T., Cantini, F., et al. (2013). Identication of a novel zinc metalloprotease through a global analysis of *Clostridium dicile* extracellular proteins. *PLoS One* 8:e81306. doi: 10.1371/journal.pone.0081306

Canzar, S., Elbassioni, K., Jones, M., and Mestre, J. (2016). Resolving conicting predictions from multimapping reads. *J. Comput. Biol.* 23, 203217. doi: 10.1089/cmb.2015.0164

Carvalho, D. M., de Sá, P. H., Castro, T. L. P., Carvalho, R. D., Pinto, A., Gil, D. J. P., et al. (2014). Reference genes for RT-qPCR studies in *Corynebacterium pseudotuberculosis* identied through analysis of RNA-seq data. *Antonie van Leeuwenhoek* 106, 605614. doi: 10.1007/s10482-014-0231-3

Chapman, J. R., and Waldenström, J. (2015). With reference to reference genes: a systematic review of endogenous controls in gene expression studies. *PLoS One* 10:e0141853. doi: 10.1371/journal.pone.0141853

Chi, C., Shen, Y., Yin, L., Ke, X., Han, D., and Zuo, Y. (2016). Selection and validation of reference genes for gene expression analysis in *Vigna angularis* using quantitative real-time RT-PCR. *PLoS One* 11:e0168479. doi: 10.1371/journal.pone.0168479

Delorenzo, D. M., and Moon, T. S. (2018). Selection of stable reference genes for RT-qPCR in *Rhodococcus opacus* PD630. *Sci. Rep.* 8:6019. doi: 10.1038/s41598-018-24486-w

Derveaux, S., Vandesompele, J., and Hellemans, J. (2010). How to do successful gene expression analysis using real-time PCR. *Methods* 50, 227230. doi: 10.1016/j.ymeth.2009.11.001

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 1521. doi: 10.1093/bioinformatics/bts635

Ewels, P., Magnusson, M., Lundin, S., and Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 32, 30473048. doi: 10.1093/bioinformatics/btw354

Ginzinger, D. G. (2002). Gene quantication using real-time quantitative PCR. *Exp. Hematol.* 30, 503512. doi: 10.1016/S0301-472X(02)00806-8

Gomes, A. É., Stuchi, L. P., Siqueira, N. M. G., Henrique, J. B., Vicentini, R., Ribeiro, M. L., et al. (2018). Selection and validation of reference genes for gene expression studies in *Klebsiella pneumoniae* using reverse transcription quantitative real-time PCR. *Sci. Rep.* 8:9001. doi: 10.1038/s41598-018-27420-2

Han, X., Lu, M., Chen, Y., Zhan, Z., Cui, Q., and Wang, Y. (2012). Selection of reliable reference genes for gene expression studies using real-time PCR in tung tree during seed development. *PLoS One* 7:e43084. doi: 10.1371/journal.pone.0043084

Hu, Y., Xie, S., and Yao, J. (2016). Identication of novel reference genes suitable for qRT-PCR normalization with respect to the zebrash developmental stage. *PLoS One* 11:e0149277. doi: 10.1371/journal.pone.0149277

Jo, J., Choi, S., Oh, J., Lee, S.-G., Choi, S. Y., Kim, K. K., et al. (2019). Conventionally used reference genes are not outstanding for normalization of gene expression in human cancer research. *BMC Bioinformatics* 20(Suppl. 10):245. doi: 10.1186/s12859-019-2809-2

Kirk, D. G., Palonen, E., Korkeala, H., and Lindström, M. (2014). Evaluation of normalization reference genes for RT-qPCR analysis of *spo0A* and four sporulation sigma factor genes in *Clostridium botulinum* group I strain ATCC 3502. *Anaerobe* 26, 1419. doi: 10.1016/j.anaerobe.2013.12.003

Kopylova, E., Noé, L., and Touzet, H. (2012). SortMeRNA: fast and accurate ltering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* 28, 32113217. doi: 10.1093/bioinformatics/bts611

Kyuma, T., Kimura, S., Hanada, Y., Suzuki, T., Sekimizu, K., and Kaito, C. (2015). Ribosomal RNA methyltransferases contribute to *Staphylococcus aureus* virulence. *FEBS J.* 282, 25702584. doi: 10.1111/febs.13302

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078 2079. doi: 10.1093/bioinformatics/btp352

Liang, P., Guo, Y., Zhou, X., and Gao, X. (2014). Expression proling in *Bemisia tabaci* under insecticide treatment: indicating the necessity for custom reference gene selection. *PLoS One* 9:e87514. doi: 10.1371/journal.pone.0087514

Liao, Y., Smyth, G. K., and Shi, W. (2014). FeatureCounts: an ecient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923930. doi: 10.1093/bioinformatics/btt656

Liao, Y., Smyth, G. K., and Shi, W. (2019). The R package Rsubread is easier, faster, cheaper and better for alignment and quantication of RNA sequencing reads. *Nucleic Acids Res.* 47:e47. doi: 10.1093/nar/gkz114

Lioy, V. S., Goussard, S., Guerineau, V., Yoon, E. J., Courvalin, P., Galimand, M., et al. (2014). Aminoglycoside resistance 16S rRNA methyltransferases block endogenous methylation, aect translation eciency and tness of the host. *RNA* 20, 382391. doi: 10.1261/rna.042572.113

Liu, D., Hao, K., Wang, W., Peng, C., Dai, Y., Jin, R., et al. (2017). *Rv2629* overexpression delays *Mycobacterium smegmatis* and *Mycobacteria tuberculosis* entry into log-phase and increases pathogenicity of *Mycobacterium smegmatis* in mice. *Front. Microbiol.* 8:2231. doi: 10.3389/fmicb.2017.02231

Liu, J., Tan, Y., Yang, X., Chen, X., and Li, F. (2013). Evaluation of *Clostridium ljungdahlii* DSM 13528 reference genes in gene expression studies by qRT-PCR. *J. Biosci. Bioeng.* 116, 460464. doi: 10.1016/j.jbiosc.2013.04.011

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15:550. doi: 10.1186/s13059-014-0550-8

Lupien, A., Gingras, H., Leprohon, P., and Ouellette, M. (2015). Induced tigecycline resistance in *Streptococcus pneumoniae* mutants reveals mutations in ribosomal proteins and rRNA. *J. Antimicrob. Chemother.* 70, 29732980. doi: 10.1093/jac/dkv211

Marr, M. T., and Roberts, J. W. (2000). Function of transcription cleavage factors GreA and GreB at a regulatory pause site. *Mol. Cell* 6, 12751285. doi: 10.1016/S1097-2765(00)00126-X

Metcalf, D., Sharif, S., and Weese, J. S. (2010). Evaluation of candidate reference genes in *Clostridium dicile* for gene expression normalization. *Anaerobe* 16, 439443. doi: 10.1016/j.anaerobe.2010.06.007

Patakova, P., Branska, B., Sedlar, K., Vasylkivska, M., Jureckova, K., Kolek, J., et al. (2019). Acidogenesis, solventogenesis, metabolic stress response and life cycle changes in *Clostridium beijerinckii* NRRL B-598 at the transcriptomic level. *Sci. Rep.* 9:1371. doi: 10.1038/s41598-018-37679-0

Patakova, P., Kolek, J., Jureckova, K., Branska, B., Sedlar, K., Vasylkivska, M., et al. (2020). Deeper below the surfacetranscriptional changes in selected genes of *Clostridium beijerinckii* in response to butanol shock. *Microbiologyopen* 10:e1146. doi: 10.1002/mbo3.1146

Peng, S., Stephan, R., Hummerjohann, J., and Tasara, T. (2014). Evaluation of three reference genes of *Escherichia coli* for mRNA expression level normalization in view of salt and organic acid stress exposure in food. *FEMS Microbiol. Lett.* 355, 7882. doi: 10.1111/1574-6968.12447

Pombo, M. A., Zheng, Y., Fei, Z., Martin, G. B., and Rosli, H. G. (2017). Use of RNA-seq data to identify and validate RT-qPCR reference genes for studying the tomato-*Pseudomonas* pathosystem. *Sci. Rep.* 7:44905. doi: 10.1038/srep44905

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* 41, 590596. doi: 10.1093/nar/gks1219

Roberts, A., Schaeer, L., and Pachter, L. (2013). Updating RNA-Seq analyses after re-annotation. *Bioinformatics* 29, 16311637. doi: 10.1093/bioinformatics/btt197

Rocha, D. J. P., Santos, C. S., and Pacheco, L. G. C. (2015). Bacterial reference genes for gene expression studies by RT-qPCR: survey and analysis. *Antonie van Leeuwenhoek* 108, 685693. doi: 10.1007/s10482-015-0524-1

Sedlar, K., Kolek, J., Gruber, M., Jureckova, K., Branska, B., Csaba, G., et al. (2019). A transcriptional response of *Clostridium beijerinckii* NRRL B-598 to a butanol shock. *Biotechnol. Biofuels* 12, 116. doi: 10.1186/s13068-019-1584-7

Sedlar, K., Kolek, J., Provaznik, I., and Patakova, P. (2017). Reclassication of non-type strain *Clostridium pasteurianum* NRRL B-598 as Clostridium beijerinckii NRRL B-598. *J. Biotechnol.* 244, 13. doi: 10.1016/j.jbiotec.2017.01.003

Sedlar, K., Koscova, P., Vasylkivska, M., Branska, B., Kolek, J., Kupkova, K., et al. (2018). Transcription proling of butanol producer *Clostridium beijerinckii* NRRL B-598 using RNA-Seq. *BMC Genomics* 19:415. doi: 10.1186/s12864-018-4805-8

Silver, N., Best, S., Jiang, J., and Thein, S. L. (2006). Selection of housekeeping genes for gene expression studies in human reticulocytes using real-time PCR. *BMC Mol. Biol.* 7:33. doi: 10.1186/1471-2199-7-33

Song, Y., Wang, Y., Guo, D., and Jing, L. (2019). Selection of reference genes for quantitative real-time PCR normalization in the plant pathogen *Puccinia helianthi* Schw. *BMC Plant Biol.* 19:20. doi: 10.1186/s12870-019-1629-x

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., et al. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* 3:research0034. doi: 10.1186/gb-2002-3-7-research0034

Vasylkivska, M., Jureckova, K., Branska, B., Sedlar, K., Kolek, J., Provaznik, I., et al. (2019). Transcriptional analysis of amino acid, metal ion, vitamin and

carbohydrate uptake in butanol-producing *Clostridium beijerinckii* NRRL B-598. *PLoS One* 14:e0224560. doi: 10.1371/journal.pone.0224560

Wagner, G. P., Kin, K., and Lynch, V. J. (2012). Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory Biosci.* 131, 281285. doi: 10.1007/s12064-012-0162-3

Wang, C., Ye, F., Kumar, V., Gao, Y. G., and Zhang, L. H. (2014). BswR controls bacterial motility and biolm formation in *Pseudomonas aeruginosa* through modulation of the small RNA rsmZ. *Nucleic Acids Res.* 42, 45634576. doi: 10.1093/nar/gku106

Wang, Y., Li, X., Mao, Y., and Blaschek, H. P. (2011). Single-nucleotide resolution analysis of the transcriptome structure of *Clostridium beijerinckii* NCIMB 8052 using RNA-Seq. *BMC Genomics* 12:479. doi: 10.1186/1471-2164-12-479

Wang, Y., Li, X., Mao, Y., and Blaschek, H. P. (2012). Genome-wide dynamic transcriptional proling in *Clostridium beijerinckii* NCIMB 8052 using single-nucleotide resolution RNA-Seq. *BMC Genomics* 13:102. doi: 10.1186/1471-2164-13-102

Wang, Y., Zhang, Z. T., Seo, S. O., Lynn, P., Lu, T., Jin, Y. S., et al. (2016). Bacterial genome editing with CRISPR-Cas9: deletion, integration, single nucleotide modication, and desirable clean mutant selection in *Clostridium beijerinckii* as an example. *ACS Synth. Biol.* 5, 721732. doi: 10.1021/acssynbio.6b00060

Xie, F., Xiao, P., Chen, D., Xu, L., and Zhang, B. (2012). miRDeepFinder: a miRNA analysis tool for deep sequencing of plant small RNAs. *Plant Mol. Biol.* doi: 10.1007/s11103-012-9885-2 **Q.

Xin, X., Cheng, C., Du, G., Chen, L., and Xue, C. (2020). Metabolic engineering of histidine kinases in *Clostridium beijerinckii* for enhanced butanol production. *Front. Bioeng. Biotechnol.* 8:214. doi: 10.3389/fbioe.2020.00214

Zytnicki, M. (2017). mmquant: how to count multi-mapping reads? *BMC Bioinformatics* 18:411. doi: 10.1186/s12859-017-1816-4

Conict of Interest: The authors declare that the research was conducted in the absence of any commercial or nancial relationships that could be construed as a potential conict of interest.

5.12 Article XII

SEDLAR, Karel, NYKRYNOVA, Marketa, BEZDICEK, Matej, BRANSKA, Barbora, LENGEROVA, Martina, PATAKOVA, Petra, and SKUTKOVA, Helena. Diversity and Evolution of Clostridium beijerinckii and Complete Genome of the Type Strain DSM 791T. *Processes*. 2021. 9(7), p. 1196. (IF = 3.352, Q2 in ENGINEERING, CHEMICAL)

DOI: https://doi.org/10.3390/pr9071196

processes

MDPI

Article

Diversity and Evolution of *Clostridium beijerinckii* and Complete Genome of the Type Strain DSM 791^T

Karel Sedlar [1,*], Marketa Nykrynova [1], Matej Bezdicek [2,3], Barbora Branska [4], Martina Lengerova [2], Petra Patakova [4] and Helena Skutkova [1]

1 Department of Biomedical Engineering, Faculty of Electrical Engineering and Communication, Brno University of Technology, Technicka 10, 616 00 Brno, Czech Republic; nykrynova@vut.cz (M.N.); skutkova@feec.vutbr.cz (H.S.)
2 Department of Internal Medicine–Hematology and Oncology, University Hospital Brno, Cernopolni 9, 613 00 Brno, Czech Republic; bezdicek.matej@fnbrno.cz (M.B.); Lengerova.Martina@fnbrno.cz (M.L.)
3 Department of Internal Medicine–Hematology and Oncology, Masaryk University, Jihlavska 20, 625 00 Brno, Czech Republic
4 Department of Biotechnology, University of Chemistry and Technology Prague, Technicka 5, 166 28 Prague, Czech Republic; barbora.branska@vscht.cz (B.B.); petra.patakova@vscht.cz (P.P.)
* Correspondence: sedlar@vut.cz

Abstract: *Clostridium beijerinckii* is a relatively widely studied, yet non-model, bacterium. While 246 genome assemblies of its various strains are available currently, the diversity of the whole species has not been studied, and it has only been analyzed in part for a missing genome of the type strain. Here, we sequenced and assembled the complete genome of the type strain *Clostridium beijerinckii* DSM 791^T, composed of a circular chromosome and a circular megaplasmid, and used it for a comparison with other genomes to evaluate diversity and capture the evolution of the whole species. We found that strains WB53 and HUN142 were misidentified and did not belong to the *Clostridium beijerinckii* species. Additionally, we filtered possibly misassembled genomes, and we used the remaining 237 high-quality genomes to define the pangenome of the whole species. By its functional annotation, we showed that the core genome contains genes responsible for basic metabolism, while the accessory genome has genes affecting final phenotype that may vary among different strains. We used the core genome to reconstruct the phylogeny of the species and showed its great diversity, which complicates the identification of particular strains, yet hides possibilities to reveal hitherto unreported phenotypic features and processes utilizable in biotechnology.

Keywords: butanol; ABE; IBE; core genome; accessory genome; pan genome

Citation: Sedlar, K.; Nykrynova, M.; Bezdicek, M.; Branska, B.; Lengerova, M.; Patakova, P.; Skutkova, H. Diversity and Evolution of *Clostridium beijerinckii* and Complete Genome of the Type Strain DSM 791^T. *Processes* **2021**, *9*, 1196. https://doi.org/10.3390/pr9071196

Academic Editor: Hoon Kim

Received: 4 June 2021
Accepted: 7 July 2021
Published: 10 July 2021

1. Introduction

Clostridium beijerinckii belongs to the group of authentic *Clostridium* spp., also referred to as Cluster I "Sensu stricto" [1]. As a Gram-positive, spore forming, rod shaped anaerobe capable of solventogenesis, *C. beijerinckii* represents an industrially relevant microorganism. From that perspective, butanol seems to be the main subject of research interest. Butanol is produced within either an acetone-butanol-ethanol (ABE) [2] or isopropanol-butanol-ethanol (IBE) [3] fermentation pathway that covers a bi-phasic process in which acid formation is later followed by the formation of solvents. Moreover, solventogenesis is usually coupled with sporulation. Nevertheless, both processes do not seem to be closely linked in all strains [4]. Thus, evolutionary and comparative studies of various clostridial strains are required to help reveal hidden aspects of the production of valuable chemicals by microbial cell factories [5].

Although the evolution and taxonomy of the clostridia were revisited several times [1,6–8], additional studies are needed as the clostridia still represent a polyphyletic group with uncertain phylogenetic affinities and reidentifications of particular strains and

reclassifications of various clostridial species are quite common [9–11]. The most recent change affecting *C. beijerinckii* is the reclassification of *C. diolis* as *C. beijerinckii* [12]. These taxonomic readjustments were expected as the previous taxonomy was built upon phenotypic differences that do not necessarily reflect genetic heterogeneity [13]. A massive reduction in DNA sequencing costs over the past 20 years resulted in many genomes of non-model bacteria being sequenced. This applies also to *C. beijerinckii* strains, and, currently, there are 241 genome assemblies of various *C. beijerinckii* strains and five genome assemblies of three different *C. diolis* strains in the GenBank database (May 2021). Unfortunately, most genomes are assembled only in the form of draft assemblies, thus leading to gaps in knowledge. Although draft assemblies of type strains *C. beijerinckii* DSM 791^T and *C. diolis* DSM 15410^T were sufficient to reveal that *C. bejerinckii* and *C. diolis* are heterotypic synonyms [12], it was the complete genome assembly that helped to reveal hitherto unreported features of *C. diolis* DSM 15410^T. A thorough analysis proved its ability to produce isopropanol [14] that was hidden for 18 years since the description of the species in 2002 [15]. There are also other strains of *C. beijerinckii* that are studied without the knowledge of a genome sequence; for example, the strain *C. beijerinckii* F-6, a butanol-tolerant hydrogen producer for which only a 16S RNA gene sequence is known [16,17].

Following reclassification there are now two type strains: *C. beijerinckii* DSM 791^T (=ATCC 25752, E. McCoy A-67, L.S. McClung 1671, NCIMB 9362) and *C. beijerinckii*, formerly *C. diolis*, DSM 15410^T (=DSM 5431, SH1, 88-273, ATCC BAA-557). Although the reclassification means the loss of type strain designation, this has not been universally agreed by various sources. Thus, the strain *C. beijerinckii* DSM15410^T can be found as the type strain in the German Collection of Microorganisms and Cell Cultures (DSZM) catalogue and must be searched as *C. diolis* DSM 15410 in the GenBank database. Here, we refer to all *C. beijerinckii/diolis* strains as *C. beijerinckii* strains for two reasons. First, the species name *C. beijerinckii* was proposed several decades before *C. diolis* [15,18]. Second, the number of assemblies is considerably higher for the *C. beijerickii* species, and it contains several well-studied strains, e.g., *C. beijerinckii* NCIMB 8052 [19–22] (formerly *C. acetobutylicum* [23]), *C. beijerinckii* NRRL B-598 [24–27] (formerly *C. pasteurianum* [9]), and *C. beijerinckii* DSM 6423 [28,29]. Despite that, only a draft genome sequence of the type strain *C. beijerinckii* DSM 791^T has been available until now. In this paper, we sequenced and assembled the complete genome of *C. beijerinckii* DSM 791^T that contains a chromosome and a single megaplasmid. The presence of the plasmid has never been reported before as the type strain has not been studied in detail. Additionally, we performed its annotation, and included the classification of protein coding sequences (CDSs) into clusters of orthologous groups (COG), a prediction of the operon structure, and the identification of prophage DNA and CRISPR arrays. Eventually, we compared its main genomic features to the other type strain (DSM 15410^T) and performed an extensive comparative study of all currently available *C. beijerinckii* genomes to define its pangenome. Although some reports comparing genomes of various *C. beijerinckii* strains have been published [9,12,14,17,30], they are limited to comparing only a few genomes simultaneously, and, sometimes, they have used only 16S rRNA gene sequences. This is the first report to show the diversity of the *C. beijerinckii* species and the first study to use the maximum information available, thanks to the definition of the core genome from 237 genomes.

2. Materials and Methods

2.1. Bacterial Strain

The strain *Clostridium beijerinckii* DSM 791^T was obtained from the German public collection of microorganisms at the Leibnitz Institute DSMZ-German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany.

2.2. Cultivation

The strain inoculum was prepared from a cryopreserved culture stored at −80 °C in 30% glycerol. Cells were cultivated at 37 °C in liquid tryptone-yeast extract-acetate (TYA)

medium overnight and subsequently transferred on solidified TYA medium with agar. The selected colony was transferred into a liquid medium and cultivated for 24 h. Cells from 10 mL of suspension were pelleted by centrifugation (6000× *g*), washed with sterile demi water, pelleted again, and stored at −80 °C prior to DNA isolation. Up to the final centrifugation and wash step, all operations were performed under a nitrogen atmosphere in an anaerobic chamber Concept 400 (Ruskinn). The composition of TYA media was as follows: 20 g·L^{-1} glucose; 2 g·L^{-1} yeast extract; 6 g·L^{-1} tryptone; 0.5 g·L^{-1} KH_2PO_4; 3 g·L^{-1} ammonium acetate; 0.3 g·L^{-1} $MgSO_4.7H_2O$; 0.01 g·L^{-1} $FeSO_4$; (agar 20 g·L^{-1}); pH was adjusted to 6.8 prior to sterilization (121 °C, 20 min).

2.3. DNA Extraction and Sequencing

For long-read sequencing, genomic high molecular weight DNA was extracted using the MagAttract HMW DNAKit (Qiagene, Venlo. NL). The extracted DNA purity and proper length were checked using the NanoDrop (Thermo Fisher Scientific, Waltham, MA, USA) and Agilent 4200 TapeStation (Agilent technologies, Santa Clara, CA, USA), respectively. Library preparation for Oxford Nanopore sequencing was performed using the Ligation sequencing 1D Kit (Oxford Nanopore Technologies, Oxford, UK). The library was sequenced using the R9.4.1 flowcell and the MinION platform (Oxford Nanopore Technologies).

For short-read sequencing, genomic DNA was purified using the GenElute Bacterial Genomic DNA Kit (SIGMA-ALDRICH, St. Louis, MI, USA). The extracted DNA purity was checked by NanoDrop (Thermo Fisher Scientific). The sequencing library was prepared using the KAPA HyperPlus kit and was carried out using the Miseq Reagent Kit v2 (500 cycles) and the Illumina MiSeq platform (Illumina, San Diego, CA, USA).

2.4. Genome Assembly

Basecalling of raw nanopore squiggles was performed with Guppy v3.4.4 and the quality of reads was checked with MinIONQC [31]. Similarly, the initial quality assessment of Illumina raw reads was conducted through a combination of FastQC v0.11.5 and MultiQC v1.7 [32]. The adapter and quality trimming was performed using Trimmomatic v1.36 [33]. Initial genome assemblies were constructed using long reads with a Flye v2.8.1 assembler [34] and short reads with a plasmidSPAdes v3.11.1 assembler [35]. Contigs from both assemblies were compared with NUCmer v3.1 [36] and selected contigs were further polished. The first step of polishing was done by four rounds of mapping long reads with minimap2 v2.17 [37] and polishing with racon v1.4.20 [38]. The second step consisted of two rounds of polishing with medaka v1.2.5, again in combination with long reads. Finally, the third step of polishing was done by two rounds of mapping short reads with BWA v0.7.17 [39] and polishing with pilon v1.24 [40], while handling files of mapped reads with SAMtools v1.7 [41]. Resulting contigs were manually examined for circularity by concatenating their ends and mapping short and long reads with BWA and minimap2, respectively. Missing or duplicated bases were manually added or trimmed. Eventually, replication origins of contigs were predicted. Chromosomal replication origin (*oriC*) was predicted with Ori-finder [42] and the sequence was rearranged according to its position, so the *dnaA* gene is the first gene in the chromosomal sequence. Similarly, replication origin in plasmid (*oriV*) was predicted using BLAST [43] searches against the database of replication origins DoriC [44] and the sequence was rearranged according to its position, so the *repB* gene is the first gene in the plasmid sequence.

2.5. Genome Annotation and Analysis

A genome annotation was added by the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) [45]. An operon prediction was completed using the Operon-mapper [46]. The functional annotation of the protein coding genes was performed by assigning clusters of orthologous group (COG) categories from the eggNOG database with the eggNOG-mapper [47]. Circular genome maps of a chromosome and a megaplasmid were prepared

with a DNAplotter [48] integrated in Artemis [49]. The genome was searched for prophage DNA with PhiSpy v4.2.12 [50] and for clustered regularly interspaced short palindromic repeats (CRISPR) arrays with CRISPRDetect v2.3 [51].

Entrez was used to search the GenBank database for *C. beijerinckii/diolis* genomes [52]. A circular graph showing whole-genome alignments was produced with BRIG v0.95 [53]. Nucleotide sequences of reference genes from the strain *C. beijerinckii* DSM 791^T were localized in other strains using BLAST [43] searches and a 90% sequence similarity in Ridom SeqSphere+ v7.6.1. Outliers were detected with R package graphics [54] using the boxplot function and setting a whiskers range as 10 times the interquartile range. Digital DNA to DNA hybridization (dDDH) values were calculated using the type strain genome server (TYGS) [55]. The core and accessory genomes and unique genes were identified with BPGA v1.3.0 [56] using amino acid sequences and a 90% sequence similarity. All sequences of unique genes and reference sequences of pangenome genes were uploaded to FAIRDOMHub [57]. The phylogenetic tree was reconstructed using the concatenated core genome sequences with the neighbor-joining method in BPGA. A reduction of leaf nodes in the tree was done by collapsing branches where the whole length was shorter than 1% of the longest branch in the tree. A final visualization of trees was conducted through Evolview v3 [58].

3. Results

3.1. Genome Sequencing and Assembly

Oxford Nanopore sequencing produced 690,277 reads with a N50 length of 21,578 bp. Moreover, 700 reads exceeded the length of 100 kbp, while five were even longer than 200 kbp. The Illumina sequencing produced more than 2.1 million additional paired reads of 250 bp in length. The final genome assembly consisted of two circular contigs. While the first one represented a circular chromosome of length 5,876,902 bp, the second corresponded to a 73,345 bp long megaplasmid. Both sequences have been deposited at the DDBJ/EMBL/GenBank under accession numbers CP073653 for the chromosome and CP073654 for the megaplasmid. Coverage of the assembly after the filtering steps was approximately 1063× and the assembly was reconstructed with the contribution of more than 2.1 million paired Illumina reads (more than 99% of all Illumina reads and more than 99% of all Illumina sequenced bases after quality trimming) and more than 640,000 of Oxford Nanopore reads (almost 90% of all Oxford Nanopore reads and more than 93% of all Oxford Nanopore sequenced bases). Average coverage, considering only short reads, was 88× for chromosome and 192× for plasmid.

3.2. The Characteristics of the C. beijerinckii DSM 791 Genome

The guanine–cytosine (GC) content of the genome was calculated as 29.87%. While the GC content of the chromosome was almost 29.90%, the GC content of the plasmid was slightly lower, reaching only 27.84%, see Table 1. The complete genome contained 5279 annotated open reading frames (ORFs) divided into 3291 operons, see Table 1 for separate statistics for chromosome and plasmid. The majority of ORFs consisted of protein coding genes, but 134 pseudogenes were also found. The sequences of 60 pseudogenes were found to be incomplete, 56 were frameshifted, 34 contained internal stops, and 13 suffered from multiple problems. The positions of particular features within the chromosome and plasmid are shown in Figure 1. Additionally, protein coding genes and pseudogenes were assigned COG categories. Unfortunately, 687 CDSs were not assigned any COG and an additional 990 CDSs were assigned to group S with an unknown function. Nevertheless, the remaining 3454 CDSs (out of all 5131 protein coding genes and pseudogenes) were divided into the remaining COG categories, see Supplementary Table S1.

Table 1. Genome features of *Clostridium beijerinckii* DSM 791^T.

Feature	Chromosome	Plasmid
Length (bp)	5,876,902	73,345
GC content (%)	29.90	27.84
Total number of ORF	5209	70
Total number of operons	3237	54
Protein coding genes	4929	68
Pseudogenes	132	2
rRNA genes (5S, 16S, 23S)	17, 16, 16	0, 0, 0
tRNA	93	0
ncRNA	6	0

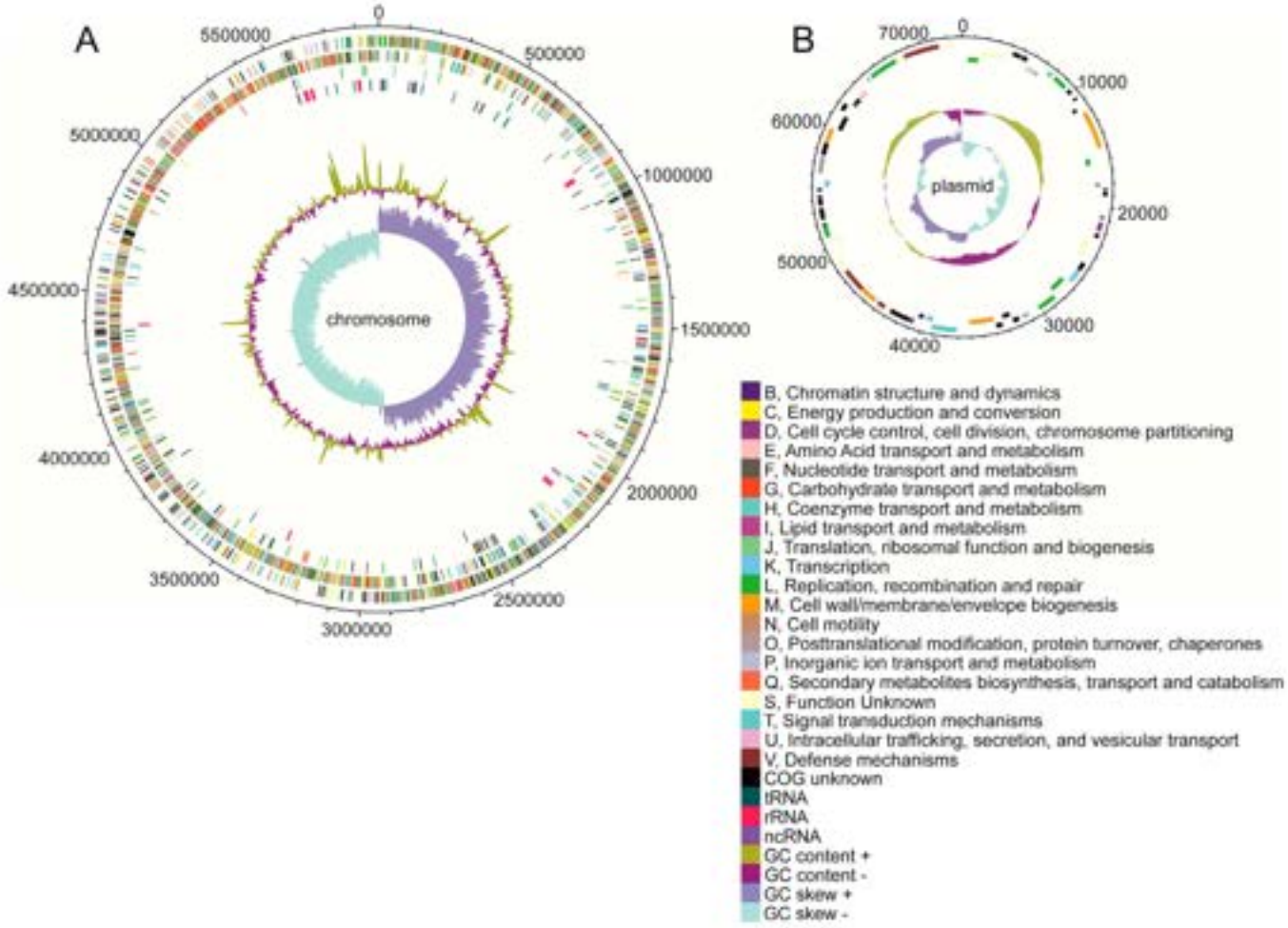


Figure 1. Circular maps of *C. beijerinckii* DSM 791 (**A**) chromosome and (**B**) plasmid. The outermost and the second outermost circles represent CDSs on the forward and reverse strands, respectively. The third circle represents pseudogenes, and the colors represent the COG functional classification. The fourth circle represents RNA genes, while the colors distinguish between tRNA, rRNA, and ncRNA. The inner shaded area represents (from outside in) GC content and GC skew, plotted using a 10-kb window with a step of 200 bp while the colors distinguish between above and below average values.

ORFs were searched for prophage genes, which resulted in 10 identified prophages of lengths ranging from 6583 bp to 42,566 bp, see Supplementary Table S2. All prophages were located on the chromosome. While the lowest number of genes in a prophage was eight, two prophages consisted of 51 genes. The cumulative length of prophages was 246,958 bp, which is less than 4.2% of the genome.

Only two CRISPR arrays were found (see Supplementary Table S3). Both arrays were extremely short, only 172 bp and 153 bp long, with two spacer units. Moreover, no *cas* or *cas*-like genes were found in their neighborhoods.

3.3. Diversity of C. beijerinckii Strains

A search for *C. beijerijcki/diolis* genomes in the GenBank database obtained 246 genome assemblies. After deduplication of multiple assemblies for the same strains, 242 assemblies were preserved, from which 11 represented complete genomes or complete chromosomal sequences, see Supplementary Table S4.

A comparison of complete chromosomal sequences to the reference, *C. beijerinckii* DSM 791^T chromosome, is shown in Figure 2. While the majority of genomes mapped to the reference with almost 100% identity in whole length, there were two genomes with lower similarities. While some parts of the DSM 6423 genome mapped with 95% identities, the majority of the WB53 genome mapped with considerably lower identities, not exceeding 90%. Those results were further supported by dDDH analysis that showed dDDH values between particular strains and the reference ranging from 71.9% to 86.4%, except for strains DSM 6423 and WB53 where the dDDH value to the strain DSM 791^T was 67.4% and 20.3%, see Supplementary Table S5.

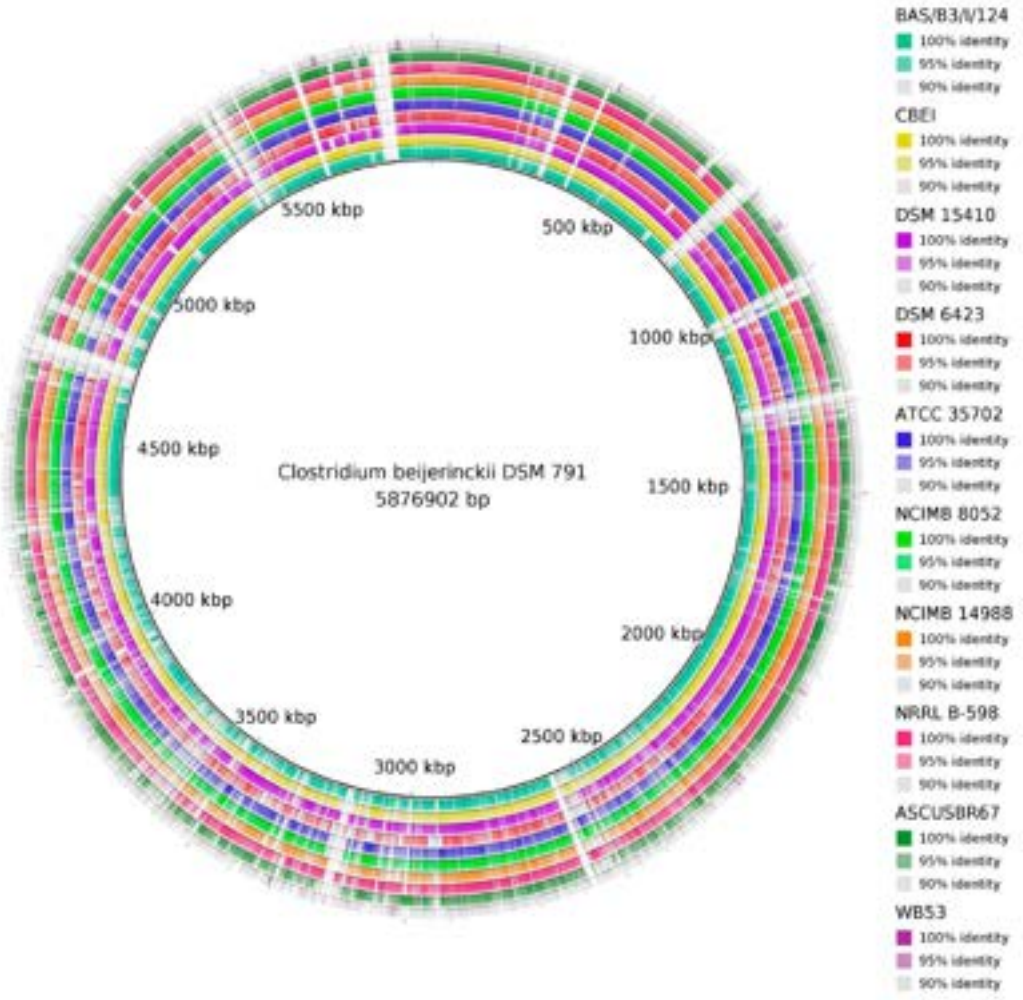


Figure 2. Whole-genome comparison of the type strain *C. beijerinckii* DSM 791^T to other *C. beijerinckii* complete genomes with the visualization of genomes percentage identity.

A whole-genome comparison was further applied to the whole dataset of 242 genomes. Genes of the reference strain *C. beijerinckii* DSM 791^T were searched in other genomes,

using nucleotide BLAST and a 90% sequence similarity, to find missing genes or genes with missing stop codon, see Supplementary Table S6. The median value of missing or corrupted genes was 993 and seven outliers were detected. While three outliers exceeded the upper fence, four outliers lay below the lower fence. Those exceeding the limit were genomes of strains ASCUSBR67, HUN142, and WB53, with 2599, 1811, and 4778 missing or corrupted genes, respectively. The dDDH value between the type strain and the strain HUN142 was 48.5%. Outliers below the lower fence were genomes of strains DJ079, DJ317, NBRC 109359, and NCTC13035, with 116, 96, 125, and 144 missing or corrupted genes, respectively.

3.4. Pangenome of C. beijerinckii

After removing unannotated genomes and genomes of strains that do not belong to the *C. beijerinckii* species, the remaining 237 genomes were used to define the core genome. In total, 2308 genes were present in all genomes, and formed the core genome. Additionally, 12,202 genes were found in at least two genomes and formed the accessory genome. Together, they presented the pangenome of *C. beijerinckii* containing 14,510 genes. Moreover, 5929 genes were unique, i.e., they were found in only one genome. The number of unique genes for particular strains ranged from zero to 516 (found in the strain DJ015), see Supplementary Table S7. While the median value of unique genes was three, type strains DSM 791^T and DSM 15410^T had one and 148 unique genes. In addition, genes exclusively absent were also counted. Such genes were found in every strain except for one. The number of exclusively missing genes ranged from zero to 86 (missing in the strain DJ032). The median value was zero and type strains DSM 791^T and DSM 15410^T were missing six genes and one gene, respectively.

A functional annotation of genes showed a different composition of the core genome and the accessory genome (see Figure 3). The core genome contained a larger proportion of genes connected to metabolism and energy production and conversion, except for genes in group (Q) "Secondary metabolites biosynthesis, transport and catabolism", where the accessory genome and unique genes had higher relative abundances. Similarly, genes connected to repair and defense mechanisms were more abundant in the accessory genome and among unique genes. While only 6.28% of genes were not assigned any COG in the core genome, their abundance in the accessory genome and among unique genes exceeded 35%, see Supplementary Table S8.

Amino acid sequences of unique genes for particular *Clostridium beijerinckii* strains as well as reference sequences of genes present in the *Clostridium beijerinckii* pangenome were uploaded to the FAIRDOMHub under the project "Clostridium beijerinckii pan-genome" https://fairdomhub.org/projects/242 (accessed on 1 June 2021).

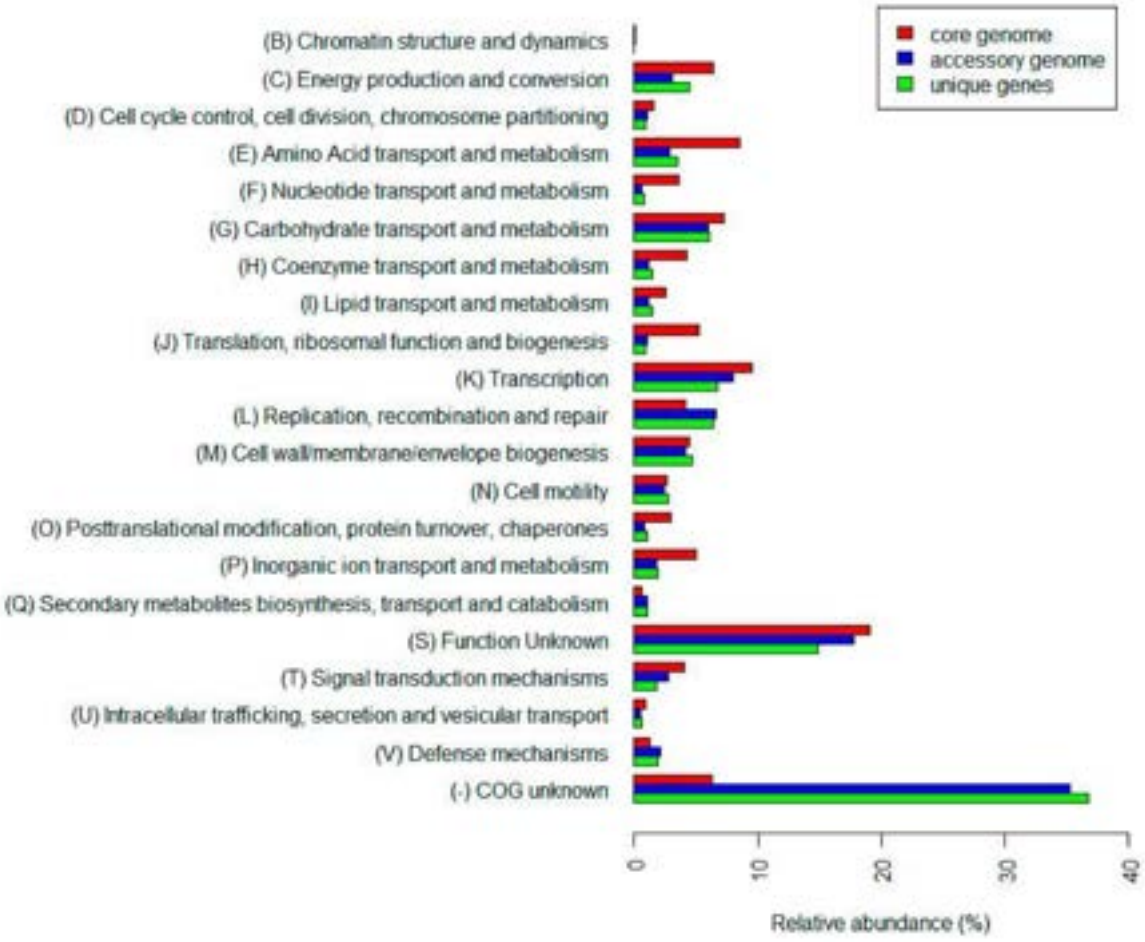


Figure 3. Relative abundances of genes in particular clusters of orthologous groups. Color coding distinguishes between gene from core (red) and accessory (blue) genomes and genes unique for particular strains (green).

3.5. *Phylogeny*

Finally, the phylogenetic tree of *C. beijerinckii* strains was reconstructed using the core genome (see Figure 4). Evolutionary closely related strains were collapsed into 16 clusters. While most clusters contained only units of strains, cluster 1 covered 145 strains, see Supplementary Table S9. The complete tree is showed in Supplementary Figure S2. While the type strain DSM 791^T had four closely related strains and formed a cluster, the other type strain DSM 15410^T formed an individual leaf node for which the strain NRRL B-598 was the closest strain.

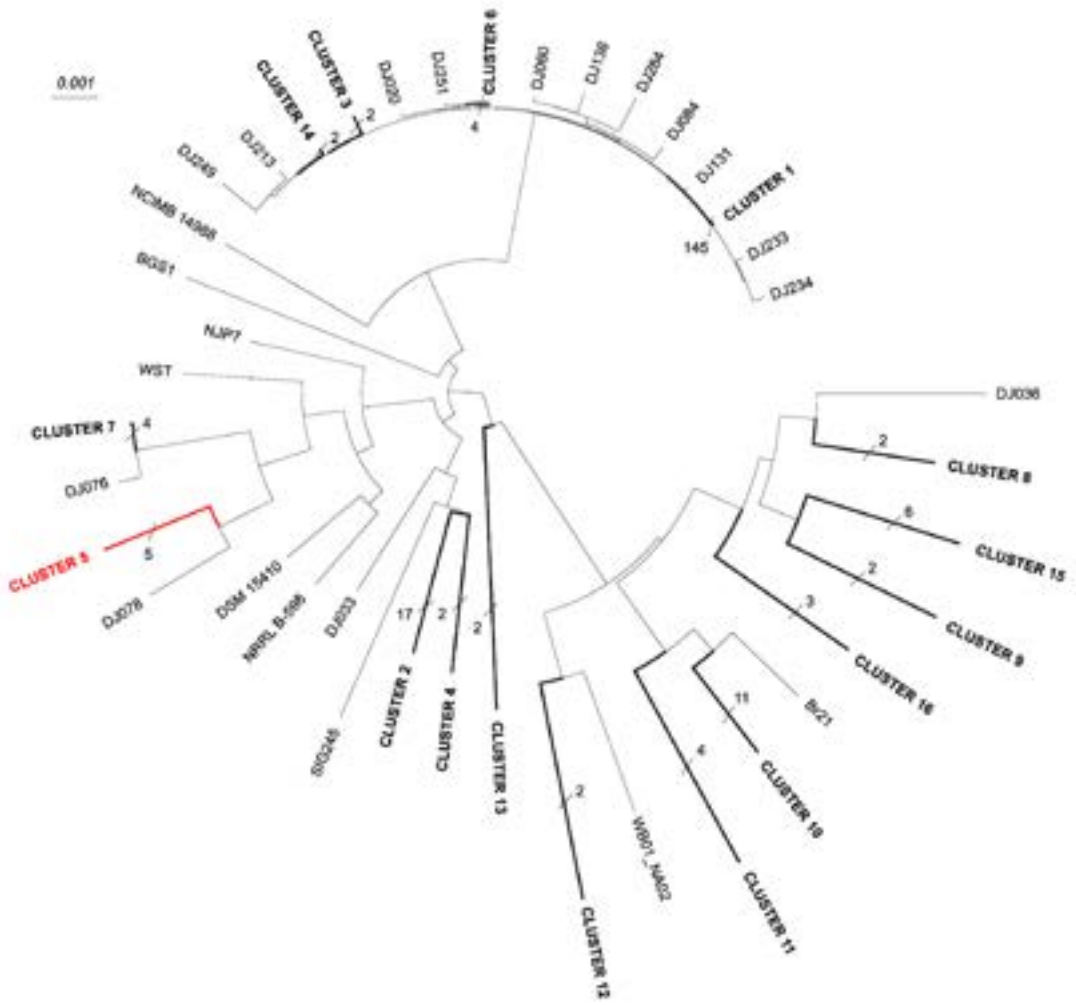


Figure 4. Phylogenetic reconstruction of *Clostridium beijerinckii* species. The phylogeny was reconstructed using 2308 genes of the core genome, present in every strain. Branches that were shorter than 1% of the maximum branch length were collapsed. The number of strains hidden under collapsed branches is written to each collapsed branch and strains are listed in Supplementary Table S9. The branch containing the type strain *C. beijerinckii* DSM 791^T is highlighted in red.

4. Discussion

The number of exceedingly long reads presented ideal data for a complete genome assembly that allowed for the identification of possible plasmids, while a number of high-quality short reads were ideal for polishing the final assembly. The presence of the megaplasmid was evident from the initial Oxford Nanopore assemblies by Flye as two long circular contigs, regardless of whatever input parameters were produced. However, the length of the shorter contig varied considerably between assemblies that were produced with different input parameters. Thus, the length of plasmid was initially predicted by comparing Flye contigs to contigs produced by plasmidSPAdes. The final confirmation was found in the different coverage of chromosome and plasmid in the final assembly as the read coverage is one possible way to distinguish between chromosome and plasmid [35]. Since almost all high-quality Illumina reads mapped unambiguously to the final genome assembly, the proposed assembly seems to be correct and of a high quality. Although the prediction of replication origin *oriC* in the chromosome was unambiguous, the prediction of replication origin *oriV* in the plasmid was complicated as no standardized algorithm is currently available [44]. The first from two putative replication initiators was frameshifted and the second, while having a complete coding sequence, had a noticeably short upstream intergenic region. As both genes were located relatively close to each other, the start of the sequence was set according to the first one.

The calculated GC content matched our presumptions as clostridia form a low GC content group of Gram-positive bacteria. While the GC content of the chromosome was the same as for other strains with complete genomes, e.g., NRRL B-598 [27], NCIMB 8052 [20], DSM 15410^T [14], the GC content of the plasmid was slightly lower. Another difference between the chromosome and the plasmid was found in abundances of genes in particular COG categories, see Supplementary Table S1. While the plasmid carried zero genes from group (C) "Energy production and conversion", the same group contained 6.28% genes of the chromosome. Similarly, the chromosome carried larger percentage of genes in groups (E) "Amino Acid transport and metabolism", (F) "Nucleotide transport and metabolism", (G) "Carbohydrate transport and metabolism", (H) "Coenzyme transport and metabolism", (I) "Lipid transport and metabolism", and (J) "Translation, ribosomal function and biogenesis", i.e., all groups directly or indirectly connected to metabolism. Thus, the plasmid in the strain *C. beijerinckii* DSM 791^T does not contribute to the production of metabolites utilizable in industrial biotechnology. In other clostridia, plasmids sometimes carry genes necessary for solventogenesis. For example, *sol* operon in *C. acetobutylicum* ATCC 824 is located on a plasmid that can be lost during the degeneration process, resulting in a loss of ability to produce solvents [59]. On the contrary, the plasmid in *C. beijerinckii* DSM 791^T carried a larger percentage of genes in groups (L) "Replication, recombination and repair" and (V) "Defense mechanisms", which means that the plasmid possessed a machinery capable of protecting the cell. The latter group contained five genes from which three belonged to the cluster involved in the production of the bacteriocin circularin A. The cluster contained genes *cfgR* (KEC93_26215), *cfgK* (KEC93_26220), *cfg02* (KEC93_26225), *cfg01* (KEC93_26230), *cirA* (KEC93_26235), *cirB* (KEC93_26240), *cirC* (KEC93_26245), *cirD* (KEC93_26250), *cirG* (KEC93_26255), *cirH* (KEC93_26260), and *cirI* (KEC93_26265). Such a structure of the gene cluster has been already experimentally proved for *C. beijerinckii* DSM 791^T. However, it was not revealed that it is carried by a plasmid [60]. In our sequence, gene *cirE*, which is necessary for bacteriocin production, was not annotated by PGAP. This was probably due to its short length of only 150 bp. Nevertheless, the gene is present in the sequence between *cirD* and *cirG* as its nucleotide sequence had 100% similarity to that already sequenced and experimentally proved by Kemperman et al. [60] (GenBank accession no. AJ566621.1). Circularin A has a wide activity range that inhibits *C. tyrobutyricum*, lactococci, enterococci, and some *Lactobacillus* strains. It is also highly resistant to digestion by sequence-specific endoproteinases [61]. For potential industrial use, the ability of the strain to kill other bacteria is highly advantageous, as contamination of fermentation processes by bacteria presents a compelling problem that makes the fermentation more expensive [62]. Moreover, the strain *C. beijerinckii* DSM 791^T has already been proved to be a robust 1,3-propanediol producer [63].

The number and cumulative length of putative prophages were approximately half compared with the strain DSM 15410^T [14]. However, this prediction is questionable as two additional tools, PHASTER [64] and Prophage Hunter [65], predicted six and 30 prophages, respectively (data not shown). Although bacteriophage infection presents a serious problem for ABE fermentation, it has been addressed by only a few studies to date. Its analysis is not trivial and usually requires a separate study [66,67]. Thus, the question of phages in *C. beijerinckii* DSM 791^T remains open. Unlike the strain DSM 15410^T [14], CRISPR arrays in the strain DSM 791^T had no *cas* or *cas* like genes in their neighborhoods. This suggests that while a culture of the strain DSM 791^T might be resistant to other bacteria thanks to bacteriocin production, it might be defenseless against phage contaminations, as a CRISPR-associated system (Cas) forms somewhat of a bacterial immune system that provides protection from foreign genetic material [68].

Clostridium beijerinckii is a widely studied species, as the number of genome assemblies available suggests. Nevertheless, only a small fraction of them present high-quality complete genomes. While this analysis showed that the species is extremely diverse, and different strains may contain different genes, it is evident that some strains have been misidentified. This applies primarily to the strain that was WB53 isolated from a woodchip

bioreactor [69]. All analyses, whole-genome alignment, dDDH, and a search for reference genes, presented strong evidence that the strain WB53 does not belong to the *C. beijerinckii* species and was omitted from following analyses to define the core genome. Similarly, the remaining two outliers arising from an analysis of the reference genes were discarded. Although the dDDH value for the strain ASCUSBR67 exceeded 70%, suggesting that it belongs to the *C. beijerinckii* species [70], its genome lacks the annotation that was needed since amino acid sequences were used in the following analysis. Unlike WB53, the strain ASCUSBR67 had a majority of corrupted, not missing genes (see Supplementary Table S4). This means that its assembly needs polishing by high-quality NGS data for further analyses, including genome annotation [71]. Yet, it is likely that the strain belongs to *C. beijerinckii*. The last outlier, the genome of the strain HUN142, had most of the reference genes missing. Together with a dDDH value below 70%, we can conclude that this strain does not belong to *C. beijerinckii*. It should be noted that when calculating the dDDH value, the d_6 formula was used for complete genomes as it preserves the maximum amount of information, while the d_4 formula was used for draft genomes as a more robust way to compare incomplete data [72]. Genomes of outliers below the lower fence in a reference genes' analysis were preserved for following analyses as these represented strains that are evolutionary extremely close to the type strain *C. beijerinckii* DSM 791^T.

Similar to ASCUSBR67, the study discarded other unannotated genomes for which amino acid sequences of protein coding genes are unavailable. Those were genomes of strains WB and G117. On the other hand, the study preserved the genome of the strain DSM 6423 [29], thedDDH value of which was below the 70% threshold. First, when considering the confidence interval, a dDDH value of 70% was reached. Second, its number of missing or corrupted reference genes did not differ substantially from other strains. Since clostridia present a diverse group of organisms, this study contends that following only a dDDH value for their delineation may be cumbersome.

It is not surprising that the core genome had a larger proportion of annotated genes as it contains housekeeping genes that maintain basic cellular functions; additionally, their orthologues are known as they are present in other related species [73]. The core genome can be used to improve a phylogenetic analysis and to correctly assign bacterial species [74]. This is highly convenient for clostridia where the reidentification of strains is still quite common. Thus, the use of the core genome defined within this study is suggested as a supplement to a dDDH analysis when assigning new strains to the *C. beijerinckii* species. From a biotechnological point of view, the core genome is not of the main interest as it contains primarily critical functions, the alteration or deletion of which are often not possible. This is the reason why the core genome lacks genes responsible for various biotechnologically relevant phenotypic manifestations, as the production of solvents varies among strains. The core genome contains master regulators, for example *spo0A* (Gene3468# in the core genome) that orchestrates except for sporulation also solvent production. On the other hand, particular genes coding enzymes necessary to form the final products, e.g., *adh* (Gene3611# in the accessory genome) that is responsible for isopropanol production or *dhaT* (Gene1153# in the accessory genome) that is responsible for 1,3-propandiol production, are parts of the accessory genome. Therefore, they are not present in every strain. Nevertheless, some of the enzymes still present parts of the core genome, e.g., *adh* (Gene4950# in the core genome) that is responsible for ethanol production or *bdh* (Gene1521# in the core genome) that is responsible for butanol production. This was expected as *C. beijerinckii* covers ABE and IBE fermentation strains, meaning that while the ability to produce acetone or isopropanol varies among strains, ethanol and butanol are produced by all strains. Nevertheless, other metabolic regulations may suppress their production, for example ethanol production is negligible in the strain *C. beijerinckii* NRRL B-598 under standard cultivation conditions [24]. Additionally, solvent production may be interrupted during the so-called acid crash phenomenon induced by cultivation conditions or genome mutations [75–77].

The core genome can find additional utilization in the identification of versatile reference genes for RT-qPCR in *C. beijerinckii*, as such genes should be present in every strain. However, experimental validation is always needed, as a reference gene-coding peptidase T (Gene4499# in the core genome), proposed as a reference gene for the strain NCIMB 8052 [20], was later proved to be not utilizable for the strain NRRL B-598 [78], where genes *greA* (Gene2564# in the core genome), *znp* (Gene1191# in the core genome) and others performed better.

The fact that approximately one third of the accessory genome and unique genes was not assigned any COG suggests that orthologues for these genes are not present in more studied species. These genes might provide desirable phenotypic features, but advances in the field of functional annotation of non-model organisms are required to reveal these hidden properties. The accessory genome also contains several sequences of hypothetical proteins, the sequences of which are formed from repetitive subsequences. Even though these sequences were found in two or more genomes, they are probably not real proteins, and they might prevent the accessory genome from being used in BLAST searches. We suggest the use of the corrected accessory genome for BLAST searches where 76 sequences with a single kind of amino acid forming more than 25% of a sequence are discarded. The corrected Fasta file was uploaded to the FAIRDOMHub.

Although genes coding bacteriocin circularin A have been described only for the type strain *C. beijerinckii* DSM 791^T [60], they are present also in other strains, as the type strain has only one unique gene. Nevertheless, the presence of the whole cluster of all genes was not found in any other genome, nor in the closely related genomes that formed cluster 5 in the phylogenetic tree. Apart from the type strain, the cluster contained strains NBRC 10935, NCTC13035, DJ317, and DJ079, which were found to be closely related in the preceding analysis of reference genes as four outliers below the lower fence.

Various strains of *C. beijerinckii* are extremely diverse, except for one large cluster containing evolutionary closely related strains. The majority of them (DJ strains) were uploaded to the GenBank at the same time by the DOE Joint Genome Institute. Unfortunately, the study describing these strains is missing. However, not all of the DJ strains clustered together and some of them even formed particular leaf nodes. From the type strain, the evolutionary-distant strain DSM 6423 was clustered with an additional five strains into cluster 15. Although both type strains have different genome characteristics, they are evolutionary closer to each other than to DSM 6423. The closest neighbor to the IBE fermenting type strain DSM 15410^T is the strain NRRL B-598, a typical ABE fermentation representative. This is surprising because evolutionary-distant DSM 6423 is also an IBE strain. The absence of a distinguished cluster of IBE strains is further supported by the position of the individual leaf node with the strain *C. beijerinckii* BGS1, which is another isopropanol producer [30]. These relations only confirm that the evolution of *Clostridium beijerinckii* is not trivial, and novel strains should be identified using the whole core genome rather than particular genes or a dDDH analysis.

5. Conclusions

In this study, we sequenced and assembled the first complete genome sequence of the type strain *Clostridium beijerinckii* DSM 791^T. We discovered that the genome of the type strain is composed of a circular chromosome and a circular megaplasmid that carries a complete cluster of genes to produce bacteriocin circularin A, which is unique among *C. beijerinckii* strains. We used the genome sequence for whole-genome comparisons, and we found out that at least two strains currently assigned as *C. beijerinckii*, WB53 and HUN142, do not belong to the species. Moreover, we proved that some of the genome assemblies, e.g., the genome sequence of the strain *C. beijerinckii* ASCUSBR67, are of lower quality and should be polished before they can be used for comparative analyses. By collecting 237 genomes that met the quality criteria, we defined for the first time the pangenome of the *Clostridium beijerinckii* species. We used the core genome to reconstruct a phylogeny of the whole species, using the maximum sequence information available. As we demonstrated,

phylogeny of the species is not trivial, and we suggest use of the core genome when performing comparative analysis and identification of novel strains. The accessory genome contained a large percentage of genes with unknown function. This means, therefore, that many unique properties of the *C. beijerinckii* species might be still unreported.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/pr9071196/s1, Supplementary tables and figures: Figure S1: Boxplot of sum of missing and corrupted genes, Figure S2: Complete phylogenetic tree of *Clostridium beijerinckii* strains, Table S1: List of COG categories and number of genes present in each COG in *C. beijerinckii* DSM 791^T genome, Table S2: Phage DNA within the *C. beijerinckii* DSM 791^T genome, Table S3: CRISPR arrays in the *C. beijerinckii* DSM 791^T genome, Table S4: List of *C. beijerinckii* genomes used for analysis, Table S5: Result of dDDH analysis for complete *C. beijerinckii* genomes, Table S6: Numbers of genes in reference genome that are missing in particular strains, Table S7: List of unique and uniquely missing genes in *C. beijerinckii* strains, Table S8: List of COG categories and number of genes present in each COG in *C. beijerinckii* core and accessory genomes and unique genes, Table S9: List of strains in each of collapsed branches in Figure 4.

Author Contributions: Conceptualization, K.S., M.L., P.P. and H.S.; methodology, K.S., M.N., M.B. and B.B., formal analysis, K.S. and M.N.; investigation, M.B. and B.B.; resources, K.S., M.L. and P.P.; data curation, K.S.; writing—original draft preparation, K.S., M.B. and B.B.; writing—review and editing, K.S.; visualization, K.S., M.N. and H.S.; supervision, K.S. and P.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The genome assembly referred in this paper is the version GCA_018223745.1. Sequences of the chromosome and the plasmid were uploaded to GenBank under accession numbers CP073653.1 and CP073654.1. The whole-genome sequencing data have been deposited in the NCBI Sequence Read Archive (SRA) under the project accession number PRJNA724001. Sequences of pangenome and unique gene were uploaded to FAIRDOMHub under the project "Clostridium beijerinckii pan-genome" https://fairdomhub.org/projects/242 (accessed on 1 June 2021) under the Creative Commons Attribution-NonCommercial 4.0 license.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Cruz-Morales, P.; Orellana, C.; Moutafis, G.; Moonen, G.; Rincon, G.; Nielsen, L.; Marcellin, E. Revisiting the Evolution and Taxonomy of Clostridia, a Phylogenomic Update. *Genome Biol. Evol.* **2019**, *11*, 2035–2044. [CrossRef]
2. Schiel-Bengelsdorf, B.; Montoya, J.; Linder, S.; Dürre, P. Butanol fermentation. *Environ. Technol.* **2013**, *34*, 1691–1710. [CrossRef] [PubMed]
3. Vieira, C.F.D.S.; Filho, F.M.; Filho, R.M.; Mariano, A.P. Acetone-free biobutanol production: Past and recent advances in the Isopropanol-Butanol-Ethanol (IBE) fermentation. *Bioresour. Technol.* **2019**, *287*, 121425. [CrossRef]
4. Patakova, P.; Kolek, J.; Sedlar, K.; Koscova, P.; Branská, B.; Kupkova, K.; Paulova, L.; Provaznik, I. Comparative analysis of high butanol tolerance and production in clostridia. *Biotechnol. Adv.* **2018**, *36*, 721–738. [CrossRef]
5. Charubin, K.; Bennett, R.K.; Fast, A.G.; Papoutsakis, E.T. Engineering Clostridium organisms as microbial cell-factories: Challenges & opportunities. *Metab. Eng.* **2018**, *50*, 173–191. [CrossRef] [PubMed]
6. Finegold, S.M.; Song, Y.; Liu, C. Taxonomy—General Comments and Update on Taxonomy of Clostridia and Anaerobic cocci. *Anaerobe* **2002**, *8*, 283–285. [CrossRef] [PubMed]
7. Keis, S.; Bennett, C.F.; Ward, V.K.; Jones, D.T. Taxonomy and Phylogeny of Industrial Solvent-Producing Clostridia. *Int. J. Syst. Bacteriol.* **1995**, *45*, 693–705. [CrossRef]
8. Johnson, J.L.; Francis, B.S. Taxonomy of the Clostridia: Ribosomal Ribonucleic Acid Homologies among the Species. *J. Gen. Microbiol.* **1975**, *88*, 229–244. [CrossRef]
9. Sedlar, K.; Kolek, J.; Provaznik, I.; Patakova, P. Reclassification of non-type strain Clostridium pasteurianum NRRL B-598 as Clostridium beijerinckii NRRL B-598. *J. Biotechnol.* **2017**, *244*, 1–3. [CrossRef] [PubMed]
10. Lawson, P.A.; Rainey, F.A. Proposal to restrict the genus Clostridium Prazmowski to Clostridium butyricum and related species. *Int. J. Syst. Evol. Microbiol.* **2016**, *66*, 1009–1016. [CrossRef]
11. Moon, C.D.; Pacheco, D.M.; Kelly, W.J.; Leahy, S.; Li, D.; Kopečný, J.; Attwood, G. Reclassification of Clostridium proteoclasticum as Butyrivibrio proteoclasticus comb. nov., a butyrate-producing ruminal bacterium. *Int. J. Syst. Evol. Microbiol.* **2008**, *58*, 2041–2045. [CrossRef]

12. Kobayashi, H.; Tanizawa, Y.; Sakamoto, M.; Nakamura, Y.; Ohkuma, M.; Tohno, M. Reclassification of Clostridium diolis Biebl and Spröer 2003 as a later heterotypic synonym of Clostridium beijerinckii Donker 1926 (Approved Lists 1980) emend. Keis et al. 2001. *Int. J. Syst. Evol. Microbiol.* **2020**, *70*, 2463–2466. [CrossRef]
13. Ackermann, M. A functional perspective on phenotypic heterogeneity in microorganisms. *Nat. Rev. Genet.* **2015**, *13*, 497–508. [CrossRef]
14. Sedlar, K.; Vasylkivska, M.; Musilova, J.; Branska, B.; Provaznik, I.; Patakova, P. Phenotypic and genomic analysis of isopropanol and 1,3-propanediol producer Clostridium diolis DSM 15410. *Genomics* **2021**, *113*, 1109–1119. [CrossRef]
15. Biebl, H.; Spröer, C. Taxonomy of the Glycerol Fermenting Clostridia and Description of Clostridium diolis sp. nov. *Syst. Appl. Microbiol.* **2002**, *25*, 491–497. [CrossRef]
16. Wu, J.; Dong, L.; Zhou, C.; Liu, B.; Xing, D.; Feng, L.; Wu, X.; Wang, Q.; Cao, G. Enhanced butanol-hydrogen coproduction by Clostridium beijerinckii with biochar as cell's carrier. *Bioresour. Technol.* **2019**, *294*, 122141. [CrossRef] [PubMed]
17. Wu, J.; Dong, L.; Zhou, C.; Liu, B.; Feng, L.; Wu, C.; Qi, Z.; Cao, G. Developing a coculture for enhanced butanol production by Clostridium beijerinckii and Saccharomyces cerevisiae. *Bioresour. Technol. Rep.* **2019**, *6*, 223–228. [CrossRef]
18. Skerman, V.B.D.; Sneath, P.H.A.; McGowan, V. Approved Lists of Bacterial Names. *Int. J. Syst. Evol. Microbiol.* **1980**, *30*, 225–420. [CrossRef]
19. Wang, Y.; Li, X.; Mao, Y.; Blaschek, H.P. Genome-wide dynamic transcriptional profiling in Clostridium beijerinckii NCIMB 8052 using single-nucleotide resolution RNA-Seq. *BMC Genom.* **2012**, *13*, 102. [CrossRef] [PubMed]
20. Wang, Y.; Li, X.; Mao, Y.; Blaschek, H.P. Single-nucleotide resolution analysis of the transcriptome structure of Clostridium beijerinckii NCIMB 8052 using RNA-Seq. *BMC Genom.* **2011**, *12*, 479. [CrossRef] [PubMed]
21. Zhang, Y.; Ezeji, T.C. Transcriptional analysis of Clostridium beijerinckii NCIMB 8052 to elucidate role of furfural stress during acetone butanol ethanol fermentation. *Biotechnol. Biofuels* **2013**, *6*, 66. [CrossRef] [PubMed]
22. Shi, Z.; Blaschek, H.P. Transcriptional Analysis of Clostridium beijerinckii NCIMB 8052 and the Hyper-Butanol-Producing Mutant BA101 during the Shift from Acidogenesis to Solventogenesis. *Appl. Environ. Microbiol.* **2008**, *74*, 7709–7714. [CrossRef]
23. Wilkinson, S.R.; Young, M. Physical map of the Clostridium beijerinckii (formerly Clostridium acetobutylicum) NCIMB 8052 chromosome. *J. Bacteriol.* **1995**, *177*, 439–448. [CrossRef]
24. Kolek, J.; Sedlar, K.; Provaznik, I.; Patakova, P. Dam and Dcm methylations prevent gene transfer into Clostridium pasteurianum NRRL B-598: Development of methods for electrotransformation, conjugation, and sonoporation. *Biotechnol. Biofuels* **2016**, *9*, 1–14. [CrossRef]
25. Sedlar, K.; Kolek, J.; Gruber, M.; Jureckova, K.; Branska, B.; Csaba, G.; Vasylkivska, M.; Zimmer, R.; Patakova, P.; Provaznik, I. A transcriptional response of Clostridium beijerinckii NRRL B-598 to a butanol shock. *Biotechnol. Biofuels* **2019**, *12*, 1–16. [CrossRef] [PubMed]
26. Vasylkivska, M.; Branska, B.; Sedlar, K.; Jureckova, K.; Provaznik, I.; Patakova, P. Phenotypic and Genomic Analysis of Clostridium beijerinckii NRRL B-598 Mutants With Increased Butanol Tolerance. *Front. Bioeng. Biotechnol.* **2020**, *8*, 598392. [CrossRef]
27. Sedlar, K.; Kolek, J.; Skutkova, H.; Branska, B.; Provaznik, I.; Patakova, P. Complete genome sequence of Clostridium pasteurianum NRRL B-598, a non-type strain producing butanol. *J. Biotechnol.* **2015**, *214*, 113–114. [CrossRef]
28. Diallo, M.; Hocq, R.; Collas, F.; Chartier, G.; Wasels, F.; Wijaya, H.S.; Werten, M.W.; Wolbert, E.J.; Kengen, S.W.; van der Oost, J.; et al. Adaptation and application of a two-plasmid inducible CRISPR-Cas9 system in Clostridium beijerinckii. *Methods* **2020**, *172*, 51–60. [CrossRef] [PubMed]
29. De Gérando, H.M.; Wasels, F.; Bisson, A.; Clement, B.; Bidard, F.; Jourdier, E.; López-Contreras, A.M.; Ferreira, N.L. Genome and transcriptome of the natural isopropanol producer Clostridium beijerinckii DSM6423. *BMC Genom.* **2018**, *19*, 242. [CrossRef] [PubMed]
30. Zhang, C.; Li, T.; He, J. Characterization and genome analysis of a butanol–isopropanol-producing Clostridium beijerinckii strain BGS1. *Biotechnol. Biofuels* **2018**, *11*, 1–11. [CrossRef] [PubMed]
31. Lanfear, R.; Schalamun, M.; Kainer, D.; Wang, W.; Schwessinger, B. MinIONQC: Fast and simple quality control for MinION sequencing data. *Bioinformatics* **2019**, *35*, 523–525. [CrossRef]
32. Ewels, P.; Magnusson, M.; Lundin, S.; Käller, M. MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **2016**, *32*, 3047–3048. [CrossRef]
33. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [CrossRef] [PubMed]
34. Kolmogorov, M.; Yuan, J.; Lin, Y.; Pevzner, P.A. Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* **2019**, *37*, 540–546. [CrossRef] [PubMed]
35. Antipov, D.; Hartwick, N.; Shen, M.; Rayko, M.; Lapidus, A.; Pevzner, P.A. plasmidSPAdes: Assembling plasmids from whole genome sequencing data. *Bioinformatics* **2016**, *32*, 3380–3387. [CrossRef] [PubMed]
36. Kurtz, S.; Phillippy, A.; Delcher, A.L.; Smoot, M.; Shumway, M.; Antonescu, C.; Salzberg, S.L. Versatile and open software for comparing large genomes. *Genome Biol.* **2004**, *5*, R12. [CrossRef]
37. Li, H. Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* **2018**, *34*, 3094–3100. [CrossRef] [PubMed]
38. Vaser, R.; Sović, I.; Nagarajan, N.; Šikić, M. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.* **2017**, *27*, 737–746. [CrossRef]

39. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **2009**, *25*, 1754–1760. [CrossRef] [PubMed]
40. Walker, B.J.; Abeel, T.; Shea, T.; Priest, M.; Abouelliel, A.; Sakthikumar, S.; Cuomo, C.A.; Zeng, Q.; Wortman, J.; Young, S.K.; et al. Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *PLoS ONE* **2014**, *9*, e112963. [CrossRef]
41. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [CrossRef]
42. Gao, F.; Zhang, C.-T. Ori-Finder: A web-based system for finding oriC s in unannotated bacterial genomes. *BMC Bioinforma.* **2008**, *9*, 1–6. [CrossRef] [PubMed]
43. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [CrossRef]
44. Luo, H.; Gao, F. DoriC 10.0: An updated database of replication origins in prokaryotic genomes including chromosomes and plasmids. *Nucleic Acids Res.* **2019**, *47*, D74–D77. [CrossRef]
45. Tatusova, T.; DiCuccio, M.; Badretdin, A.; Chetvernin, V.; Nawrocki, P.; Zaslavsky, L.; Lomsadze, A.; Pruitt, K.D.; Borodovsky, M.; Ostell, J. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res.* **2016**, *44*, 6614–6624. [CrossRef] [PubMed]
46. Taboada, B.; Estrada, K.; Ciria, R.; Merino, E. Operon-mapper: A web server for precise operon identification in bacterial and archaeal genomes. *Bioinformatics* **2018**, *34*, 4118–4120. [CrossRef] [PubMed]
47. Huerta-Cepas, J.; Szklarczyk, D.; Heller, D.; Hernández-Plaza, A.; Forslund, S.K.; Cook, H.V.; Mende, D.R.; Letunic, I.; Rattei, T.; Jensen, L.J.; et al. eggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* **2019**, *47*, D309–D314. [CrossRef] [PubMed]
48. Carver, T.; Thomson, N.; Bleasby, A.; Berriman, M.; Parkhill, J. DNAPlotter: Circular and linear interactive genome visualization. *Bioinform.* **2008**, *25*, 119–120. [CrossRef]
49. Rutherford, K.; Parkhill, J.; Crook, J.; Horsnell, T.; Rice, P.; Rajandream, M.-A.; Barrell, B. Artemis: Sequence visualization and annotation. *Bioinform.* **2000**, *16*, 944–945. [CrossRef]
50. Akhter, S.; Aziz, R.; Edwards, R.A. PhiSpy: A novel algorithm for finding prophages in bacterial genomes that combines similarity- and composition-based strategies. *Nucleic Acids Res.* **2012**, *40*, e126. [CrossRef]
51. Biswas, A.; Staals, R.; Morales, S.; Fineran, P.; Brown, C.M. CRISPRDetect: A flexible algorithm to define CRISPR arrays. *BMC Genom.* **2016**, *17*, 1–14. [CrossRef]
52. Sayers, E.W.; Cavanaugh, M.; Clark, K.; Ostell, J.; Pruitt, K.D.; Mizrachi, I.K. GenBank. *Nucleic Acids Res.* **2019**, *48*, D84–D86. [CrossRef]
53. Alikhan, N.-F.; Petty, N.K.; Ben Zakour, N.L.; Beatson, S.A. BLAST Ring Image Generator (BRIG): Simple prokaryote genome comparisons. *BMC Genom.* **2011**, *12*, 402. [CrossRef]
54. Murrell, P. R Graphics. *Wiley Interdiscip. Rev. Comput. Stat.* **2009**, *1*, 216–220. [CrossRef]
55. Meier-Kolthoff, J.P.; Göker, M. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. *Nat. Commun.* **2019**, *10*, 1–10. [CrossRef] [PubMed]
56. Chaudhari, N.M.; Gupta, V.K.; Dutta, C. BPGA- an ultra-fast pan-genome analysis pipeline. *Sci. Rep.* **2016**, *6*, 24373. [CrossRef] [PubMed]
57. Wolstencroft, K.; Krebs, O.; Snoep, J.L.; Stanford, N.J.; Bacall, F.; Golebiewski, M.; Kuzyakiv, R.; Nguyen, Q.; Owen, S.; Soiland-Reyes, S.; et al. FAIRDOMHub: A repository and collaboration environment for sharing systems biology research. *Nucleic Acids Res.* **2017**, *45*, D404–D407. [CrossRef] [PubMed]
58. Subramanian, B.; Gao, S.; Lercher, M.J.; Hu, S.; Chen, W.-H. Evolview v3: A webserver for visualization, annotation, and management of phylogenetic trees. *Nucleic Acids Res.* **2019**, *47*, W270–W275. [CrossRef]
59. Cornillot, E.; Nair, R.V.; Papoutsakis, E.T.; Soucaille, P. The genes for butanol and acetone formation in Clostridium acetobutylicum ATCC 824 reside on a large plasmid whose loss leads to degeneration of the strain. *J. Bacteriol.* **1997**, *179*, 5442–5447. [CrossRef]
60. Kemperman, R.; Jonker, M.; Nauta, A.; Kuipers, O.P.; Kok, J. Functional Analysis of the Gene Cluster Involved in Production of the Bacteriocin Circularin A by Clostridium beijerinckii ATCC 25752. *Appl. Environ. Microbiol.* **2003**, *69*, 6174–6178. [CrossRef]
61. Kemperman, R.; Kuipers, A.; Karsens, H.; Nauta, A.; Kuipers, O.; Kok, J. Identification and Characterization of Two Novel Clostridial Bacteriocins, Circularin A and Closticin 574. *Appl. Environ. Microbiol.* **2003**, *69*, 1589–1597. [CrossRef]
62. Thieme, N.; Panitz, J.C.; Held, C.; Lewandowski, B.; Schwarz, W.H.; Liebl, W.; Zverlov, V. Milling byproducts are an economically viable substrate for butanol production using clostridial ABE fermentation. *Appl. Microbiol. Biotechnol.* **2020**, *104*, 8679–8689. [CrossRef]
63. Wischral, D.; Zhang, J.; Cheng, C.; Lin, M.; de Souza, L.M.G.; Pessoa, F.L.P.; Pereira, N.; Yang, S.-T. Production of 1,3-propanediol by Clostridium beijerinckii DSM 791 from crude glycerol and corn steep liquor: Process optimization and metabolic engineering. *Bioresour. Technol.* **2016**, *212*, 100–110. [CrossRef]
64. Arndt, D.; Grant, J.R.; Marcu, A.; Sajed, T.; Pon, A.; Liang, Y.; Wishart, D.S. PHASTER: A better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* **2016**, *44*, W16–W21. [CrossRef]
65. Song, W.; Sun, H.-X.; Zhang, C.; Cheng, L.; Peng, Y.; Deng, Z.; Wang, D.; Wang, Y.; Hu, M.; Liu, W.; et al. Prophage Hunter: An integrative hunting tool for active prophages. *Nucleic Acids Res.* **2019**, *47*, W74–W80. [CrossRef] [PubMed]

66. Pyne, M.E.; Liu, X.; Moo-Young, M.; Chung, D.A.; Chou, C.P. Genome-directed analysis of prophage excision, host defence systems, and central fermentative metabolism in Clostridium pasteurianum. *Sci. Rep.* **2016**, *6*, 26228. [CrossRef]
67. Schüler, M.; Stegmann, B.; Poehlein, A.; Daniel, R.; Dürre, P. Genome sequence analysis of the temperate bacteriophage TBP2 of the solvent producer Clostridium saccharoperbutylacetonicum N1-4 (HMT, ATCC 27021). *FEMS Microbiol. Lett.* **2020**, *367*. [CrossRef] [PubMed]
68. Sorek, R.; Lawrence, C.M.; Wiedenheft, B. CRISPR-Mediated Adaptive Immune Systems in Bacteria and Archaea. *Annu. Rev. Biochem.* **2013**, *82*, 237–266. [CrossRef] [PubMed]
69. Anderson, E.L.; Jang, J.; Venterea, R.T.; Feyereisen, G.W.; Ishii, S. Isolation and characterization of denitrifiers from woodchip bioreactors for bioaugmentation application. *J. Appl. Microbiol.* **2020**, *129*, 590–600. [CrossRef]
70. Wayne, L.G.; Brenner, D.J.; Colwell, R.R.; Grimont, P.A.D.; Kandler, O.; Krichevsky, M.I.; Moore, L.H.; Moore, W.E.C.; Murray, R.G.E.; Stackebrandt, E.; et al. Report of the Ad Hoc Committee on Reconciliation of Approaches to Bacterial Systematics. *Int. J. Syst. Bacteriol.* **1987**, *37*, 463–464. [CrossRef]
71. Chen, Z.; Erickson, D.L.; Meng, J. Polishing the Oxford Nanopore long-read assemblies of bacterial pathogens with Illumina short reads to improve genomic analyses. *Genomics* **2021**, *113*, 1366–1377. [CrossRef]
72. Meier-Kolthoff, J.P.; Auch, A.F.; Klenk, H.-P.; Göker, M. Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinforma.* **2013**, *14*, 60. [CrossRef] [PubMed]
73. Saha, J.; Saha, B.K.; Sarkar, M.P.; Roy, V.; Mandal, P.; Pal, A. Comparative Genomic Analysis of Soil Dwelling Bacteria Utilizing a Combinational Codon Usage and Molecular Phylogenetic Approach Accentuating on Key Housekeeping Genes. *Front. Microbiol.* **2019**, *10*, 2896. [CrossRef]
74. Chung, M.; Munro, J.B.; Tettelin, H.; Hotopp, J.C.D. Using Core Genome Alignments to Assign Bacterial Species. *mSystems* **2018**, *3*, e00236-18. [CrossRef] [PubMed]
75. Seo, S.; Janssen, H.; Magis, A.; Wang, Y.; Lu, T.; Price, N.D.; Jin, Y.; Blaschek, H.P. Genomic, Transcriptional, and Phenotypic Analysis of the Glucose Derepressed Clostridium beijerinckii Mutant Exhibiting Acid Crash Phenotype. *Biotechnol. J.* **2017**, *12*, 1700182. [CrossRef]
76. Patakova, P.; Branská, B.; Sedlar, K.; Vasylkivska, M.; Jureckova, K.; Kolek, J.; Koscova, P.; Provaznik, I. Acidogenesis, solventogenesis, metabolic stress response and life cycle changes in Clostridium beijerinckii NRRL B-598 at the transcriptomic level. *Sci. Rep.* **2019**, *9*, 1–21. [CrossRef]
77. Branska, B.; Vasylkivska, M.; Raschmanova, H.; Jureckova, K.; Sedlar, K.; Provaznik, I.; Patakova, P. Changes in efflux pump activity of Clostridium beijerinckii throughout ABE fermentation. *Appl. Microbiol. Biotechnol.* **2021**, *105*, 877–889. [CrossRef] [PubMed]
78. Jureckova, K.; Raschmanova, H.; Kolek, J.; Vasylkivska, M.; Branska, B.; Patakova, P.; Provaznik, I.; Sedlar, K. Identification and Validation of Reference Genes in Clostridium beijerinckii NRRL B-598 for RT-qPCR Using RNA-Seq Data. *Front. Microbiol.* **2021**, *12*, 640054. [CrossRef] [PubMed]

5.13 Article XIII

HERMANKOVA, Kristyna, KOURILOVA, Xenie, PERNICOVA, Iva, BEZDICEK, Matej, LENGEROVA, Martina, OBRUCA, Stanislav, and **SEDLAR, Karel**. Complete Genome Sequence of the Type Strain Tepidimonas taiwanensis LMG 22826T, a Thermophilic Alkaline Protease and Polyhydroxyalkanoate Producer. *Genome Biology and Evolution*. 2021. 13(12), p. evab280. 2021 IF = 4.065, Q2 in EVOLUTIONARY BIOLOGY)

DOI: https://doi.org/10.1093/gbe/evab280

GBE

Complete Genome Sequence of the Type Strain *Tepidimonas taiwanensis* LMG 22826^T, a Thermophilic Alkaline Protease and Polyhydroxyalkanoate Producer

Kristyna Hermankova[1], Xenie Kourilova[2], Iva Pernicova[2], Matej Bezdicek[3], Martina Lengerova[3], Stanislav Obruca[2], and Karel Sedlar [1,4,*]

[1]Department of Biomedical Engineering, Faculty of Electrical Engineering and Communication, Brno University of Technology, Czech Republic

[2]Department of Food Chemistry and Biotechnology, Faculty of Chemistry, Brno University of Technology, Czech Republic

[3]Department of Internal Medicine—Hematology and Oncology, University Hospital Brno, Czech Republic

[4]Department of Informatics, Institute of Bioinformatics, Ludwig-Maximilians-Universität München, Germany

*Corresponding author: E-mail: sedlar@vut.cz.

Accepted: 9 December 2021

Abstract

Tepidimonas taiwanensis is a moderately thermophilic, Gram-negative, rod-shaped, chemoorganoheterotrophic, motile bacterium. The alkaline protease producing type strain *T. taiwanensis* LMG 22826^T was recently reported to also be a promising producer of polyhydroxyalkanoates (PHAs)—renewable and biodegradable polymers representing an alternative to conventional plastics. Here, we present its first complete genome sequence which is also the first complete genome sequence of the whole species. The genome consists of a single 2,915,587-bp-long circular chromosome with GC content of 68.75%. Genome annotation identified 2,764 genes in total while 2,634 open reading frames belonged to protein-coding genes. Although functional annotation of the genome and division of genes into Clusters of Orthologous Groups (COGs) revealed a relatively high number of 694 genes with unknown function or unknown COG, the majority of genes were assigned a function. Most of the genes, 406 in total, were involved in energy production and conversion, and amino acid transport and metabolism. Moreover, particular key genes involved in the metabolism of PHA were identified. Knowledge of the genome in connection with the recently reported ability to produce bioplastics from the waste stream of wine production makes *T. taiwanensis* LMG 22826^T, an ideal candidate for further genome engineering as a bacterium with high biotechnological potential.

Key words: *Tepidimonas taiwanensis*, polyhydroxyalkanoates, hybrid assembly, Oxford Nanopore Technologies, functional annotation, alkaline protease.

Introduction

The majority of current plastics, for example, polyethylene, polyvinyl chloride, polystyrene, and nylon, are made from petroleum. Although their production is cheap, the environmental burden and the resources for their production will be depleted in the future. Therefore, a search for alternatives is needed. A solution has re-emerged in bio-based plastics (Kawashima et al. 2019). A promising group of bioplastics is now presented by polyesters of hydroxyalkanoic acids, that is, polyhydroxyalkanoates (PHA). These environmentally friendly alternatives to petroleum-based polymers are accumulated naturally by numerous prokaryotic microorganisms (Muhammadi et al. 2015; Sabapathy et al. 2020). Unfortunately, less than 1% of the total plastic production comes from the bioplastics industry (Shogren et al. 2019). The main obstacle preventing wider utilization of PHA in viable industrial processes is the cost of the carbon resources and the cost of the fermentation and downstream processing. A promising strategy which might help to reduce the cost of PHA is the use of inexpensive or waste carbon substrates that do not compete with the human food chain (Koller 2018) as well as the employment of extremophilic microorganisms

Significance

Next-generation industrial biotechnology is a concept that relies on the biosynthetic potential of extremophilic microorganisms, and which benefits from the robustness of such processes against contamination by common microflora. In this context, *Tepidimonas taiwanensis* is a very interesting, moderately thermophilic bacterium with great biotechnological potential. It has been reported that it is a potent producer of alkaline proteases, which are enzymes used in large quantities in numerous fields, including but not limited to the detergent industry. Furthermore, it was recently reported that *T. taiwanensis* is also capable of producing polyhydroxyalkanoates, a "green" alternative to petrochemical polymers, from inexpensive waste substrates such as grape pomace. In this work, we describe the complete genome sequence of the bacterium, which is an important step on the route toward complete utilization of the biosynthetic potential of the bacterium that can be further expanded by approaches of genetic engineering and synthetic biology.

(Obruca et al. 2018). Although some pivotal work has been completed and the fact that microorganisms use PHA to store unused energy and carbon in the cytoplasm in the form of intracellular granules is known (Obruca et al. 2018), additional knowledge that can be mined from various genomes of PHA producers is of high importance.

The type strain *Tepidimonas taiwanensis* LGM 22826^T (=BCRC 17406^T, I1-1^T) is a thermophilic, Gram-negative bacterium that was isolated from a hot spring in the Pingtung area in southern Taiwan (Chen et al. 2006). The rod-shaped cells are motile via a single polar flagellum. The bacterium was originally investigated for its strong alkaline protease activity, which is usable in different industries (Gupta et al. 2002). Nevertheless, other important features of the strain remained hidden, which may be due to the missing high-quality complete genome assembly and functional annotation of the genome. Only recently was its ability to utilize glucose and fructose to produce PHA reported (Kourilova et al. 2021). As it is a thermophile, PHA production takes place within the temperature range 45–55 °C, which reduces the risk of microbial contamination. Therefore, the strain presents an ideal organism for utilization under unsterile conditions, known as the "next-generation industrial biotechnology" concept (Chen and Jiang 2018). In this article, we present its first high-quality complete genome sequence. We annotated the genome, identified key genes in PHA metabolism and in coding extracellular proteases, and searched for prophage DNA and CRISPR arrays.

Results and Discussion

Genome Assembly and Properties

The complete genome sequence of *T. taiwanensis* LMG 22826^T was reconstructed using more than 415,000 Oxford Nanopore Technologies (ONT) reads with average length of 17 kb and polished by an additional 2.3 million high-quality (average Phred score $Q \approx 35$) Illumina PE reads. The overall coverage of the final assembly consisting of a single circular chromosome was 2785x. The genome has been deposited at the DDBJ/EMBL/GenBank under accession No CP083911.1.

More than 2.9 Mb long, the genome of *T. taiwanensis* consists of 2,764 genes, some of them organized into 569 predicted operons that comprise two or more structural genes. From 2,700 coding genes, 66 were marked as pseudogenes, which in most cases are made of incomplete gene sequences according to NCBI Prokaryotic Genome Annotation Pipeline (PGAP). All rRNA genes are present in three copies and 16S rRNA copies have similarity >99%. Further analysis of tRNAs encoded in the genome with tRNA-scanSE (Chan and Lowe 2019) revealed the differences between numbers of different isoacceptor tRNAs which can correlate with codon usage bias as has been reported (Rocha 2004). For example, three of four possible types of alanine amino acid isoacceptors are encoded in the genome in the ratio 1:1:3, so the abundance of the codon corresponding to the third isoacceptor is expected to be higher. In addition, the tRNA analysis revealed a high number, precisely 42, of tRNA isoacceptors that can be affected by relatively high GC content (Kanaya et al. 1999), in this case, almost 69%. All genome features of the *T. taiwanensis* genome are summarized in Table 1. Using the complete genome sequence, *Tepidimonas thermarum* was found to be the closest species to *T. taiwanensis*. Whole-genome sequence-based phylogeny of the ten most closely related species is available under supplementary figure S1, Supplementary Material online. Genome similarity of *T. taiwanensis* to these ten species

Table 1

Genomic Features of *Tepidimonas taiwanensis* LMG 22826^T

Feature	Chromosome
Length [bp]	2,915,587
GC content [%]	68.75
Genes	2,764
Operons	569
CDSs	2,700
Pseudogenes	66
ncRNAs	3
rRNAs (5S, 16S, 23S)	3, 3, 3
tRNAs	52

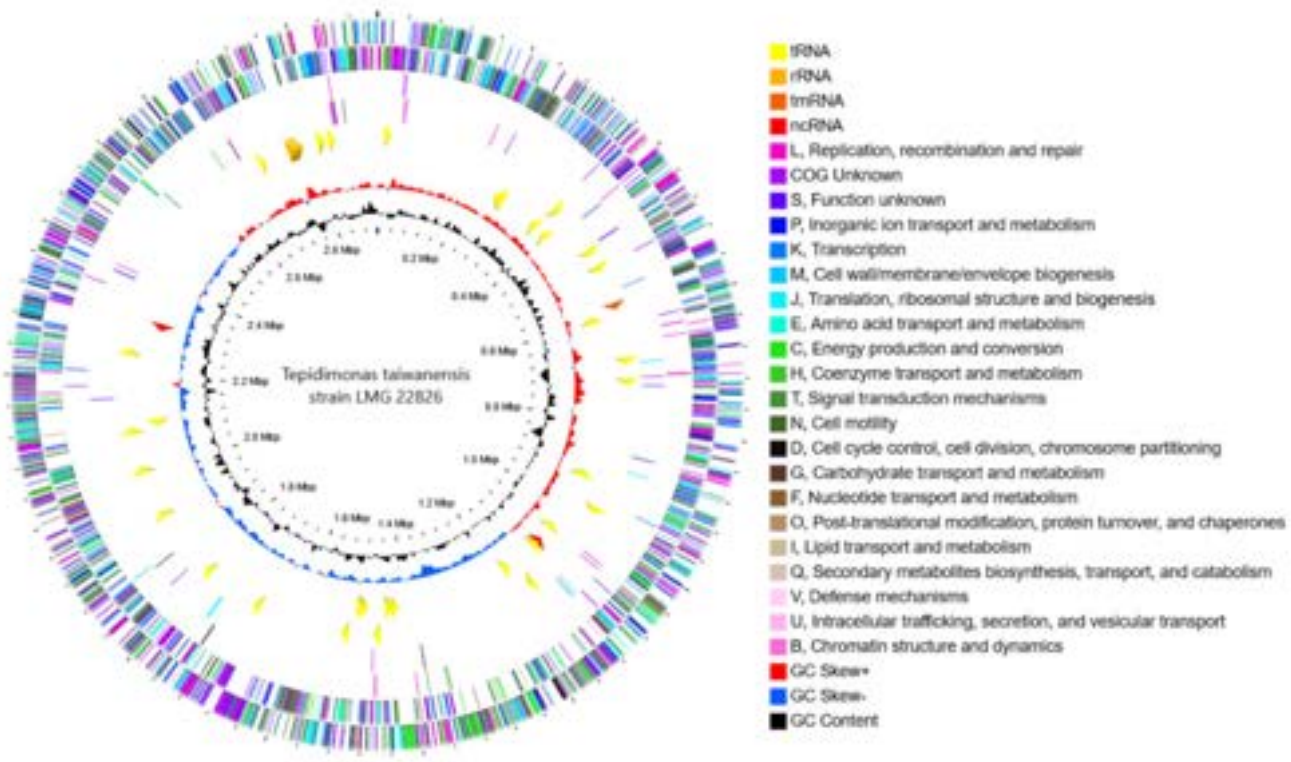


FIG. 1.—A genome map of the *T. taiwanensis* LMG 22826^T. The first two outermost circles represent CDS on the forward and backward strands, respectively. The next two circles contain pseudogenes on the forward and backward strands. Colors represent the functional classification of a COG. The fifth and the sixth circles consist of various types of RNA genes. Two inner circles show GC skew and GC content, respectively.

expressed as digital DNA to DNA hybridization reached values from 19.8% to 25.6%.

Functional Annotation

Protein-coding sequences (CDSs) were divided into 21 categories according to Cluster of Orthologous Groups (COGs). The most abundant known gene function is the "Energy production and conversion group of genes (C)" with 7.59% from all CDSs. Additionally, many genes belong to "Amino acid transport and metabolism group (E)," which makes up 7.44%. The high number of genes in these two groups corresponded to housekeeping functions of cells but was also related to industrially utilizable features, for example, the ability to produce PHA. Its production by the strain *T. taiwanensis* LMG 22826^T was proved recently, and the presence of *phaC* gene was confirmed by PCR (Kourilova et al. 2021). Here, we identified loci of *phaC* (LCC91_05560) and neighboring *phaR* (LCC91_04215) genes that are necessary for PHA production as well as a locus of *phaZ* (LCC91_05500) coding PHA depolymerase.

Unfortunately, almost 17% of coding genes were assigned to "Unknown function (S)" category and 8.85% genes were not recognized at all. All categories with gene counts are available under supplementary table S1, Supplementary Material online.

The arrangement of all genes in the *T. taiwanensis* LMG 22826^T genome is shown in figure 1 where every COG and every type of RNA is distinguished by a different color.

Although the strain was reported to produce extracellular alkaline proteases, due to the lack of genome sequence the enzymes were never identified. KEGG searches revealed 21 orthologues for alkaline proteases, see supplementary table S2, Supplementary Material online. Three of them, *sppA* (LCC91_01535), *degP/htrA* (LCC91_09805), and *prpL* (LCC91_10460), coded for extracellular proteases. Moreover, their predicted molecular weights 36.4, 52.8, and 69.9 kDa matched experimental evidence provided by zymography (Chen et al. 2006). These enzymes might be of great industrial importance. For example, the optimum enzymatic activity of protease IV coded by *sppA* was reported to be at pH 10 and temperature of 45 °C (Engel et al. 1998), which makes this enzyme utilizable in the detergent industry for production of washing powder.

The genome was searched for CRISPR arrays and, as a consensus from three prediction methods, four large arrays were reported in the *T. taiwanensis* LMG 22826^T genome. The largest array consisted of 42 spacers and was 2,590 bp long. Moreover, *cas* genes, such as *cas1* or *cas2*, whose proteins are responsible for spacers acquisition into CRISPR arrays (Yosef et al. 2012), were found in the genome. Unfortunately, neither of these genes was the gene that encodes the Cas9

protein well known for its high utilization in genetic engineering. The summary of CRISPR arrays is included under supplementary in table S3, Supplementary Material online.

Finally, the presence of prophage DNA was checked. PHASTER found three prophages: two of them were labeled as incomplete and one as intact prophage. The sequence that was labeled as intact prophage consists of 70 proteins, and 46 of them match to phage proteins such as phage tail protein or phage virion protein. Eight of these proteins correspond to *Escherichia* phage *vB_EcoM_ECO1230-10*, which has not been reported as a phage able to survive life conditions of thermophilic bacteria such as *T. taiwanensis*. Although Prophage Hunter did not label any of the phages as active, the previously mentioned phage sequence achieved the highest score and corresponded to *Acidithiobacillus* phage AcaML1, which has been found in thermophilic, acidophilic bacterium *Acidithiobacillus caldus* (Covarrubias et al. 2018), so it is possible to presume this phage has the ability to survive the life conditions of *T. taiwanensis*. Overall, the reliable statement of whether the prophage is active would need further analysis. The summary table of present prophage DNA from PHASTER tool is available under supplementary table S4, Supplementary Material online.

Materials and Methods

Growth Conditions, DNA Extraction, and Sequencing

Bacterial cultures of *T. taiwanensis* LMG 22826^T were purchased from Belgian-coordinated collections of microorganisms. First, the bacterial culture was cultivated in complex medium (nutrient broth with 1% peptone, HiMedia) for 24 h at 50 °C with constant shaking (180 rpm). Subsequently, *T. taiwanensis* LMG 22826^T was cultivated in a mineral salt medium composed of $Na_2HPO_4 \cdot 12\ H_2O$ (9.0 g/l), KH_2PO_4 (1.5 g/l), NH_4Cl (1.0 g/l), $MgSO_4 \cdot 7\ H_2O$ (0.2 g/l), $CaCl_2 \cdot 2\ H_2O$ (0.02 g/l), $Fe^{(III)}NH_4$ citrate (0.0012 g/l), yeast extract (0.5 g/l), 1 ml/l of microelements solution containing EDTA (50.0 g/l), $FeCl_3 \cdot 6\ H_2O$ (13.8 g/l), $ZnCl_2$ (0.84 g/l), $CuCl_2 \cdot 2\ H_2O$ (0.13 g/l), $CoCl_2 \cdot 6\ H_2O$ (0.1 g/l), $MnCl_2 \cdot 6\ H_2O$ (0.016 g/l) and H_3BO_3 (0.1 g/l) dissolved in distilled water, and a glucose (20.0 g/l) as a carbon substrate. The parameters of cultivation conditions on mineral salt medium corresponded to the cultivation conditions on complex medium.

Genomic DNA was extracted using MagAttract HMW DNA kit (Qiagene, NL). The DNA purity was checked using NanoDrop (Thermo Fisher Scientific, USA), the concentration was measured using Qubit 2.0 Fluorometr (Thermo Fisher Scientific, USA), and the proper length of the isolated DNA was confirmed using Agilent 4200 TapeStation (Agilent Technologies, USA). The sequencing library for Oxford Nanopore sequencing was prepared using Ligation Sequencing 1D Kit (Oxford Nanopore Technologies, UK). The sequencing was performed using the R9.4.1 Flow Cell and the MinION platform (Oxford Nanopore Technologies, UK). The sequencing library for short read sequencing was prepared using KAPA HyperPlus kit and was carried out using Miseq Reagent kit v2 (500 cycles) and Illumina MiSeq platform (Illumina, USA).

Genome Assembly

Long ONT reads were basecalled with Guppy basecaller v3.4.4 (https://nanoporetech.com/, last accessed November 8, 2021) and the quality was checked with MinIONQC v1.4.1 (Lanfear et al. 2019). Similarly, Illumina paired-end reads quality was checked with FastQC v0.11.5 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/, last accessed November 17, 2021) and low-quality ends and adapters were trimmed with Trimmomatic v0.39 (Bolger et al. 2014).

Nanopore reads were assembled using Flye v2.8.1 (Kolmogorov et al. 2019). The final assembly was then polished with four rounds of Racon v1.4.20 (Vaser et al. 2017) using Minimap2 v2.17 (Li 2018) to map long reads. After that, two rounds of polishing with Medaka v1.2.5 (https://nanoporetech.github.io/medaka/, last accessed November 9, 2021) were applied. In addition, Illumina reads were mapped to the polished assembly with BWA aligner v0.7.17 (Li and Durbin 2009) and additional corrections were performed using three rounds of Pilon v1.24 (Walker et al. 2014) polishing. The final step was to rearrange the genome according to the predicted replication origin (*oriC*) with Ori-Finder 2 (Luo et al. 2014) so that the genome starts with the *dnaA* gene.

Genome Annotation and Analysis

The genome of *T. taiwanensis* LMG 22826^T was annotated with the PGAP v5.3 (Tatusova et al. 2016). The prediction of operons was performed using Operon-mapper (Taboada et al. 2018). Protein-coding genes were assigned to COGs according to eggNOG database (Huerta-Cepas et al. 2019) through eggNOG mapper (Cantalapiedra et al. 2021). Genome-based phylogeny and comparison to the most closely related species was done with the type strain genome server (Meier-Kolthoff and Göker 2019). Furthermore, the genome sequence was searched for prophage DNA with the PHASTER tool (Arndt et al. 2016), Prophage Hunter (Song et al. 2019), and the occurrences of CRISPR arrays and cas genes were inspected through CRISPRDetect (Biswas et al. 2016) and CRISPRCasFinder (Couvin et al. 2018). The results were compared with PGAP annotation and only arrays predicted by at least two tools were reported. Finally, the genome sequence was processed with CGView (Stothard and Wishart 2005) to construct the circular genome map. Identification of orthologous genes coding key enzymes was made by manual BLAST (Altschul et al. 1990) searches and by browsing the KEGG database (Kanehisa and Goto 2000). Molecular weights were predicted from the primary structure of protein sequence using ExPASy (Gasteiger et al. 2003).

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgment

This study was funded by the Czech Science Foundation (GACR) (Project No. GA19-20697S).

Data Availability

The whole-genome sequence has been deposited at DDBJ/ENA/GenBank under the accession number CP083911.1. The NCBI BioProject and BioSample accession numbers are PRJNA764859 and SAMN21531155, respectively. The raw reads have been deposited at NCBI SRA database under the accession numbers SRR16956434 (paired-end Illumina) and SRR16956433 (Oxford Nanopore Technologies).

Literature Cited

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol. 215(3):403–410.

Arndt D, et al. 2016. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res. 44(W1):W16–W21.

Biswas A, Staals RHJ, Morales SE, Fineran PC, Brown CM. 2016. CRISPRDetect: a flexible algorithm to define CRISPR arrays. BMC Genomics 17(1):356.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30(15):2114–2120.

Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. 2021. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. Mol Biol Evol. 38(12):5825–5829.

Chan PP, Lowe TM. 2019. tRNAscan-SE: searching for tRNA genes in genomic sequences. Methods Mol Biol. 1962:1.

Chen GQ, Jiang XR. 2018. Next generation industrial biotechnology based on extremophilic bacteria. Curr Opin Biotechnol. 50:94–100.

Chen TL, Chou YJ, Chen WM, Arun B, Young CC. 2006. *Tepidimonas taiwanensis* sp. nov., a novel alkaline-protease-producing bacterium isolated from a hot spring. Extremophiles 10(1):35–40.

Couvin D, et al. 2018. CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. Nucleic Acids Res. 46(W1):W246–W251.

Covarrubias PC, et al. 2018. Occurrence, integrity and functionality of AcaML1–like viruses infecting extreme acidophiles of the *Acidithiobacillus* species complex. Res Microbiol. 169(10):628–637.

Engel LS, Hill JM, Caballero AR, Green LC, O'Callaghan RJ. 1998. Protease IV, a unique extracellular protease and virulence factor from *Pseudomonas aeruginosa*. J Biol Chem. 273(27):16792–16797.

Gasteiger E, et al. 2003. ExPASy: the proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res. 31(13):3784–3788.

Gupta R, Beg QK, Khan S, Chauhan B. 2002. An overview on fermentation, downstream processing and properties of microbial alkaline proteases. Appl Microbiol Biotechnol. 60:4. 60:381–395.

Huerta-Cepas J, et al. 2019. EggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. 47(D1):D309–D314.

Kanaya S, Yamada Y, Kudo Y, Ikemura T. 1999. Studies of codon usage and tRNA genes of 18 unicellular organisms and quantification of *Bacillus subtilis* tRNAs: gene expression level and species-specific diversity of codon usage based on multivariate analysis. Gene 238(1):143–155.

Kanehisa M, Goto S. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 28(1):27–30.

Kawashima N, Yagi T, Kojima K. 2019. How do bioplastics and fossil-based plastics play in a circular economy? Macromol Mater Eng. 304(9):1900383.

Koller M. 2018. A review on established and emerging fermentation schemes for microbial production of polyhydroxyalkanoate (PHA) biopolyesters. Fermentation 4(2):30.

Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol. 37(5):540–546.

Kourilova X, et al. 2021. Biotechnological conversion of grape pomace to poly(3-hydroxybutyrate) by moderately thermophilic bacterium *Tepidimonas taiwanensis*. Bioengineering 8(10):141.

Lanfear R, Schalamun M, Kainer D, Wang W, Schwessinger B. 2019. MinIONQC: fast and simple quality control for MinION sequencing data. Bioinformatics 35(3):523–525.

Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34(18):3094–3100.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25(14):1754–1760.

Luo H, Zhang CT, Gao F. 2014. Ori-Finder 2, an integrated tool to predict replication origins in the archaeal genomes. Front Microbiol. 5:482.

Meier-Kolthoff JP, Göker M. 2019. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. Nat Commun. 10:1–10.

Muhammadi S, Afzal M, Hameed S. 2015. Bacterial polyhydroxyalkanoates-eco-friendly next generation plastic: production, biocompatibility, biodegradation, physical properties and applications. Green Chem Lett Rev.8:56–77.

Obruca S, Sedlacek P, Koller M, Kucera D, Pernicova I. 2018. Involvement of polyhydroxyalkanoates in stress resistance of microbial cells: biotechnological consequences and applications. Biotechnol Adv. 36(3):856–870.

Rocha EPC. 2004. Codon usage bias from tRNA's point of view: redundancy, specialization, and efficient decoding for translation optimization. Genome Res. 14(11):2279–2286.

Sabapathy PC, et al. 2020. Recent developments in polyhydroxyalkanoates (PHAs) production – A review. Bioresour Technol. 306:123132.

Shogren R, Wood D, Orts W, Glenn G. 2019. Plant-based materials and transitioning to a circular economy. Sust Prod Consump. 19:194–215.

Song W, et al. 2019. Prophage Hunter: an integrative hunting tool for active prophages. Nucleic Acids Res. 47(W1):W74–W80.

Stothard P, Wishart DS. 2005. Circular genome visualization and exploration using CGView. Bioinformatics 21(4):537–539.

Taboada B, Estrada K, Ciria R, Merino E. 2018. Operon-mapper: a web server for precise operon identification in bacterial and archaeal genomes. Bioinformatics 34(23):4118–4120.

Tatusova T, et al. 2016. NCBI prokaryotic genome annotation pipeline. Nucleic Acids Res. 44(14):6614–6624.

Vaser R, Sović I, Nagarajan N, Šikić M. 2017. Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res. 27(5):737–746.

Walker BJ, et al. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9(11):e112963.

Yosef I, Goren MG, Qimron U. 2012. Proteins and DNA elements essential for the CRISPR adaptation process in *Escherichia coli*. Nucleic Acids Res. 40(12):5569–5576.

Associate editor: Howard Ochman

5.14 Article XIV

PATAKOVA, Petra, BRANSKA, Barbora, VASYLKIVSKA, Maryna, JURECKOVA, Katerina, MUSILOVA, Jana, PROVAZNIK, Ivo, and **SEDLAR, Karel**. Transcriptomic studies of solventogenic clostridia, Clostridium acetobutylicum and Clostridium beijerinckii. *Biotechnology Advances*. 2022. 58, p. 107889. (published online in 2021, 2021 IF = 17.681, D1 in BIOTECHNOLOGY & APPLIED MICROBIOLOGY)

DOI: https://doi.org/10.1016/j.biotechadv.2021.107889

Biotechnology Advances 58 (2022) 107889

Contents lists available at ScienceDirect

Biotechnology Advances

journal homepage: www.elsevier.com/locate/biotechadv

Research review paper

Transcriptomic studies of solventogenic clostridia, *Clostridium acetobutylicum* and *Clostridium beijerinckii*

Petra Patakova [a,*], Barbora Branska [a], Maryna Vasylkivska [a], Katerina Jureckova [b], Jana Musilova [b], Ivo Provaznik [b], Karel Sedlar [b]

[a] University of Chemistry and Technology Prague, Technicka 5, 16628 Prague 6, Czech Republic
[b] Brno University of Technology, Technicka 10, 61600 Brno, Czech Republic

ARTICLE INFO

Keywords:
Transcriptomic analysis
RNA-Seq
Microarray
Clostridium acetobutylicum
Clostridium beijerinckii
ABE fermentation
Sporulation
Lignocellulose inhibitors

ABSTRACT

Solventogenic clostridia are not a strictly defined group within the genus *Clostridium* but its representatives share some common features, i.e. they are anaerobic, non-pathogenic, non-toxinogenic and endospore forming bacteria. Their main metabolite is typically 1-butanol but depending on species and culture conditions, they can form other metabolites such as acetone, isopropanol, ethanol, butyric, lactic and acetic acids, and hydrogen. Although these organisms were previously used for the industrial production of solvents, they later fell into disuse, being replaced by more efficient chemical production. A return to a more biological production of solvents therefore requires a thorough understanding of clostridial metabolism. Transcriptome analysis, which reflects the involvement of individual genes in all cellular processes within a population, at any given (sampling) moment, is a valuable tool for gaining a deeper insight into clostridial life. In this review, we describe techniques to study transcription, summarize the evolution of these techniques and compare methods for data processing and visualization of solventogenic clostridia, particularly the species *Clostridium acetobutylicum* and *Clostridium beijerinckii*. Individual approaches for evaluating transcriptomic data are compared and their contributions to advancements in the field are assessed. Moreover, utilization of transcriptomic data for reconstruction of computational clostridial metabolic models is considered and particular models are described. Transcriptional changes in glucose transport, central carbon metabolism, the sporulation cycle, butanol and butyrate stress responses, the influence of lignocellulose-derived inhibitors on growth and solvent production, and other respective topics, are addressed and common trends are highlighted.

1. Introduction

Maintaining acceptable living conditions on Earth for future generations is a major challenge for contemporary science. As humanity, we have unexpected and surprising allies in meeting this challenge, but we have not yet used their help enough. These are microorganisms, some of which have experience beyond the memory of the human species with a dramatically changing climate. Such a group of evolutionarily ancient bacteria includes clostridia, which have the potential to produce various chemicals such as solvents, acids or hydrogen, and their production would not need to be based on oil, one of the pillars of our current civilization; we could process wastes of many types to recreate useful materials. Acetogenic clostridial species, capable of using CO_2 and/or CO as a carbon source, such as *Clostridium ljungdahlii*, are nowadays of great interest (for a recent review of their metabolic pathways see Katsyv and Müller (2020)).

In this paper, we described the transcriptome analysis of native butanol producing clostridia such as *Clostridium acetobutylicum*, *Clostridium beijerinckii* and *Clostridium saccharoperbutylacetonicum* that perform acetone-butanol-ethanol/isopropanol-butanol-ethanol (ABE/IBE) fermentations. In addition to native butanol producers, engineered butanol producing microorganisms have been developed, such as *Clostridium tyrobutyricum* (Bao et al., 2020; Li et al., 2020c) or *Clostridium cellulovorans* (Bao et al., 2021), transcriptomic studies of which are not included in this review. The ABE/IBE processes have a history of industrial use (Jones and Woods, 1986) but now, despite on-going research, are seldom used at an industrial scale. Recently, the entire ABE/IBE technology stream, from substrate preparation to product isolation, including an overview of the companies involved in industrial process development, was reviewed (Veza et al., 2021). The

* Corresponding author.
E-mail address: petra.patakova@vscht.cz (P. Patakova).

https://doi.org/10.1016/j.biotechadv.2021.107889
Received 27 August 2021; Received in revised form 10 December 2021; Accepted 14 December 2021
Available online 17 December 2021

fermentative decomposition of carbohydrates by solventogenic clostridia occurs as a two-step process. In the first phase, associated with cell multiplication, acetic acid, butyric acid, hydrogen and carbon dioxide are formed as the main products. In the second phase, solvents such as butanol, acetone or isopropanol, as well as fermentation gases, are produced from these carbohydrates and the acids. The metabolic switch depends on pH, acid concentration, population density and other factors; for reviews of ABE/IBE metabolic pathways description and regulation, see Gheshlaghi et al. (2009); Li et al. (2020b); Yang et al. (2018). Although the metabolic switch is the focus of research in this field, it is still not thoroughly understood and novel findings such as the potential influence of acetylation and butylation of cellular proteins on the shift have been reported recently (Xu et al., 2018). Solventogenesis can, but need not necessarily take place simultaneously with sporulation in clostridia. There are also gaps in our complete understanding of solventogenic clostridial sporulation, despite progress made in recent years (Al-Hinai et al., 2015; Diallo et al., 2021).

Transcriptomic studies complement our knowledge of the genomes of individual strains and give us a much better idea of the course of individual processes (e.g. nutrient transport, metabolite production, cell movement or sporulation) at the level of the cell population. Transcriptomics data for individual genes clearly tell us which genes are actually used by cells and which are present in their genome but are not used. This review is focused on the acquisition and use of transcriptomic data of solventogenic clostridia, in particular *C. acetobutylicum* and *C. beijerinckii* because other solventogenic clostridia have not previously been used for a transcriptomic study with the exception of *C. saccharoperbutylacetonicum* (Li et al., 2020a). Such a review has never been compiled, although some published review articles have focused on omics insights into microbial butanol/alcohol tolerance (Arsov et al., 2021; Horinouchi et al., 2018; Sandoval and Papoutsakis, 2016). The first part of this review is focused on a brief introduction of techniques for obtaining transcriptomes, with the specific aim of techniques used for studying transcription in solventogenic clostridia. Data processing, visualization and evaluation techniques are summarized and linked to particular studies in which they were used. The use of data for reconstruction of biological computational models by means of systems biology is also reviewed. We also present topics studied by transcriptomic techniques in solventogenic clostridia, in particular saccharide consumption, central metabolic pathways, sporulation, metabolite (s) stress and the effect of lignocellulose-derived inhibitors. This review provides a comparison between *C. acetobutylicum* and *C. beijerinckii* species, shows the possibilities of using transcriptomic data and may become a useful tool for the evaluation of new transcriptomic studies not only of solventogenic clostridia.

2. Transcriptomic techniques

Measuring the expression of genetic information in various organisms is a standard research procedure that reveals microbial behavior by providing information on regulation of their genes. It can even help infer novel, hitherto unannotated genes, or to explain the functions of poorly described genes. Transcriptomes of solventogenic clostridia started to be studied in early 2000s, after the advent of high-throughput techniques that allowed to perform transcriptomic experiments even in non-model organisms at reasonable costs. Nevertheless, one low-throughput technique, reverse transcriptase quantitative PCR (RT-qPCR) (Becker-André and Hahlbrock, 1989) has survived to the present day. The technique focuses only on selected genes by using specific primers. Selected mRNA is transcribed into complementary DNA (cDNA) by reverse transcriptase and quantitative PCR is performed. Although the technique is limited by the need to design a primer, it is relatively cheap and is still used in studies where only a limited number of genes are tracked. In solventogenic clostridia, RT-qPCR is usually utilized in studies of mutant strains where expression of one or several key genes is altered, such as overexpression of the *spo0A* gene in *C. acetobutylicum* (Alsaker et al., 2004) or in *C. beijerinckii* (Kolek et al., 2017), or for verification of results gathered by high-throughput techniques (see chapter below). For accurate quantification of expression and its comparison between several samples, RT-qPCR needs a suitable reference gene (Radonić et al., 2004). Even though some reference genes in the bacterial domain are defined, only a limited number of them have been confirmed by two or more studies (Rocha et al., 2015). In solventogenic clostridia, suitable reference genes may vary even between two strains of the same species, as in the case of *C. beijerinckii* NRRL B-598, where *zmp* and *greA* were found to be the most suitable reference genes, and NCIMB 8052 with *pepT* being selected for RT-qPCR (Jureckova et al., 2021; Wang et al., 2011).

Currently, clostridial transcriptomics heavily relies upon utilization of high-throughput sequencing of transcriptomes, i.e. RNA-Seq (Wang et al., 2009). Before RNA-Seq, transcriptome profiling was primarily performed with microarrays (Schena et al., 1995). To measure expression with microarrays, isolated transcripts were transcribed into more stable cDNA and a fluorescence marker was added to each fragment. These fragments were later hybridized on a chip with complementary oligonucleotides. Thus, microarrays can only measure expression of genes whose sequences are known. Nevertheless, since the chip can contain hundreds of oligonucleotides, microarray analysis belongs to the category of high-throughput techniques. The first completed genome among solventogenic clostridia was that of *C. acetobutylicum* ATCC 824, assembled in 2001 (Nölling et al., 2001). Only two years later, expression of this strain and its mutants became widely studied (Alsaker et al., 2004; Alsaker and Papoutsakis, 2005; Tomas et al., 2003a). At that time, genome-wide transcription in other solventogenic clostridia was not possible due to their missing genomes and therefore the inability to propose a suitable microarray. Following advances in next-generation sequencing (NGS) techniques, transcriptomes of other solventogenic clostridia were studied, mainly by RNA-Seq that allowed the capture of genome-wide gene expression by sequencing complete transcriptomes without the need for any *a priori* information. This technique advanced studies of transcriptomes in other clostridial species and strains, primarily *C. beijerinckii* and *C. saccharoperbutylacetonicum*, see Supplement Table 1. RNA-Seq itself covers two main steps. The cornerstone of the first step is library preparation, which includes processes such as fragmentation, RNA conversion into cDNA, and addition of adaptors. High-throughput sequencing of short fragments is then performed. The output of the experiment, fragments or more usually referred to as reads, are then analyzed using bioinformatics. The schematic overview of currently used transcriptomic techniques is shown in Fig 1, while particular techniques and data processing used within particular clostridial studies are described in the following subchapters.

Analysis of expression starts with isolation of RNA which is transcribed into more stable cDNA. In the case of RNA-Seq, highly abundant rRNA that does not contribute to quantification of expression is removed by ribodepletion. Data processing differs between particular techniques,

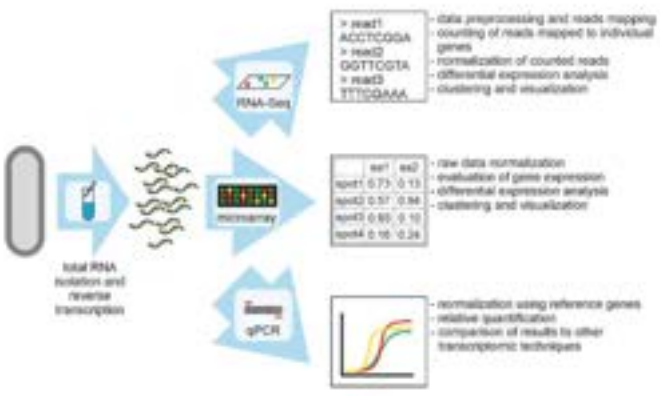


Fig. 1. Overview of transcriptomic techniques used for quantification of expression in clostridia.

while RNA-Seq provides information on transcription of all genes, data processing is the most complicated.

2.1. Data processing

2.1.1. Microarray

Transcriptome profiling performed with microarrays is based on array scanning, followed by data extraction and analysis. Some of the most widely used scanner devices in clostridial studies are: Agilent DNA Microarray scanner (Ren et al., 2012; Servinsky et al., 2010; Wang et al., 2013a; Zhang et al., 2017; Jiao et al., 2016; Liu et al., 2016b; Alsaker et al., 2005), GenePix 4000B (Grimmler et al., 2011; Grimmler et al., 2010; Hillmann et al., 2009), GSI Lumonics scanner (Tomas et al., 2003a, 2003b; Tummala et al., 2003), Axon 4000B (Shi and Blaschek, 2008; Zhang and Ezeji, 2013), and Tecan MS200 (Yoo et al., 2017; Nguyen et al., 2016; Yoo et al., 2016). Next, the output, in the form of an image, is processed using software such as NimbleScan (Servinsky et al., 2010), GenePix Pro (Grimmler et al., 2011; Grimmler et al., 2010; Hillmann et al., 2009; Shi and Blaschek, 2008; Zhang and Ezeji, 2013), Agilent's Feature Extraction Software (Wang et al., 2013a; Jiao et al., 2016; Liu et al., 2016b), Genesis (Schwarz et al., 2012), Molecular Annotation System (Zhang et al., 2017), or ScanArray (Tomas et al., 2003a, 2003b) with the aim of extracting raw data.

Raw data normalization is needed to eliminate differences between samples that originate from technical aspects of microarray handling that may bias biological differences in a given experimental setup (Jaksik et al., 2015). There are several methods: quantile normalization using the whole set of genes as the reference set (Liu et al., 2019) used in studies by Ren et al. (2012); Liu et al. (2016a) and Servinsky et al. (2010); setting the arithmetic mean of the ratios equal to 1 (Grimmler et al., 2011; Grimmler et al., 2010; Hillmann et al., 2009); segmental nearest-neighbor logarithmic expression method (SNN-LERM) (Yang et al., 2003) used by Alsaker and Papoutsakis (2005); Wang et al. (2013a); Alsaker et al. (2010) and Alsaker et al. (2005).

Subsequently, additional analyses are processed based on the study goal. Usually, gene expression is evaluated using logarithmic transformation. In the article of Ren et al. (2012), fold changes in expression of selected genes were investigated by dividing the normalized signal intensities of the same gene at each time point with a threshold for differential expression equal to a 2-fold change. Servinsky et al. (2010) generated and subsequently $\log_2$-transformed gene calls and considered mean expression levels of arrays. In order to select significantly regulated genes, several studies (Schwarz et al., 2012; Janssen et al., 2012; Hönicke et al., 2014; Grimmler et al., 2011; Liu et al., 2016b) describe computation of logarithmic expression ratios, setting a threshold ≥ 1.6 and ≤ -1.6; ≥ 1.0 and ≤ -1.0; or ≥ 1.3 and ≤ -1.3 to define upregulated or repressed genes, respectively. In several studies (Wang et al., 2013a; Zhang et al., 2017; Dash et al., 2014), Significant Analysis of Microarrays (SAM) was used to analyze data by pairwise and point-by-point comparison with the aim of revealing differentially expressed genes for metabolite stress or gene expression profiles. In the next study (Alsaker et al., 2010), differential-gene expression was identified using a segmental nearest-neighbor approach. Tummala et al. (2003) used principal component analysis (PCA) in order to clean data for clustering.

Finally, clusters or groups of expression similarities are computed by average-linkage hierarchical clustering, k-means clustering, and self-organizing maps (SOMs). Two former methods are usually processed using software Cluster (Ren et al., 2012; Tomas et al., 2003a, 2003b; Alsaker et al., 2005) or TIGR MeV (Alsaker and Papoutsakis, 2005; Wang et al., 2013a; Alsaker et al., 2010). SOMs are generated by an unsupervised machine learning technique usually using GeneCluster software (Alsaker and Papoutsakis, 2005; Tomas et al., 2003a, 2003b; Tummala et al., 2003).

Following visualization of gene clusters in heat plots or Eisen plots is usually carried out using Java TreeView (Alsaker et al., 2005; Liu et al., 2016a; Ren et al., 2012; Tomas et al., 2003a, 2003b; Tummala et al., 2003; Zhang and Ezeji, 2013) or TIGR MeV (Alsaker and Papoutsakis, 2005; Wang et al., 2013a; Alsaker et al., 2010; Shi and Blaschek, 2008).

2.1.2. RNA-Seq

The RNA-Seq technique is based on sequencing using high-throughput devices, usually Illumina. The most frequently used Illumina's platforms are the following: HiSeq 2000 (Liu et al., 2016b; Tan et al., 2015; Venkataramanan et al., 2015; Venkataramanan et al., 2013; Wang et al., 2013b; Wu et al., 2015), HiSeq 2500 (Diallo et al., 2020; Luo et al., 2020; Ralston and Papoutsakis, 2018; Venkataramanan et al., 2015), HiSeq 4000 (Herman et al., 2017; Li et al., 2020a; Sedlar et al., 2018), and NextSeq 500 (Germane et al., 2018; Patakova et al., 2019; Sedlar et al., 2018, 2019; Servinsky et al., 2018). Other differences between clostridial studies are given by library preparation, the read length varying from 50 bp to 250 bp, the sequencing depth ranging from 4.7 to 32 million reads per sample; reads, which are single-end or paired-end, or even strands specific (see Supplement Table I). In addition, differences are also caused by bioinformatics data processing, performed in four basic steps: data preprocessing and reads mapping; counting of reads mapped to individual genes; normalization of counted reads; differential expression analysis of genes.

Preprocessing step covers adapters and quality trimming of the reads and usually is performed by Trimmomatic (Bolger et al., 2014) software (Germane et al., 2018; Patakova et al., 2019; Ralston and Papoutsakis, 2018; Sedlar et al., 2018, 2019; Servinsky et al., 2018). Remaining contamination of 16S and 23S rRNA after ribodepletion can be removed by the SortMeRNA (Kopylova et al., 2012) tool (Patakova et al., 2019; Sedlar et al., 2018, 2019). Cleansed reads are further used for mapping to a reference genome or for *de novo* assembly of transcripts(Hölzer and Marz, 2019). Selection of the right alignment tool is an important step in transcriptomic analysis, but one universal aligner does not exist and the right choice always depends on specific sequencing data and the goal of the study (Baruzzo et al., 2017; Musich et al., 2021). In solventogenic clostridia, most studies map reads directly to the reference genome by various tools such as STAR (Dobin et al., 2013) aligner (Patakova et al., 2019; Sedlar et al., 2018, 2019), Bowtie (Langmead et al., 2009) aligner (Diallo et al., 2020; Tan et al., 2015), Bowtie2 (Jin et al., 2015; Li et al., 2020a; Liu et al., 2020; Liu et al., 2013; Ralston and Papoutsakis, 2018) or TopHat (Trapnell et al., 2009) software (Venkataramanan et al., 2015; Venkataramanan et al., 2013).

Following quantification of expression proved to be a very important step in RNA-Seq data analysis, because counting can underestimate or overestimate the level of gene expression (Corchete et al., 2020). There are two popular quantification tools used in clostridial studies: HTSeq (Anders et al., 2015) python package (Li et al., 2020b; Liu et al., 2020) and similar R/Bioconductor featureCounts function (Liao et al., 2014) in the Subread package (Patakova et al., 2019; Sedlar et al., 2018, 2019) both performing well in benchmark tests (Corchete et al., 2020; Liao et al., 2014).

Following normalization of raw count deals with differences in library sizes and lengths of the genes, expressed as RPKM (Reads per Kilobase Million) (Liu et al., 2016b; Patakova et al., 2019; Sedlar et al., 2018, 2019; Servinsky et al., 2018; Wang et al., 2011, 2012; Wang et al., 2013b; Wu et al., 2015) or TPM (Transcripts per Million) (Jureckova et al., 2021; Sandoval-Espinola et al., 2017). Both approaches are of different nature and only TPM, allows comparing expression between different samples (Wagner et al., 2012). Alternatively, a popular normalization technique is a built-in function in the R/Bioconductor DESeq2 package (Love et al., 2014), which is based on the negative binomial distribution and is used in a majority of clostridial studies (Diallo et al., 2020; Li et al., 2020a; Luo et al., 2020; Mäté de Gérando et al., 2018; Patakova et al., 2019; Ralston and Papoutsakis, 2018; Sedlar et al., 2018, 2019; Servinsky et al., 2018; Venkataramanan et al., 2013, 2015).

The goal of transcriptomic studies is usually differential expression analysis, which determines and statistically tests changes in gene

expression between different samples or conditions and identifies upregulated and downregulated genes. Commonly used algorithms for differential expression analysis are R/Bioconductor packages edgeR (Robinson et al., 2010) and DESeq2 (Diallo et al., 2020; Jureckova et al., 2021; Li et al., 2020a; Maté de Gérando et al., 2018; Patakova et al., 2021; Patakova et al., 2019; Ralston and Papoutsakis, 2018; Sedlar et al., 2018, 2019; Servinsky et al., 2018; Vasylkivska et al., 2019; Venkataramanan et al., 2013, 2015). Both tools are model-based algorithms using negative-binomial distribution, yet their performance differs based on parameters of RNA-Seq experiments such as sequencing depth or number of replicates (Zhang et al., 2014). Similarly to microarrays, gene expression profiles are often visually represented by heatmaps (Branska et al., 2021; Patakova et al., 2019; Vasylkivska et al., 2019; Wang et al., 2012).

2.1.3. RT-qPCR

RT-qPCR data in transcriptomics are often used for validation of the data from other genome-wide techniques, in solventogenic clostridia see (Diallo et al., 2020; Grimmler et al., 2010, 2011; Jiao et al., 2016; Jones et al., 2008; Ren et al., 2012; Schwarz et al., 2012; Servinsky et al., 2010; Shi and Blaschek, 2008; Tan et al., 2015; Wang et al., 2013a; Wang et al., 2011, 2012; Wu et al., 2015; Yang et al., 2020b; Zhang et al., 2017; Zhang and Ezeji, 2013). Validation is conducted by RT-qPCR experiments of selected genes together with appropriate reference genes, which are used for normalization of expression of tested genes. Relative quantification of expression can be performed, for example, by the mathematical model of Dr. Pfaffl (Pfaffl, 2001) as in the case of several studies (Diallo et al., 2020; Grimmler et al., 2010, 2011), by the delta-delta Ct method (Vandesompele et al., 2002) in the study of (Schwarz et al., 2012) or by the comparative C_T method (Schmittgen and Livak, 2008) in the study by (Zhang and Ezeji, 2013). Estimated fold changes are then compared with data from other transcriptomic techniques and their validity is proved by similar trends in gene regulation (Tan et al., 2015) or by high positive correlations between data (Jiao et al., 2016; Ren et al., 2012; Shi and Blaschek, 2008; Wang et al., 2011, 2012; Zhang et al., 2017; Zhang and Ezeji, 2013).

2.2. Data interpretation and comparison of results among studies

For several considerably different lab techniques, various machines to measure expression and a wide range of bioinformatics tools to process the data, comparisons of results from different studies are cumbersome. Moreover, the simple fact that the sequence of a particular gene can differ between two bacteria must be considered. The solution can be found in functional annotation, i.e. every gene is assigned a putative function. The very basic, yet widely used annotation is to assign a gene to a Cluster of Orthologous Genes (COG), sometimes also referred to as a Cluster of Orthologous Groups of proteins (Tatusov et al., 2000). The transcriptional response can then be summarized by counting up- or down-regulated genes in each category. Such a strategy was utilized in clostridial studies for microarrays (Alsaker and Papoutsakis, 2005) as well as for RNA-Seq (Sedlar et al., 2018). Responses of different strains or species are done by simply comparing absolute or relative abundancies of regulated genes in particular categories. Annotation can be done by searching for annotated orthologous genes in various databases, such as the recently updated COG database (Galperin et al., 2021) or EggNOG (Huerta-Cepas et al., 2019). Despite the fact that COG annotation is usually performed when a novel clostridial genome is published (Nölling et al., 2001; Sedlar et al., 2021a, 2021b), its utilization for comparison of various transcriptomic studies is not widely adopted. The reason probably lies in the design of particular studies. While there are several transcriptomic studies involving solventogenic clostridia, they measure transcription to different stimuli and under different conditions; see Supplementary Table I. While a comparison of these results using COG is technically possible, it is often not meaningful. Nevertheless, this procedure will be of great use when more transcriptomic studies are published.

The simple counting of genes in different categories might be very limiting since the total number of genes may vary among different studies, especially when microarrays are used. The solution is to calculate if a functional group is statistically significantly represented in data with the hypergeometric test (= one-sided Fisher's exact test). This can be used to confirm some presumption, e.g. some COG categories are over-represented among putative housekeeping genes in *C. beijerinckii* NCIMB 8052 (Wang et al., 2011). Similar, so called enrichment analyses, are more common in Gene Ontology (GO) (Harris et al., 2004). While COG provides only coarse-grained annotation, GO offers general as well as a number of more specific terms to be used. A similar analysis was used to summarize the transcriptional response of *C. beijerinckii* NRRL B-598 to butanol shock and to characterize putative housekeeping genes utilizable as reference genes for RT-qPCR (Jureckova et al., 2021; Sedlar et al., 2019). GO annotation can be constructed in a similar manner to COG annotation, by searching for annotated orthologues using, for example, Blast2GO (Conesa et al., 2005) or QuickGO (Binns et al., 2009). There are also other ways to perform a functional annotation, for example annotation using the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa and Goto, 2000), used to find enriched biological pathways in *C.acetobutylicum* (Liu et al., 2020) and to construct a biological model of *C. beijerinckii* (Milne et al., 2011).

2.3. Use of transcriptomes in biological models

Novel high-throughput techniques were game changers for the field of biological modeling. Modeling of biochemical systems that would explain and fully describe the metabolism of living organisms has its roots in the Michaelis-Menten equation (Srinivasan, 2021), which forms the basis of every kinetic model. However, these kinetics presume that the concentration of an enzyme remains constant during the whole time of the reaction. This is only achievable for a limited set of reactions and very short modeling time, and not for large genome scale models (GSM) containing tens or hundreds of reactions. As enzymes are products of proteosynthesis, their concentrations are changing continuously among others according to changing expression of genes encoding these enzymes. This regulation was never initially considered in GSM of solventogenic clostridia (Senger and Papoutsakis, 2008a, 2008b). Only later, was transcriptomic regulation taken into account for the kinetic modelling of clostridial ABE fermentation (Millat et al., 2013a). As the activity of various enzymes changes during acidogenic and solventogenic phases, the authors presumed that the total concentration of enzymes was not constant, but rather a function of pH. Nevertheless, the model was fitted using only fermentation data and no transcriptomic verification was done. This model was further augmented and the pH shift was linked with cell growth (Millat et al., 2013b). Regarding gene expression, this model also considered antagonistic expression of *adhE1* and *adhE2* genes encoding alcohol/aldehyde dehydrogenases. Similar to its previous version, no transcriptomic data were used. This applies also to other kinetic model that combines metabolic reactions, gene regulation, and environmental cues without experimental verification of gene expression (Liao et al., 2015). While the previously described models were proposed for the strain *C. acetobutylicum* ATCC 824, another GSM, iCM925, was derived from the *C. beijerinckii* NCIMB 8052 genome (Milne et al., 2011). In that case, due to missing transcriptomic data for the strain in that time, no changes in gene expression were considered in the model.

Thanks to the genome sequence and a couple of transcriptomic datasets available, *Clostridium acetobutylicum* ATCC 824 became a model organism for solventogenic clostridia and other kinetic models were proposed for the strain. The model iCac802 considers 802 genes, 1,137 metabolites, and 1,462 reactions (Dash et al., 2014), and considerably expands the previous model iCac490 (McAnulty et al., 2012). Transcriptomic regulation is fully incorporated into iCac802 thanks to a method called CoreReg (Dash et al., 2014). The method weights fluxes of

metabolites by changes in expression of genes participating in corresponding enzymatic reactions. Transcriptomic data used for adjusting the model were gathered using a set of microarrays examining transcriptional changes in response to chemical stress (Wang et al., 2013a). Another model, iCac967, considers changes in the transcriptome (Yoo et al., 2015). This model spans 967 genes and includes 1,058 metabolites participating in 1,231 reactions. An analysis of the transcriptome was carried out using microarrays under three different stable metabolic states: acidogenic, solventogenic, and alcohologenic. The authors performed precise quantification of particular mRNAs and supplemented the study with proteomics analyses. The model was manually adjusted according to the calculated numbers of particular molecules of mRNA per cell.

Besides GSMs, transcriptomic data can be utilized to infer additional types of models. As RNA-Seq technology measures transcriptoin of all genes at once, and other omics data are nowadays easily accessible, inference of gene regulatory networks (GRNs) capturing regulation among particular genes, became very popular (Mercatelli et al., 2020). Although some clostridial GRNs were already inferred, almost none of them spanned solventogenic clostridia (Ihekwaba et al., 2016; Ohtani and Shimizu, 2016; Saujet et al., 2014). The only exception was in the gene coexpression network of *C. acetobutylicum*, which was used to reveal novel metabolic mechanisms (Liu et al., 2020). Additionally, microarray data for *C. acetobutylicum* ATCC 824, which were used to adjust the GSM, were also used to propose a stress response network (SRN) (Wang et al., 2013a). This network consists of eight core, stress-related regulons and can be considered as a part of the *C. acetobutylicum* GRN. Moreover, since the stress-related genes are highly conserved, the proposed SRN is widely applicable to other solventogenic clostridia.

3. Transcriptomic insight into solventogenic clostridia physiology

A scheme of glucose transport, glycolysis and the main acidogenic and solventogenic pathways are shown in Fig. 2 for better orientation in the text. Carbohydrate uptake in bacteria occurs via either the phosphotransferase system (PTS) or via non-PTS mechanisms such as ABC transporters or symporters (Mitchell, 2016). If carbohydrate is transported by PTS, its phosphorylation is carried out by the PTS system, but for non-PTS uptake, another kinase activity is needed for phosphorylation. The central metabolic pathway of ABE fermentation was described

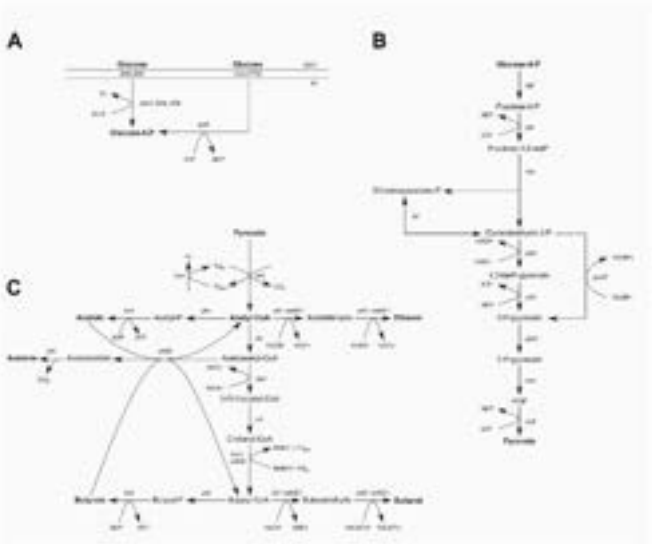


Fig. 2. Glucose transport (A), glycolysis (B), main acidogenic and solventogenic pathways, hydrogen production (C).

long ago and branches that lead to the most common products are known. However, transcriptomic studies enabled the identification or verification of the major genes involved in the ABE metabolic pathway, their coordinated or differing forms of regulation, suggestions or verification of putative operon structures, implied metabolic fluxes, and the involvement of alternative routes employing various isoenzymes. This is obviously different for *C. acetobutylicum* and *C. beijerinckii;* however, from a global point of view, during standard ABE fermentation, gene expression of the main metabolic pathways follows a similar pattern.

PTS - phosphotransferase system; EI - PTS enzyme EI; EI−P – phosphorylated PTS enzyme EI. Gene products: *ack* - acetate kinase; *adc* - acetoacetate decarboxylase; *adhE1* - NADH-dependent bifunctional alcohol/aldehyde dehydrogenase; *ald* - aldehyde dehydrogenase; *adh* – alcohol dehydrogenase; *bcd* - acyl-CoA dehydrogenase (butyrylCoA dehydrogenase); *bdh* - butanol dehydrogenase; *buk* - butyrate kinase; *crt* – crotonase; *ctfA* and *ctfB* - CoA transferase subunit alpha and beta, respectively; *EIIA, EIIB, EIIC, EIID* - PTS enzyme EII A, B, C, D subunits, respectively; *eno* – enolase; *etfA* and *etfB* - electron transfer flavoprotein subunit alpha and beta; *fba* - fructose-1,6-bisphosphate aldolase; *gap* - glyceraldehyde-3-phosphate dehydrogenase; *gapN* - NADP-dependent glyceraldehyde-3-phosphate dehydrogenase; *glcK* – glucokinase; *gpm* - phosphoglycerate mutase; *hbd* - 3-hydroxybutyryl-CoA dehydrogenase; *hyd* - hydrogenase; *pfk* - 6-phosphofructokinase; *pfo* - pyruvate:ferredoxin oxidoreductase; *pgi* - glucose-6-phosphate isomerase; *pgk* - phosphoglycerate kinase; *pta* - phosphate acetyltransferase; *ptb* - phosphate butyryltransferase; *ptsH* - histidine-containing protein HPr; *pyk* - pyruvate kinase; *thl* -thiolase; *tpi* - triose-phosphate isomerase. Fd – ferredoxin.

Clostridia are known for their great metabolic flexibility and exceptional ability to grow on different kinds of substrates. The ability to utilize gaseous substrates such as mixtures of CO_2, H_2 and CO are connected usually with *C. ljungdahlii* (for the transcriptomic studies comparing autotrophic and heterotrophic growth of the species see Aklujkar et al. (2017); Tan et al. (2013)) and similar clostridia, however, it was also confirmed at the transcriptional level in *C. beijerinckii* SA-1, a mutant of *C. beijerinckii* NCIMB 8052 with a butanol tolerant phenotype (Sandoval-Espinola et al., 2017).

It is common for solventogenesis to occur simultaneously with sporulation. If neither solventogenesis nor sporulation occurs, or both processes are significantly impaired, we can speak of so-called culture degeneration. The permanently degenerated phenotype is exhibited by mutants of production strains due to various causes, for example after loss of the megaplasmid in the case of *C. acetobutylicum* ATCC 824 (Cornillot et al., 1997), or after multiple sub-culturing without sporulation in the case of *C. beijerinckii* (Lv et al., 2016). A short characterization of strains, described in the text, is given in Table 1. For a phylogenetic tree depicting distances among most of them, see Cruz-Morales et al. (2019), and for comprehensive and current characterization of *C. beijerinckii* strains, see Sedlar et al. (2021a). In some transcriptomic studies, mutants of parent strains with deactivation or overexpression of one or multiple genes were used and these are not included in Table 1.

In the transcriptomic studies of solventogenic clostridia, not only different techniques and data evaluations were used, described in section 2, but also different experimental and cultivation designs. Cultivation was performed either in batch or continuous chemostat modes where both options have their advantages and disadvantages. The cell cycle, including sporulation and the two metabolic phases, acidogenesis and solventogenesis, as well as the metabolite(s) stress response elicited by both acids and solvents, can be monitored during batch culture. However, transition of the population from one phase to another and the onset of sporulation often overlap, making it difficult to evaluate both processes independently. In addition, both acidic and solvent stresses are frequently coupled with a drop in population viability, which complicates obtaining mRNA from live cells. A typical example of such a study was performed using *C. beijerinckii* NRRL B-598 (Patakova et al.,

Table 1
Characterization of strains used for transcriptomic studies and mentioned in the text

Strain	Genome and mutant generation (if relevant)	Characteristics	Reference (where relevant)
C. acetobutylicum ATCC 824	3.94 Mbp chromosome 0.19 Mbp pSOL1 megaplasmid	Wild type strain having *sol* operon and certain sporulation genes on megaplasmid The most frequently studied *C. acetobutylicum* strain	-
C. acetobutylicum ATCC 55025	4.07 Mbp chromosome 0.19 Mbp pSOL1 megaplasmid	Asporogenic strain derived from *C. acetobutylicum* ATCC 4259	(Xu et al., 2017)
C. acetobutylicum M5	Mutant without megaplasmid MNNG treatment	The strain derived from ATCC 824 Degenerative phenotype	(Clark et al., 1989; Comfiet et al., 1997)
C. acetobutylicum EA 2018	Mutant of natural isolate MNNG treatment	Non- sporulating butanol hyper-producer	(Hu et al., 2011)
C. acetobutylicum CGMCC 5234 (B3)	Mutant UV irradiation mutagenesis	Production of extracellular polysaccharide	(Ying et al., 2012; Zhuang et al., 2016)
C. beijerinckii NCIMB 8052	6.0 Mbp	The most frequently studied *C. beijerinckii* strain	-
C. beijerinckii DG 8052	Mutant Multiple sub-culturing without sporulation	The strain derived from NCIMB 8052 Degenerative phenotype	(Lv et al., 2016)
C. beijerinckii BA 101	Mutant MNNG + 2-deoxyglucose treatment	The strain derived from NCIMB 8052; butanol hyper-producer, hyperamylolytic	(Annous and Blaschek, 1991)
C. beijerinckii BA 105	Mutant MNNG + 2-deoxyglucose treatment	The strain derived from NCIMB 8052 Glucose derepressed mutant	(Annous and Blaschek, 1991)
C. beijerinckii SA-1	Mutant Directed evolution	The strain derived from NCIMB 8052, butanol tolerant phenotype	(Noar et al., 2014)
C. beijerinckii NRRL B-598	6.19 Mbp	Originally identified as *C. puniceum*	(Sedlar et al., 2015, 2017)
C. beijerinckii DSM 6423	6.38 Mbp 2 small plasmids, 1 double stranded linear phage	IBE fermentation strain	(Mate de Gerando et al., 2018)
C. saccharoperbutylacetonicum N1-4	6.66 Mbp Chromosome 0.14 megaplasmid	ABE fermentation strain	(Gu et al., 2019)

ABE - acetone, butanol, ethanol; IBE - isopropanol, butanol, ethanol; MNNG - N-Methyl-N'-nitro-N-nitrosoguanidine; UV - ultraviolet.

2019). On the other hand, continuous phosphorus-limited chemostat cultures with pH regulation were used for separate studies of acidogenesis (pH 5.7) and solventogenesis (pH 4.5) in *C. acetobutylicum* ATCC 824 (for a typical example of such study, see Grimmler et al., 2011). In this continuous cultivation, sporulation was suppressed, which actually facilitated the study of both metabolic processes and excluded change of pH as an additional stress. As the continuous culture with separate metabolic phases was performed using only *C. acetobutylicum*, there is an open question whether it would be possible to perform the experiment under similar conditions with *C. beijerinckii*. In addition, sampling plans are not comparable within the studies. In some cases, mutant(s) behavior was also compared with that of the parent strain. For an overview of transcriptomic studies in solventogenic clostridia, as well as transcriptomic techniques and databases where the data are stored, see Supplement Table 1.

3.1. Glucose transport and phosphorylation

Glucose transport mainly occurs via the PTS in solventogenic clostridia (Mitchell, 2016). In *C. acetobutylicum* ATCC 824, two sets of PTS EII transporters belonging to the glucose family PTS EII proteins, were identified. While the genes encoding one of the sets were expressed constitutively, expression of genes encoding the second set was induced only in the presence of glucose (Servinsky et al., 2010). Down-regulation of PTS EII genes was reported for a mutant strain with *ccpA* (gene encoding the pleiotropic regulator of carbon metabolism CcpA) inactivation, suggesting that CcpA may be necessary for efficient expression of PTS EII genes in the strain (Ren et al., 2012). Moreover, a role of glucose PTS EII genes in carbon catabolite repression (CCR) was suggested (Wu et al., 2015; Xiao et al., 2011). Differential upregulation of genes encoding the glucose family PTS EII genes was reported in the strain when medium was supplemented with zinc or sodium sulfide (Jin et al., 2015; Wu et al., 2015). On the other hand, genetic manipulation of the strain, such as inactivation of *ccpA* or deletion of *buk* and *ptb* genes, (Ren et al., 2012; Yoo et al., 2017) or acid stress (Alsaker et al., 2010), resulted in their downregulation.

In the case of non-PTS uptake, *glcK* encoding glucokinase exhibited high expression throughout the cultivation in *C. acetobutylicum* ATCC 824 (Jones et al., 2008). Alteration of its expression observed in different studies was triggered either by acid/ solvent stress (downregulation) (Alsaker et al., 2010; Alsaker and Papoutsakis, 2005; Janssen et al., 2010, 2012) or by medium supplementation with zinc or sodium sulfide (upregulation) (Jin et al., 2015; Wu et al., 2015).

Although glucose family PTS transporters take part in sugar uptake during exponential and early stationary phases in *C. beijerinckii* NCIMB 8052 (Wang et al., 2012), it seems that for the strain, as well as for some other *C. beijerinckii* strains, glucose uptake by PTS mainly occurs via mannose family PTS EII proteins (Vasylkivska et al., 2019; Wang et al., 2012). Moreover, different, independent studies (Liu et al., 2016b; Patakova et al., 2019; Wang et al., 2012) show that the switch from PTS to non-PTS uptake can be observed for glucose utilization by different *C. beijerinckii* strains, with a prevalence of PTS during acidogenesis and ATP-dependent glucose phosphorylation during solventogenesis. The same was true for the butanol-overproducing mutant strain *C. beijerinckii* BA101 (offspring of *C. beijerinckii* NCIMB 8052 obtained by random mutagenesis), however, expression of genes encoding mannose family PTS EII proteins in the mutant was reported to be much lower than in the wild type (Shi and Blaschek, 2008). This corresponds to additional evidence that in *C. beijerinckii* BA101, PTS transport is partly impaired (Lee and Blaschek, 2001) and non-PTS uptake might play an important role in the overproducing phenotype (Lee et al., 2005). On the other hand, in *C. beijerinckii* BA105 (offspring of *C. beijerinckii* NCIMB 8052 obtained by random mutagenesis, exhibiting a glucose derepressed phenotype),

elevated expression of genes encoding mannose family PTS EII proteins was observed throughout the fermentations, significantly higher than that of the wild type (Seo et al., 2017). Such expression of PTS EII genes in the mutant resulted in increased glucose uptake and acid production and, ultimately, contributed to an acid crash when pH was not regulated (Seo et al., 2017).

3.2. Carbon catabolite repression (CCR)

Carbon catabolite repression (CCR) usually means preferential use of glucose from the mixture of saccharides and this preference is regulated by transcriptional activation or repression, or by translational control in bacteria. Effective utilization of all carbohydrates, i.e. mainly glucose, xylose, arabinose and galactose present in hydrolysates of lignocellulosic biomass, an abundant waste substrate, can be a key point for more efficient ABE/IBE processes. Mixtures of saccharides can also be found in seaweed or sorted municipal waste hydrolysates, which represent other frequently studied sources of carbon. As diauxic utilization of most of the saccharides in the presence of glucose is one of the bottlenecks for clostridia fermenting these substrates, a study of CCR regulatory mechanisms is very important.

Although most transcriptomic studies of solventogenic clostridia used d-glucose as a sole carbon source (see Supplement Table I), there are a few exceptions. For example, transcriptomic data of *C. acetobutylicum* ATCC 824 exhibiting a typical diauxic growth pattern in defined minimal medium containing d-glucose and d-xylose revealed that among the genes subjected to catabolite repression by glucose were one of the two putative operons involved in d-xylose degradation, genes encoding sugar transporters, xylanases and endoglucanases (Grimmler et al., 2010). It was shown that disruption of *ccpA* in *C. acetobutylicum* ATCC 824 eliminates catabolite repression of xylose utilization in the presence of glucose (Ren et al., 2010). Disruption of *ccpA* in a mutant strain resulted in higher expression of genes encoding transporters and enzymes catalyzing metabolism of non-glucose carbohydrates, including pentoses (Ren et al., 2012). Moreover, *ccpA* was shown to be involved in the regulation of key genes of solventogenesis and sporulation in the strain (Ren et al., 2012).

In *C. acetobutylicum* ATCC 824, CCR was also observed for growth in a mixture of arabinose and xylose, with the former being utilized preferentially, suggesting that the strain has a hierarchy in pentose utilization (Aristilde et al., 2015). Transcriptional analysis of samples taken after addition of d-arabinose to a culture growing on d-xylose revealed modulation of transcription of genes involved in xylose utilization and other ccpA-controlled genes (Servinsky et al., 2018). It was suggested that the catabolite repression phosphocarrier, histidine protein Crh, may take part in the process and that arabinose can activate the protein. Moreover, a shift in pentose metabolism from expression of pentose phosphate pathway genes to phosphoketolase pathway genes was observed in this study, resulting in increased acetate production.

Expression of genes encoding different metabolic pathways was also observed for *C. beijerinckii* DSM 6423 cultivated on d-glucose or l-rhamnose as a sole carbon source (Diallo et al., 2018). IBE was produced on glucose medium, and 1,2-propanediol, propanol and propionate were produced on rhamnose. Interestingly, the same study reports simultaneous utilization of substrates when *C. beijerinckii* DSM 6423 was grown on a mixture of d-glucose and l-rhamnose, mimicking the hydrolysate of the seaweed *Ulva latuca*.

In *C. beijerinckii* NCIMB 8052, preferential utilization of glucose was observed during growth on glucose and glucitol (Tangney et al., 1998), as well as on a combination of glucose, xylose, cellobiose and arabinose (Zhang et al., 2012a), but when glucose and fructose were used as carbon sources, no CCR was observed (Mitchell, 1996). It seems that CCR in *C. beijerinckii* species might not be as strong as in *C. acetobutylicum* ATCC 824 as, for example, addition of glucose to xylose-containing growth medium for the culture of *C. beijerinckii* NCIMB 8052 resulted in the co-utilization of substrates (Birgen et al., 2018). Simultaneous utilization of glucose and xylose was also observed for *C. beijerinckii* SE-2 isolated from sludge (Zhang et al., 2016). *C. saccharoperbutylacetonicum* N1-4 also did not exhibit CCR when grown in a mixture of saccharides (Noguchi et al., 2013). Unfortunately, transcriptional analyses of these strains growing on the mixed substrates are still not available.

3.3. Nutritional needs based on transcriptomic data

Uptake and utilization of non-carbohydrate components of a culture medium are equally important for ABE/IBE processes. While bacteria are able to synthesize most of their required nutrients, for example, vitamins or amino acids, their uptake from the culture medium is more energetically advantageous for the cell (Jaehme and Slotboom, 2015). Different types of membrane transporters, such as electrochemical potential-driven transporters, primary active transporters and group translocators, take part in nutrient uptake (Vasylkivska et al., 2019).

Nutritional requirements of solventogenic *Clostridium* strains change during each stage of growth and production, as well as under stress conditions. This is reflected in the expression patterns of genes encoding uptake and utilization of different nutrients (Heluane et al., 2011; Vasylkivska et al., 2019). According to the differential expression of genes connected to nutrient transport in *C. beijerinckii* NRRL B-598, glutamine, methionine, zinc, niacin, thiamine and biotin seem to play important roles in butanol production (Vasylkivska et al., 2019). For *Clostridium acetobutylicum* ATCC 824 vitamin B_{12}, riboflavin, tryptophan, glutamine, asparagine, cysteine and histidine may play important roles under butanol stress. These finding were based on transcriptomic analysis of differentially expressed genes connected to the transport and metabolism of nutrients (Alsaker et al., 2010; Heluane et al., 2011). A new medium composition based on these findings was proposed to support growth and butanol production in continuous culture. This medium was tested on *C. beijerinckii* SA-1 and under test conditions, stable butanol production and prolonged vegetative state were achieved (Heluane et al., 2011).

3.4. Central metabolic pathway

Glycolytic genes are quite steadily expressed without apparent temporal regulation (Jones et al., 2008) or slight declination towards late solventogenesis (Patakova et al., 2019; Shi and Blaschek, 2008; Wang et al., 2012) with a high number of transcripts (Máté de Gérando et al., 2018). In a biofilm-growing culture (Liu et al., 2016a) or zinc-supplemented culture (Wu et al., 2015), these genes were upregulated, and this was simultaneously accompanied by a higher rate of glucose utilization.

Within the part of glycolysis pathway that serves for transformation of glyceraldehyde-3-P to phosphoenolpyruvate, an alternative route for direct 3-phosphoglycerate formation from glyceraldehyde 3-phosphate producing NADPH can be found (see Fig. 2). It is hypothesized that this by-pass might represent the possible source of NADPH for NADPH-dependent butanol dehydrogenase. Upregulation of the *gapN* (NADPH-dependent glyceraldehyde-3-phosphate dehydrogenase) was detected e.g. in a biofilm growing *C. acetobutylicum* ATCC 824. However, comparing the level of *gapN* and *gap* gene transcripts suggests only limited flux through GapN, accounting for less than 5 % of NADPH required (Yoo et al., 2015), which seems insufficient to provide NADPH for butyraldehyde reduction using butanol dehydrogenase. In *C. acetobutylicum* ATCC 824, two bifunctional aldehyde/alcohol dehydrogenases, AdhE1 and AdhE2, are believed to contribute to solvent formation. First, *adhE1* gene, which is part of the *sol* operon, was verified in transcriptomic studies to encode a major butyryl-CoA dehydrogenase, playing a role in solventogenesis (Grimmler et al., 2011; Janssen et al., 2010; Yoo et al., 2016). Nevertheless, this "bifunctional" dehydrogenase exhibited high butyraldehyde dehydrogenase activity, but only marginal butanol dehydrogenase ability (Yoo et al., 2015), suggesting the need for an alternative butanol dehydrogenase (Bdh) for butyraldehyde

reduction. Both *adhE1* and *adhE2* seem to be regulated in an antagonistic manner, with *adhE2* being the major gene expressed in acidogenesis (Grimmler et al., 2011) and alcohologenesis (Yoo et al., 2016).

Under solventogenesis, more than one butanol dehydrogenase might contribute to the final step in butanol formation and at least three of them are possible: BdhA, BdhB, and BdhC (Yoo et al., 2017) with the highest flux attributed to BdhB. The increase in expression of *bdhB* coincided with the onset of solventogenesis (Janssen et al., 2010; Jones et al., 2008) as well as *bdhA* (Alsaker and Papoutsakis, 2005). Both *bdhA* and *bdhB* were down-regulated (together with *adhE1*, *ctfAB*, and *adc*) in nonsolventogenic *C. acetobutylicum* with inactivated *spo0A* (Tomas et al., 2003a). However, another study (Cooksley et al., 2012) disputes the importance of these two butanol dehydrogenases because their mutations did not alter solvent formation. The contribution of Bdh in *C. beijerinckii* is even less clear. Only aldehyde dehydrogenase activity is assumed for protein encoded by the *ald* situated in the *sol* operon. There are several genes coding for Bdhs with elevated expression of some of them from the beginning of fermentation (Máté de Gérando et al., 2018; Wang et al., 2012) through late acidogenesis (Shi and Blaschek, 2008) to later solventogenesis (Jiao et al., 2016; Patakova et al., 2019).

In both species, the core solventogenic genes are identically up-regulated with the onset of solventogenesis. Unlike *C. acetobutylicum*, in *C. beijerinckii*, the expression of solvent-related genes was observed from the early stages of fermentation (Máté de Gérando et al., 2018; Patakova et al., 2019), and might be accompanied by low but detectable solvent production. Expression profile of acetate (*pta*, *ack*) and butyrate (*ptb*, *buk*) synthesis-related genes was apparently influenced by the chosen sampling points and the reference transcriptome. Both *pta* plus *ack* and *ptb* plus *buk* form bicistronic operons with coordinated expressions in these pairs. In *C. beijerinckii* DSM 6423, no differential expression of any of these genes was recorded throughout the IBE fermentation (Máté de Gérando et al., 2018). A similar pattern was recorded for *ack-pta* in *C. acetobutylicum* 824 (Jones et al., 2008) but not for *ptb-buk*, which peaked during late exponential phase and declined afterwards. Alsaker and Papoutsakis recorded a transient up-regulation of *ptb-buk* in late exponential phase and both *pta-ack* and *ptb-buk* in early stationary phase (Alsaker and Papoutsakis, 2005). A local expression maximum at late acidogenesis, followed by a decrease after entering solventogenesis, was observed in *C. beijerinckii* (Wang et al., 2012). The importance and interconnected coexpression of *ptb* and *buk* in solventogenic clostridia were verified by difficulties in attempts to obtain the respective mutants. Viable *ptb* mutants contained additional mutations in the thiolase gene (Cooksley et al., 2012) and the single deletion of *buk* was unsuccessful without simultaneous disruption of *ptb* (Yoo et al., 2017). Mutation of *ack* did not influence butyrate formation (Cooksley et al., 2012) suggesting a certain degree of regulatory independence of acetate and butyrate formation. The *ptb-buk* branch of the ABE pathway was also shown to take over part of the metabolic activity in the stationary phase when solventogenesis was disrupted due to the absence of the pSOL megaplasmid (Tomas et al., 2003a). Increased acid formation by the degenerate strain *C. beijerinckii* DG-8052, or the *C. beijerinckii* BA105 mutant exhibiting the so-called 'acid crash' phenotype, was accompanied by up-regulation of both operons (Seo et al., 2017; Zhang et al., 2017). In the latter strain, the authors speculate that the increased expression of these genes is caused by the mutation in gene encoding AbrB regulator, which was shown to affect the expression of many genes in the central metabolic pathway of *C. acetobutylicum* ATCC824 by suppression of Spo0A phosphorylation (Xue et al., 2016).

Finally, higher expression of most genes involved in butyryl-CoA synthesis from acetyl-CoA was observed in both acidogenesis (Grimmler et al., 2011; Janssen et al., 2010; Patakova et al., 2019) and solventogenesis (Alsaker and Papoutsakis, 2005) or at the onset of solventogenesis (Shi and Blaschek, 2008). Nevertheless, these changes were relatively mild and depended on the experimental design and sampling setup. The stability of expression of these genes was confirmed by no obvious change even between the parental strain *C. beijerinckii* NCIMB 8052 and its mutant DG-8052 exhibiting a degenerate phenotype with impaired solventogenesis (Zhang et al., 2017) or only mild upregulation in a better-growing and producing culture of *C. acetobutylicum* supplemented with zinc (Wu et al., 2015).

Hydrogen formation is an interesting, useful, but underestimated feature of solventogenic clostridia, which they use for maintaining redox balance by regeneration of the reduced ferredoxin complex. While in *C. acetobutylicum*, the use of hydrogenases for this purpose is the only possibility, in *C. beijerinckii* there is also the possibility of proton translocation by the Rnf complex (Poehlein et al., 2017). Addition of methyl viologen, an artificial electron carrier substituting for ferredoxin, to a culture of *C. acetobutylicum* inhibited hydrogen production, increased the butanol:acetone ratio and altered the transcription of *sol* operon genes (Hönicke et al., 2012). Recently, [NiFe] hydrogenese of *C. acetobutylicum* was proposed to be involved in hydrogen cycling (generation/uptake) and solvent formation via the production of NADH and regeneration of ferredoxin under solventogenesis (Germane et al., 2018). In *C. beijerinckii* NRRL B-598, a similar role of [NiFe] hydrogenese is probably based on the transcriptional profile of respective genes during ABE fermentation (Patakova et al., 2019).

3.5. Transcriptional regulators

There are a few *C. acetobutylicum* transcriptomic studies of global metabolic regulators in which authors compared the wild-type strain ATCC 824 with a mutant containing a deactivated regulator gene. In these studies, *ccpA* (Ren et al., 2012), *csrA* (Tan et al., 2015), *cap0037* (Nguyen et al., 2016), *fur* (Vasileva et al., 2012), *perR* (Hillmann et al., 2009) and *spo0A* (Tomas et al., 2003a) genes were deactivated. In all cases, transcription of central metabolic genes as well as those regulating nutrient transport and other functions was significantly altered in mutant strains in comparison with the parent strain. Moreover, differential expression of genes encoding transcriptional regulators LexA, Rex, PerR, ArgR, HisR, CymR and PurR and their regulons under acid and butanol stress was mapped (Wang et al., 2013a). Expression of genes encoding some transcriptional regulator(s) have also been described in transcriptional studies performed not only for *C. acetobutylicum* but also for *C. beijerinckii* which were not specifically targeting transcriptional regulators.

Studies of small regulatory RNAs are also limited to *C. acetobutylicum*, and their transcriptomic analysis is scarce, with only a few visionary articles being published. Recently, a small RNA (sr8384) was found and its significant effect on *C. acetobutylicum* growth was described (Yang et al., 2020b). Further, a key role of the small RNome in the stress response was described (Venkataramanan et al., 2013, 2015) and involvement of small antisense RNAs in final gene transcription was revealed (Ralston and Papoutsakis, 2018). This area of work, not previously examined at the transcriptomic level in *C. beijerinckii*, awaits for more in depth research at all levels.

3.6. Sporulation

Solventogenesis in ABE fermentation takes place frequently, but not necessarily simultaneously, with sporulation. Sporulation of solventogenic clostridia is a similar process to that of the *Bacillus* bacteria (Riley et al., 2021), but differs in certain aspects, i.e. accumulation of stock polysaccharide, granulose, prior to sporulation, exploitation of species-specific orphan histidine kinases for phosphorylation of the master sporulation regulator Spo0A (instead of the phosphorelay cascade known in bacilli), and other smaller variations (Al-Hinai et al., 2015; Diallo et al., 2021). Recently, the effect of orphan histidine kinases, involved in Spo0A phosphorylation/dephosphorylation, on sporulation and butanol production has been studied in the asporogenic *C. acetobutylicum* ATCC 55025 (Du et al., 2021a, 2021b) and its mutant JB200 (Xu et al., 2015), in *C. beijerinckii* NCIMB 8052 (Xin et al., 2020) and *C. beijerinckii* BA 101 (Seo et al., 2021). Unlike bacilli, solventogenic

clostridia usually do not sporulate without excess carbohydrate substrate, and supposed signals for sporulation include a combination of species-specific low pH of the culture medium, accumulation of organic acids (mainly acetic and butyric) and a certain population density. Recently, the level of acetylation and butyrylation of proteins have been revealed to be a significant signal for the metabolic shift from acidogenesis to solventogenesis, but also for sporulation (Xu et al., 2018). The population density is sensed via quorum sensing mechanisms, which were mainly studied in *C. acetobutylicum* (Kotte et al., 2020; Steiner et al., 2012) and *C. saccharoperbutylacetonicum* (Feng et al., 2020), however, they are probably also functioning in *C. beijerinckii* (Patakova et al., 2019, 2021).

For the first time, the complete transcriptional view of sporulation was published in 2008 for *C. acetobutylicum* (Jones et al., 2008), although some partial studies were published previously (Alsaker et al., 2004; Alsaker and Papoutsakis, 2005; Tomas et al., 2003a). In *C. beijerinckii*, sporulation was first described in strain NCIMB 8052 (Wang et al., 2012), and later in strains DSM 6423 (Mâté de Gérando et al., 2018) and NRRL B-598 (Patakova et al., 2019). Studies have also been published describing mutant strains obtained after random mutagenesis, i.e. *C. acetobutylicum* EA 2018 (Hu et al., 2011), *C. acetobutylicum* M5 (Alsaker and Papoutsakis, 2005), *C. beijerinckii* BA 105 (Seo et al., 2017), *C. beijerinckii* BA 101 (Shi and Blaschek, 2008) and *C. beijerinckii* DG 8052 (Jiao et al., 2016; Zhang et al., 2017), in which transcription of the sporulation gene set is mentioned but was not the focus of the work. A summary of interventions that have resulted in significant changes in sporulation and were studied at the transcriptional level is shown in Table 2. The most frequently studied gene is probably *spo0A*, encoding the global transcriptional regulator.

Unfortunately, in many transcriptional studies of solventogenic clostridia, the phenomenon of sporulation is not mentioned and, therefore, remained hidden regardless of whether the studied strains sporulated or not. On the other hand, some sporulation changes have been studied at non-transcriptional levels such as metabolic engineering of histidine kinase in *C. beijerinckii* NCIMB 8052 (Xin et al., 2020). For the future, overexpression of *spo0A* in *C. beijerinckii* NRRL B-598 (Kolek et al., 2017), no sporulation in the same strain during the cultivation in RCM culture medium (Branska et al., 2018; Kolek et al., 2016); deletion of the autolysin-encoding gene in *C. acetobutylicum* ATCC 824 (Liu et al., 2015b) and others await transcriptional studies that would enable an in-depth view of sporulation in particular strains.

Solventogenesis need not be coupled with sporulation. Even in the

Table 2
Interventions affecting sporulation described at the transcriptional level

Intervention	Strain	Main transcriptional changes	Morphology changes	Reference
Change in genome				
spoIIE deletion (SpoIIE – phosphatase involved in *SigF* phosphorylation)	*C. beijerinckii* NCIMB 8052	↑solventogenesis; granulose formation, *groEL*, *dnaJ*, *ftsZ*, *minD*, *refZ* ↓granulose degradation, rubrerythrins	no sporulation but granulose accumulation mislocation of asymetric septum, cells with two septa, elongated cells	(Diallo et al., 2020)
pks deletion; (type I single module PKS)	*C. acetobutylicum* ATCC 824	↑amino acid transport, sulfur metabolism, stress response ↓carbohydrate transport and metabolism	Decrease in sporulation No granulose accumulation	(Herman et al., 2017)
spo0A deletion (Spo0A – master sporulation regulator)	*C. acetobutylicum* ATCC 824	↑chemotaxis/motility, glycosylation, *clvB* ↓ solventogenesis, *sigF*, carbohydrate metabolism	no sporulation filamentous morphology	(Tomas et al., 2003a)
Loss of megaplasmid[a]	*C. acetobutylicum* ATCC 824	↑ chemotaxis/motility, electron transport ↓ solventogenesis, *sigF*, *spo0A*, carbohydrate metabolism, *fts* family genes	no sporulation	(Alsaker and Papoutsakis, 2005; Tomas et al., 2003a)
spo0A overexpression	*C. acetobutylicum* ATCC 824	↑ fatty acids biosynthesis, heat shock genes ↓ flagella formation	Increase in sporulation	(Alsaker et al., 2004)
*cac0139*KO (histidine kinase knock-out)	*C. acetobutylicum* ATCC 55025	↑ butanol production ↓ sporulation initiation	no sporulation	(Du et al., 2021b)
*cac0437*KO (histidine kinase knock-out)	*C. acetobutylicum* ATCC 55025	↑ early autolysis ↓ butanol production granulose formation	no sporulation	(Du et al., 2021b)
Other change				
Butanol addition	*C. beijerinckii* NRRL B-598	↑ one of three Agr quorum system, heat shock protein genes, *cfa* ↓ fatty acid biosynthesis; polyketide biosynthesis genes	No sporulation, no granulose accumulation	(Patakova et al., 2021; Sedlar et al., 2019)
Biofilm growth after 150 h	*C. acetobutylicum* CGMCC 5234	↓*sigK*, spore coat proteins, small acid soluble proteins	No sporulation, no granulose accumulation; filaments formation	(Liu et al., 2016a)
Biofilm growth after 34 h	*C. acetobutylicum* CGMCC 5234	↑sporulation genes, granulose formation	sporulation	(Liu et al., 2016a)
	C. beijerinckii DG 8052[b]	↑chemotaxis, ↓*pfk*, *pyk*, *spo0A*, sporulation sigma factors	No sporulation, no granulose accumulation	(Zhang et al., 2017)
$CaCO_3$ addition	*C. beijerinckii* DG 8052[b]	↑ solventogenesis ↓ *spo0A*, sporulation sigma factors	No sporulation, no granulose accumulation	(Jiao et al., 2016)

↑ upregulation, ↓ downregulation

Gene products: *abrB* - repressor of sigma factor H, binds with Spo0A-P which results in derepression; *cfa* - cyclopropane-fatty-acid-synthetase; *dnaJ* - molecular chaperone DnaJ; *fts*, *ftsZ* - cell division proteins; *groEL* - chaperonin GroEL; *minD* - MinD/ParA family protein; *pfk* - 6-phosphofructokinase; *pks* - polyketide synthase; *pyk* - pyruvate kinase; *refZ* - forespore capture DNA-binding protein RefZ, regulator of FtsZ; *sigF* - sigma factor F; *sigK* - sigma factor K, active in mother cell; *spo0A* - master sporulation regulator, potentially also repressor of chemotaxis/motility; *spoIIE* - phosphatase, dephosphorylation of SpoIIAA, assymetric division regulator, involved in formation *sigF*, *sigE*, *sigG*.

[a] Wild type strain *C.acetobutylicum* ATCC 824 contains a megaplasmid on which are sol operon genes. However, *spo0A* gene resides on the chromosome.

[b] Strain derived from *C.beijerinckii* NCIMB 8052 exhibiting permanent degenerative phenotype (no sporulation, no solventogenesis).

iconic *C. acetobutylicum* ATCC 824, their simultaneous occurrence was observed only when cultivation was performed in batch mode and pH was not regulated. However if the same strain was cultivated in phosphorous-limited chemostat culture, acidogenic and solventogenic non-sporulating cultures were obtained by maintaining pH at 5.7 and 4.5, respectively. The observed phenomenon of abolished sporulation under continuous cultivation has never been explained. The phenomenon of separate butanol formation without sporulation was also found in the asporogenic *C.acetobutylicum* ATCC 55025 and its mutant JB200 (Xu et al., 2015, 2017). The cause of the inability to sporulate in *C.acetobutylicum* ATCC 55025 was recently found to be a point mutation in histidine kinase gene *cac0437* involved in dephosphorylation of Spo0A-P, while overexpression of the histidine kinase gene *cac0139*, involved in Spo0A phosphorylation, restored the sporulation (Du et al., 2021a). In *C. beijerinckii* NRRL B-598 (described in this study as *C. pasteurianum* because the strain was reclassified later (Sedlar et al., 2017)) under continuous cultivation, suppressed sporulation was not observed (Lipovsky et al., 2016), possibly because pH was not regulated. However, sporulation was also not suppressed in this strain even under pH regulation and preferential production of acids (Drahokoupil and Patáková, 2020). However, butanol formation without sporulation was observed in reinforced clostridial medium supplemented with glucose in this strain (Branska et al., 2018; Kolek et al., 2016).

3.7. Stress response

Accumulation of metabolites, especially during batch culture, results in stress caused by unfavorable environmental conditions; these are sensed by different mechanisms including intramembrane histidine kinases, part of a two-component signal transduction system (Mascher, 2006). The stress response might be quick or slow, untargeted or targeted on a specific stressor and its aim is to enhance chances of survival for a single cell or population. The mechanisms of stress response include re-modelling transcription, damage repair and minimization. While in *Bacillus subtilis*, the recently reviewed stress response (Bonilla, 2020), and the response goal is the same as in solventogenic clostridia, its regulation and means of achievement are different. An alternative sigma factor B, which is believed to govern the stress response in G^+ bacteria such as *Bacillus subtilis* and *Staphylococcus aureus*, is missing in solventogenic clostridia (Alsaker et al., 2010; Patakova et al., 2019) and its regulation, as well as components, are different. The most frequent and prompt stress response, which is shared between different cell kingdoms, is production of heat-shock proteins or chaperones. Further cell envelope re-modelling and changes in regulation of metabolic protein production represent other common types of stress responses.

ABE fermentation of solventogenic clostridia is the process responsible for production of two kinds of toxic products, acids and solvents. Among them, butyric acid and butanol exhibit the highest toxicity, however other metabolites or medium components such as lignocellulose derived inhibitors, or environmental factors such as pH may strengthen the overall stress. While butyrate stress can be studied independently of butanol, butanol was usually added to the medium with a certain concentration of butyrate. In addition, it is necessary to consider that metabolite stress responses are a natural part of almost all transcriptomic studies and that their significance increases with an increased concentration of the toxic metabolite(s). This means that the stress response can be studied not only after deliberate addition of a stressor but also as a natural part of all ABE/IBE fermentations. It is estimated that about 15-20 % of genes in the total genome of *C. acetobutylicum* and *C. beijerinckii* species (Alsaker et al., 2010; Patakova et al., 2019) are differentially expressed under stress.

3.7.1. Butyrate (acetate) stress response

Transcriptomic analysis of the acetate stress response after deliberate acetate addition has only been studied in *C. acetobutylicum* (Alsaker et al., 2010) and, therefore, it is more common to consider overall acid stress, which is affected not only by the sum of organic acids present, but also by the environmental pH. Overall acid stress was studied in acidogenic chemostat cultures of *C. acetobutylicum* (Grimmler et al., 2011; Janssen et al., 2010) at a constant pH of 5.7, or during acidogenesis in batch cultures of *C. beijerinckii* under unregulated pH (Máté de Gérando et al., 2018; Patakova et al., 2019; Wang et al., 2011, 2012). Separate butyrate stress was studied after its deliberate addition to acidogenic continuous cultures of *C. acetobutylicum* at a pH above 5.0 (Venkataramanan et al., 2015) or to acidogenic batch cultures of *C. acetobutylicum* (Alsaker et al., 2010). If we compare *C. acetobutylicum* and *C. beijerinckii* acidic stress studies, we can see that *C. acetobutylicum* is more acid stress tolerant and can survive at a lower pH than *C. beijerinckii*.

Understanding the stress response elicited by production and accumulation of organic acids is very important for construction of tolerant microbial producers in industrial microbiology. Regarding this point of view, microbial acid stress response was reviewed recently (Guan and Liu, 2020; Liu et al., 2015a) and many mechanisms, including maintenance of pH homeostasis, enhancement of proton pump activity, alterations in cell membrane composition, changes in metabolic regulation, and protection and repair of macromolecules were confirmed by transcriptomic studies as functioning in solventogenic clostridia. For their brief overview, see Table 3. The main event that follows the accumulation of acids in the culture medium during ABE fermentation is the transition to solvent formation. However addition of butyrate to a final concentration of 40 mM to the medium prior to inoculation with *C. beijerinckii* (Wang et al., 2013b), led to an early trigger of solventogenesis and uncoupling of sporulation and solventogenesis. This study showed similar effects on solventogenesis as earlier research (e.g. Lee et al. (2008)) focused on feeding mixed substrates (butyrate/glucose or others) as carbon sources (these studies were partially summarized by Patakova et al. (2013)).

The most frequently described acid stress response is upregulation of heat shock protein (HSP) genes, especially HSP genes from class I, namely operons *dnaKJ* and *groESL*, in both *C. acetobutylicum* (Alsaker et al., 2010; Grimmler et al., 2011; Wang et al., 2013a) and *C. beijerinckii* (Patakova et al., 2019). This stress response is probably a general reaction of bacteria to acid stress and the chaperones and their gene organization seem to be conserved (Guan and Liu, 2020). Interestingly, in *Acetobacter pasteurianum*, an industrial acetic acid (vinegar) producer, the same chaperone genes were upregulated during acetic acid production from ethanol (Yang et al., 2019). The importance of chaperone production is underlined by their finding among extracellular polymeric substances (EPS) in biofilms of *C. acetobutylicum* (Liu et al., 2018a), with the culture exhibiting higher stress resistance. The production of heat shock proteins is closely related to the oxidative stress response, probably regulated by the peroxide sensor PerR in *C. acetobutylicum*. However, the described response of this regulator to acidic stress is ambiguous, see Table 3. In this regard, a single transcriptional study of oxygen stress on *C. acetobutylicum* and its Δ*perR* mutant (Hillmann et al., 2009), confirmed the influence of PerR on transcription of reverse rubrerythrins (*rbr3A-rbr3B*), desulfoferrodoxin (*dfx*), rubredoxin (*rd*), NADH-dependent rubredoxin oxidoreductase (NROR), and the oxygen-reducing FDPs FprA1 and FprA2 under O_2 stress.

A classical microbiological model that explains suppression of bacterial growth in acidic foods by organic acid preservatives (e.g. acetates, lactates or propionates) (Taylor et al., 2014) shows that at low pH, undissociated acids that can pass through the cell membrane by diffusion, are dissociated in cells. This results in active pumping of H^+ by F_0-F_1-ATPases coupled with consumption of ATP (Liu et al., 2015b) and a slow-down of growth. Although the pump was described in *C. acetobutylicum* (Extembrink et al., 2000), and its functioning in the metabolic switch was assumed (Grupe and Gottschalk, 1992), it was not specifically identified in the transcriptomic studies. However, upregulation of genes encoding ATP-binding cassette (ABC) transporters was described both for *C. acetobutylicum* (Alsaker et al., 2010; Grimmler

Table 3
Common selected examples of acid stress response found in solventogenic clostridia at the transcriptomic level

Functional group or metabolic process description	Change in genes/gene clusters transcription	COG[a]	Studied species	Reference
Heat shock proteins	↑ *dnaK, dnaJ, groES, groEL, hsp90, hsp18, clpC, htrA*	O	*C. acetobutylicum* *C. beijerinckii*	(Alsaker et al., 2010; Grimmler et al., 2011; Wang et al., 2013a) (Patakova et al., 2019)
Solventogenesis	↑ *ald, ctfA, ctfB*	C	*C. acetobutylicum* *C. beijerinckii*	(Alsaker et al., 2010; Grimmler et al., 2011; Janssen et al., 2010) (Máté de Gérando et al., 2018; Patakova et al., 2019; Sedlar et al., 2018; Wang et al., 2012)
Oxidative stress	↑ *perR* ↓*perR*	L	*C. acetobutylicum* *C. acetobutylicum*	(Alsaker et al., 2010) (Wang et al., 2013a)
Sporulation	↑ *spo0A* and early sporulation genes	D	*C. beijerinckii*	(Máté de Gérando et al., 2018; Patakova et al., 2019; Sedlar et al., 2018; Wang et al., 2012; Wang et al., 2013b)
Efflux	↑ ABC transporters	Q	*C. acetobutylicum* *C. beijerinckii*	(Alsaker et al., 2010; Grimmler et al., 2011; Janssen et al., 2010Wang et al., 2013a) (Branska et al., 2021; Vasylkivska et al., 2019)
Fatty acids biosynthesis	↑*fab* operon ↑*cfa* ↓*fab* operon	I	*C. acetobutylicum* *C. beijerinckii*	(Wang et al., 2013a) (Patakova et al., 2019)
Ribosomal RNA processing	↑ non-coding RNA	A	*C. acetobutylicum*	(Borden et al., 2010; Ralston and Papoutsakis, 2018; Venkataramanan et al., 2013, 2015)
Motility, chemotaxis	↓ motility, chemotaxis genes	N	*C. acetobutylicum* *C. beijerinckii*	(Grimmler et al., 2011) (Máté de Gérando et al., 2018; Wang et al., 2012)

↑ upregulation, ↓ downregulation
Gene products: *ald* - aldehyde dehydrogenase; *cfa* - cyclopropane-fatty-acid-synthetase; *clpC* - ATP-dependent Clp protease ATP-binding subunit ClpC; *ctfA* and *ctfB* - acetyl-CoA–acetoacetyl-CoA transferase subunit alpha and beta, respectively; *dnaJ* - molecular chaperone DnaJ; *dnaK* - molecular chaperone DnaK; *groEL* - chaperonin GroEL; *groES* - chaperonin GroES; *hsp18* - small heat shock family protein; *hsp90* – heat-shock protein, part of the CIRCE (controlling inverted repeat of chaperone expression) heat shock regulon; *htrA* - high-temperature-requirement protein A family protease; *perR* - peroxide repressor; *spo0A* - master sporulation regulator
[a] COG categories – see https://www.ncbi.nlm.nih.gov/research/cog#

et al., 2011; Janssen et al., 2010; Wang et al., 2013a) and *C. beijerinckii* (Branska et al., 2021; Vasylkivska et al., 2019).

3.7.2. Butanol stress response

Butanol stress is more frequently studied than acid stress in solventogenic clostridia, see recent reviews Arsov et al. (2021); Gao et al. (2019); Patakova et al. (2018); Vasylkivska and Patakova (2020), and has a more pronounced effect on the clostridial population. An increase in butanol concentration in the culture medium slows overall energy metabolism and nutrient transport, assembly of flagella and chemotaxis and may even negatively affect sporulation. On the other hand, it results in re-modeling of cell membranes, capsule formation, an increase in heat shock protein production and an intracellular accumulation of protective compounds such as glycerol. An overview of the most frequently described processes and cell functions that are up/down regulated under butanol stress is shown in Fig. 3. Moreover, some mechanisms of butanol stress are similar to those described for acid stress, such as upregulation of heat shock genes in both *C. acetobutylicum* and *C. beijerinckii* (Alsaker et al., 2010; Janssen et al., 2012; Máté de Gérando et al., 2018; Patakova et al., 2019; Ralston and Papoutsakis, 2018; Schwarz et al., 2012; Tomas et al., 2003a, 2004). Further in this section, will be described transcriptional changes that are unique and significant for butanol stress:

- There is one unexplained and unintuitive stress response that was described in both *C. acetobutylicum* and *C. beijerinckii* after deliberate butanol addition - upregulation of butanol synthesis genes (Alsaker et al., 2010; Sedlar et al., 2019; Tomas et al., 2003a, 2004). This probably indicates that solventogenesis, and especially butanol formation, are seen as a stress response by clostridial cells and are regulated by the same transcription regulator(s) as the stress response.
- Efflux pumps seem to contribute to butanol tolerance and their synthesis is activated under butanol stress in both *C. acetobutylicum* (Alsaker et al., 2010; Yang et al., 2020a) and *C. beijerinckii* (Branska et al., 2021; Patakova et al., 2021)
- In *C. saccharoperbutylacetonicum*, a secondary metabolite synthesized by the non-ribosomal peptide, polyketide synthase, was involved in glycerol accumulation after butanol shock (Li et al., 2020a). Glycerol accumulation was also described in *C. acetobutylicum* (Alsaker et al., 2010; Schwarz et al., 2012).
- Upregulation of fatty acid biosynthesis and phospholipid genes was found both in *C. acetobutylicum* (Janssen et al., 2012; Schwarz et al., 2012) and *C. beijerinckii* (Patakova et al., 2019, 2021).
- Upregulation of amino acid biosynthetic genes by butanol was reported for *C. acetobutylicum* (Alsaker et al., 2010; Janssen et al., 2012)).
- Biofilm growth and EPS excretion have protective effects on *C. acetobutylicum* (Liu et al., 2018a; Zhang et al., 2021). Under biofilm growth, sporulation was abolished and heat shock proteins (GroEL), as well as oxygen response proteins (ruberythrin), were found in the EPS.
- Overexpression of *spo0A* gene (Alsaker et al., 2004) in *C. acetobutylicum* resulted in increased butanol tolerance.

There is no doubt that transcriptomics enables insights into the population processes on a specific level and that to obtain a complete picture it is also necessary to employ other techniques. In this regard, it is fair to admit that comparisons of proteomic and transcriptomic analyses do not always match. Especially under butanol stress, but also under acid stress, differences between the transcriptome and the proteome in *C. acetobutylicum* were significant (Venkataramanan et al., 2015) and were attributed to post-transcriptional regulation, in particular, specific ribosomal translation machinery upregulated under stress and also to regulation of small RNAs (Venkataramanan et al., 2013). Both studies (Venkataramanan et al., 2013, 2015) evaluated population reactions to deliberately added butanol (or acid) and further studies under production conditions are necessary. When proteomic analysis of *C. acetobutylicum* was performed during ABE fermentation (Jang et al., 2014; Yoo et al., 2015) the conclusions deduced from transcriptomic analyses were mostly confirmed. A comparison of produced versus added butanol in *C. beijerinckii* (Patakova et al., 2019, 2021; Sedlar et al., 2018, 2019) shows significant transcriptional differences under both circumstances and highlights a drastic drop in population viability, determined by flow cytometry and double fluorescent staining (Branska

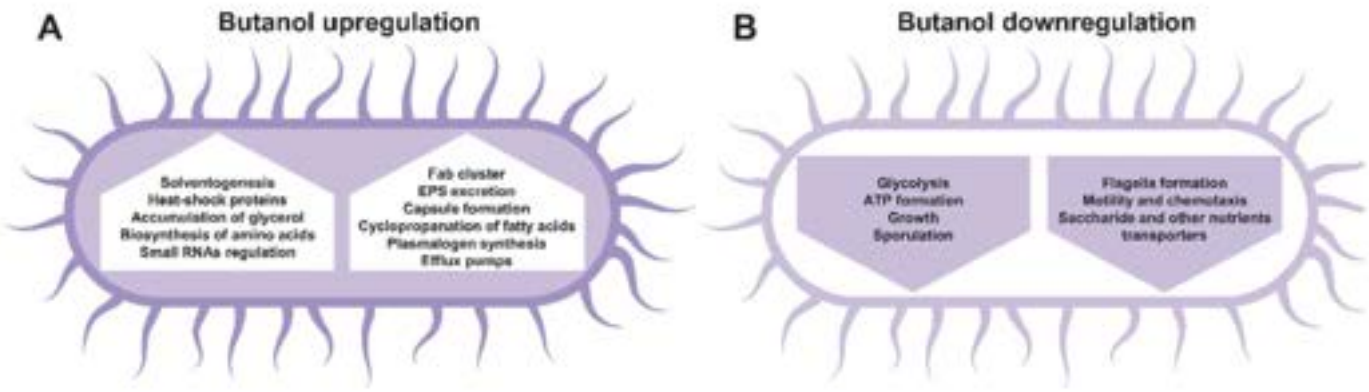


Fig. 3. Butanol stress induced upregulation (A) and downregulation (B) of genes coding for phenomena and metabolic functions
Fab – fatty acid biosynthesis, EPS – extracellular polymeric substances.

Table 4
Transcriptional response to the most common inhibitors

Inhibitor and challenge design	Strain	Growth and fermentation pattern	Highlighted up-/down regulated genes or groups	Reference
Vanillic acid + p-hydroxybenzoic acid (0.2 + 0.2 g/L) addition at OD ≈ 3.5 (12 h) sampling after 12 h	*C. acetobutylicum* ATCC 824	decrease in growth, ABE production and lower rate of glucose consumption; reduced decrease in acids	↑ABC transporters; glycolysis; HSPs; chemotaxis proteins ↓fructose/mannose-specific transporter subunit IIA; phosphoglycerate mutase; sporulation	(Luo et al., 2020)
Vanillin + p-hydroxybenzaldehyde (0.2 + 0.2 g/L) addition at OD ≈ 3.5 (12 h) sampling after 12 h		decrease in growth, ABE production and lower rate of glucose consumption; reduced decrease in acids	↑some ABC transporters; glycolysis; HSPs; chemotaxis proteins ↓some ABC transporters; PTS; phosphoglycerate mutase; *ack*, *buk*; sporulation	
Wheat straw hydrolysate (2.52 g/L HMF; 1.96 g/L furfural; 0.13 g/L syringaldehyde; 0.16 g/L vanillin, 0.37 g/L ferulic acid) comparison of expression profile in medium with and without addition of Na_2S sampling 0 and 3 h after addition	*C. acetobutylicum* CICC 8012	in presence of Na_2S increased ABE production, carbohydrate consumption and product yield	↑cellobiose and xylose transport; xylose metabolism; cellobiose hydrolysis; most glycolytic genes; solvent and acid formation ↓pyruvate kinase	(Jin et al., 2015)
Ferulic acid (0.5 g/L) addition at the beginning of cultivation from pre-culture sampling 10 min, 4 h, 9.5-12.5 h	*C. beijerinckii* NCIMB 8052	delayed growth, decreased ABE production	↑**Instantaneous response**: arginine and proline metabolism; efflux related genes. **Mid-term**: class I and III HSPs; nitrogenase-related genes; iron-sulphur cluster assembly; ABC transporters, sulphur metabolism. **Long-term**: DNA topological changes (SASPs) ↓**Instantaneous response**: ATPase activity. **Mid-term**: chemotaxis; flagellar activity.	(Lee et al., 2015)
Ferulic acid (0.5 g/L) Addition at 8 h sampling 12 h from challenge beginning		not evaluated	↑SDR family genes; hydrogenases/ dehydrogenases, FMN reductases, NADH-quinone oxidoreductase; TCS - chemotaxis; HSPs ↓ATP synthase; TCS- QS	(Liu et al., 2017)
Furfural (2 g/L) addition in acidogenic phase		unaltered growth; in a short-term view: decreased rate in solvents production; end of fermentation: higher ABE concentration compared to control	↑redox; ABC transporters, cobalt transport ↓PTS; TCS - chemotaxis	(Zhang and Ezeji, 2013)
Furfural (3 g/L) addition in solventogenic phase		unaltered growth, decrease in ABE production, acids accumulation	↑redox; iron-sulphur cluster assembly; ABC transporters; HSPs ↓PTS; TCS - chemotaxis; TCS - QS; nitrogen assimilation; cell cycle progression and development	
Comparison of expression profile of mutant and parent strain NCIMB 8052 in P2 media without inhibitors	*C. beijerinckii* 3304::*int* (higher tolerance to ferulic and vanillic acid)	enhanced growth and butanol production in the presence of ferulic and vanillic acid compared to wild type	↑ABC transporters ↓phosphate ABC transporter, acid and solvents formation	(Liu et al., 2018a)
Ferulic acid (0.5 g/L) Addition at 8 h sampling 12 h from challenge beginning	*C. beijerinckii* NCIMB 46/93::*int* (mutant - FA tolerant)	not evaluated	↑TCS - chemotaxis ↓ATP synthase; TCS - QS; HSPs	

ABC - ATP-binding cassette; HSPs - heat-shock proteins; PTS - phosphotransferase system; SASPs – small acid soluble proteins; SDR - Short-chain dehydrogenase/ reductase family protein; TCS - two-component signal transduction system; QS - quorum sensing;

et al., 2018; Kolek et al., 2016) after butanol addition, which actually interferes with the assessment.

Transcriptional differences between butanol produced and added butanol in *C. beijerinckii* (Patakova et al., 2021) include: suppression of sporulation after butanol addition, differential regulation of heat shock proteins in both situations, upregulation of the *fab* operon during butanol production but not after butanol addition, differential regulation of Agr-quorum sensing systems. Differences within the *C. acetobutylicum* butanol stress studies group (for a general overview of all studies, see Supplement Table I) are also remarkable and probably reflect specific physiological states of the culture and population density at the time when butanol was added, the amount of butanol added and other environmental factors such as pH, total acid concentrations, redox potential etc. Taken together, it can be noted that the natural butanol stress response is probably more correctly captured in undisturbed ABE fermentation batch studies.

3.7.3. Response to lignocellulose derived inhibitors

One of the requirements for sustainable biotechnological production of bulk chemicals is the selection of cheap, renewable, and easily available substrates. Lignocellulose biomass meets all requirements and *C. beijerinckii* or *C. saccharoperbutylacetonicum* are very promising species for its utilization (Branska et al., 2020; Yao et al., 2017). Unfortunately, processing lignocellulose into fermentable sugars inevitably leads to the simultaneous production of substances that could deteriorate fermentation parameters (Amiri and Karimi, 2018; Baral and Shah, 2014). The number and spectrum of lignocellulose-derived inhibitors are wide and are dependent both on processing technology and on the composition and characteristics of the specific biomass (Jönsson and Martín, 2016). Generally, there are four groups of substances considered as potential fermentation inhibitors: furans, phenols, carboxylic acids, and inorganic salts (for a comprehensive list of inhibitors see Klinke et al. (2004)). Most studies proved a negative effect of these inhibitors on ABE fermentation (Zhang and Ezeji, 2014), predominantly on suppression of solvent production, while surprisingly, growth might be unaltered. In a few cases, an opposite stimulatory effect was observed (Ezeji et al., 2007; Qureshi et al., 2012). Mechanisms underlying these phenomena are poorly understood and up to now, only a limited number of studies offer insights into the transcriptional response to the most common inhibitors (see Table 4). As the spectrum of individual inhibitors is diverse, the cellular response to their addition is also broad, although some common features can be discerned. Many of these are shared with the general stress response of the cell, such as up-regulation of genes for HSPs. Increased expression of HSPs belonging to class I, including the *dnaKJ* and *groESL* operons negatively regulated by HrcA, is most commonly observed. These HSP I family proteins seem to predominate over HSPs III, which is in agreement with the response to the presence of inhibitors, but the increase in expression is less pronounced (Lee et al., 2015).

Another common feature is the increased expression of genes encoding redox proteins (Liu et al., 2017; Zhang et al., 2015) that are involved in antioxidant activity, suggesting oxidative stress caused by the generation of reactive oxygen species (Allen et al., 2010) in the presence of phenolic and furan inhibitors. In cultures challenged with furfural, a specific group of oxidoreductases was up-regulated, including genes for aldo/keto reductase (AKR) and short-chain dehydrogenase/reductase (SDR) (Zhang and Ezeji, 2013), which are probably involved in the transformation of furfural to less toxic furfuryl alcohol (Zhang et al., 2015) or other inhibitors (Okonkwo et al., 2019). The apparent upregulation of NADH-quinone oxidoreductase in wild type strain *C. beijerinckii* NCIMB 8052 is also of interest. This increased expression was not detected in the more ferulic acid-tolerant mutant, *C. beijerinckii* int:4693 (Liu et al., 2017). NADH-quinone oxidoreductase is known to participate in oxidative phosphorylation, nevertheless, it might have other functions such as participation in the reoxidation of NADH, thus maintaining the redox state of cells (Spero et al., 2015). Inactivation of one of the subunits (nuoG) of this electron transport chain complex in *C. beijerinckii* significantly influenced gene expression, with 1538 and 929 genes being differentially expressed in acidogenesis and solventogenesis respectively, as well as increased butanol production (Liu et al., 2016b) in relation to the parent strain NCIMB 8052. Keeping redox balance, coping with oxidative stress and rapid transformation of inhibitors (Bui et al., 2015; Liu et al., 2018b; Zhang and Ezeji, 2013) to less toxic substances (Liu et al., 2017; Zhang et al., 2012b) seems to be one of the survival strategies. The decrease in redox potential achieved by the addition of Na_2S improved growth performance and promoted the expression of solvent production genes in *C. acetobutylicum* grown in non-detoxified lignocellulose hydrolysate (Jin et al., 2015), which confirms the culture's need to maintain a certain redox potential and deal with oxidative stress that inhibitors are likely to represent. When the *C. beijerinckii* culture was challenged with ferulic acid, an increase in the expression of genes related to proline metabolism was observed (Lee et al., 2015). The external addition of proline, as well as the overexpression of its related biosynthetic pathway, was shown to effectively cope with lignocellulose-derived stress in *C. acetobutylicum* (Liao et al., 2019), possibly due to the participation of proline in the mitigation of oxidative stress (Wang et al., 2015).

Unsurprisingly, together with up-regulation of proteins involved in redox reactions, genes responsible for their synthesis showed higher levels of expression. This was especially true for genes encoding the iron-sulphur cluster assembly (Lee et al., 2015; Zhang and Ezeji, 2013) with higher expression when the culture was challenged during solventogenesis (Zhang and Ezeji, 2013).

A common feature of a clostridium culture response to lignocellulose-derived inhibitors is the enhanced expression of transport-related genes. This was more apparent when the culture was challenged by furfural in solventogenic phase, while ATP-binding cassette transporters (ABC) were widely represented (Zhang and Ezeji, 2013). On the other hand, genes related to the phosphotransferase system (PTS) were, in many cases, repressed (Luo et al., 2020; Zhang and Ezeji, 2013). The activity of transmembrane transporters is considered to be involved in resistance to various substances. Liu et al., 2018 (Liu et al., 2018a) submitted to transcriptomic analysis, a mutant strain that was tolerant to ferulic acid and which had inactivated a hypothetical membrane transporter. Comparison of mutant and wild type showed significantly increased expression of genes encoding transferases belonging to the ABC family in the mutant strain, presumably to cover the missing functionality, especially in acidogenic phase, where there was a greater need for them (Branska et al., 2021). Another transcriptomic response observed in cultures challenged by lignocellulose inhibitors was altered regulation of genes belonging to a two-component signal transduction system (Lee et al., 2015; Liu et al., 2018a; Liu et al., 2017; Zhang and Ezeji, 2013), giving bacteria a tool to sense stimuli and respond to signals. This two-component system includes, among others, genes related to chemotaxis and motility, whose expression is probably attenuated facing the inhibitor challenge; nevertheless, its up-regulation was also observed (Luo et al., 2020).

One of the most recently published studies of the effect of lignocellulosic inhibitors on the transcriptomic profile of solventogenic clostridia uses protein-protein interaction analysis and shows some differences in the effect of different inhibitors by detecting interconnected gene sets related to selected phenolic inhibitors (Liu et al., 2020). While cultures challenged with syringaldehyde, vanillin, and ferulic acid revealed only a limited number of subnets in the coexpression network, the response to the challenge with coumaric acid led to the identification of nine dense regions with high internal connectivity. Although the analysis showed different responses in clusters of *C. acetobutylicum* genes to selected inhibitors, similarities were apparent with other studies, indicating that, in some respects, the action of all inhibitors investigated induces a similar cellular response. For example, the influence on purine and pyrimidine metabolism-related genes under vanillin stress was also described for furfural in *C. beijerinckii* (Zhang and

Ezeji, 2013) and many other genes found within enriched groups are responsible for other described activities. An interesting group of genes up-regulated in cultures stressed by artificial addition of inhibitors are genes related to the metabolism of cellulose, xylose, and pectin. This feature is also connected with the entry of unchallenged cultures into stationary or solventogenic phases in medium containing a sufficient quantity of glucose (Alsaker and Papoutsakis, 2005; Grimmler et al., 2011; Janssen et al., 2010; Yoo et al., 2015), although this phenomenon has not yet been satisfactorily explained.

4. Future outline

Transcriptional studies of solventogenic clostridia animate and provide proof of function to genomic data. Their importance in the advancement of the field is undeniable, and one can only hope that as the cost of RNA-seq analysis decreases, their use will increase. So far, the main weakness in their comprehensive evaluation is huge fragmentation. Basic phenomena, e.g., glucose transport, central metabolism, the general stress response, have been studied many times and individual studies confirm others and demonstrate unequivocal common trends in transcription of central metabolism, glucose transport and stress response genes. However, other important processes, e.g. hydrogen production, oxygen stress, catabolism of carbohydrates other than glucose, quorum sensing or sporulation, as well as regulation of all other phenomena, have been scarcely studied. Transcriptomic studies have been carried out in only two species of solventogenic clostridia, *C. acetobutylicum* and *C. beijerinckii*, while others such as *C. saccharopebutylacetonicum* or *C. pasteurianum* have been left aside. For these reasons, construction of more general metabolic, sporulation or other regulatory nets from available transcriptional data is not yet possible. Nevertheless, advances in bioinformatics tools and constant development of databases for functional annotation, driven by genome and transcriptome sequencing of various bacteria, will certainly make possible the reanalysis of older data for comparative studies with newly gathered data. Novel techniques will be used to infer additional knowledge about regulatory elements, e.g. sRNA, as the necessary information is hidden in the majority of current datasets, but as yet, is impossible to be visualized by current computational tools.

Funding

This work was supported by the Technology Agency of the Czech Republic, project BIOCIRTEC, TN01000048. PhD study of Jana Musilova is supported by the Brno Ph.D. Talent Scholarship—Funded by the Brno City Municipality.

Declaration of Competing Interest

None

Acknowledgement

Many thanks to John Brooker for language editing of the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.biotechadv.2021.107889.

References

Aklujkar, M., Leang, C., Shrestha, P.M., Shrestha, M., Lovley, D.R., 2017. Transcriptomic profiles of *Clostridium ljungdahlii* during lithotrophic growth with syngas or H_2 and CO_2 compared to organotrophic growth with fructose. Sci. Rep. 7, 13135. https://doi.org/10.1038/s41598-017-12712-w.

Al-Hinai, M.A., Jones, S.W., Papoutsakis, E.T., 2015. The *Clostridium* sporulation programs: diversity and preservation of endospore differentiation. Microbiol. Mol. Biol. Rev. 79, 19–37. https://doi.org/10.1128/MMBR.00025-14.

Allen, S.A., Clark, W., McCaffery, J.M., Cai, Z., Lanctot, A., Slininger, P.J., Liu, Z.L., Gorsich, S.W., 2010. Furfural induces reactive oxygen species accumulation and cellular damage in *Saccharomyces cerevisiae*. Biotechnol. Biofuels. 3, 2. https://doi.org/10.1186/1754-6834-3-2.

Alsaker, K.V., Papoutsakis, E.T., 2005. Transcriptional program of early sporulation and stationary-phase events in *Clostridium acetobutylicum*. J. Bacteriol. 187, 7103–7118. https://doi.org/10.1128/jb.187.20.7103-7118.2005.

Alsaker, K.V., Spitzer, T.R., Papoutsakis, E.T., 2004. Transcriptional analysis of spo0A overexpression in *Clostridium acetobutylicum* and its effect on the cell's response to butanol stress. J. Bacteriol. 186, 1959–1971. https://doi.org/10.1128/JB.186.7.1959-1971.2004.

Alsaker, K.V., Paredes, C.J., Papoutsakis, E.T., 2005. Design, optimization and validation of genomic DNA microarrays for examining the *Clostridium acetobutylicum* transcriptome. Biotechnol. Bioprocess Eng. 10, 432–443. https://doi.org/10.1007/BF02989826.

Alsaker, K.V., Paredes, C., Papoutsakis, E.T., 2010. Metabolite stress and tolerance in the production of biofuels and chemicals: gene-expression-based systems analysis of butanol, butyrate, and acetate stresses in the anaerobe *Clostridium acetobutylicum*. Biotechnol. Bioeng. 105, 1131–1147. https://doi.org/10.1002/bit.22628.

Amiri, H., Karimi, K., 2018. Pretreatment and hydrolysis of lignocellulosic wastes for butanol production: challenges and perspectives. Bioresour. Technol. 270, 702–721. https://doi.org/10.1016/J.BIORTECH.2018.08.117.

Anders, S., Pyl, P.T., Huber, W., 2015. HTSeq–a Python framework to work with high-throughput sequencing data. Bioinformatics 31, 166–169. https://doi.org/10.1093/bioinformatics/btu638.

Annous, B.A., Blaschek, H.P., 1991. Isolation and characterization of *Clostridium acetobutylicum* mutants with enhanced amylolytic activity. Appl. Environ. Microbiol. 57, 2544–2548.

Aristilde, L., Lewis, I.A., Park, J.O., Rabinowitz, J.D., 2015. Hierarchy in pentose sugar metabolism in *Clostridium acetobutylicum*. Appl. Environ. Microbiol. 81, 1452–1462. https://doi.org/10.1128/AEM.03199-14.

Arsov, A., Petrov, K., Petrova, P., 2021. How to outwit nature: omics insight into butanol tolerance. Biotechnol. Adv. 46, 107658 https://doi.org/10.1016/J.BIOTECHADV.2020.107658.

Bao, T., Feng, J., Jiang, W., Fu, H., Wang, J., Yang, S.T., 2020. Recent advances in n-butanol and butyrate production using engineered *Clostridium tyrobutyricum*. World J. Microbiol. Biotechnol. 369 (36), 138. https://doi.org/10.1007/S11274-020-02914-2.

Bao, T., Hou, W., Wu, X., Lu, L., Zhang, X., Yang, S.T., 2021. Engineering *Clostridium cellulovorans* for highly selective n-butanol production from cellulose in consolidated bioprocessing. Biotechnol. Bioeng. 118, 2703–2718. https://doi.org/10.1002/BIT.27789.

Baral, N.R., Shah, A., 2014. Microbial inhibitors: formation and effects on acetone-butanol-ethanol fermentation of lignocellulosic biomass. Appl. Microbiol. Biotechnol. 98, 9151–9172. https://doi.org/10.1007/s00253-014-6106-8.

Baruzzo, G., Hayer, K.E., Kim, E.J., Di Camillo, B., FitzGerald, G.A., Grant, G.R., 2017. Simulation-based comprehensive benchmarking of RNA-seq aligners. Nat. Methods 14, 135–139. https://doi.org/10.1038/nmeth.4106.

Becker-André, M., Hahlbrock, K., 1989. Absolute mRNA quantification using the polymerase chain reaction (PCR). A novel approach by a PCR aided transcript titration assay (PATTY). Nucleic Acids Res. 17, 9437–9446. https://doi.org/10.1093/nar/17.22.9437.

Binns, D., Dimmer, E., Huntley, R., Barrell, D., O'Donovan, C., Apweiler, R., 2009. QuickGO: a web-based tool for Gene Ontology searching. Bioinformatics 25, 3045–3046. https://doi.org/10.1093/bioinformatics/btp536.

Birgen, C., Markussen, S., Wentzel, A., Preisig, H.A., 2018. The effect of feeding strategy on butanol production by *Clostridium beijerinckii* NCIMB 8052 using glucose and xylose. Chem. Eng. Trans. 65, 283–288. https://doi.org/10.3303/CET1865048.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

Bonilla, C.Y., 2020. Generally stressed out bacteria: Environmental stress response mechanisms in Gram-positive bacteria. Integr. Comp. Biol. 60, 126–133. https://doi.org/10.1093/ICB/ICAA002.

Borden, J.R., Jones, S.W., Indurthi, D., Chen, Y., Terry Papoutsakis, E., 2010. A genomic-library based discovery of a novel, possibly synthetic, acid-tolerance mechanism in *Clostridium acetobutylicum* involving non-coding RNAs and ribosomal RNA processing. Metab. Eng. 12, 268–281. https://doi.org/10.1016/J.YMBEN.2009.12.004.

Branska, B., Pechacova, Z., Kolek, J., Vasylkivska, M., Patakova, P., 2018. Flow cytometry analysis of *Clostridium beijerinckii* NRRL B-598 populations exhibiting different phenotypes induced by changes in cultivation conditions. Biotechnol. Biofuels. 11, 99. https://doi.org/10.1186/s13068-018-1096-x.

Branska, B., Fortova, L., Dvorakova, M., Lisa, H., Patakova, P., Zhang, J., Melzoch, M., 2020. Chicken feather and wheat straw hydrolysate for direct utilization in biobutanol production. Renew. Energy 145, 1941–1948. https://doi.org/10.1016/J.RENENE.2019.07.094.

Branska, B., Vasylkivska, M., Raschmanova, H., Jureckova, K., Sedlar, K., Provaznik, I., Patakova, P., 2021. Changes in efflux pump activity of *Clostridium beijerinckii* throughout ABE fermentation. Appl. Microbiol. Biotechnol. 105, 877–889. https://doi.org/10.1007/s00253-020-11072-2.

Bui, L.M., Lee, J.Y., Geraldi, A., Rahman, Z., Lee, J.H., Kim, S.C., 2015. Improved n-butanol tolerance in *Escherichia coli* by controlling membrane related functions. J. Biotechnol. 204, 33–44. https://doi.org/10.1016/j.jbiotec.2015.03.025.

Clark, S.W., Bennett, G.N., Rudolph, F.B., 1989. Isolation and characterization of mutants of *Clostridium acetobutylicum* ATCC 824 deficient in acetoacetyl-coenzyme A:acetate/butyrate:coenzyme A-transferase (EC 2.8.3.9) and in other solvent pathway

enzymes. Appl. Environ. Microbiol. 55, 970–976. https://doi.org/10.1128/aem.55.4.970-976.1989.
Conesa, A., Götz, S., García-Gómez, J.M., Terol, J., Talón, M., Robles, M., 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21, 3674–3676. https://doi.org/10.1093/bioinformatics/bti610.
Cooksley, C.M., Zhang, Y., Wang, H., Redl, S., Winzer, K., Minton, N.P., 2012. Targeted mutagenesis of the *Clostridium acetobutylicum* acetone–butanol–ethanol fermentation pathway. Metab. Eng. 14, 630–641. https://doi.org/10.1016/J.YMBEN.2012.09.001.
Corchete, L.A., Rojas, E.A., Alonso-López, D., De Las Rivas, J., Gutiérrez, N.C., Burguillo, F.J., 2020. Systematic comparison and assessment of RNA-seq procedures for gene expression quantitative analysis. Sci. Rep. 10, 19737. https://doi.org/10.1038/s41598-020-76881-x.
Cornillot, E., Nair, R.V., Papoutsakis, E.T., Soucaille, P., 1997. The genes for butanol and acetone formation in *Clostridium acetobutylicum* ATCC 824 reside on a large plasmid whose loss leads to degeneration of the strain. J. Bacteriol. 179, 5442–5447. https://doi.org/10.1128/jb.179.17.5442-5447.1997.
Cruz-Morales, P., Orellana, C.A., Moutafis, G., Moonen, G., Rincon, G., Nielsen, L.K., Marcellin, E., 2019. Revisiting the evolution and taxonomy of Clostridia, a phylogenomic update. Genome Biol. Evol. 11, 2035–2044. https://doi.org/10.1093/gbe/evz096.
Dash, S., Mueller, T.J., Venkataramanan, K.P., Papoutsakis, E.T., Maranas, C.D., 2014. Capturing the response of *Clostridium acetobutylicum* to chemical stressors using a regulated genome-scale metabolic model. Biotechnol. Biofuels. 7, 144. https://doi.org/10.1186/s13068-014-0144-4.
Diallo, M., Simons, A.D., van der Wal, H., Collas, F., Houweling-Tan, B., Kengen, S.W.M., López-Contreras, A.M., 2018. L-Rhamnose metabolism in *Clostridium beijerinckii* strain DSM 6423. Appl. Environ. Microbiol. 85 https://doi.org/10.1128/AEM.02656-18 e02656-18.
Diallo, M., Kist, N., Monot, M., Collas, F., Martin-Verstraete, I., van der Oost, J., Kengen, S.W.M., López-Contreras, A.M., 2020. Transcriptomic and phenotypic analysis of a *spoIIE* mutant in *Clostridium beijerinckii*. Front. Microbiol. 11, 2225. https://doi.org/10.3389/fmicb.2020.556064.
Diallo, M., Kengen, S.W.M., López-Contreras, A.M., 2021. Sporulation in solventogenic and acetogenic clostridia. Appl. Microbiol. Biotechnol. 105, 3533–3557. https://doi.org/10.1007/S00253-021-11289-9.
Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., Gingeras, T.R., 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21. https://doi.org/10.1093/bioinformatics/bts635.
Drahokoupil, M., Patáková, P., 2020. Production of butyric acid at constant pH by a solventogenic strain of *Clostridium beijerinckii*. Czech J. Food Sci. 38, 185–191. https://doi.org/10.17221/95/2020-CJFS.
Du, G., Zhu, C., Wu, Y., Kang, W., Xu, M., Yang, S.-T., Xue, C., 2021a. Effects of orphan histidine kinases on clostridial sporulation progression and metabolism. Biotechnol. Bioeng. https://doi.org/10.1002/BIT.27968.
Du, G., Zhu, C., Xu, M., Wang, L., Yang, S.T., Xue, C., 2021b. Energy-efficient butanol production by *Clostridium acetobutylicum* with histidine kinase knockouts to improve strain tolerance and process robustness. Green Chem. 23, 2155–2168. https://doi.org/10.1039/D0GC03993D.
Extembrink, T., Hujer, S., Winzer, K., Dürre, P., 2000. Sequence analysis of the *atp* operon of *Clostridium acetobutylicum* DSM 792 enioding the F_0F_1 ATP synthase. DNA Seq. 11, 109–118. https://doi.org/10.3109/10425170009033977.
Ezeji, T., Qureshi, N., Blaschek, H.P., 2007. Butanol production from agricultural residues: impact of degradation products on *Clostridium beijerinckii* growth and butanol fermentation. Biotechnol. Bioeng. 97, 1460–1469. https://doi.org/10.1002/bit.21373.
Feng, J., Zong, W., Wang, P., Zhang, Z.T., Gu, Y., Dougherty, M., Borovok, I., Wang, Y., 2020. RRNPP-Type quorum-sensing systems regulate solvent formation, sporulation and cell motility in *Clostridium saccharoperbutylacetonicum*. Biotechnol. Biofuels. 13, 84. https://doi.org/10.1186/s13068-020-01723-x.
Galperin, M.Y., Wolf, Y.I., Makarova, K.S., Alvarez, R.V., Landsman, D., Koonin, E.V., 2021. COG database update: focus on microbial diversity, model organisms, and widespread pathogens. Nucleic Acids Res. 49, D274–D281. https://doi.org/10.1093/nar/gkaa1018.
Gao, Y., Zhang, M. Miao, Guo, X. Peng, Li, W. Jian, Lu, D., 2019. The biological mechanisms of butanol tolerance and the application of solvent-tolerant bacteria for environmental protection. J. Chem. Technol. Biotechnol. 34, 157. https://doi.org/10.1002/jctb.6255.
Germane, K.L., Liu, S., Gerlach, E.S., Savage, A.M., Renberg, R.L., Zu, T.N.K., Dong, H., Walck, S.D., Servinsky, M.D., Sund, C.J., 2018. Hydrogen-cycling during solventogenesis in *Clostridium acetobutylicum* American Type Culture Collection (ATCC) 824 requires the [NiFe]-hydrogenase for energy conservation. Fermentation 4, 55. https://doi.org/10.3390/fermentation4030055.
Gheshlaghi, R., Scharer, J.M., Moo-Young, M., Chou, C.P., 2009. Metabolic pathways of clostridia for producing butanol. Biotechnol. Adv. 27, 764–781. https://doi.org/10.1016/J.BIOTECHADV.2009.06.002.
Grimmler, C., Held, C., Liebl, W., Ehrenreich, A., 2010. Transcriptional analysis of catabolite repression in *Clostridium acetobutylicum* growing on mixtures of d-glucose and d-xylose. J. Biotechnol. 150, 315–323. https://doi.org/10.1016/j.jbiotec.2010.09.938.
Grimmler, C., Janssen, H., Krauße, D., Fischer, R.-J.R.-J., Bahl, H., Dürre, P., Liebl, W., Ehrenreich, A., 2011. Genome-wide gene expression analysis of the switch between acidogenesis and solventogenesis in continuous cultures of *Clostridium acetobutylicum*. Microb. Physiol. 20, 1–15. https://doi.org/10.1159/000320973.
Grupe, H., Gottschalk, G., 1992. Physiological events in *Clostridium acetobutylicum* during the shift from acidogenesis to solventogenesis in continuous culture and presentation of a model for shift induction. Appl. Environ. Microbiol. 58, 3896–3902.
Gu, Y., Feng, J., Zhang, Z.T., Wang, S., Goo, L., Wang, Yifen, Wang, Yi, 2019. Curing the endogenous megaplasmid in *Clostridium saccharoperbutylacetonicum* N1-4 (HMT) using CRISPR-Cas9 and preliminary investigation of the role of the plasmid for the strain metabolism. Fuel 236, 1559–1566. https://doi.org/10.1016/J.FUEL.2018.09.030.
Guan, N., Liu, L., 2020. Microbial response to acid stress: mechanisms and applications. Appl. Microbiol. Biotechnol. 104, 51–65. https://doi.org/10.1007/s00253-019-10226-1.
Harris, M.A., Clark, J., Ireland, A., Lomax, J., Ashburner, M., Foulger, R., Eilbeck, K., Lewis, S., Marshall, B., Mungall, C., Richter, J., Rubin, G.M., Blake, J.A., Bult, C., Dolan, M., Drabkin, H., Eppig, J.T., Hill, D.P., Ni, L., Ringwald, M., Balakrishnan, R., Cherry, J.M., Christie, K.R., Costanzo, M.C., Dwight, S.S., Engel, S., Fisk, D.G., Hirschman, J.E., Hong, E.L., Nash, R.S., Sethuraman, A., Theesfeld, C.L., Botstein, D., Dolinski, K., Feierbach, B., Berardini, T., Mundodi, S., Rhee, S.Y., Apweiler, R., Barrell, D., Camon, E., Dimmer, E., Lee, V., Chisholm, R., Gaudet, P., Kibbe, W., Kishore, R., Schwarz, E.M., Sternberg, P., Gwinn, M., Hannick, L., Wortman, J., Berriman, M., Wood, V., de la Cruz, N., Tonellato, P., Jaiswal, P., Seigfried, T., White, R., 2004. The Gene Ontology (GO) database and informatics resource. Nucleic Acids Res. 32, D258–D261. https://doi.org/10.1093/nar/gkh036.
Heluane, H., Dagher, M.R.E., Bruno-Barcena, J.M., 2011. Meta-analysis and functional validation of nutritional requirements of solventogenic clostridia growing under butanol stress conditions and costilization of D-glucose and D-xylose. Appl. Environ. Microbiol. 77, 4473–4485. https://doi.org/10.1128/aem.00116-11.
Herman, N.A., Kim, S.J., Li, J.S., Cai, W., Koshino, H., Zhang, W., 2017. The industrial anaerobe *Clostridium acetobutylicum* uses polyketides to regulate cellular differentiation. Nat. Commun. 8, 1514. https://doi.org/10.1038/s41467-017-01809-5.
Hillmann, F., Döring, C., Riebe, O., Ehrenreich, A., Fischer, R.J., Bahl, H., 2009. The role of PerR in O_2-affected gene expression of *Clostridium acetobutylicum*. J. Bacteriol. 191, 6082–6093. https://doi.org/10.1128/JB.00351-09.
Hölzer, M., Marz, M., 2019. *De novo* transcriptome assembly: a comprehensive cross-species comparison of short-read RNA-Seq assemblers. Gigascience 8, giz039. https://doi.org/10.1093/gigascience/giz039.
Hönicke, D., Janssen, H., Grimmler, C., Ehrenreich, A., Lütke-Eversloh, T., 2012. Global transcriptional changes of *Clostridium acetobutylicum* cultures with increased butanol: acetone ratios. New Biotechnol. 29, 485–493. https://doi.org/10.1016/j.nbt.2012.01.001.
Hönicke, D., Lütke-Eversloh, T., Liu, Z., Lehmann, D., Liebl, W., Ehrenreich, A., 2014. Chemostat cultivation and transcriptional analyses of *Clostridium acetobutylicum* mutants with defects in the acid and acetone biosynthetic pathways. Appl. Microbiol. Biotechnol. 98, 9777–9794. https://doi.org/10.1007/s00253-014-6040-9.
Horinouchi, T., Maeda, T., Furusawa, C., 2018. Understanding and engineering alcohol-tolerant bacteria using OMICS technology. World J. Microbiol. Biotechnol. 34, 157. https://doi.org/10.1007/s11274-018-2542-4.
Hu, S., Zheng, H., Gu, Y., Zhao, J., Zhang, W., Yang, Y., Wang, S., Zhao, G., Yang, S., Jiang, W., 2011. Comparative genomic and transcriptomic analysis revealed genetic characteristics related to solvent formation and xylose utilization in *Clostridium acetobutylicum* EA 2018. BMC Genomics 12, 93. https://doi.org/10.1186/1471-2164-12-93.
Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S.K., Cook, H., Mende, D.R., Letunic, I., Rattei, T., Jensen, L.J., Von Mering, C., Bork, P., 2019. EggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. 47, D309–D314. https://doi.org/10.1093/nar/gky1085.
Ihekwaba, A.E.C., Mura, I., Walshaw, J., Peck, M.W., Barker, G.C., 2016. An integrative approach to computational modelling of the gene regulatory network controlling *Clostridium botulinum* type A1 toxin production. PLoS Comput. Biol. 12, e1005205 https://doi.org/10.1371/journal.pcbi.1005205.
Jaehme, M., Slotboom, D.J., 2015. Diversity of membrane transport proteins for vitamins in bacteria and archaea. Biochim. Biophys. Acta, Gen. Subj. 1850, 565–576. https://doi.org/10.1016/j.bbagen.2014.05.006.
Jaksik, R., Iwanaszko, M., Rzeszowska-Wolny, J., Kimmel, M., 2015. Microarray experiments and factors which affect their reliability. Biol. Direct 10, 46. https://doi.org/10.1186/s13062-015-0077-2.
Jang, Y.S., Han, M.J., Lee, J., Im, J.A., Lee, Y.H., Papoutsakis, E.T., Bennett, G., Lee, S.Y., 2014. Proteomic analyses of the phase transition from acidogenesis to solventogenesis using solventogenic and non-solventogenic *Clostridium acetobutylicum* strains. Appl. Microbiol. Biotechnol. 98, 5105–5115. https://doi.org/10.1007/s00253-014-5738-z.
Janssen, H., Döring, C., Ehrenreich, A., Voigt, B., Hecker, M., Bahl, H., Fischer, R.-J., 2010. A proteomic and transcriptional view of acidogenic and solventogenic steady-state cells of *Clostridium acetobutylicum* in a chemostat culture. Appl. Microbiol. Biotechnol. 87, 2209–2226. https://doi.org/10.1007/S00253-010-2741-X.
Janssen, H., Grimmler, C., Ehrenreich, A., Bahl, H., Fischer, R.J., 2012. A transcriptional study of acidogenic chemostat cells of *Clostridium acetobutylicum* - Solvent stress caused by a transient n-butanol pulse. J. Biotechnol. 161, 354–365. https://doi.org/10.1016/j.jbiotec.2012.03.027.
Jiao, S., Zhang, Y., Wan, C., Lv, J., Du, R., Zhang, R., Han, B., 2016. Transcriptional analysis of degenerate strain *Clostridium beijerinckii* DG-8052 reveals a pleiotropic response to $CaCO_3$-associated recovery of solvent production. Sci. Rep. 6, 38818. https://doi.org/10.1038/srep38818.

Jin, Y., Pang, Y., Huang, M., Sun, J., Huang, Y., Gao, X., Li, R., He, K., Zhao, H., 2015. Combination of RNA sequencing and metabolite data to elucidate improved toxic compound tolerance and butanol fermentation of *Clostridium acetobutylicum* from wheat straw hydrolysate by supplying sodium sulfide. Bioresour. Technol. 198, 77–86. https://doi.org/10.1016/j.biortech.2015.08.139.

Jones, D.T., Woods, D.R., 1986. Acetone-butanol fermentation revisited. Microbiol. Rev. 50, 484–524.

Jones, S.W., Paredes, C.J., Tracy, B., Cheng, N., Sillers, R., Senger, R.S., Papoutsakis, E.T., 2008. The transcriptional program underlying the physiology of clostridial sporulation. Genome Biol. 9, R114. https://doi.org/10.1186/gb-2008-9-7-r114.

Jönsson, L.J., Martín, C., 2016. Pretreatment of lignocellulose: formation of inhibitory by-products and strategies for minimizing their effects. Bioresour. Technol. 199, 103–112. https://doi.org/10.1016/j.biortech.2015.10.009.

Jureckova, K., Raschmanova, H., Kolek, J., Vasylkivska, M., Branska, B., Patakova, P., Provaznik, I., Sedlar, K., 2021. Identification and validation of reference genes in *Clostridium beijerinckii* NRRL B-598 for RT-qPCR using RNA-Seq data. Front. Microbiol. 12, 476. https://doi.org/10.3389/fmicb.2021.640054.

Kanehisa, M., Goto, S., 2000. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 28, 27–30. https://doi.org/10.1093/nar/28.1.27.

Katsyv, A., Müller, V., 2020. Overcoming energetic barriers in acetogenic C1 conversion. Front. Bioeng. Biotechnol. 8, 621166 https://doi.org/10.3389/FBIOE.2020.621166/BIBTEX.

Klinke, H.B., Thomsen, A.B., Ahring, B.K., 2004. Inhibition of ethanol-producing yeast and bacteria by degradation products produced during pre-treatment of biomass. Appl. Microbiol. Biotechnol. 66, 10–26. https://doi.org/10.1007/s00253-004-1642-2.

Kolek, J., Branska, B., Drahokoupil, M., Patakova, P., Melzoch, K., 2016. Evaluation of viability, metabolic activity and spore quantity in clostridial cultures during ABE fermentation. FEMS Microbiol. Lett. 363, fnw031. https://doi.org/10.1093/femsle/fnw031.

Kolek, J., Diallo, M., Vasylkivska, M., Branska, B., Sedlar, K., Lopez-Contreras, A.M., Patakova, P., López-Contreras, A.M., Patakova, P., 2017. Comparison of expression of key sporulation, solventogenic and acetogenic genes in *C. beijerinckii* NRRL B-598 and its mutant strain overexpressing spo0A. Appl. Microbiol. Biotechnol. 101, 8279–8291. https://doi.org/10.1007/s00253-017-8555-3.

Kopylova, E., Noe, L., Touzet, H., 2012. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics 28, 3211–3217. https://doi.org/10.1093/bioinformatics/bts611.

Kotte, A.K., Severn, O., Bean, Z., Schwarz, K., Minton, N.P., Winzer, K., 2020. RRNPP-type quorum sensing affects solvent formation and sporulation in *Clostridium acetobutylicum*. Microbiol. (United Kingdom) 166, 579–592. https://doi.org/10.1099/mic.0.000916.

Langmead, B., Trapnell, C., Pop, M., Salzberg, S.L., 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10, R25. https://doi.org/10.1186/gb-2009-10-3-r25.

Lee, J., Blaschek, H.P., 2001. Glucose uptake in *Clostridium beijerinckii* NCIMB 8052 and the solvent-hyperproducing mutant BA101. Appl. Environ. Microbiol. 67, 5025–5031.

Lee, J., Mitchell, W.J., Tanguey, M., Blaschek, H.P., 2005. Evidence for the presence of an alternative glucose transport system in *Clostridium beijerinckii* NCIMB 8052 and the solvent-hyperproducing mutant BA101. Appl. Environ. Microbiol. 71, 3384–3387. https://doi.org/10.1128/AEM.71.6.3384-3387.2005.

Lee, S.-M., Cho, M.O., Park, C.H., Chung, Y.-C., Kim, J.H., Sang, B.-I., Um, Y., 2008. Continuous butanol production using suspended and immobilized *Clostridium beijerinckii* NCIMB 8052 with supplementary butyrate. Energy Fuel 22, 3459–3464. https://doi.org/10.1021/EF800076J.

Lee, S., Lee, J.H., Mitchell, R.J., 2015. Analysis of *Clostridium beijerinckii* NCIMB 8052's transcriptional response to ferulic acid and its application to enhance the strain tolerance. Biotechnol. Biofuels. 8, 68. https://doi.org/10.1186/S13068-015-0252-9.

Li, J.S., Barber, C.C., Herman, N.A., Cai, W., Zafrir, E., Du, Y., Zhu, X., Skyrud, W., Zhang, W., 2020a. Investigation of secondary metabolism in the industrial butanol hyper-producer *Clostridium saccharoperbutylacetonicum* N1-4. J. Ind. Microbiol. Biotechnol. 47, 319–328. https://doi.org/10.1007/S10295-020-02266-8.

Li, S., Huang, L., Ke, C., Pang, Z., Liu, L., 2020b. Pathway dissection, regulation, engineering and application: lessons learned from biobutanol production by solventogenic clostridia. Biotechnol. Biofuels. 13, 39. https://doi.org/10.1186/s13068-020-01674-3.

Li, W., Cheng, C., Cao, G., Yang, S.T., Ren, N., 2020c. Comparative transcriptome analysis of *Clostridium tyrobutyricum* expressing a heterologous uptake hydrogenase. Sci. Total Environ. 749, 142022 https://doi.org/10.1016/J.SCITOTENV.2020.142022.

Liao, Y., Smyth, G.K., Shi, W., 2014. FeatureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30, 923–930. https://doi.org/10.1093/bioinformatics/btt656.

Liao, C., Seo, S.O., Celik, V., Liu, H., Kong, W., Wang, Y., Blaschek, H., Jin, Y.S., Lu, T., 2015. Integrated, systems metabolic picture of acetone-butanol-ethanol fermentation by *Clostridium acetobutylicum*. Proc. Natl. Acad. Sci. U. S. A. 112, 8505–8510. https://doi.org/10.1073/PNAS.1423143112.

Liao, Z., Guo, X., Hu, J., Suo, Y., Fu, H., Wang, J., 2019. The significance of proline on lignocellulose-derived inhibitors tolerance in *Clostridium acetobutylicum* ATCC 824. Bioresour. Technol. 272, 561–569. https://doi.org/10.1016/j.biortech.2018.10.038.

Lipovsky, J., Patakova, P., Paulova, L., Pokorny, T., Rychtera, M., Melzoch, K., 2016. Butanol production by *Clostridium pasteurianum* NRRL B-598 in continuous culture compared to batch and fed-batch systems. Fuel Process. Technol. 144, 139–144. https://doi.org/10.1016/j.fuproc.2015.12.020.

Liu, J., Tan, Y., Yang, X., Chen, X., Li, F., 2013. Evaluation of *Clostridium ljungdahlii* DSM 13528 reference genes in gene expression studies by qRT-PCR. J. Biosci. Bioeng. 116, 460–464. https://doi.org/10.1016/j.jbiosc.2013.04.011.

Liu, Y., Tang, H., Lin, Z., Xu, P., 2015a. Mechanisms of acid tolerance in bacteria and prospects in biotechnology and bioremediation. Biotechnol. Adv. 33, 1484–1492. https://doi.org/10.1016/J.BIOTECHADV.2015.06.001.

Liu, Z., Qiao, K., Tian, L., Zhang, Q., Liu, Z.Y., Li, F.L., 2015b. Spontaneous large-scale autolysis in *Clostridium acetobutylicum* contributes to generation of more spores. Front. Microbiol. 6, 950. https://doi.org/10.3389/fmicb.2015.00950.

Liu, D., Xu, J., Wang, Y., Chen, Y., Shen, X., Niu, H., Guo, T., Ying, H., 2016a. Comparative transcriptomic analysis of *Clostridium acetobutylicum* biofilm and planktonic cells. J. Biotechnol. 218, 1–12. https://doi.org/10.1016/j.jbiotec.2015.11.017.

Liu, J., Guo, T., Wang, D., Shen, X., Liu, D., Niu, H., Liang, L., Ying, H., 2016b. Enhanced butanol production by increasing NADH and ATP levels in *Clostridium beijerinckii* NCIMB 8052 by insertional inactivation of Cbei_4110. Appl. Microbiol. Biotechnol. 100, 4985–4996. https://doi.org/10.1007/s00253-016-7299-9.

Liu, J., Guo, T., Yang, T., Xu, J., Tang, C., Liu, D., Ying, H., 2017. Transcriptome analysis of *Clostridium beijerinckii* adaptation mechanisms in response to ferulic acid. Int. J. Biochem. Cell Biol. 86, 14–21. https://doi.org/10.1016/j.biocel.2017.02.009.

Liu, D., Yang, Z., Chen, Y., Zhuang, W., Niu, H., Wu, J., Ying, H., 2018. *Clostridium acetobutylicum* grows vegetatively in a biofilm rich in heteropolysaccharides and cytoplasmic proteins. Biotechnol. Biofuels. 11, 315. https://doi.org/10.1186/S13068-018-1316-4.

Liu, J., Lin, Q., Chai, X., Luo, Y., Guo, T., 2018a. Enhanced phenolic compounds tolerance response of *Clostridium beijerinckii* NCIMB 8052 by inactivation of Cbei_3304. Microb. Cell Factories 17, 35. https://doi.org/10.1186/s12934-018-0884-0.

Liu, J., Liu, Z., Chai, X., Luo, Y., Guo, T., Ying, H., 2018b. Regulation of *p*-coumaric acid tolerance in *Clostridium beijerinckii* by disturbing the intracellular electron transport chain. Process Biochem. 68, 43–52. https://doi.org/10.1016/J.PROCBIO.2018.02.016.

Liu, X., Li, N., Liu, S., Wang, J., Zhang, N., Zheng, X., Leung, K.S., Cheng, L., 2019. Normalization methods for the analysis of unbalanced transcriptome data: a review. Front. Bioeng. Biotechnol. 7, 358. https://doi.org/10.3389/fbioe.2019.00358.

Liu, H., Zhang, J., Yuan, J., Jiang, X., Jiang, L., Li, Z., Yin, Z., Du, Y., Zhao, G., Liu, B., Huang, D., 2020. Gene coexpression network analysis reveals a novel metabolic mechanism of *Clostridium acetobutylicum* responding to phenolic inhibitors from lignocellulosic hydrolysates. Biotechnol. Biofuels. 13, 163. https://doi.org/10.1186/s13068-020-01802-z.

Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550. https://doi.org/10.1186/s13059-014-0550-8.

Luo, H., Zheng, P., Bilal, M., Xie, F., Zeng, Q., Zhu, C., Yang, R., Wang, Z., 2020. Efficient bio-butanol production from lignocellulosic waste by elucidating the mechanisms of *Clostridium acetobutylicum* response to phenolic inhibitors. Sci. Total Environ. 710, 136399 https://doi.org/10.1016/j.scitotenv.2019.136399.

Lv, J., Jiao, S., Du, R., Zhang, R., Zhang, Y., Han, B., 2016. Proteomic analysis to elucidate degeneration of *Clostridium beijerinckii* NCIMB 8052 and role of Ca^{2+} in strain recovery from degeneration. J. Ind. Microbiol. Biotechnol. 43, 741–750. https://doi.org/10.1007/s10295-016-1754-6.

Mascher, T., 2006. Intramembrane-sensing histidine kinases: a new family of cell envelope stress sensors in Firmicutes bacteria. FEMS Microbiol. Lett. 264, 133–144. https://doi.org/10.1111/J.1574-6968.2006.00444.X.

Maté de Gérando, H., Wasels, F., Bisson, A., Clement, B., Bidard, F., Jourdier, E., López-Contreras, A.M., Lopes Ferreira, N., 2018. Genome and transcriptome of the natural isopropanol producer *Clostridium beijerinckii* DSM 6423. BMC Genomics 19, 242. https://doi.org/10.1186/s12864-018-4636-7.

McAnulty, M.J., Yen, J.Y., Freedman, B.G., Senger, R.S., 2012. Genome-scale modeling using flux ratio constraints to enable metabolic engineering of clostridial metabolism in silico. BMC Syst. Biol. 6, 42. https://doi.org/10.1186/1752-0509-6-42.

Mercatelli, D., Scalambra, L., Triboli, L., Ray, F., Giorgi, F.M., 2020. Gene regulatory network inference resources: a practical overview. Biochim. Biophys. Acta - Gene Regul. Mech. 2020, 194430. https://doi.org/10.1016/j.bbagrm.2019.194430.

Millat, T., Janssen, H., Bahl, H., Fischer, R.J., Wolkenhauer, O., 2013a. Integrative modelling of pH-dependent enzyme activity and transcriptomic regulation of the acetone-butanol-ethanol fermentation of *Clostridium acetobutylicum* in continuous culture. Microb. Biotechnol. 6, 526–539. https://doi.org/10.1111/1751-7915.12033.

Millat, T., Janssen, H., Thorn, G.J., King, J.R., Bahl, H., Fischer, R.J., Wolkenhauer, O., 2013b. A shift in the dominant phenotype governs the pH-induced metabolic switch of *Clostridium acetobutylicum* in phosphate-limited continuous cultures. Appl. Microbiol. Biotechnol. 97, 6451–6466. https://doi.org/10.1007/s00253-013-4860-7.

Milne, C.B., Eddy, J.A., Raju, R., Ardekani, S., Kim, P.J., Senger, R.S., Jin, Y.S., Blaschek, H.P., Price, N.D., 2011. Metabolic network reconstruction and genome-scale model of butanol-producing strain *Clostridium beijerinckii* NCIMB 8052. BMC Syst. Biol. 5, 130. https://doi.org/10.1186/1752-0509-5-130.

Mitchell, W.J., 1996. Carbohydrate uptake and utilization by *Clostridium beijerinckii* NCIMB 8052. Anaerobe 2, 379–384. https://doi.org/10.1006/anae.1996.0048.

Mitchell, W.J., 2016. Sugar uptake by the solventogenic clostridia. World J. Microbiol. Biotechnol. 32, 32. https://doi.org/10.1007/s11274-015-1981-4.

Munich, B., Cadle-Davidson, L., Osier, M.V., 2021. Comparison of short-read sequence aligners indicates strengths and weaknesses for biologists to consider. Front. Plant Sci. 12 https://doi.org/10.3389/fpls.2021.657240.

Nguyen, N.-P.-T., Linder, S., Flitsch, S.K., Schiel-Bengelsdorf, B., Dürre, P., Soucaille, P., 2016. Cap0037, a novel global regulator of *Clostridium acetobutylicum* metabolism. MBio 7. https://doi.org/10.1128/mbio.01218-16 e01218-16.

Noar, J., Makwana, S.T., Bruno-Bárcena, J.M., 2014. Complete genome sequence of solvent-tolerant *Clostridium beijerinckii* strain SA-1. Genome Announc. 2, 1310–1324. https://doi.org/10.1128/genomeA.01310-14.

Noguchi, T., Tashiro, Y., Yoshida, T., Zheng, J., Sakai, K., Sonomoto, K., 2013. Efficient butanol production without carbon catabolite repression from mixed sugars with *Clostridium saccharoperbutylacetonicum* N1-4. J. Biosci. Bioeng. 116, 716–721. https://doi.org/10.1016/J.JBIOSC.2013.05.030.

Nölling, J., Breton, G., Omelchenko, M.V., Makarova, K.S., Zeng, Q., Gibson, G., Lee, Hong Mei, Dubois, J., Qiu, D., Hitti, J., Aldredge, T., Ayers, M., Bashirzadeh, R., Bochner, H., Boivin, M., Bross, S., Bush, D., Butler, C., Caron, A., Caruso, A., Cook, R., Daggett, P., Deloughery, C., Egan, J., Elliston, D., Engelstein, M., Ezedi, J., Gilbert, K., Goyal, A., Guerin, J., Ho, T., Holtham, K., Joseph, P., Keagle, P., Kozlovsky, J., LaPlante, M., LeBlanc, G., Lumm, W., Majeski, A., McDougall, S., Mank, P., Mao, J.I., Nocco, D., Patwell, D., Phillips, J., Pothier, B., Prabhakar, S., Richterich, P., Rice, P., Rosetti, D., Rossetti, M., Rubenfield, M., Sachdeva, M., Snell, P., Spadafora, R., Spitzer, L., Shimer, G., Thomann, H.U., Vicaire, R., Wall, K., Wang, Y., Weinstock, K., Wong, Lai Peng, Wonsey, A., Xu, Q., Zhang, L., Wolf, Y.I., Tatusov, R.L., Sabathe, F., Doucette-Stamm, L., Soucaille, P., Daly, M.J., Bennett, G. N., Koonin, E.V., Smith, D.R., 2001. Genome sequence and comparative analysis of the solvent-producing bacterium *Clostridium acetobutylicum*. J. Bacteriol. 183, 4823–4838. https://doi.org/10.1128/JB.183.16.4823-4838.2001.

Ohtani, K., Shimizu, T., 2016. Regulation of toxin production in *Clostridium perfringens*. Toxins (Basel). 8, 207. https://doi.org/10.3390/toxins8070207.

Okonkwo, C.C., Ujor, V., Ezeji, T.C., 2019. Chromosomal integration of aldo-keto-reductase and short-chain dehydrogenase/reductase genes in *Clostridium beijerinckii* NCIMB 8052 enhanced tolerance to lignocellulose-derived microbial inhibitory compounds. Sci. Rep. 9, 7634. https://doi.org/10.1038/s41598-019-44061-1.

Patakova, P., Linhova, M., Rychtera, M., Paulova, L., Melzoch, K., 2013. Novel and neglected issues of acetone–butanol–ethanol (ABE) fermentation by clostridia: *Clostridium* metabolic diversity, tools for process mapping and continuous fermentation systems. Biotechnol. Adv. 31, 58–67. https://doi.org/10.1016/J.BIOTECHADV.2012.01.010.

Patakova, P., Kolek, J., Sedlar, K., Koscova, P., Branska, B., Kupkova, K., Paulova, L., Provaznik, I., 2018. Comparative analysis of high butanol tolerance and production in clostridia. Biotechnol. Adv. 36, 721–738. https://doi.org/10.1016/j.biotechadv.2017.12.004.

Patakova, P., Branska, B., Sedlar, K., Vasylkivska, M., Jureckova, K., Kolek, J., Koscova, P., Provaznik, I., 2019. Acidogenesis, solventogenesis, metabolic stress response and life cycle changes in *Clostridium beijerinckii* NRRL B-598 at the transcriptomic level. Sci. Rep. 9, 1371. https://doi.org/10.1038/s41598-018-37679-0.

Patakova, P., Kolek, J., Jureckova, K., Branska, B., Sedlar, K., Vasylkivska, M., Provaznik, I., 2021. Deeper below the surface—transcriptional changes in selected genes of *Clostridium beijerinckii* in response to butanol shock. Microbiologyopen 10, e1146. https://doi.org/10.1002/mbo3.1146.

Pfaffl, M.W., 2001. A new mathematical model for relative quantification in real-time RT–PCR. Nucleic Acids Res. 29, E45. https://doi.org/10.1093/nar/29.9.e45.

Poehlein, A., Solano, J.D.M., Flitsch, S.K., Krabben, P., Winzer, K., Reid, S.J., Jones, D.T., Green, E., Minton, N.P., Daniel, R., Dürre, P., 2017. Microbial solvent formation revisited by comparative genome analysis. Biotechnol. Biofuels. 10, 58. https://doi.org/10.1186/s13068-017-0742-z.

Qureshi, N., Bowman, M.J., Saha, B.C., Hector, R., Berhow, M.A., Cotta, M.A., 2012. Effect of cellulosic sugar degradation products (furfural and hydroxymethyl furfural) on acetone–butanol–ethanol (ABE) fermentation using *Clostridium beijerinckii* P260. Food Bioprod. Process. 90, 533–540. https://doi.org/10.1016/J.FBP.2011.09.002.

Radonić, A., Thulke, S., Mackay, I.M., Landt, O., Siegert, W., Nitsche, A., 2004. Guideline to reference gene selection for quantitative real-time PCR. Biochem. Biophys. Res. Commun. 313, 856–862. https://doi.org/10.1016/j.bbrc.2003.11.177.

Ralston, M.T., Papoutsakis, E.T., 2018. RNAseq-based transcriptome assembly of *Clostridium acetobutylicum* for functional genome annotation and discovery. AIChE J. 64, 4271–4280. https://doi.org/10.1002/AIC.16396.

Ren, C., Gu, Y., Hu, S., Wu, Y., Wang, P., Yang, Y., Yang, C., Yang, S., Jiang, W., 2010. Identification and inactivation of pleiotropic regulator CcpA to eliminate glucose repression of xylose utilization in *Clostridium acetobutylicum*. Metab. Eng. 12, 446–454. https://doi.org/10.1016/j.ymben.2010.05.002.

Ren, C., Gu, Y., Wu, Y., Zhang, W., Yang, C., Yang, S., Jiang, W., 2012. Pleiotropic functions of catabolite control protein CcpA in butanol-producing *Clostridium acetobutylicum*. BMC Genomics 13, 349. https://doi.org/10.1186/1471-2164-13-349.

Riley, E.P., Schwarz, C., Derman, A.I., Lopez-Garrido, J., 2021. Milestones in *Bacillus subtilis* sporulation research. Microb. Cell 8, 1–16. https://doi.org/10.15698/MIC2021.01.739.

Robinson, M.D., McCarthy, D.J., Smyth, G.K., 2010. EdgeR: a bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140. https://doi.org/10.1093/bioinformatics/btp616.

Rocha, D.J.P., Santos, C.S., Pacheco, L.G.C., 2015. Bacterial reference genes for gene expression studies by RT-qPCR: survey and analysis. Antonie Van Leeuwenhoek 108, 685–693. https://doi.org/10.1007/s10482-015-0524-1.

Sandoval, N.R., Papoutsakis, E.T., 2016. Engineering membrane and cell-wall programs for tolerance to toxic chemicals: Beyond solo genes. Curr. Opin. Microbiol. https://doi.org/10.1016/j.mib.2016.06.005.

Sandoval-Espinola, W.J., Chinn, M.S., Thon, M.R., Bruno-Bárcena, J.M., 2017. Evidence of mixotrophic carbon-capture by n-butanol-producer *Clostridium beijerinckii*. Sci. Rep. 7, 12759. https://doi.org/10.1038/s41598-017-12962-8.

Saujet, L., Pereira, F.C., Henriques, A.O., Martin-Verstraete, I., 2014. The regulatory network controlling spore formation in *Clostridium difficile*. FEMS Microbiol. Lett. 358, 1–10. https://doi.org/10.1111/1574-6968.12540.

Schena, M., Shalon, D., Davis, R.W., Brown, P.O., 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science (80-.)). 270, 467–470. https://doi.org/10.1126/science.270.5235.467.

Schmittgen, T.D., Livak, K.J., 2008. Analyzing real-time PCR data by the comparative CT method. Nat. Protoc. 3, 1101–1108. https://doi.org/10.1038/nprot.2008.73.

Schwarz, K.M., Kuit, W., Grimmler, C., Ehrenreich, A., Kengen, S.W.M., 2012. A transcriptional study of acidogenic chemostat cells of *Clostridium acetobutylicum* - Cellular behavior in adaptation to n-butanol. J. Biotechnol. 161, 366–377. https://doi.org/10.1016/j.jbiotec.2012.03.018.

Sedlar, K., Kolek, J., Skutkova, H., Branska, B., Provaznik, I., Patakova, P., 2015. Complete genome sequence of *Clostridium pasteurianum* NRRL B-598, a non-type strain producing butanol. J. Biotechnol. 214, 113–114. https://doi.org/10.1016/j.jbiotec.2015.09.022.

Sedlar, K., Kolek, J., Provaznik, I., Patakova, P., 2017. Reclassification of non-type strain *Clostridium pasteurianum* NRRL B-598 as *Clostridium beijerinckii* NRRL B-598. J. Biotechnol. 244, 1–3. https://doi.org/10.1016/j.jbiotec.2017.01.003.

Sedlar, K., Koscova, P., Vasylkivska, M., Branska, B., Kolek, J., Kupkova, K., Patakova, P., Provaznik, I., 2018. Transcription profiling of butanol producer *Clostridium beijerinckii* NRRL B-598 using RNA-Seq. BMC Genomics 19, 415. https://doi.org/10.1186/s12864-018-4805-8.

Sedlar, K., Kolek, J., Gruber, M., Jureckova, K., Branska, B., Csaba, G., Vasylkivska, M., Zimmer, R., Patakova, P., Provaznik, I., 2019. A transcriptional response of *Clostridium beijerinckii* NRRL B-598 to a butanol shock. Biotechnol. Biofuels. 12, 243. https://doi.org/10.1186/s13068-019-1584-7.

Sedlar, K., Nykrynova, M., Bezdicek, M., Branska, B., Lengerova, M., Patakova, P., Skutkova, H., 2021a. Diversity and evolution of *Clostridium beijerinckii* and complete genome of the type strain DSM 791T. Processes 9, 1196. https://doi.org/10.3390/pr9071196.

Sedlar, K., Vasylkivska, M., Musilova, J., Branska, B., Provaznik, I., Patakova, P., 2021b. Phenotypic and genomic analysis of isopropanol and 1,3-propanediol producer *Clostridium diolis* DSM 15410. Genomics 113, 1109–1119. https://doi.org/10.1016/j.ygeno.2020.11.007.

Senger, R.S., Papoutsakis, E.T., 2008a. Genome-scale model for *Clostridium acetobutylicum*: Part I. Metabolic network resolution and analysis. Biotechnol. Bioeng. 101, 1036–1052. https://doi.org/10.1002/bit.22010.

Senger, R.S., Papoutsakis, E.T., 2008b. Genome-scale model for *Clostridium acetobutylicum*: Part II. Development of specific proton flux states and numerically determined sub-systems. Biotechnol. Bioeng. 101, 1053–1071. https://doi.org/10.1002/bit.22009.

Seo, S.-O., Jansen, H., Magis, A., Wang, Y., Lu, T., Price, N.D., Jin, Y.-S., Blaschek, H.P., 2017. Genomic, transcriptional, and phenotypic analysis of the glucose derepressed *Clostridium beijerinckii* mutant exhibiting acid crash phenotype. Biotechnol. J. 12, 1700182. https://doi.org/10.1002/biot.201700182.

Seo, S.O., Lu, T., Jin, Y.S., Blaschek, H.P., 2021. A comparative phenotypic and genomic analysis of *Clostridium beijerinckii* mutant with enhanced solvent production. J. Biotechnol. 329, 49–55. https://doi.org/10.1016/J.JBIOTEC.2021.02.002.

Servinsky, M.D., Kiel, J.T., Dupuy, N.F., Sund, C.J., 2010. Transcriptional analysis of differential carbohydrate utilization by *Clostridium acetobutylicum*. Microbiology 156, 3478–3491. https://doi.org/10.1099/mic.0.037085-0.

Servinsky, M.D., Rentsch, R.L., Perisin, M.A., Gerlach, E.S., Liu, S., Sund, C.J., 2018. Arabinose-induced catabolite repression as a mechanism for pentose hierarchy control in *Clostridium acetobutylicum* ATCC 824. mSystems 3. https://doi.org/10.1128/msystems.00064-18 e00064-18.

Shi, Z., Blaschek, H.P., 2008. Transcriptional analysis of *Clostridium beijerinckii* NCIMB 8052 and the hyper-butanol-producing mutant BA101 during the shift from acidogenesis to solventogenesis. Appl. Environ. Microbiol. 74, 7709–7714. https://doi.org/10.1128/aem.01948-08.

Spero, M.A., Aylward, F.O., Currie, C.R., Donohue, T.J., 2015. Phylogenomic analysis and predicted physiological role of the proton-translocating NADH: Quinone oxidoreductase (complex I) across bacteria. MBio 6. https://doi.org/10.1128/MBIO.00389-15 e00389-15.

Srinivasan, B., 2021. A guide to the Michaelis-Menten equation: steady state and beyond. FEBS J. febs.16124 https://doi.org/10.1111/FEBS.16124.

Steiner, E., Scott, J., Minton, N.P., Winzer, K., 2012. An agr quorum sensing system that regulates granulose formation and sporulation in *Clostridium acetobutylicum*. Appl. Environ. Microbiol. 78, 1113–1122. https://doi.org/10.1128/AEM.06376-11.

Tan, Y., Liu, J., Chen, X., Zheng, H., Li, F., 2013. RNA-seq-based comparative transcriptome analysis of the syngas-utilizing bacterium *Clostridium ljungdahlii* DSM 13528 grown autotrophically and heterotrophically. Mol. BioSyst. 9, 2775–2784. https://doi.org/10.1039/C3MB70232D.

Tan, Y., Liu, Z.-Y., Liu, Z., Zheng, H.-J., Li, F.-L., 2015. Comparative transcriptome analysis between *carA*-disruption *Clostridium acetobutylicum* and its parent strain. Mol. BioSyst. 11, 1434–1442. https://doi.org/10.1039/C4MB00600C.

Tangney, M., Brehm, J.K., Minton, N.P., Mitchell, W.J., 1998. A gene system for glucitol transport and metabolism in *Clostridium beijerinckii* NCIMB 8052. Appl. Environ. Microbiol. 64, 1612–1619. https://doi.org/10.1128/aem.64.5.1612-1619.1998.

Tatusov, R.L., Galperin, M.Y., Natale, D.A., Koonin, E.V., 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res. 28, 33–36. https://doi.org/10.1093/NAR/28.1.33.

Taylor, T.M., Ravishankar, S., Bhargava, K., Juneja, V.K., 2014. Chemical preservatives and natural antimicrobial compounds. In: Davidson, P.M., Taylor, T.M., Schmidt, S. E. (Eds.), Food Microbiology: Fundamentals and Frontiers. John Wiley & Sons, Ltd, pp. 765–801. https://doi.org/10.1128/9781555818463.CH30.

Tomas, C.A., Alsaker, K.V., Bonarius, H.P.J., Hendriksen, W.T., Yang, H., Beamish, J.A., Paredes, C.J., Papoutsakis, E.T., 2003a. DNA array-based transcriptional analysis of asporogenous, nonsolventogenic *Clostridium acetobutylicum* strains SKO1 and M5. J. Bacteriol. 185, 4539–4547. https://doi.org/10.1128/JB.185.15.4539-4547.2003.

Tomas, C.A., Welker, N.E., Papoutsakis, E.T., 2003b. Overexpression of *groESL* in *Clostridium acetobutylicum* results in increased solvent production and tolerance, prolonged metabolism, and changes in the cell's transcriptional program. Appl. Environ. Microbiol. 69, 4951–4965. https://doi.org/10.1128/AEM.69.8.4951-4965.2003.

Tomas, C.A., Beamish, J., Papoutsakis, E.T., 2004. Transcriptional analysis of butanol stress and tolerance in *Clostridium acetobutylicum*. J. Bacteriol. 186, 2006–2018. https://doi.org/10.1128/JB.186.7.2006-2018.2004.

Trapnell, C., Pachter, L., Salzberg, S.L., 2009. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25, 1105–1111. https://doi.org/10.1093/bioinformatics/btp120.

Tummala, S.B., Junne, S.G., Paredes, C.J., Papoutsakis, E.T., 2003. Transcriptional analysis of product-concentration driven changes in cellular programs of recombinant *Clostridium acetobutylicum* strains. Biotechnol. Bioeng. 84, 842–854. https://doi.org/10.1002/bit.10851.

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., Speleman, F., 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol. 3 https://doi.org/10.1186/gb-2002-3-7-research0034. RESEARCH0034.

Vasileva, D., Janssen, H., Hönicke, D., Ehrenreich, A., Bahl, H., 2012. Effect of iron limitation and fur gene inactivation on the transcriptional profile of the strict anaerobe *Clostridium acetobutylicum*. Microbiol. (United Kingdom) 158, 1918–1929. https://doi.org/10.1099/mic.0.056978-0.

Vasylkivska, M., Patakova, P., 2020. Role of efflux in enhancing butanol tolerance of bacteria. J. Biotechnol. 320, 17–27. https://doi.org/10.1016/j.jbiotec.2020.06.008.

Vasylkivska, M., Jureckova, K., Branska, B., Sedlar, K., Kolek, J., Provaznik, I., Patakova, P., 2019. Transcriptional analysis of amino acid, metal ion, vitamin and carbohydrate uptake in butanol-producing *Clostridium beijerinckii* NRRL B-598. PLoS One 14, e0224560. https://doi.org/10.1371/journal.pone.0224560.

Venkataramanan, K.P., Jones, S.W., McCormick, K.P., Kunjeti, S.G., Ralston, M.T., Meyers, B.C., Papoutsakis, E.T., 2013. The *Clostridium* small RNome that responds to stress: the paradigm and importance of toxic metabolite stress in *C. acetobutylicum*. BMC Genomics 14, 849. https://doi.org/10.1186/1471-2164-14-849.

Venkataramanan, K.P., Min, L., Hou, S., Jones, S.W., Ralston, M.T., Lee, K.H., Papoutsakis, E.T., 2015. Complex and extensive post-transcriptional regulation revealed by integrative proteomic and transcriptomic analysis of metabolite stress response in *Clostridium acetobutylicum*. Biotechnol. Biofuels. 8, 81. https://doi.org/10.1186/S13068-015-0260-9.

Veza, I., Muhamad Said, M.F., Latiff, Z.A., 2021. Recent advances in butanol production by acetone-butanol-ethanol (ABE) fermentation. Biomass Bioenergy 144, 105919. https://doi.org/10.1016/J.BIOMBIOE.2020.105919.

Wagner, G.P., Kin, K., Lynch, V.J., 2012. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. Theory Biosci. 131, 281–285. https://doi.org/10.1007/s12064-012-0162-3.

Wang, Z., Gerstein, M., Snyder, M., 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet. 10, 57–63. https://doi.org/10.1038/nrg2484.

Wang, Y., Li, X., Mao, Y., Blaschek, H.P., 2011. Single-nucleotide resolution analysis of the transcriptome structure of *Clostridium beijerinckii* NCIMB 8052 using RNA-Seq. BMC Genomics 12, 479. https://doi.org/10.1186/1471-2164-12-479.

Wang, Y., Li, X., Mao, Y., Blaschek, H.P., 2012. Genome-wide dynamic transcriptional profiling in *Clostridium beijerinckii* NCIMB 8052 using single-nucleotide resolution RNA-Seq. BMC Genomics 13, 102. https://doi.org/10.1186/1471-2164-13-102.

Wang, Q., Venkataramanan, K.P., Huang, H., Papoutsakis, E.T., Wu, C.H., 2013a. Transcription factors and genetic circuits orchestrating the complex, multilayered response of *Clostridium acetobutylicum* to butanol and butyrate stress. BMC Syst. Biol. 7, 120. https://doi.org/10.1186/1752-0509-7-120.

Wang, Y., Li, X., Blaschek, H.P., 2013b. Effects of supplementary butyrate on butanol production and the metabolic switch in *Clostridium beijerinckii* NCIMB 8052: genome-wide transcriptional analysis with RNA-Seq. Biotechnol. Biofuels. 6, 138. https://doi.org/10.1186/1754-6834-6-138.

Wang, X., Bai, X., Chen, D.-F., Chen, F.-Z., Li, B.-Z., Yuan, Y.-J., 2015. Increasing proline and myo-inositol improves tolerance of *Saccharomyces cerevisiae* to the mixture of multiple lignocellulose-derived inhibitors. Biotechnol. Biofuels. 8, 1. https://doi.org/10.1186/S13068-015-0329-5.

Wu, Y.D., Xue, C., Chen, L.J., Wan, H.H., Bai, F.W., 2015. Transcriptional analysis of micronutrient zinc-associated response for enhanced carbohydrate utilization and earlier solventogenesis in *Clostridium acetobutylicum*. Sci. Rep. 5, 16598. https://doi.org/10.1038/srep16598.

Xiao, H., Gu, Y., Ning, Y., Yang, Y., Mitchell, W.J., Jiang, W., Yang, S., 2011. Confirmation and elimination of xylose metabolism bottlenecks in glucose phosphoenolpyruvate-dependent phosphotransferase system-deficient *Clostridium acetobutylicum* for simultaneous utilization of glucose, xylose, and arabinose. Applied and Environmental Microbiology 77, 7886–7895. https://doi.org/10.1128/AEM.00644-11.

Xin, X., Cheng, C., Du, G., Chen, L., Xue, C., 2020. Metabolic engineering of histidine kinases in *Clostridium beijerinckii* for enhanced butanol production. Front. Bioeng. Biotechnol. 8, 214. https://doi.org/10.3389/fbioe.2020.00214.

Xu, M., Zhao, J., Yu, L., Tang, I.C., Xue, C., Yang, S.T., 2015. Engineering *Clostridium acetobutylicum* with a histidine kinase knockout for enhanced n-butanol tolerance and production. Appl. Microbiol. Biotechnol. 99, 1011–1022. https://doi.org/10.1007/s00253-014-6249-7.

Xu, M., Zhao, J., Yu, L., Yang, S.T., 2017. Comparative genomic analysis of *Clostridium acetobutylicum* for understanding the mutations contributing to enhanced butanol tolerance and production. J. Biotechnol. 263, 36–44. https://doi.org/10.1016/j.jbiotec.2017.10.010.

Xu, J.-Y., Xu, Z., Liu, X., Tan, M., Ye, B.-C., 2018. Protein acetylation and butyrylation regulate the phenotype and metabolic shifts of the endospore-forming *Clostridium acetobutylicum*. Mol. Cell. Proteomics 17, 1156–1169. https://doi.org/10.1074/MCP.RA117.000372.

Xue, Q., Yang, Y., Chen, J., Chen, L., Yang, S., Jiang, W., Gu, Y., 2016. Roles of three AbrBs in regulating two-phase *Clostridium acetobutylicum* fermentation. Appl. Microbiol. Biotechnol. 100, 9081–9089. https://doi.org/10.1007/s00253-016-7638-x.

Yang, H., Haddad, H., Tomas, C., Alsaker, K., Papoutsakis, E.T., 2003. A segmental nearest neighbor normalization and gene identification method gives superior results for DNA-array analysis. Proc. Natl. Acad. Sci. 100, 1122–1127. https://doi.org/10.1073/PNAS.0237337100.

Yang, Y., Nie, X., Jiang, Y., Yang, C., Gu, Y., Jiang, W., 2018. Metabolic regulation in solventogenic clostridia: regulators, mechanisms and engineering. Biotechnol. Adv. 36, 905–914. https://doi.org/10.1016/J.BIOTECHADV.2018.02.012.

Yang, H., Yu, Y., Fu, C., Chen, F., 2019. Bacterial acid resistance toward organic weak acid revealed by RNA-Seq transcriptomic analysis in *Acetobacter pasteurianus*. Front. Microbiol. 10, 1616. https://doi.org/10.3389/FMICB.2019.01616/FULL.

Yang, Y., Lang, N., Zhang, L., Wu, H., Jiang, W., Gu, Y., 2020a. A novel regulatory pathway consisting of a two-component system and an ABC-type transporter contributes to butanol tolerance in *Clostridium acetobutylicum*. Appl. Microbiol. Biotechnol. 104, 5011–5023. https://doi.org/10.1007/S00253-020-10555-6.

Yang, Y., Zhang, H., Lang, N., Zhang, L., Chai, C., He, H., Jiang, W., Gu, Y., 2020b. The Small RNA sr8384 is a crucial regulator of cell growth in solventogenic clostridia. Appl. Environ. Microbiol. 86 https://doi.org/10.1128/AEM.00665-20 e00665-20.

Yao, D., Dong, S., Wang, P., Chen, T., Wang, J., Yue, Z.-B., Wang, Y., 2017. Robustness of *Clostridium saccharoperbutylacetonicum* for acetone-butanol-ethanol production: effects of lignocellulosic sugars and inhibitors. Fuel 208, 549–557. https://doi.org/10.1016/J.FUEL.2017.07.004.

Ying, H., Liu, D., Chen, Y., Li, A., Zhou, T., Wu, J., Lin, X., Chen, X., Xie, J., Bai, J., 2012. *Clostridium acetobutylicum* and Application Thereof. US9249433B2.

Yoo, M., Bestel-Corre, G., Croux, C., Riviere, A., Meynial-Salles, I., Soucaille, P., 2015. A quantitative system-scale characterization of the metabolism of *Clostridium acetobutylicum*. MBio 6, e01808–e01815. https://doi.org/10.1128/mBio.01808-15.

Yoo, M., Croux, C., Meynial-Salles, I., Soucaille, P., 2016. Elucidation of the roles of *adhE1* and *adhE2* in the primary metabolism of *Clostridium acetobutylicum* by combining in-frame gene deletion and a quantitative system-scale approach. Biotechnol. Biofuels. 9, 92. https://doi.org/10.1186/s13068-016-0507-0.

Yoo, M., Croux, C., Meynial-Salles, I., Soucaille, P., 2017. Metabolic flexibility of a butyrate pathway mutant of *Clostridium acetobutylicum*. Metab. Eng. 40, 138–147. https://doi.org/10.1016/j.ymben.2017.01.011.

Zhang, Y., Ezeji, T.C., 2013. Transcriptional analysis of *Clostridium beijerinckii* NCIMB 8052 to elucidate role of furfural stress during acetone butanol ethanol fermentation. Biotechnol. Biofuels. 6, 66. https://doi.org/10.1186/1754-6834-6-66.

Zhang, Y., Ezeji, T.C., 2014. Elucidating and alleviating impacts of lignocellulose-derived microbial inhibitors on *Clostridium beijerinckii* during fermentation of *Miscanthus giganteus* to butanol. J. Ind. Microbiol. Biotechnol. 41, 1505–1516. https://doi.org/10.1007/S10295-014-1493-5.

Zhang, W.L., Liu, Z.Y., Liu, Z., Li, F.L., 2012a. Butanol production from corncob residue using *Clostridium beijerinckii* NCIMB 8052. Lett. Appl. Microbiol. 55, 240–246. https://doi.org/10.1111/j.1472-765X.2012.03283.x.

Zhang, Y., Han, B., Ezeji, T.C., 2012b. Biotransformation of furfural and 5-hydroxymethyl furfural (HMF) by *Clostridium acetobutylicum* ATCC 824 during butanol fermentation. New Biotechnol. 29, 345–351. https://doi.org/10.1016/J.NBT.2011.09.001.

Zhang, Z.H., Jhaveri, D.J., Marshall, V.M., Bauer, D.C., Edson, J., Narayanan, R.K., Robinson, G.J., Lundberg, A.E., Bartlett, P.F., Wray, N.R., Zhao, Q.-Y., 2014. A comparative study of techniques for differential expression analysis on RNA-Seq data. PLoS One 9, e103207. https://doi.org/10.1371/journal.pone.0103207.

Zhang, Y., Ujor, V., Wick, M., Ezeji, T.C., 2015. Identification, purification and characterization of furfural transforming enzymes from *Clostridium beijerinckii* NCIMB 8052. Anaerobe 33, 124–131. https://doi.org/10.1016/J.ANAEROBE.2015.03.005.

Zhang, J., Zhu, W., Xu, H., Li, Y., Hua, D., Jin, F., Gao, M., Zhang, X., 2016. Simultaneous glucose and xylose uptake by an acetone/butanol/ethanol producing laboratory *Clostridium beijerinckii* strain SE-2. Biotechnol. Lett. 38, 611–617. https://doi.org/10.1007/s10529-015-2028-5.

Zhang, Y., Jiao, S.Y., Lv, J., Du, R.J., Yan, X.N., Wan, C.X., Zhang, R.J., Han, B., 2017. Sigma factor regulated cellular response in a non-solvent producing *Clostridium beijerinckii* degenerated strain: a comparative transcriptome analysis. Front. Microbiol. 8, 23. https://doi.org/10.3389/fmicb.2017.00023.

Zhang, H., Yang, P., Wang, Z., Li, M., Zhang, J., Liu, D., Chen, Y., Ying, H., 2021. *Clostridium acetobutylicum* biofilm: advances in understanding the basis. Front. Bioeng. Biotechnol. 9, 658568 https://doi.org/10.3389/fbioe.2021.658568.

Zhuang, W., Yang, J., Wu, J., Liu, D., Zhou, J., Chen, Y., Ying, H., 2016. Extracellular polymer substances and the heterogeneity of *Clostridium acetobutylicum* biofilm induced tolerance to acetic acid and butanol. RSC Adv. 6, 33695–33704. https://doi.org/10.1039/c5ra24923f.

5.15 Article XV

OBRUČA, Stanislav, DVOŘÁK, Pavel, SEDLÁČEK, Petr, KOLLER, Martin, **SEDLÁŘ, Karel**, PERNICOVÁ, Iva, and ŠAFRÁNEK, David. Polyhydroxyalkanoates synthesis by halophiles and thermophiles: towards sustainable production of microbial bioplastics. *Biotechnology Advances*. 2022. 58, p. 107906. (2022 IF not announced yet, 2021 IF = 17.681, D1 in BIOTECHNOLOGY & APPLIED MICROBIOLOGY)

DOI: https://doi.org/10.1016/j.biotechadv.2022.107906

Article XV: Polyhydroxyalkanoates synthesis by halophiles and thermophiles

Biotechnology Advances 58 (2022) 107906

Contents lists available at ScienceDirect

Biotechnology Advances

journal homepage: www.elsevier.com/locate/biotechadv

Research review paper

Polyhydroxyalkanoates synthesis by halophiles and thermophiles: towards sustainable production of microbial bioplastics

Stanislav Obruča [a,*], Pavel Dvořák [b], Petr Sedláček [a], Martin Koller [c,d], Karel Sedlář [e], Iva Pernicová [a], David Šafránek [f]

[a] Faculty of Chemistry, Brno University of Technology, Purkynova 118, 612 00 Brno, Czech Republic
[b] Department of Experimental Biology, Section of Microbiology, Faculty of Science, Masaryk University, Kamenice 753/5, Brno 62500, Czech Republic
[c] Institute of Chemistry, NAWI Graz, University of Graz, Heinrichstrasse 28/III, 8010 Graz, Austria
[d] ARENA Arbeitsgemeinschaft für Ressourcenschonende & Nachhaltige Technologien, Inffeldgasse 21b, 8010 Graz, Austria
[e] Department of Biomedical Engineering, Faculty of Electrical Engineering and Communication, Brno University of Technology, Technicka 12, Brno, 616 00, Czech Republic
[f] Faculty of Informatics, Masaryk University, Botanicka 68a, Brno 60200, Czech Republic

ARTICLE INFO

Keywords:
Polyhydroxyalkanoates
Bacteria
Archaea
Halophiles
Thermophiles
Stress robustness
Adaptation
Biotechnological production of polyhydroxyalkanoates
Metabolic engineering and synthetic biology

ABSTRACT

Polyhydroxyalkanoates (PHA) are microbial polyesters produced by numerous prokaryotes. These materials are generally considered to be renewable and biodegradable alternatives to petrochemical polymers in numerous applications. PHA are accumulated by microbial cells in form of intracellular granules primarily as storage compounds; nevertheless, numerous recent reports also highlight the importance of PHA for the stress robustness of bacteria. Therefore, in this review, we focus on summarizing current knowledge on PHA accumulation in halophiles and thermophiles – prokaryotic microorganisms adapted to high salinity and high temperature, respectively. Utilization of extremophiles for PHA production brings numerous benefits stemming especially from the enhanced robustness of the process against contamination by common mesophilic microflora as a basement of the Next-Generation Industrial Biotechnology concept. Further, recent advances and future perspectives in metabolic engineering and synthetic biology of halophiles and thermophiles for PHA production improvement are also summarized and suggested. Facts and ideas gathered in this review hold a promise that biotechnological production of PHA by extremophiles can be sustainable and economically feasible enabling PHA to enter the market massively and compete with non-biodegradable petrochemical polymers in suitable applications.

1. Introduction

During the long history of our planet, life has managed to penetrate all, even the least hospitable environmental niches. Nature has developed plenty of diverse mechanisms how to push the boundaries of life, some of which certainly remain hidden from our knowledge still. Prokaryotes from the domains Bacteria and Archaea, witnessing the history of evolution since its very beginning, represent the most suitable study subjects in our quest for understanding of how life at the extremes has evolved. Prospering in the harshest habitats imaginable, from extremely hot hydrothermal vents to deep-cold Antarctic soils, from the Dead sea ultra-salty water to acidic mine drainages, they display impressive examples of how the terrestrial life is constrained, providing crucial implications not only for considering the origin of life on our planet, but also for the search for extraterrestrial life. Furthermore, as we humans have been employing microbes for centuries to improve our lives, research of these extreme-thriving microorganisms also opens up new horizons in various fields of science and technology.

Numerous prokaryotes – including the extremophilic ones – accumulate intracellular granules composed of polyesters of hydroxyalkanoic acids – polyhydroxyalkanoates (PHA). These granules are also termed „carbosonomes"to stress out their complex structure and multiple biological functions. The PHA polymer represents the core of the carbonosomes, while the surface is covered by numerous PHA granules-associated proteins representing an interface between the hydrophobic polymer and hydrophilic cytoplasm. Moreover, these proteins are also

* Corresponding author.
E-mail address: obruca@fch.vut.cz (S. Obruča).

https://doi.org/10.1016/j.biotechadv.2022.107906
Received 14 September 2021; Received in revised form 15 December 2021; Accepted 7 January 2022
Available online 13 January 2022

involved in PHA synthesis, degradation and regulation of these processes (Jendrossek, 2009). PHA may represent a substantial portion of a prokaryotic cell, with PHA contents reaching up to 90% of cell dry mass (CDM); nevertheless, PHA accumulating cells regulate their diameters to control the volumetric fraction of the PHA up to 40 vol% (Mravec et al., 2016). The primary biological function of PHA is the storage of carbon and energy; nevertheless, PHA are also involved in other biological processes such as fuelling sporulation in *Bacilli* and related species (Sadykov et al., 2017), or establishment of symbiosis between prokaryotes and plants (Alves et al., 2019) or between insects and prokaryotes (Kim et al., 2013). Moreover, numerous recent publications emphasize that the accumulation of PHA enhances the robustness of bacteria against various stressors such as osmotic pressure fluctuations (Obruca et al., 2017; Sedlacek et al., 2019), oxidative pressure (Batista et al., 2018; Koskimäki et al., 2016), UV irradiation (Slaninova et al., 2018; Tribelli et al., 2020) or high (Alves et al., 2020; Gonçalves et al., 2019) and low temperature (Nowroth et al., 2016) or even repeated freezing and subsequent thawing cycles (Obruca et al., 2016a). The current knowledge regarding the interconnection of PHA and stress survival and robustness of bacteria was also recently summarized by several reviews (Obruca et al., 2018, 2020, 2021; Müller-Santos et al., 2020). Apart from their biological importance, PHA attract the attention of the scientific community and industry due to their properties.

1.1. Characterization of polyhydroxyalkanoates

Generally, PHA are versatile polymers that are considered to be biodegradable, biocompatible, renewable and sustainable alternatives to petrochemical polymers (Koller et al., 2017). The mechanical and technological properties of PHA are generally dependant upon monomer composition of the polymer. Since prokaryotes are capable of introducing more than 150 various hydroxyalkanoic acids into the polymer chain (Park et al., 2012), PHA materials reveal a great variety of properties and functionalities. Generally, PHA are classified according to the number of carbon atoms per monomer unit: short-chain-length (scl-) PHA contain 3–5 carbon atoms per monomer unit and reveal thermoplastic properties, most of them are crystalline, rigid and fragile polymeric materials with high melting temperature and low glass transition temperature. Scl-PHA are the most abundant type of PHA among prokaryotes. Medium-chain-length (mcl-) PHA consist of monomers with 6–14 carbon atoms per monomer unit; these materials are elastic, possess low crystallinity and tensile strength and substantially lower melting and glass transition temperature as compared to scl-PHA. Predominantly *Pseudomonas* species are capable of mcl-PHA biosynthesis. Some bacteria, e.g., *Aeromonas caviae* or *Aeromonas hydrophila* are reported to synthesize hybride-type copolymers consisting of scl- and mcl-PHA building blocks. Generally, PHA monomers are usually chiral 3-hydroxyalkanoic acids; nevertheless, even chiral and achiral 2-, 4- and 5-hydroxy acids can be incorporated into the polymer chain. Importantly, due to the strict stereospecificity of PHA synthase (PhaC), the key enzyme responsible for the biosynthesis of the PHA polymer chain, all the monomer units are in *R* configuration (Mozejko-Ciesielska et al., 2019).

PHA synthases are, based on their substrate preference and subunit composition, divided into four classes. Class I accommodates enzymes consisting of a single subunit (PhaC) with molecular weight between 60 and 70 kDa and preferring scl-PHA monomers, the PHA synthase of *Cupriavidus necator* (formerly *Alcaligenes eutrophus*, *Wautersia eutropha* and *Ralstonia eutropha*) represents the model enzyme of Class I synthases. Moreover, enzymes belonging to Class II synthases also contain only one subunit (PhaC1 or PhaC2); however, these enzymes catalyze the polymerization of mcl-PHA monomers, therefore, they are very common among *Pseudomonads*. Unlike Class I and Clas II synthases, Class III synthases are heterodimers requiring two subunits – PhaC and and PhaE (Mw of about 40 kDa) for full activity. The PHA synthase from *Allochromatium vinosum* (originally named *Chromatium vinosum*) is a representative of the bacterial Class III PHA synthases, but also numerous *Haloarchaea* possess special types of Class III PHA synthases. Class IV of PHA synthases includes enzymes containing two subunits – PhaC and PhaR (Mw of PhaR is about 20 kDa) and are typical for *Bacilli* and related species; these enzymes are specific for scl-PHA polymerization. Such synthases of Class IV reveal a unique characteristic – in presence of alcohol, the PhaR subunit catalyzes alcohol cleavage of the PHA chain. The major purpose of this activity is probably the regulation of Mw of produced polymer. This phenomenon can be used to modify the carboxy terminus of PHA chain and incorporate active groups such as benzyl-, thiol- or hydroxy- group which might be beneficial for further modification and functionalization of the polymer (Zou et al., 2017). However, this long-established classification of PHA synthases in four Classes is currently in status of getting adapted; new synthases not fitting in one pof the four described Classes were recently isolated, such as a potentially new class of PHA synthase, which was recently identified from Antarctic bacterial isolates (Tan et al., 2020).

The substrates for PHA synthases are generated in three major pathways. The first pathway, which leads to scl-PHA synthesis, is based on the condensation of two acetyl-CoA molecules, which is catalyzed by 3-ketothiolase (PhaA; formerly known as β-ketothiolase) and results in the formation of acetoacetyl-CoA. The second step is a stereospecific reduction of acetoacetyl-CoA by NAD(P)H-dependant acetoacetyl-CoA reductases (PhaB); this "pseudofermentation" regenerates the oxidized form of $NAD(P)H^+$, and generates *R*-3-hydroxybutyryl-CoA, which is finally converted by the action of PHA synthase to the polymer chain of poly(3-hydroxybutyrate) (PHB). When mcl-PHA are synthesized, particular 3-hydroxyacyl-CoA acids can be obtained either from fatty acids *de-novo* synthesis or from β-oxidation of fatty acids. Further, it should be stated that PHA metabolism is of cyclic nature since PHA are intracellularly simultaneously synthesized and also hydrolyzed. Therefore, the PHA metabolism is also termed the "PHA cycle" (Prieto et al., 2016). The main metabolic pathways of PHA synthesis as well as the morphology of PHA granules, their properties and localization in bacterial cells are shown in Fig. 1.

1.2. PHA - polymers for numerous purposes

Apart from their complex biological and evolutionary role, PHA have continuously been researched also as highly promising polymers with a wide range of potential applications. As the (micro)plastic pollution became one of the major environmental concerns, the production of eco-friendly bio-based and biodegradable alternatives to conventional petroleum plastics has emerged among the most intensively addressed technological issues. Current global production of bioplastics (2.11 million tons in 2020) is expected to increase by more than 35% by 2025, still representing less than 1% of the total annual production of plastics (368 million tons in 2020 according to Bioplastics market data, 2020). Among the currently identified bioplastics, PHA still represents a minor contributor to the market (1.7% of the total amount of bioplastics produced in 2020); nevertheless, the market share is expected to increase significantly to 11.5% by 2025. The foreseen growth is promoted mainly by the outstanding versatility of material properties that can be provided by various members of PHA family. Depending on their monomer composition (at least 150 monomers are known currently (Steinbüchel, 1995)) and polymerization degree, PHA can be produced with a large variety of mechanical (from brittle to flexible, highly elastic), thermal (wide range of glass and melting temperatures as summarized in (Muhammadi et al., 2015)), chemical and physical properties. In combination with their high biodegradability, industrial- and home compostability, and non-toxicity, this makes PHA a promising candidate for replacing (or being blended with) the most important conventional petroleum-based plastics, such as polyethylene or polypropylene (Chen, 2010).

Over the past years, the range of end uses that PHA has been proposed or tested for, as well as the list of PHA producing and/or

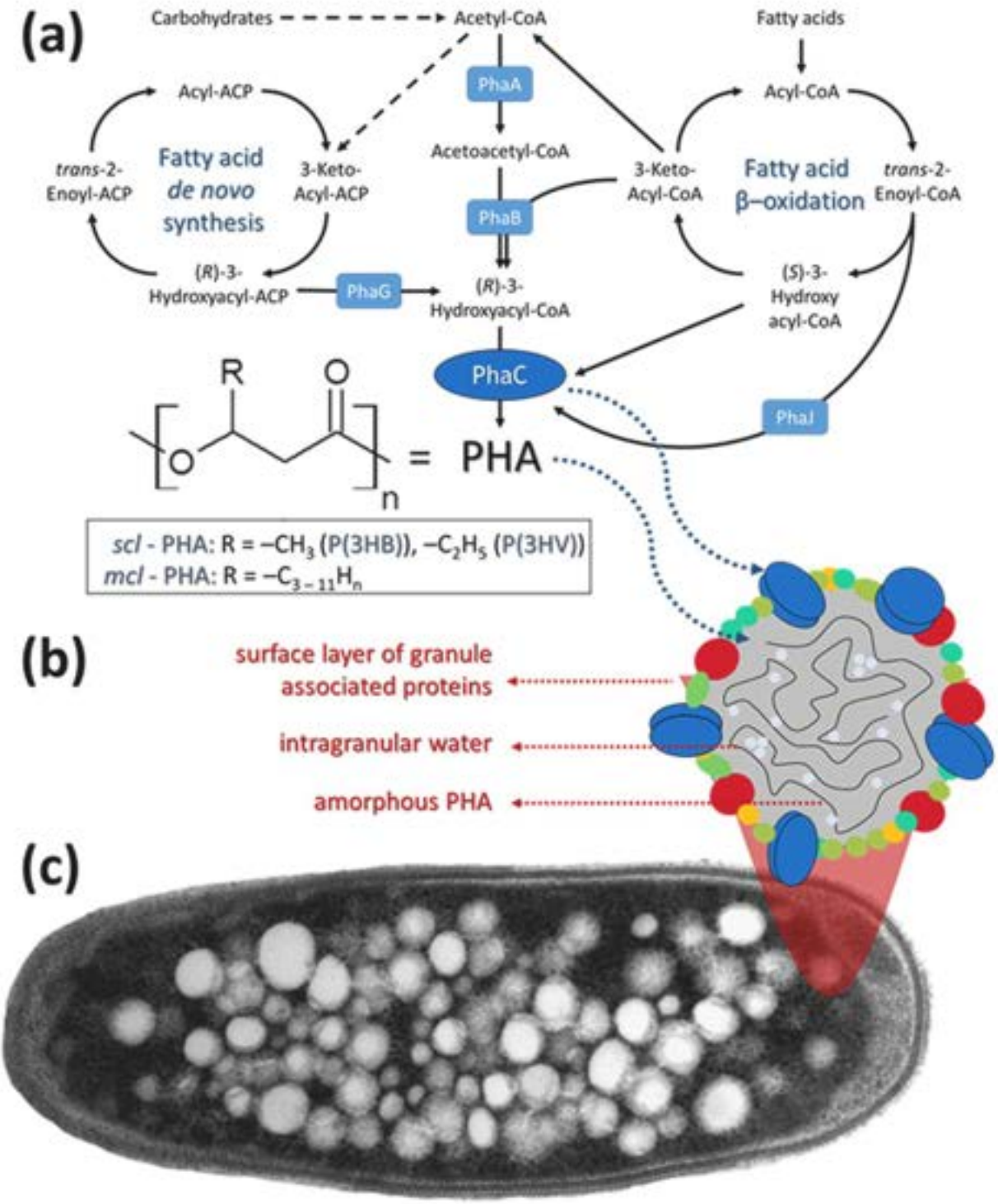


Fig. 1. The main metabolic pathways of PHA synthesis, morphology of PHA granules and their localization in cells of halophilic bacterium *Halomonas hydrothermalis*.

researching companies has steadily grown. A detailed overview of the currently operating PHA manufacturers together with their production scales and application specialization is listed elsewhere (Chen, 2009; Kourmentza et al., 2020). Probably the most often proposed use of PHA-based plastics is in the packaging industry, in particular in the production of packaging materials with a short lifespan, including food utensils and daily consumables, where the biodegradability and good gas barrier properties represent the main benefit of PHA. As the inherent brittleness and rigidity of PHB limits the usability of the PHB films, the efforts in packaging application have been targeted mainly to copolymers (e.g., the copolyester poly(3-hydroxybutyrate-co—3-hydroxyvalerate) (poly (3HB-*co*-3HV)) and mcl-PHAs and PHA-containing polymer blends or resins (Israni and Shivakumar, 2019; Vandewijngaarden et al., 2016; Zhang and Thomas, 2011)). Among the other low-end uses utilizing primarily fast and complete decomposition of PHA, agricultural applications (including production of biodegradable mulch films, growth bags or pesticide controlled-release systems (Grillo et al., 2011)), and production of single-use disposable items (such as bottles, cups, shopping bags, cosmetic containers) should be emphasized.

Nevertheless, specific biological and material properties of PHA

make them beneficial also for various high-end applications. PHA are biocompatible, i.e., non-harmful to human blood and tissues, rendering them attractive for a wide range of biomedical uses, targeted mainly to tissue engineering (production of medical implant materials from PHA) and drug delivery systems development (Ali and Jamil, 2016). Furthermore, PHA have also been concerned as the bulk materials for the synthesis of fine chemicals. Due to the chiral purity (all monomers are in (*R*)-configuration) and to the content of easily modifiable hydroxyl and carboxyl groups, monomers of PHA represent attractive precursors or intermediates for the synthesis of various fine chemicals such as antibiotics, vitamins, drugs, or pheromones (Ruth et al., 2007) and also for the production of novel tailor-made polyesters (Lakshmanan et al., 2019). Last but not least, also the versatile application potential of PHA granule-associated proteins biosynthesized together with PHA *in vivo*, was proposed by several authors (Chen, 2009; Mezzina et al., 2015, 2016).

Nevertheless, mainly in the field of low-cost applications, relatively high production costs compared to other (bio)plastics still represent the main limitation hindering the further expansion of PHA in the market. Therefore, strong efforts are currently targeted on searching for novel trends in the biotechnological production of PHA that would increase the economical feasibility and competitiveness of PHA uses.

1.3. Current trends in PHA production

There are numerous approaches to PHA production. The most traditional scenario includes the employment of an axenic microbial culture cultivated aseptically in submerged culture using defined media with pure carbon substrates, such as glucose or saccharose. However, the most important factor which prevents PHA from entering the market in the low-end application (e.g., packaging, single-use items, etc.) is the high price of PHA as compared to petrochemical polymers. The cost of carbon source in the above-mentioned scenario represents about 40% of the final cost of PHA (Koutmentza et al., 2017), therefore, there are attempts to omit expensive carbon substrates and utilize cheap, abundant, and human-food-chain non-competing resources such as waste streams of agriculture or food industry, which can be used as the carbon source for heterotrophic axenic culture (Koller, 2019a; Medeiros Garcia Alcántara et al., 2020). Apart from organic substrates, it is also possible to use CO_2 as a carbon source when photoautotrophic microbial cultures, for instance, cyanobacteria are used (Panuschka et al., 2019). More recent endeavors to save substrate costs involve the use of syngas, which can be produced as a homogenously composed carbon source from various organic waste streams by microbial specialists like *Rhodospirillum rubrum* (Amstutz and Zinn, 2020). In addition, several methanotrophs are reported to readily use methane, present, e.g., in biogas or natural gas, as C1 feedstock for PHA biosynthesis (Rodríguez et al., 2020).

Further, also the operation of the fermentation is expensive, especially the maintenance of aseptic conditions is energetically demanding. In this perspective, utilization of mixed microbial consortia which are cultivated without any sterility precautions seems to be a very promising approach. The cultivation is operated in a special open cultivation scenario which enables enrichment of mixed microbial culture in PHA producers. The most frequently used strategy - feast/famine cycling - is based on the periodic alternation of intervals of carbon substrate excess (feast phase) and relatively long periods of carbon substrate deprivation (famine phase). The microbes which are not capable of storing PHA are seriously disadvantaged during the famine period and are, therefore, eliminated from the system (Di Caprio, 2021; Oliveira et al., 2017). The bottleneck of PHA production by mixed microbial cultures are low volumetric productivity and inconstancy of properties of produced materials (monomer composition, molecular weight and polydispersity index), which might be limiting for some applications.

An alternative route towards sustainable production of PHA but also other microbial metabolites was recently proposed by Prof. Chen as a concept of "Next-Generation Industrial Biotechnology" (NGIB). This concept is based on the employment of extremophiles as chassis instead of traditionally used mesophiles. Due to the application of extreme conditions which are required by extremophilic culture, the process is protected from contamination by commonly omnipresent mesophilic microflora which enables reduction or even complete omission of sterility precautions, and the process can be operated in highly productive continuous or semi-continuous scenarios in cheaper ceramic, cement or plastic bioreactors. NGIB should enable sustainable and economically feasible production of PHA competitive with cost-effective chemical production of petrochemical polymers (Chen and Jiang, 2018; Tan et al., 2021; Yu et al., 2019).

PHA production has already been described for numerous prokaryotic microbes including various extremophiles (Koller, 2017; Kumar et al., 2020; Obulisamy and Mehariya, 2021), nevertheless, due to high PHA yields, range of utilizable substrates and other advantages (discussed further in the text) the most promising extremophiles concerning PHA production seem to be found among halophiles and thermophiles. Therefore, this review focuses on the description and discussion of the current status of PHA synthesis by these microorganisms and provides also potential outcomes and perspectives.

2. Salt above gold – PHA production by halophiles

'Salt above gold' is not only the title of a famous Czech folkstory, but also a leitmotif surprisingly widespread in nature. Halophilicity, hence, the adaptability of life forms to high salinity, is described for all domains of life, such as for eukaryotes (e.g., halophytes or marine microalgae) or for halophilic prokaryotes (bacteria and archaea) (Oren, 2002). Hence, the group of halophilic species is phylogenetically highly diversified, and have one feature in common: they prefer or even require hypersaline environments for optimal growth, where NaCl usually acts as the main salt constituent. In this context, a steadily growing number of halophilic archaea and bacteria, identified in diverse saline habitats, are currently investigated as potential PHA-producing cell factories (reviewed by Koller et al., 2020; Stanley et al., 2021).

2.1. Adaptation of microbes to high salinity environments

To cope with extremely saline (hypertonic) conditions indeed is a challenging task for organisms, and requires specific adaptation strategies. Therefore, nature developed several fundamentally different approaches to support halophiles to manage high salt loads. The existence of these manifold approaches in turn substantiates the broad occurrence of halophilicity in nature (reviewed by Fathima et al., 2017).

The presumably most frequently occurring strategy involves the intracellular accumulation of soluble compatible osmotic compounds, such as the organic kosmotropics ectoine (first described as an osmoprotectant in the purple sulfur bacterium *Ectothiorhodospira halochloris* (Lippert et al., 1993), nowadays known to occur in both Gram-positive and Gram-negative bacteria), the glycine-derivative betaine (prevents loss of intracellular water) (Deole and Hoff, 2020), the osmoprotectant trehalose (described to be pivotal for yeast fermentations (Majara et al., 2018) and adaptation of bacteria (Dinnbier et al., 1988), or diverse amino acids.

Generally, the accumulation of these compatible organic solutes serves to stabilize the intracellular osmotic pressure; importantly, this strategy does not necessarily imply that involved organisms also need to adapt their proteome (enzymatic machinery) to the high salt loads. The intracellular accumulation of these compounds is based either on *de novo* osmolyte biosynthesis by the accumulating cells, or cells are supplied by them from the environment and transport them into the cell's interior. After the decline of the environmental salinity, these compatible solutes are excreted by the cells in order to balance osmotic pressure inside and outside the cells, which in turn allows for convenient isolation and purification of these often technologically relevant products (Wood

et al., 2001): because of their wide-ranging protective function and chaperons-like role, they are used in biotechnology, cosmetics (products for skin care), food industry, or the medicinal and pharmaceutical field (Sauer and Galinski, 1998). Regarding PHA-accumulating microbes, this formation of osmolytes is a typical feature of adaption among halotolerant and halophilic eubacteria, such as *Halomonas* spp. (Cummings and Gilmour, 1995; Shivanand and Mugeraya, 2011).

Such accumulation of compatible solutes considerably differs from the alternative "salt-in" strategy, which is based on intracellular salt accumulation, often potassium chloride, as a tool to balance the inner and outer osmotic pressure of cells. This "salt-in" strategy is the preferred approach followed by extremely halophilic haloarchaea, such as well-known PHA producers like *Hfx. mediterranei*. This strategy was first described in 1970 by Ginzburg *et al.* for a "*Halobacterium* sp.", a Dead Sea isolate; in this organism, changes of the intracellular ions composition during different growth conditions were observed (Ginzburg et al., 1970). As reported later, this strain, at the same time a halophile and thermophile, belongs to the genus *Haloarcula*, and was finally classified as strain *Haloarcula marismortui* Volcani (Oren et al., 1990). In contrast to the above-described accumulation of organic solutes, the "salt-in" strategy indispensably goes in parallel with an adaptation of enzymes and other universally conserved macromolecules (nucleic acids) by harnessing them with salt-resistance features; generated enzyme variants are termed "extremozymes" or "halozymes" (Danson and Hough, 1998). Mechanistically, this enzyme adaptation is mainly characterized by the formation of a surplus of acidic amino acids on the enzyme's surface, which allows the enzyme to work even under conditions near saturating salinity (Kennedy et al., 2001). For instance, *Hfx. mediterranei* glucose dehydrogenase (EC 1.1.1.47), an enzyme catalyzing the reaction of glucose to β-D-glucono-1,5-lactone, is reported to possess no flexible side chains on its surface, thus, it occurs as a highly ordered, multilayered solvation shell well adapted to avoid attachment of ions and thus to prevent damage (formation of inactive enzyme inclusion bodies) by salt (Bonete et al., 1996; Britton et al., 2006).

Besides the proteome, also the genome of extremely halophiles differs significantly from that of osmo-mesophiles and often requires specific editing. This is manifested, for example, by the high number of salt-resistant genes found in such halophiles (Das et al., 2015). As an example, Bolhuis *et al.* described that the genome of the square-shaped PHA-producing haloarchaeon *Haloquadratum walsbyi*, which is found in NaCl-saturated and $MgCl_2$-rich aquatic environments, features several exceptional adaptive particularities, enabling this strain to thrive well in highly saline environments. The probably most remarkable of these genomic particularities is the presence of a gene encoding a high molecular mass halomucin, similar to halomucins protecting animal tissue from desiccation and harsh chemical conditions. These halomucins with a negative overall charge reach from the inside to the outside of cells, thus forming a "water shield" sheltering the cell (Bolhuis et al., 2006). The genome of *Halobacterium* sp. NRC-1 also reveals several remarkable physical adaptations to high salinity, such as the encoding of predominately acidic enzymes, which is recognized as a crucial strategy to prevent salting-out of proteins in the hypersaline cytoplasm (Kennedy et al., 2001; Ng et al., 2000); such pronounced negative surface charge of folded proteins to overcome salting-out of hypersaline cytoplasm was also described for the haloarchaeon *Haloarcula marismortui* (Baliga et al., 2004).

In addition, the change of pigment patterns also constitutes a tool for organisms to adapt to changing salinity. As amply described in the literature, many eukaryotic microalgae adapt their pigment pattern under conditions of fluctuating salinity. This was reported by Masojídek *et al.* for the green microalga (chlorophyta) *Chlorococcum* spp., which drastically increased the intracellular carotenoids-to-chlorophylls ratio during exposure of the culture to nitrogen deficiency, high salinity and illumination stress (Masojídek et al., 2000). In addition to eukaryotes, this pigment pattern change was also reported for extremely halophilic prokaryotes like the haloarchaeal PHA producer *Hfx. mediterranei*. The hyperproduction of the C50-carotenoid bacterioruberin, which is typically produced as a reaction to extreme UV-irradiation and excessive oxidative stress, for instance, provoked by H_2O_2 addition (Giani and Martínez-Espinosa, 2020), is also produced at suboptimal salt concentrations (D'Souza et al., 1997; Fang et al., 2010; Montero-Lobato et al., 2018). In this case, bacterioruberin is suspected to adapt the flexibility of the cell membrane and thus to maintain cells' viability (D'Souza et al., 1997). Indeed, it was demonstrated by Chen *et al.* that for *Hfx. mediterranei*, when cultivated on extruded rice bran and starch under optimal salinity of brined medium, there exists an indirect relationship between the level of salinity and pigment formation, and a direct correlation of salinity and PHA biosynthesis; these authors demonstrated that the pigment bacterioruberin, a compound of potential technological application, for instance, as colorant on cosmetic products, can be produced by *Hfx. mediterranei* at high throughput (more than 0.5 g/L pigment) by adapting the salt concentration of the cultivation medium. Importantly, mentioned inexpensive carbon sources were successfully used as feedstocks for product formation, both for PHA and pigments (Chen et al., 2015).

2.2. Role of PHA in the adaptation of halophiles to hypertonic conditions

In the context of halophilic PHA production strains, it is of interest to reveal how biosynthesis of PHA and above-described compatible solutes is interlinked, and how these anabolic processes relate to the organisms' strategy to cope with high external salinity. In 2006, Quillaguamán and colleagues noticed the sudden change of the morphology of PHA-accumulating cultures of the halophilic eubacterium *Halomonas boliviensis* LC1, a salt lake isolate, at suddenly increasing salinity; the authors proposed that the switch towards the accumulation of organic solutes serves to withstand the osmotic pressure before the onset of increased PHA biosynthesis (Quillaguamán et al., 2005). Mothes *et al.* reported the parallel production of PHA and ectoines by the halophilic bacterium *Halomonas elongata*. At a salinity of 10 wt% NaCl, the strain accumulated 0.5 g PHA per g biomass plus up to 14 wt% of ectoines, which substantiates the theory of the parallel formation of PHA and compatible organic solutes (Mothes et al., 2008). Guzmán *et al.* studied the parallel PHA and ectoines biosynthesis by the halophilic bacterium *Halomonas boliviensis* in two fed-batch cultivation experiments. While the first cultivation was performed at 45 g/L of NaCl without nitrogen or phosphate limitation to obtain a high concentration of active biomass, the salinity in the second cultivation setup was increased to 75 g/L NaCl in order to increase ectoine accumulation; moreover, in this second cultivation, nitrogen- and phosphate sources were supplied exclusively during the initial growth phase and were later depleted to boost PHB formation. Intracellular PHB fraction and volumetric PHB productivity amounted to about 0.96 g/g and 1 g/(L·h), respectively, while the ectoine concentration and content were reported with 4.3 g/L and 0.072 g/g, respectively (Guzmán et al., 2009).

Importantly, it was shown that the presence of PHA granules in cells protects them against damage caused by hyperosmotic shock (a sudden increase of salinity). This was for the first time demonstrated by Soto *et al.*, who exposed the halotolerant Gram-negative PHA production strain *Pseudomonas* sp. CT13 and its PHA-negative mutant to hyperosmotic up-shock. It turned out that the presence of PHA granules boosts the intracellular level of 3-hydroxybutyrate (3HB), the monomer of the homopolyester PHB. 3HB, in turn, acts as a kind of "chaperone" by inhibiting protein agglomeration, which is a typical lethal consequence for PHA-negative cells when exposed to high salt concentration and/or elevated temperature. As demonstrated by increased PHA productivity at increased salinity, PHA and 3HB constitute compatible solutes, supporting the bacteria to counterattack hyperosmotic stress (Soto et al., 2012). More recently, Obruca *et al.* substantiated these outcomes by illuminating the indeed expedient chaperoning efficiency of 3HB. These authors showed that 3HB's protective effect was competitive with the effect described for well-known compatible solutes such as ectoines or

trehalose. From the mechanistic point of view, this effect can be understood by the fact that the presence of PHA granules, which are water-insoluble and of a high degree of polymerization, does not increase the intracellular osmotic pressure; however, they act as a storage of water-soluble compatible osmolytes (3HB), which are mobilized by the cells under conditions of sudden hyperosmotic pressure and other environmental stress factors (Obruca et al., 2016b). Later, it was detected by the same research team that the presence of PHA granules in bacterial cells prevents massive damage of the cytoplasmic membrane ("plasmolysis"), thereby supporting cells to maintain integrity under hyperosmotic conditions (Obruca et al., 2017). These studies evidence that significant amounts of PHA are commonly found among halophilic species due to their proficiency to overcome the negative effects of osmotic up-shock.

This adaptive role of PHA in halophilic microbes is fundamentally different from the effect of salinity on PHA biosynthesis for microbes industrially used for PHA production, such as the best described PHA producer *Cupriavidus necator* (used for PHA production at, e.g., PHB Industrial S.A. Brazil, Biomer, Germany, or Bio-On, Italy), turned out to be highly sensitive against increased salinity; as reported by Mozumder *et al.*, cell propagation and PHA biosynthesis by *C. necator* were completely stopped at sodium concentrations in the nutrient broth of 8.9 g/L and 10.5 g/L, respectively. In fed-batch cultivation setups carried out on a bioreactor scale, these studies showed that longer biomass growth phases under nutritionally balanced conditions, characterized by the permanent supply of NaOH for maintaining a neutral pH-value, causes sodium accumulation in cells, which results in decreased PHA production rates in the subsequent product formation phase (Mozumder et al., 2015).

2.3. Pros and cons of PHA production by halophiles

Occurring both among eubacteria and haloarchaea, such halophilic organisms are expected to render PHA production more cost-efficient in a not too distant future. This expectation is mainly based on the fact that halophilic PHA production strains can be cultivated at restricted sterility precautions, or even without any sterilization of the bioreactor equipment or the nutrient broth, which in turn saves energy and time during the upstream processing.

Also PHA recovery, hence, the downstream processing step, can be facilitated when using extreme halophilic PHA production strains. Due to their high intracellular osmotic pressure, cells of extremely halophilic species can conveniently be disintegrated by exposing them to hypotonic media (deionized water), which offers a convenient approach to recover intact PHA granules of remarkable purity by centrifugation. For mesophilic PHA production strains, recovery of PHA as an intracellular product requires cumbersome cell disintegration by chemical (use of strong oxidants), enzymatic (hydrolases) or mechanical methods, or solvent-intensive extraction of the product from biomass (reviewed by Koller, 2020).

A particular aspect of many PHA-producing extreme halophiles is the presence of certain metabolic pathways to produce PHA-building blocks different from 3HB, predominantly 3-hydroxyvalerate (3HV). Starting from different simple raw materials like carbohydrates, the intracellular carbon flux in such metabolic specialists gets shifted towards production of PHA copolyesters of improved processability (decreased crystallinity, broader temperature window between melting and decomposition temperature) when compared to PHB homopolyesters; for technically used, mesophilic PHA production strains like *C. necator*, PHA copolyester biosynthesis requires feeding the cells with precursor compounds chemically related to 3HV, such as propionic, levulinic or valeric acid. These precursor co-substrates are typically expensive and/or toxic, and are not needed in the case of copolyester biosynthesis by extreme halophiles like the haloarchaea *Hfx. mediterranei*, *Haloarcula hispanica*, *Halobacterium noricense*, *Halogeometricum borinquense* strain E3, *Halogranum amylolyticum*, or *Natrinema ajinwuensis* (reviewed by Koller, 2019b), or, as shown only recently, also for some moderately halophilic eubacteria like *Halomonas pacifica* or *Halomonas salifodiane* (El-malek et al., 2020). For *Hfx. mediterranei*, this 3HV formation from unrelated substrates was well elucidated by Han *et al.*, who discovered multiple active pathways in this strain, which supply propionyl-CoA, which in turn acts as 3HV-precursor; propionyl-CoA couples with acetyl-CoA, forming of 3HV, which gets incorporated into growing poly(3HB-*co*-3HV) copolyester chains (Han et al., 2013).

Moreover, a range of technologically relevant by-products are produced in parallel to PHA biosynthesis by various both moderate or extreme halophiles; important examples are, besides above discussed organic osmolytes (Mothes et al., 2008; rational flux tuning for fine-tuned co-production of PHA and ectoines was only recently reported by Ma *et al.* for *Halomonas bluephagenensis* (Ma et al., 2020)), pigments like bacterioruberin (Giani et al., 2019; Giani and Martínez-Espinosa, 2020), antibacterial halocins (Kaur and Tiwari, 2021), or special exopolysaccharides (EPS) with xanthan-like properties, which might be of potential medical and food-industrial use (Cui et al., 2017a; Pacholak et al., 2021).

Hence, application of such halophiles for biotechnological product formation offers several beneficial options: firstly, the high salinity of the culture media minimizes the risk of microbial contamination ("infection") by foreign organisms, which makes the cultivation batches stable and less energy-demanding (Yin et al., 2015). Secondly, the cultivation of halophiles in saline media leads to the intracellular accumulation of salt, hence, the excessive salt load gets partially removed from the medium (Rodríguez-Contreras et al., 2016). This is advantageous in the case of salt-rich waste streams to be used as a culture medium (e.g., acidically hydrolyzed lignocellulose materials, whey, etc.); halophilic production strains used in such production processes need the salt generated by the acidic hydrolysis and the subsequent neutralization as an indispensable media component, and are therefore the ideal candidates for such bioprocesses. After the process, a considerable fraction of salts, which was highly diluted in the cultivation medium, is now stored in a limited volume of generated microbial biomass, which can easily be handled; hence, less salt remains in the spent fermentation broth, which is an environmental advantage regarding its disposal (Obruca et al., 2015; Obruca et al., 2014).

Despite all these positive aspects of PHA production by halophiles, we need to consider that this still is kind of a "double-edged sword"; as the downside of the medal, highly saline cultivation media are aggressive towards bioreactor equipment and electrodes used for bioprocess control. Therefore efforts are described in the literature to generate robust bioreactor facilities to farm such strains, such as reported by Hezayen *et al.*, who constructed a corrosion-resistant 8 L composite-type bioreactor made of poly(ether ether ketone) (PEEK), silicium nitrite ceramics and tech glass, stirred by corrosion-resistant magnetic coupling, to cultivate the PHA producing haloarchaeon *Halopiger aswanensis* at a salinity of 25 wt% NaCl (Hezayen et al., 2000). Alternatively, high-quality nickel-molybdenum-based Hastelloy alloy can be used to produce salt-corrosion resistant bioreactors to farm extremely halophilic strains, as shown by Mahler *et al.*, who developed a corrosion-resistant bubble column made of this material for this purpose (Mahler et al., 2018). Of course, such high-quality materials make the bioreactor equipment more expensive; alternatives, namely open, non-sterile cultivation process, as described in the previous paragraphs, might economically outperform processes requiring such expensive equipment, especially on a large scale. Moreover, in most cases, volumetric productivities for PHA and specific growth rates are still lower for haloarchael processes when compared to established processes using *C. necator*; promising exceptions are provided by genetically engineered *Halomonas bluephagenensis* strains, which are described in the subsequent section and summarized in Table 1.

2.4. Overview of PHA producing halophiles

A comprehensive literature review on diverse PHA homo- and

Table 1
Selected PHA production processes by halophiles.

Strain	Type of organisms; Classification according to halophilicity	NaCl concentration [g/L]; Cultivation scale	Carbon source	Type of PHA accumulated	Intracellular PHA fraction [g/g]; Volumetric productivity [g/(L·h)]	Strain origin (location of isolation)	Reference
Bacillus megaterium uyuni S29	Gram-positive eubacterium Halotolerant	45; Shaking flasks	Glucose	PHB	0.41; 0.1	Uyuni salt lake (Bolivia)	Rodriguez-Contreras et al., 2016
		10; Shaking flasks	Denatured sugar beet molasses	PHB	0.6; 0.42		Schmid et al., 2019
		5; Pilot scale cultivation on 500 L (working volume) bioreactor scale	Denatured sugar beet Molasses	PHB	0.65; 0.2		Schmid et al., 2021b
Salinivibrio sp. M318	Gram-negative eubacterium; Moderate halophile	30; Fed-batch bioreactor cultivation	Mixture of waste fish oil and glycerol (Nitrogen source: waste fish sauce)	PHB; PHBHV and PHB4HB after precursor addition in batch cultivations	0.52; 0.46	Isolated in Vietnam from fermented shrimp paste	Van Thuoc et al., 2019
Salinivibrio sp. TGB10	Gram-negative eubacterium; Moderate halophile	27.5; Fed-batch bioreactor cultivation		PHB; PHBHV with precursor addition	0.82; 0.25	Salt field in Binhai District, Tianjin (PR China)	Tao et al., 2021

Strain	Type of organisms; Classification according to halophilicity	NaCl concentration [g/L]; Cultivation scale	Carbon source	Type of PHA accumulated	Intracellular PHA fraction [g/g]; Volumetric productivity [g/(L·h)]	Strain origin (location of isolation)	Reference
Yangia sp. ND199	Gram-negative eubacterium; Moderate halophile	45; Fed-batch cultivations on shaking flask scale	(crude) glycerol (other inexpensive substrates, e.g., fructose corn syrup, converted to PHBHV)	PHBHV (PHB on glutamate)	0.53 (glycerol); 0.41 (crude glycerol); 0.56 (crude glycerol + fructose corn syrup); 0.44 (glycerol); 0.25 (crude glycerol); 1.10 (crude glycerol + fructose corn syrup)	Mangrove forest soil	Van-Thuoc et al., 2015
Halomonas venusta KT832796	Gram-negative eubacterium; Moderate halophile	15; 2 L bioreactor fed-batch cultivation	Glucose	PHB	0.88; 0.25	Indian coast samples	Stanley et al., 2018
Halomonas profundus	Gram-negative eubacterium; Moderate halophile	27; 2-stage batch 5 L bioreactor setup	Glucose	PHB (PHBHV after precursor feeding)	0.8–0.9; 0.004	Isolated from a shrimp	Simon-Colin et al., 2008
Halomonas pacifica ASL10	Gram-negative eubacterium; Moderate halophile, extremely halotolerant	tolerates up to 170 g/L NaCl; Shaking flasks	Sucrose	PHBHV	0.84; 0.04	Mariout salt lake (Egypt)	El-malek et al., 2020
Halomonas salifodiane ASL11	Gram-negative eubacterium; Moderate halophile, extremely halotolerant	tolerates up to 170 g/L NaCl; Shaking flasks	Sucrose	PHBHV	0.82; 0.04	Mariout salt lake (Egypt)	El-malek et al., 2020
Halomonas sp. YLGW01	Gram-negative eubacterium Moderate halophile; extremely halotolerant	20; high PHA productiovity even at 100 g/L NaCl Shaking flasks	Fructose syrup	PHB	0.95; 0.11	Gwangalli beach in Busan, South Korea.	Park et al., 2020

Strain	Type of organisms; Classification according to halophilicity	NaCl concentration [g/L]; Cultivation scale	Carbon source	Type of PHA accumulated	Intracellular PHA fraction [g/g]; Volumetric productivity [g/(L·h)]	Strain origin (location of isolation)	Reference
Halomonas halophila	Gram-negative eubacterium; Moderate halophile	66; Shaking flasks	Glucose (other saccharides from inexpensive sources converted)	PHB	0.72; 0.05	Salt pond located at the Costa Blanca near Alicante (Spain)	Kucera et al., 2018
Halomonas hydrothermalis	Gram-negative eubacterium; Moderate halophile	40; Shaking flasks	Waste frying oil	PHB (PHBHV after precursor addition)	0.62; 0.03	Deep-sea hydrothermal-vent environments; sea at West Coast of India	Pernicova et al., 2019

(continued on next page)

Table 1 (continued)

Strain	Type of organisms; Classification according to halophilicity	NaCl concentration [g/L]; Cultivation scale	Carbon source	Type of PHA accumulated	Intracellular PHA fraction [g/g]; Volumetric productivity [g/(L·h)]	Strain origin (location of isolation)	Reference
Halomonas neptunia	Gram-negative eubacterium; Moderate halophile	60; Shaking flasks	Waste frying oil	PHB (PHBHV after precursor addition)	0.56; 0.02	Deep-sea hydrothermal-vent environments	Pernicova et al., 2019
Halomonas organivorans	Gram-negative eubacterium; Moderate halophile	81; Shaking flasks	Waste frying oil	PHB	0.12; 0.02		Pernicova et al., 2019
Halomonas elongata 2FF	Gram-negative eubacterium; Moderate halophile, extremely halotolerant	100; Shaking flasks	Glucose	PHB	0.4; 0.02	Hypersaline meromictic Fără fund lake (Romania)	Cristea et al., 2017

Strain	Type of organisms; Classification according to halophilicity	NaCl concentration [g/L]; Cultivation scale	Carbon source	Type of PHA accumulated	Intracellular PHA fraction [g/g]; Volumetric productivity [g/(L·h)]	Strain origin (location of isolation)	Reference
Halomonas bluephagenesis	Gram-negative eubacterium (genetically engineered; wild type: *H. bluephagenesis* TD01); Moderate halophile	60; 500 L pilot scale open continous cultivation for 56 h	Glucose	PHB (PHBHV after precursor addition)	0.7 (0.92 in shaking flask setups); 1.12	Aydingkol salt lake (PR China)	Fu et al., 2014
		60; 5 m^3 pilot scale cultivation for 36 h	Glucose, γ-butyrolactone, and waste corn steep liquor; waste gluconate	PHB4HB	0.604 (max.: 0.74); 1.67		Ye et al., 2018
Haloferax mediterranei	Haloarchaeon (family: Haloferacaceae); Extreme halophile	150; 10 L bioreactor fed-batch cultivation	Glucose (plus yeast extract)	PHBHV (no precursor addition!)	0.7; 0.21	Salt pond located at the Costa Blanca near Alicante Spain)	Koller et al., 2015
		20; 42 L bioreactor fed-batch cultivation	Hydrolyzed whey permeate	PHBHV (no precursor addition!) PHBHV4HB (plus precursor)	0.73; 0.10		Koller et al., 2007a

Strain	Type of organisms; Classification according to halophilicity	NaCl concentration [g/L]; Cultivation scale	Carbon source	Type of PHA accumulated	Intracellular PHA fraction [g/g]; Volumetric productivity [g/(L·h)]	Strain origin (location of isolation)	Reference
Haloferax mediterranei (cont.)	Haloarchaeon (family: Haloferacaceae); Extreme halophile	200; 10 L bioreactor fed-batch cultivation	Hydrolyzed whey permeate, spent fermentation broth of previous whey-based processes	PHBHV	70; 0.04	Salt pond located at the Costa Blanca near Alicante Spain)	Koller, 2015
		200; Shaking flasks	Waste stillage from rice-based ethanol production; recovery and re-use of medium salts	PHBHV	0.7 0.17		Bhattacharyya et al., 2014
		234; 6 L bioreactor pH-stat fed-batch cultivation	Native cornstarch treated via enzymatic reactive extrusion plus yeast extract	PHBHV	0.5 0.28		Chen et al., 2006b

Strain	Type of organisms; Classification according to halophilicity	NaCl concentration [g/L]; Cultivation scale	Carbon source	Type of PHA accumulated	Intracellular PHA fraction [g/g]; Volumetric productivity [g/(L·h)]	Strain origin (location of isolation)	Reference
Haloferax mediterranei (cont.)	Haloarchaeon (family: Haloferacaceae); Extreme halophile	220; Shaking flask	Native and dephenolized olive oil waste water	PHBHV	0.43; n.r.	Salt pond located at the Costa Blanca near Alicante Spain)	Alsafadi and Al-Mashaqbeh, 2017
		150; 10 L bioreactor fed-batch cultivation	Crude glycerol phase	PHBHV (no precursor addition!) PHBHV4HB (plus precursor)	0.75; 0.12		Hermann-Krauss et al., 2013

(continued on next page)

Table 1 (*continued*)

Strain	Type of organisms; Classification according to halophilicity	NaCl concentration [g/L]; Cultivation scale	Carbon source	Type of PHA accumulated	Intracellular PHA fraction [g/g]; Volumetric productivity [g/(L•h)]	Strain origin (location of isolation)	Reference
Haloterrigena hispanica	Haloarchaeon (family: Natrialbaceae); Extreme halophile	200; 1 L Bioreactor batch cultivation	Complex carrot-waste medium	PHB	0.0013; n.r.	Saltern crystallizer pond at Fuente de Piedra saline lake, province Malaga (Spain)	Di Donato et al., 2011
Halopiger aswanensis	Haloarchaeon (family: Natrialbaceae); Extreme halophile	250; Batch cultivation in corrosion-resistant 8 L bioreactor	Sodium acetate and butyric acid	PHB	0.53; 0.0045	Samples collected from surface of hypersaline soil collected in Aswan (Egypt)	Hezayen et al., 2000
Halogranum amylolyticum	Haloarchaeon (family: Haloferacaceae); Extreme halophile	200; 7.5 L bioreactor fed-batch cultivation	Glucose	PHBHV	0.266; 0.013	Tainan marine solar saltern (PR China)	Zhao et al., 2015
Halogeometricum borinquense strain E3	Haloarchaeon; (family: Haloferacaceae); Extreme halophile	200; Shaking flask	Glucose	PHBHV	0.74; 0.21	Marakkanam solar saltern (India)	Salgaonkar and Bragança, 2015
		20; Shaking flask	Sugarcane bagasse hydrolysates	PHBHV	0.4–0.5; 0.0113		Salgaonkar and Bragança, 2017
		200; Shaking flask	Starch and carbon-rich fibrous cassava bagasse waste	PHBHV	0.74 (starch); 0.45 (cassava waste) 0.02 (starch); 0.006 (cassava waste)		Salgaonkar et al., 2019
Natrinema ajinwuensis (=altunense)	Haloarchaeon (family: Natrialbaceae); Extreme halophile	200; Repeated batch cultivations in shaking flaks	Glucose	PHBHV	0.61; 0.21	Indian salt production pans	Mahansaria et al., 2018

copolyester production processes by to date described eubacterial and haloarchaeal biopolyester accumulating organisms of different halophilicity, using different carbonaceous feedstocks, was only recently provided (Koller, 2019b). For classification of halophiles based on their salt tolerance/requirement, literature distinguishes at least three different clusters of halophilic microorganisms; this classification is based on the minimum salt required for growth and the maximum salt level these organisms are able to tolerate (reviewed by (Rodríguez-Contreras et al., 2016)).

The first cluster of "halotolerant" microbes are insensitive towards up to 2.5 M NaCl (upper salinity limit strongly fluctuating between different species), but are not dependent on such excessive salt concentration salinity. Halotolerant *Bacilli*, such as the expedient molasses converter *Bacillus megaterium* uyuni S29 isolated from a Bolivian salt lake (Rodríguez-Contreras et al., 2016; Schmid et al., 2021a, 2021b), are the prototype PHA-producing strains among this cluster, in addition to species like *Vibrio proteolyticus* (Hong et al., 2019).

Members of the cluster of "moderately halophiles" not only tolerate, but indispensably require 1–15 wt% (0.15–2.5 M) NaCl to thrive well; regarding PHA producers among them, *Halomonas* spp., such as *H. boliviensis*, *H. campaniensis*, *H. elongata*, *H. salina*, *H. profundus*, *H. venusta*, *H. neptunia*, *H. hydrothermalis*, *H. marina*, *H. smyrnensis*, *H. pacifica*, *H. salifodiane*, or *H. bluephagenensis* (genetically engineered organism originating from the wild type ancestor strain *Halomonas* sp. TD01) are prototype organisms of this cluster, besides other representatives like *Spirulina* sp. (Jau et al., 2005), *Salinivibrio* sp. (Tao et al., 2021; Van Thuoc et al., 2019) or *Yangia* sp., an expedient converter of crude glycerol from biodiesel production (Van-Thuoc et al., 2015) (reviewed by Koller, 2019b). Technologically, genetically engineered *H. bluephagenensis* currently attracts attention as the only commercially implemented halophilic PHA production strain in several new biotech companies such as PhaBuilder© or MedPha©; these recombinants are constructed to accumulate PHA copolyesters poly(3HB-*co*-3HV); PHBHV) and poly(3-hydroxybutyrate-*co*-4-hydroxybutyrate) (poly (3HB-*co*-4HB); PHB4HB)) at high productivity at minor sterility conditions, and can be cultivated on inexpensive raw materials like corn steep liquor (Tan et al., 2021). It was shown before in pilot-scale experiments that genetically modified *H. bluephagenesis* can be readily cultivated in pilot-scale (500 L) open fermentation setups to PHA concentrations of more than 70 g/L, corresponding to volumetric productivities exceeding 1 g/(L•h), which is already approaching reported data for pilot-scale PHA production by mesophilic strains (Nonato et al., 2001). Later, the same group of authors reported the production of more than 60 g/L poly (3HB-*co*-15%-4HB) by *H. bluephagenensis* on a 5 m^3 scale using glucose and corn step liquor as substrates plus γ-butyrolactone as 4HB-related precursor substrate. The productivity of this process amounted to an impressive value of 1.67 g/(L•h). Authors calculated the production costs of poly(3HB-*co*-15%-4HB) by this process with 2.57 € per kg; according to calculations performed based on data obtained by 7.5 L scale optimization experiments, this cost could further be reduced by 0.64 €/kg (Ye et al., 2018). These setups are described by the authors to be especially energy- and freshwater saving, which opened the door for mentioned industrial endeavors for PHA production by differently modified *H. bluephagenensis* strains (Fu et al., 2014).

Importantly, also for PHA production by *Halomonas* spp., there is a strong trend towards utilization of inexpensive raw materials, as recently shown by Liu *et al.*, who cultivated both a wild type *H. elongata* A1, isolated from a Chinese salt lake and its recombinants on glucose,

reaching intracellular PHB fractions of 22.81 wt% and 90.76 wt%, respectively. Recombinant strain *H. elongata* P2, subjected to further genetic modification, was able to readily accumulate PHB even from waste straw hydrolysate at a salinity of 50 g/L (Liu et al., 2021). In this context, *Salinivibrio* sp. M318, isolated from fermented shrimp paste in Vietnam was recently cultivated by Van Thuoc *et al.* on a mixture of fish oil and glycerol as a carbon source, while the fish waste sauce was used as a nitrogen source. In this process, a volumetric PHA productivity of 0.46 g/(L•h) was obtained (Van Thuoc et al., 2019). Other inexpensive raw materials for PHA production by moderately halophiles were used, for example by Pernicova *et al.*, who described the successful conversion of waste frying oil by *H. hydrothermalis* and *H. neptunia*, isolates from Deep-sea hydrothermal-vent environments, to PHB homopolyester (Pernicova et al., 2019). Moreover, Kucera *et al.* described the conversion of a variety of inexpensive hydrolysates, such as from cheese whey, spent coffee grounds, sawdust and corn stover, as well as sugar beet molasses, to PHB by *H. halophila*, a strain featuring a salinity optimum of 60 g/L. The salinity of cultivation media was conveniently adapted by neutralizing the acidic hydrolysis cocktails of applied inexpensive raw materials (Kucera et al., 2018).

The third and phylogenetically most ancient cluster encompasses the so-called "extreme halophilic" microbes; these microbial specialists require excessive salinity of 15–30 wt% (2.5–5 M) and even more for optimum growth. Within this cluster, some literature also classifies the so-called "borderline extreme halophiles", which thrive best in media containing 1.5–4.0 M NaCl (Park et al., 2006). Prototype organisms of extreme halophiles producing PHA are haloarchaea; exclusively scl-PHA production is reported for haloarchaea, while for eubacteria, both scl- and mcl-PHA production is reported (reviewed by Koller et al., 2020). Extremely challenging habitats are those environments, where such highly adaptive survivalists are typically isolated are, e.g., the Great Salt Lake, the Dead Sea, hypersaline anoxic deep-sea basins, solar saltern crystallizers, hypersaline soil samples, salt mine boreholes, salt production pans, or even alpine dry salt rocks (reviewed by Koller, 2019b).

Extremely halophilic haloarchaeal species reported to accumulate PHA belong to the families Halobacteriaceae (example of PHA producers: *Halobacterium noricense*), Haloarculaceae (PHA producers found among the genera *Haloarcula* and *Halorhabdus*), Halococcaceae (PHA producers occurring among the genus *Halococcus*), and Halorubraceae (PHA producers among the genus *Halorubrum*), Haloferacaceae (members of the genera *Haloferax*, *Haloquadratum*, *Halogranum*, and *Halogeometricum* are reported to accumulate PHA), and Natrialbaceae (members of the genera *Halopiger*, *Haloterringena*, *Natronobacterium*, *Halobiforma*, *Natronococcus*, and *Natrinema* reported to produce PHA) (reviewed by Koller, 2019b, Koller et al., 2020). Remarkably, most of the PHA production processes by these extreme halophiles are reported only for small shaking flask setups; exceptions are the steadily growing number of described bioreactor fermentation setups for *Hfx. mediterranei* on glucose, starch, hydrolyzed whey permeate from the dairy industry, extruded rice bran, rice-based stillage, mixtures of fatty acids, olive oil by-products, or crude glycerol phase as side-product of biodiesel production. In this context, *Hfx. mediterranei* was also cultivated on a pilot scale (200 L fed-batch setups in stirred tank reactor) on hydrolyzed whey permeate; this process was described and evaluated by a techno-economic assessment, which concluded that PHA production from surplus whey by extreme halophiles is at least competitive from an economic and life cycle perspective in comparison to competing technologies for whey utilization (Koller et al., 2013). Among the scarce assessments of potential processes for PHA production from inexpensive waste- and surplus materials, the *Hfx. mediterrei* process on hydrolyzed whey permeate was estimated with a production price per kg PHA of about 3 €, based on results from 200 L pilot scale, and including a polymer recovery by hypotonic cell disrupture and recycling of the fermentation side streams (spent fermentation broth, saline cell debris) (Koller et al., 2007b). Similarly, Bhattacharyya *et al.* emphasized the need to close material cycles in *Hfx. mediterranei*-based PHA production processes based on waste stillage originating from rice-based ethanol production to improve the process economically and environmentally (Bhattacharyya et al., 2014). Notably, in all these set-ups using *Hfx. mediterranei*, the strain accumulated PHBHV copolyester, even without the addition of 3HV-related precursor compounds. When adding 4HB-related precursors (GBL), poly(3HB-*co*-3HV-*co*-4HB) terpolyester was produced (Hermann-Krauss et al., 2013). In addition to *Hfx. mediterranei*, there is a small number of literature reports available for PHA-production process with other extreme halophiles under controlled conditions on bioreactor scale, such as for *Halogranum amylolyticum* on glucose (Zhao et al., 2015), *Halopinger aswanensis* on acetate and butyrate (Hezayen et al., 2000), or the thermohalophilic organism *Haloterringena hispanica* on a complex carrot-waste medium (Di Donato et al., 2011); these processes are summarized in Table 1. However, the industrial-scale application of extreme halophiles is still waiting for its realization.

Table 1 provides an overview of selected PHA production processes using different halotolerant, moderately halophilic and extremely halophilic PHA production strains on different carbon sources; the selection of examples is focused on most recent reports and, for haloarchaea, on processes performed under controlled conditions in bioreactors. Cultivation scale, type of PHA produced, PHA productivity and intracellular PHA fraction, as well as the location of isolation of strains, are compared.

3. Some like it hot - PHA biosynthesis by thermophiles

Thermophiles are microorganisms that are naturally adapted to the high temperature of the environment; literally, they are defined as microorganisms with optimal cultivation temperature above 45 °C. They have usually been classified into the three main categories: i. moderate thermophiles with optimal growth temperature in the range 45–65 °C, ii. extreme thermophiles with optimal growth temperature between 65 and 79 °C and iii. Hyperthermophiles, which reveal the best growth at temperatures above 80 °C (Zeldes et al., 2015). Actually, from our human perspectives, thermophiles can be regarded as exotic microbes colonizing rare niches with conditions close to the edge of life, nevertheless, habitats with high temperatures are surprisingly not uncommon. They are usually associated with volcanic activity, e.g., geothermally heated lakes or deep-sea vents. Also, solar activity can substantially enhance the temperature of some niches and other constantly hot habitats might be associated with human activity, such as household or agricultural compost piles (reaching up to 70 °C) or thermal effluents from various industrial processes (Urbieta et al., 2015). Thermophiles play also an important ecological role, for instance, they were likely the first microorganisms capable of thriving on Earth (Lusk, 2019) and they are also in the focus of astrobiologists as microbes that could survive and prosper at surfaces of planets that are hotter than the Earth (Hickey and Singer, 2004).

The adaptation to a high-temperature environment is a complex process that usually includes several strategies such as accumulation of compatible solutes (serving as molecular chaperones preventing biomolecules from heat-induced denaturation), changes in phospholipids structures (ensuring membrane integrity at high temperatures), and alternation in nucleic acids (prevalence of GC pairs) and proteins structure, which enhances their stability at elevated temperatures (Urbieta et al., 2015).

Also, PHA production by thermophiles provides numerous advantages as compared to mesophiles. Similar to increased media osmolarity in halophiles, also high cultivation temperature represents a hurdle that prevents the most common microflora from contamination of the biotechnological process. Therefore, also thermophilic processes can be considered being contamination-resistant. This assumption was confirmed by Ibrahim and Steinbüchel (2010) who performed long-term cyclic cultivations of the PHA producing thermophilic microorganism *Chelatococcus* sp. MW10 using a non-sterilized medium and the cultivation

temperature of even 50 °C prevented the process from contamination (Ibrahim and Steinbüchel, 2010). Further, increased cultivation temperature provides also other advantages, for instance, it enhances the solubility of substrates (apart from oxygen), reduces the viscosity of the cultivation medium, and improves its homogeneity. Moreover, according to Arrhenius, high temperature enhances reaction rates (Peleg et al., 2012), hence, it accelerates desired metabolic processes (Krüger et al., 2018). It might seem that cultivations operated at elevated temperatures are energetically demanding since intensive heating is required, but the opposite is true. When the process is operated as high-cell density cultivation (which is the desired scenario in PHA production), the reactor is at least partially self-heated by metabolic heat of the employed microbial culture, and, importantly, no energetically challenging cooling is required. It should be pointed out that cooling is usually problematic (high energy and cooling water consumption, additional cooling systems might be needed which raises the cost of the equipment) in mesophilic processes operated in high-cell densities at large scales. Therefore, both heating and cooling demands are lowered when thermophiles are employed. Since also energy demands for sterilization are decreased, the thermophilic processes might be of beneficial energy balance (Ibrahim and Steinbüchel, 2010). In halophiles, the high salt concentration in cultivation media represents a serious problem for wastewater treatment management. An important advantage of thermophilic processes is the fact that the application of high temperature does not bring such consequences and complications.

However, unlike in halophiles, PHA production by thermophiles is not a deeply investigated field and was reported only in a very limited number of cases. Therefore, we performed a detailed literature search for reports on PHA production in thermophiles as well as in-depth bioinformatics analysis of available genomes of thermophilic microorganisms. In this case, we were looking for genes encoding for PHA synthases as the presence of *phaC* is a necessary prerequisite for the capability of PHA synthesis. The results are demonstrated in Table 2 and deeply discussed in the following text. We primarily searched through the database of non-redundant sequences standing for particular species, thus, where possible, accession numbers of PhaC synthases in NCBI RefSeq database (O'Leary et al., 2016) were used. Since not all genomes were stored in RefSeq database, in several cases accession numbers leading to GenBank database (Sayers et al., 2019) to sequences of particular strains of selected species were used. Moreover, we took all amino acid sequences from Table 2 and reconstructed a phylogenetic tree of all available PhaC synthases in thermophilic species (see Fig. 2). Two major clusters were formed in the phylogram. The smaller cluster contains three different *Aneurinibacillus* species and additional seven bacteria carrying putative Class III PhaC synthases. The larger cluster is formed by putative Class I PhaC synthases. However, there are five species (*Rubellimicrobium thermophilum*, *Pseudomonas* sp. SG4502, *Inmirania thermothiophila*, *Pseudonocardia thermophila*, and *Thermomonas hydrothermalis*) that each form a distinguished branch. These might be different, even unknown, classes of PHA synthases. Their classification is problematic, which is further supported by the low bootstrap support of these branches. Also, the inability to find PHA synthases in the genome of *Geobacilli*, *T. thermophilus*, and *T. elongatus* indicate that there might be some yet undiscovered classes of PhaC synthases. Even identification between known classes based on sequence similarity is impossible and has to be supplemented by the prediction of molecular weight (Kourilova et al., 2021b).

3.1. PHA production by Gram-negative thermophiles

PHA accumulation is likely a common feature typical for the members of the genus *Chelatococcus* which includes several thermophilic and thermotolerant strains. Ibrahim *et al.* isolated PHA producing thermophiles and obtained six gram-negative thermophilic isolates capable of PHA accumulation. Five isolates formed stable star-shaped cell aggregates, only the strain MW10 grew as free-living rod-shaped cells. By 16S rRNA sequencing, the isolates were taxonomically classified as members of the *Chelatococcus* genus with an optimal growth temperature of 50 °C and strain-dependent preference for glucose or glycerol (Ibrahim et al., 2010). The strain MW9 was further systematically classified as a type strain of a novel species - *Chelatococcus thermostellatus* and is currently available in public collections of microorganisms (DSM 28244) (Ibrahim et al., 2016). The strain MW10 was successfully employed for PHA production in the advanced high-cell-density cyclic fed-batch cultivations in a 42-L bioreactor under semi-sterile conditions which proved the viability of the thermophiles-based PHA production (Ibrahim and Steinbüchel, 2010). *Chelatococcus daeguensis* TAD1 represents another thermophilic member of the genus *Chelatococcus* which can be considered being a potential candidate for industrial production of PHA, since PHA synthesis in this bacterium does not require limitation by any nutrient and is growth-associated. Hence, this strain was capable of reaching high PHA content (about 84% of CDM) in a relatively short period (24 h) at 50 °C using glucose as the sole carbon substrate. Apart from glucose, the strain is also able to utilize cheap substrates such as glycerol or starch (Xu et al., 2014). PHA production from glycerol employing *Chelatococcus daeguensis* TAD1 was further in-depth studied and optimized; utilization of two-stage fed-batch cultivation strategy resulted in a very high volumetric productivity (0.434 g/(L·h)) which could be a basis for economically feasible PHA production (Cui et al., 2015).

Another thermophilic Gram-negative genus for which PHA production was reported on phenotype level and also confirmed on genotype level by the identification of *phaC* gene in available genome databases is the genus *Caldimonas*. This genus contains two validly published strains and both were successfully used for PHA production. *Caldimonas manganoxidans* was isolated from a hot spring, and, interestingly, the bacterium was capable not only of accumulating PHA granules but it also revealed the ability to degrade extracellular PHB (Takeda et al., 2002). Later, the PHA accumulation potential of *C. manganoxidans* was investigated in detail; it was observed that it can achieve high product titers in short periods (5.4 g/L PHB in shaken flasks within 24 h) (Hsiao et al., 2016). In a follow-up study, *C. manganoxidans* was employed for PHA production from biodiesel-derived glycerol with promising results (Hsiao et al., 2018). Also the second validly published member of the genus – *Caldimonas taiwanensis* - can be used for PHA biosynthesis. Thanks to its amylolytic activity, Sheu *et al.* used this bacterial strain for PHA production from starch. The elevated cultivation temperature provides, in this case, not only a hurdle for microbial competitors, thus preventing the process from contamination, but it also enhances solubilization of starch, decreases the viscosity of the medium and, therefore, enhances homogeneity of the batch. When valerate was used as a 3HV precursor, the bacterium accumulated a copolymer consisting of 3HB and 3HV. The concentration of valerate in media was used as a regulatory factor influencing the monomer composition of the copolymer, 3HV fraction could reach a value as high as 95 mol% (Sheu et al., 2009).

PHA synthesis is likely a common feature also in the closely related thermophilic genus *Tepidimonas* since the presence of *phaC* gene was determined in numerous genus members such as *Tepidimonas alkaliphilus*, *Tepidimonas aquatica*, *Tepidimonas fonticaldi*, *Tepidimonas charontis*, *Tepidimonas ignava*, *Tepidimonas sediminis*, *Tepidimonas taiwanensis* or *Tepidimonas thermarum* (see Table 2). Nevertheless, studies describing PHA accumulation in *Tepidimonas* spp. are still lacking. The exception is the work of Chen *et al.*, who isolated a thermophilic alkaline-protease-producing bacterium from hot spring in the Pingtung area in Southern Taiwan. The bacterium was systematically classified as novel species *T. taiwanensis* and strong PHA accumulation was observed as one of the important physiological characteristics (Chen et al., 2006b). Generally, from a biotechnological point of view, the potential of *Tepidimonas* spp. for PHA production is limited by a relatively low range of utilizable substrates since most *Tepidimonas* species utilize neither carbohydrates nor lipids; they are restricted to organic acids and amino acids (Moreira

Table 2
List of thermophilic bacteria in which PHA synthesis was identified on genotype and/or phenotype level.

	Gram staining	Optimal cultivation temperature [°C]	PHA biosynthesis identified on genotype level	PhaC accession number	PHA synthesis described on phenotype level	Reference
Amphiplicatus metriothermophilus	G+	37–65	Yes	WP_089411661.1	No	
Aneurinibacillus danicus	G+	35–55	Yes	WP_146808431.1	No	
Aneurinibacillus terranovensis	G+	20–55	Yes	WP_027415498.1	No	
Aneurinibacillus thermoaerophilus	G+	45–60	Yes	WP_057899523.1	Yes	Pernicova et al., 2020a; Xiao et al., 2015
Anoxybacillus calidus	G+	35–70	Yes	WP_181538453.1	No	
Anoxybacillus vitaminiphilus	G+	35–70	Yes	WP_111643468.1	No	
Aquabacterium tepidiphilum	G-	25–50	Yes	WP_119153972.1	Yes	Khan et al., 2019
Bacillus thermoamylovorans	G+	50	No	–	Yes	Choonut et al., 2020a, 2020b
Caldimonas manganoxidans	G-	50	Yes	WP_026330149.1	Yes	Takeda et al., 2002; Hsian et al., 2016
Caldimonas taiwanensis	G-	35–60	Yes	WP_062192707.1	Yes	Shen et al., 2009
Chelatococcus daeguensis	G-	50	Yes	WP_071923939.1	Yes	Cui et al., 2015; Xu et al., 2014
Chelatococcus thermostalatus	G-	50	No	–	Yes	Ibrahim et al., 2016
Cupriavidus sp. S-6	G-	45	Yes	CCE46061.1	Yes	Shen et al., 2012
Dichotomicrobium thermohalophilum	G-	20–65	Yes	WP_119062396.1	No	
Elioraea tepidiphila	G-		Yes	WP_019014986.1	No	
Elioraea thermophila	G-	45–60	Yes	WP_114377842.1	No	
Geobacillus kaustophilus	G+	55	No	–	Yes	Gedikli et al., 2019
Geobacillus stearothermophilus	G+	60	No	–	Yes	Gedikli et al., 2019
Hydrogenophilus thermoluteolus	G-	50	Yes	WP_197713626.1	Yes	Nguyen et al., 2019

	Gram staining	Optimal cultivation temperature [°C]	PHA biosynthesis identified on genotype level	PhaC accession number	PHA synthesis described on phenotype level	Reference
Isosphaera thermothiophila	G-	35–68	Yes	WP_123400972.1	No	
Pseudomonas sp. SG4502	G-	45	No	–	Yes	Satoh et al., 2011
Pseudonocardia thermophila	G-	50	Yes	SHJ98529.1	No	
Rubellimicrobium thermophilum	G-	28–56	Yes	WP_021097566.1	Yes	Denner et al., 2006
Rubrobacter spartanus	G+	45–55	No	–	Yes	Kourilova et al., 2021b
Rubrobacter xylanophilus	G+	40–70	Yes	WP_143527769.1	Yes	Kourilova et al., 2021b
Schlegelella aquatica	G-	30–60	No	–	Yes	Chou et al., 2006
Schlegelella thermodepolymerans	G-	37–60	Yes	WP_104356814.1	Yes	Kourilova et al., 2020, 2021a
Synechococcus sp. MA19	G-	50	Yes	AAK38139.1	Yes	Miyake et al., 1996; Nishioka et al., 2001
Tepidicella baoligensis	G-	20–60	Yes	WP_180683314.1	Yes	You et al., 2019
Tepidicella xavieri	G-	25–55	Yes	WP_133596978.1	No	
Tepidimonas alkaliphila	G-	37–55	Yes	TSE20543.1	No	
Tepidimonas aquatica	G-	35–60	Yes	WP_144324707.1	No	
Tepidimonas fonticaldi	G-	35–60	Yes	WP_143968067.1	No	
Tepidimonas charontis	G-	25–60	Yes	WP_144327474.1	No	
Tepidimonas ignava	G-	50–55	Yes	WP_132961516.1	No	
Tepidimonas sediminis	G-	37–55	Yes	WP_143892980.1	No	

	Gram staining	Optimal cultivation temperature [°C]	PHA biosynthesis identified on genotype level	PhaC accession number	PHA synthesis described on phenotype level	Reference
Tepidimonas taiwanensis	G-	35–60	Yes	TSE30381.1	Yes	Chen et al., 2006b, Kourilova et al., 2021c
Tepidimonas thermarum	G-	25–60	Yes	WP_143902344.1	No	
Tepidiphilus margaritifer	G-	25–61	Yes	WP_051240655.1	Yes	Manaia et al., 2003
Tepidiphilus succinatimandens	G-	50–55	Yes	WP_206202189.1	No	
Tepidiphilus thermophilus	G-	30–60	Yes	WP_055423697.1	No	
Thauera hydrothermalis	G-	37–55	Yes	WP_114649837.1	No	
Thermomonas hydrothermalis	G-	50–55	Yes	WP_072756679.1	No	
Thermosyntropha lipolytica	G+	52–70	Yes	WP_073089411.1	No	
Thermus thermophilus	G-	70–75	No	–	Yes	Pantazaki et al., 2003, 2009
Ureibacillus terrenus	G+	42–65	Yes	WP_141600988.1	No	
Ureibacillus thermophilus	G+	30–65	Yes	WP_208649453.1	No	
Zhizhongheella caldifontis	G-	40–50	Yes	PPE66523.1	No	

et al., 2000). Again, *T. taiwanensis* represents an interesting exception because it is, according to literature, capable of assimilation of glucose and fructose (Chen et al., 2006a). Therefore, it was recently used for PHA production from the inexpensive resource grape pomace, which is rich in these abundant sugars. At the cultivation temperature of 50 °C, the bacterial culture was capable of accumulating PHA up to 65% of CDM in flasks experiments, the final PHA titer reached a very promising value of 3.55 g/L (Kourilova et al., 2021c).

Similarly to *Tepidimonas*, also in genus *Tepidiphilus*, we identified the presence of *phaC* genes in available genomes of *Tepidiphilus margaritifer*, *Tepidiphilus succinatimandens* and *Tepidiphilus thermophilus* (see Table 2). However, reports confirming PHA accumulation on phenotype level are lacking. Only Manaia *et al.* mentioned that *Tepidiphilus margaritifer* accumulates PHA granules as physiological characteristic of the strain without any further details (Manaia et al., 2003). Similar to *Tepidimonas* spp., also by *Tepidiphilus* species organic acids and amino acids, but no sugars or lipids, are used as carbon sources, which limits their potential biotechnological applications as chassis for PHA production (Manaia et al., 2003).

Moreover, very similarly to *Tepidimonas* and *Tepidiphilus*, *phaC* genes were detected also in the only two members of the genus *Tepidicella* - *Tepidicella baoligensis* and *Tepidicella xavieri* (Table 2). Presence of PHA

granules was observed in *Tepidicella baoligensis*, but *Tepidicella* species are not capable of effective assimilation of carbohydrates or lipids (You et al., 2019), thus their biotechnological potential within the concept of NGIB is negligible.

Further, the presence of *phaC* genes was observed in several other species (*Dichotomicrobium thermohalophilum, Elioraea tepidiphila, Elioraea thermophila, Inmirania thermothiophila, Pseudonocardia thermophila, Thauera hydrothermalis, Thermomonas hydrothermalis, Thermosyntropha lipolytica, Thauera hydrothermalis, , Thermosyntropha lipolytica* or *Zhizhongheella caldifontis*), which indicates that also these bacteria could be considered as PHA accumulators, but no definitive reports on PHA accumulation are available. Investigation of their PHA production capability and biotechnological potential could be an interesting topic for further research.

In other bacteria, PHA accumulation was identified on the genotype level and also mentioned as physiological characteristics without any reports regarding the type of polymer accumulated, PHA content in biomass or gained PHA titers. For instance, *Aquabacterium tepidiphilum* contains *PhaC* (Table 2) and the presence of PHA granules was observed by Khan et al. (2019). Also in *Rubellimicrobium thermophilum*, PHA synthesis capability was observed on genotype level and further detected on phenotype level. This bacterium could be an interesting candidate for PHA production since it is capable of assimilation of a wide range of carbohydrates and it also produces carotenoids, which, analogous to PHA and pigment co-production by many halophiles, could be interesting side-products to PHA (Denner et al., 2006); nevertheless, no study on PHA production potential of the strain is available.

On the contrary, PHA production was already studied in *Hydrogenophilus thermoluteolus* TH-1, a thermophilic hydrogen-oxidizing bacterium revealing the highest growth rate among autotrophs. This bacterium contains the genetic machinery for PHA synthesis, the strain was capable of PHB production under nitrogen limitations both under both autotrophic (gas mixture consisting of H_2:O_2:CO_2 75:10:15 was used) as well as heterotrophic (various organic acids were used as substrates) conditions at 50 °C. Especially the autotrophic production potential of this bacterium seems to be very interesting (Nguyen et al., 2019).

The top position among thermophiles holds *Thermus thermophilus* HB8, which was used for PHA production at as high a temperature as 75 °C. Pantazaki *et al.* cultivated this bacterium on octanoate or gluconate; PHA represented about 35–40% of CDM, the polymer consisted of a mixture of scl- and mcl- monomers with 3-hydroxydecanoate as the major monomer unit (Pantazaki et al., 2003). In the follow-up study, *Thermus thermophilus* HB8 was used for scl- and mcl-PHA copolymers production from cheese whey under nitrogen limitation, the maximal PHA titer reached 0.57 g/L (Pantazaki et al., 2009). Nevertheless, even though the complete information on genomes of various strains of *Thermus thermophilus* are available, we were not able to discover a *phaC* gene in these genomes, hence, the only evidence for the PHA accumulating capability of *Thermus thermophilus* are publications by Pantazaki et al. (2003, 2009).

Schlegella thermodepolymerans is a Gram-negative bacterium that was isolated from activated sludge under aerobic and thermophilic conditions by the group of Professor Alexander Steinbüchel as a representative of a novel genus. This microorganism was capable of biodegradation of poly(3-hydroxybutyrate-*co*-3-mercaptopropionate), a biosynthesized polymer that had been considered being non-biodegradable. Therefore, the species name reflects its special polymer degradation capacity (Elbanna et al., 2003). In the subsequent studies, extracellular PHA depolymerase of *Schlegella thermodepolymerans* was characterized in more detail and it was observed that the bacterium is not capable of efficient hydrolysis of thioester bonds as was originally expected (Elbanna et al., 2004). The bacterium was, for a long time, studied only from the perspective of PHA biodegradation, but we have recently suggested that *S. thermodepolymerans* is also a very promising PHA producer. It possesses the complete genetic machinery for PHA biosynthesis, and the PHA portion in biomass can reach up to 87% of CDM at 55 °C. Most thermophilic PHA producers described above reveal only limited catabolic flexibility – hence, the number of substrates that can be converted to PHA is restricted. On the contrary, *S. thermodepolymerans* seems to be very flexible in this perspective; it can utilize a wide range of carbohydrates, lipids and glycerol. Surprisingly, the most preferred substrate was the pentose xylose, a substrate rather rarely accepted by microbes for PHA biosynthesis. Actually, *S. thermodepolymerans* prefers xylose even over glucose, therefore, it can be considered to be an astonishing candidate for PHA production following the NGIB principle from cheap xylose-rich lignocellulose-based resources (Kourilova et al., 2020; Musilova et al., 2021). In the follow-up study, the PHA production potential of *S. thermodepolymerans* from lignocellulose-based media was compared with the halophilic PHA producer *H. halophila* (also listed in Table 1) and the prototype xylose-converting mesophilic bacterium *Burkholderia sacchari*. Among the tested bacteria, *S. thermodepolymerans* demonstrated the highest PHA yields on substrates rich in xylose, and, compared to *H. halophila*, *S. thermodepolymerans* was also robust against microbial inhibitors present in lignocellulose hydrolysates such as ferulic acid, gallic acid, furfural or levulinic acid. This confirms that *S. thermodepolymerans* is an auspicious PHA producer deserving further attention (Kourilova et al., 2021a). Aside from *S. thermodepolymerans*, PHA granules accumulation was reported also for another moderately thermophilic member of the *Schlegelella* genus – *Schlegelella aquatica* (Chou et al., 2006), indicating that PHA synthesis might be a common feature for *Schlegelella* spp.

3.2. PHA production by Gram-positive thermophiles

Besides Gram-negative bacteria, PHA are also accumulated by numerous Gram-positive prokaryotes. Actually, Maurice Lemoigne who is considered being a discoverer of PHA, isolated and chemically described PHA from cells of Gram-positive bacterium *Bacillus megaterium* in 1926 (Lemoigne, 1926). Although mostly Gram-negative bacteria are currently considered as candidates for PHA production, Gram-positive PHA producers provide an important advantage. One of the major obstacles preventing the use of PHA in health care, cosmetics or medicine is the contamination of the polymer produced by Gram-negative bacteria by pyrogenic lipopolysaccharides - endotoxins. These contaminants are co-isolated along with PHA and induce a severe immunological response in the human body, which is extremely unsuitable in numerous uses of PHA (Singh et al., 2019). Of course, if a Gram-positive bacterium unable of endotoxin synthesis is used for PHA production, this contamination can not occur and does not represent a problem.

Very recently, Sangkharak *et al.* employed an isolate classified as *Bacillus thermoamylovorans* strain PHA005 for production of PHA from waste cooking oil; the cultivation was operated at 45 °C and the bacterium showed high PHA accumulating potential since within 48 h of cultivation in shaken flasks, the polymer content in biomass reached 87% of CDM. The accumulated polymer was composed of 85 mol% of 3HB and 15 mol% of 3HV (Sangkharak et al., 2020). In another work, the same collective of authors reported that *B. thermoamylovorans* strain PHA005 produces mcl-PHA when cultivated at 45 °C on sodium octanoate; polymer content in biomass reached up to 63% of CDM and the polymer consisted of C8-C18 monomers, 3-hydroxydecanoic acid (3HD) being the major monomer unit (Choonut et al., 2020b). As mentioned before in this article, mcl-PHA production is usually observed in *Pseudomonads*; however, mcl-PHA synthesis in *Bacilli* was also already reported before (Shahid et al., 2013), but it is a very rare feature; hence, *B. thermoamylovorans* strain PHA005 seems to be a very interesting bacterium. To our best knowledge, the genome of the strain was not published so far and PhaC encoding genes were not identified in available genomes of *B. thermoamylovorans*. Thus, the unique PHA synthetic potential of *B. thermoamylovorans* sp. PHA005 might be a strain-dependant property not observed in other members of the genus.

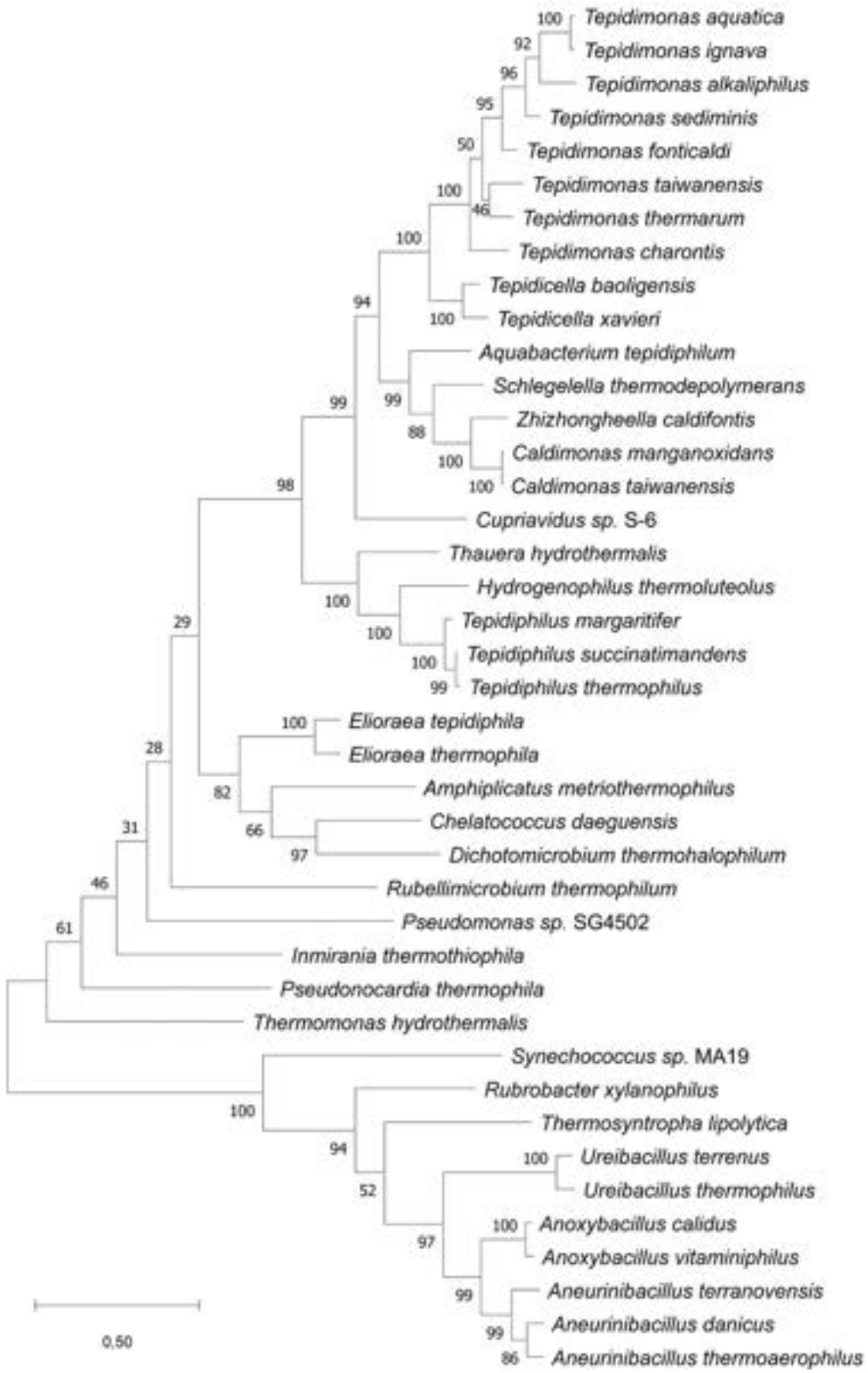


Fig. 2. The phylogeny was inferred by using the Maximum Likelihood method and Whelan And Goldman model (Whelan and Goldman, 2001). The tree with the highest log likelihood (−27,119,16) is shown. The boostrap support calculated from 500 replicates is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the JTT model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. This analysis involved 41 amino acid sequences. There were a total of 780 positions in the final dataset. Evolutionary analyses were conducted in MEGA X (Kumar et al., 2018).

As was already implied, PHA production is a very common feature among the genus *Bacillus* and related species. Historically, during the last decades, several genera were set aside from the parental genus *Bacillus*, including thermophilic or thermotolerant genera such as *Geobacillus*, *Anoxybacillus*, *Aneurinibacillus*, *Ureibacillus* or *Thermobacillus*. Therefore, it can be expected that members of at least some of these thermophilic genera could also accumulate PHA.

The genus *Geobacillus* comprises a group of Gram-positive endospore-formic thermophilic bacteria that can grow over a range of 45–75 °C. Due to their catabolic versatility and rapid growth, *Geobacillus* spp. have raised attention as microorganisms with great potential for biofuel or chemical production following the NGIB concept (Hussein et al., 2015); nevertheless, it seems that PHA synthesis capacity is not widespread among members of *Geobacillus*. Even though genomes of numerous *Geobacillus* species are available in public databases, we were not able to identify genes encoding PHA synthases in *Geobacilli*. Also, particular reports on PHA synthesis by members of the *Geobacillus* genus are scarce; PHA synthesis was observed only in one isolate, which was identified as *Geobacillus kaustophilus* (Gedikli et al., 2019). Hence, it seems that PHA accumulation is not a common feature among *Geobacillus* spp. but it is a rather rare and strain-dependent characteristic.

On the contrary, we have identified that two members of the *Ureibacillus* genus - *Ureibacillus terrenus* and *Ureibacillus thermophilus* (see Table 2) and also in two members of genus *Anoxybacillus* - *Anoxybacillus calidus* and *Anoxybacillus vitaminiphilus*, contain genes encoding for PHA synthase. Nevertheless, to our best knowledge, there are no literature reports on PHA accumulation on phenotype level by these bacteria.

In contrast, PHA synthesis in the genus *Aneurinibacillus* is very well documented. Xiao *et al.* isolated PHA accumulating Gram-positive bacterium from Gudao oilfield in China. The isolate was taxonomically classified as *Aneurinibacillus thermoaerophilus* XH2, its optimal temperature for growth as well as PHA synthesis is 55 °C. Interestingly, the accumulated PHA polymer consisted of 3HB and 3HV with a minor fraction of 3-hydroxyoctanoate and 3-hydroxy-4-phenylbutanoate, so the bacterium accumulated a very interesting PHA copolymer, nevertheless, product titers were rather low - about 0.25 g/L of PHA (Xiao et al., 2015). The authors subsequently published a complete genome sequence, therefore, PHA production capability was confirmed also on the genotype level. As expected, the isolate harbors PHA synthase belonging to class IV which is typical for *Bacillus* and related genera (Xi et al., 2016). Further, Pernicova *et al.* also focused on the isolation of PHA producing thermophiles and obtained several isolates classified as *Aneurinibacillus* sp. (Pernicova et al., 2020b). The most promising isolate – *Aneurinibacillus thermoaerophilus* H1 was capable of accumulation of PHA up to 50% of CDM, maximal PHA titers in shaking flasks were obtained with glycerol as a substrate (about 2 g/L) at 45 °C; nevertheless, the bacterium was capable of growth and PHA synthesis up to 65 °C even though enhancement cultivations at a temperature above 50 °C decreased biomass growth and PHA accumulation. The most interesting property of this bacterium was its capability of incorporating various 4HB (1,4-butanediol, gama-butyrolactone) and 3HV (propionate, valerate) precursors into the PHA structure, resulting in production of poly(3HB-*co*-4HB) copolymer and poly(3HB-*co*-4HB-*co*-3HV) terpolymers with very high 4HB (up to 93 mol%) and 3HV fractions. The monomer composition of poly(3HB-*co*-4HB) copolymer can be simply and precisely controlled by the ratio of 1,4-butanediol (4HB precursor) and glycerol (carbon substrate and 3HB precursor) in the cultivation media (Pernicova et al., 2020a). The material properties of produced copolymers were further studied in detail and it was proved that manipulation of the monomer composition enables regulation of the crucial material properties of PHA such as degree of crystallinity or melting temperature (Sedlacek et al., 2020). Since the presence of *phaC* gene was detected also in other *Aneurinibacilli* such as *Aneurinibacillus danicus* and *Aneurinibacillus terranovensis*, it is suggested that PHA accumulation is rather a common capability among the members of this genus.

The genus *Rubrobacter* accommodates non-motile, obligatory aerobic, asporogenic and Gram-positive actinobacteria. The genus members are considered to be "polyextremophiles" because they can survive under various adverse conditions. For instance, they are highly resistant to harmful effects of various radiations including UV and even ionogenic radiation, moreover, some members are also halotolerant and reveal thermophilic attributes (Castro et al., 2019). Recently, PHA production was confirmed both on genotype as well as phenotype level in two thermophilic members of the genus – *Rubrobacter spartanus* and *Rubrobacter xylanophilus*. Interestingly, the bioinformatic analysis indicated that *Rubrobacter* spp. contain two genes encoding for different PHA synthases – one gene encodes for Class I and the second for Class III PhaC, the second subunit of Class III synthase – PhaE - was identified as well. PHA content in both tested strains reached up to 50% of CDM, PHA polymer consisted of 3HB monomer, 3HV units were incorporated when the proper structural precursors (valerate, propionate, n-amyl alcohol) were supplemented. Especially *R. xylanophilus* demonstrated high robustness concerning cultivation temperature since cultivation temperature had only a very minor effect on both biomass growth and PHA accumulation in the temperature range 45–60 °C. This strain also revealed high catabolic flexibility, therefore, it might be an interesting candidate for PHA production. Importantly, *Rubrobacter* species are Gram-positive thus endotoxin none-forming bacteria but, unlike other Gram-positive PHA producers, *Rubrobacter* species are also not capable of sporulation, which is a very important attribute (Kourilova et al., 2021b). Sporulation is considered being an unfavorable characteristic of the strain to be employed in industrial PHA production because it is accompanied by mobilization of PHA storage and also a sudden loss of desired metabolic activity; moreover, sporulation shifts carbon flux from the formation of desired products (PHA) to unwanted by-products (endospores), which results in an economic loss (Sadykov et al., 2017). Thus, thanks to its thermophilic feature, catabolic flexibility, Gram-positivity and also an absence of sporulation capability, *Rubrobacter* species are also interesting chassis for industrial PHA production within the NGIB concept (Kourilova et al., 2021b).

3.3. PHA synthesis as an adaptation strategy to high temperature

In halophiles, the interconnection between adaptation to hypertonic conditions and PHA biosynthesis is well documented. Direct experimental observations are confirming the osmoprotective role of PHA granules (see the previous section) and there are also indirect hints such as the fact that the list of PHA producing halophiles is long (Table 1) and includes not only moderately halophilic bacteria but also extremely halophilic *Archaea* such as *Hfx. mediterranei* or *Hgm. borinquense* (Koller, 2019b; Pfeifer et al., 2021). On the contrary, it seems that the role of PHA in adaptation to high temperatures is not so evident. There are reports that PHA accumulation protects against heat. For instance, Zhao *et al.* observed that PHA biosynthesis capable wild-type strain of *Aeromonas hydrophila* was more resistant to heat than PHA synthase negative-mutant unable of PHA synthesis (Zhao et al., 2007). Nevertheless, potential protective mechanisms of PHA against high temperature are not clear. It is likely that PHA monomer units possessing strong chemical chaperone activity might protect bacteria from the adverse effects of high temperature (Obruca et al., 2016b; Soto et al., 2012). Furthermore, also phasins, PHA granules associated proteins, reveal chaperoning activity and might prevent cellular proteins from denaturation at high temperatures (de Almeida et al., 2011). Nevertheless, apart from *Thermus thermophilus* – the only extremely thermophilic bacterium on the list of PHA producers, PHA synthesis was reported only for thermotolerant and moderately thermophilic bacteria (see Table 2) with optimal growth temperature between 50 and 60 °C. To the best of our knowledge, PHA synthesis was observed neither in hyperthermophilic bacteria nor archaea. Hence, despite their proven protective function against slightly elevated temperatures (Alves et al., 2020; Gonçalves et al., 2019), it is likely that PHA are not the most suitable

metabolites for adaptation to extremely high temperatures. From the biotechnological point of view, this fact is not a limitation. At very high cultivation temperatures, considerably decreased oxygen solubility represents a serious obstacle practically thwarting to reach high cell density of aerobic microbial cultures. Since PHA are intracellular metabolites of the secondary metabolism, high cell density is a necessary prerequisite for gaining high product titers. Therefore, the fact that only moderately thermophilic prokaryotes produces PHA is not a drawback.

3.4. Thermophilic PHA producers as sources of genes encoding potent PHA-related enzymes

Besides direct utilization for biotechnological production of PHA, thermophilic bacteria capable of PHA production are also very interesting as sources of thermostable and highly active PHA synthases and other enzymes for PHA synthesis *in-vivo* or even *in-vitro*. Tajima *et al.* investigated *in-vitro* synthesis of PHA. These authors emphasized the importance of the utilization of thermostable enzymes for the rapid and sustainable process of *in-vitro* PHA production. In their study, they utilized thermostable acetyl-CoA synthase from the thermophilic bacterium *Pelotomaculum thermopropionicum* JCM10971, CoA transferase from *Thermus thermophilus* JCM10941 and PHA synthase from the thermotolerant bacterium *Pseudomonas* sp. SG4502. The system was capable of synthesis of PHB and a copolymer consisting of 3HB and lactic acid. Due to the thermostability of employed enzymes, the system was operated at 45 °C. Even at 37 °C, the yields obtained by thermophilic enzymes were 1.4-fold higher than those obtained by mesophilic enzymes (Tajima et al., 2016). In this context, all the bacteria listed in Table 2, even those which are no promising candidates for PHA synthesis, for instance because they can not be cultivated on biotechnologically relevant substrates or are not capable of reaching high cell densities or high PHA amounts in biomass, might be considered being interesting resources for thermostable and highly active PHA synthases and other enzymes which to be employed for PHA synthesis.

Further, discovery and in-depth characterization and understanding of thermostable and highly active PHA synthases from thermophilic prokaryotes might be used for the rational design of highly active and stable PHA synthases. Sheu *et al.* constructed a chimeric PHA synthase of the thermophilic strain *Cupriavidus* sp. S-6 and the mesophile *C. necator* H16. In details, this constructed chimeric enzyme was a PHA synthase from mesophilic *C. necator* H16 bearing 30 point mutations derived from the middle region of PHA synthase of the thermophilic strain; it demonstrated 3.45-fold higher specific activity than the parental enzyme at 30 °C and substantially higher stability since at 45 °C its half-life was 127-fold higher than that of the parental enzyme. Transgenic *Escherichia coli* harboring chimeric PHA synthase cultivated at 37 °C accumulated at 59% PHB per CDM, which was substantially more than in a strain harboring PHA synthase of mesophilic *C. necator* H16 strain (38% of CDM) (Tajima et al., 2016).

4. Making a good thing better- metabolic engineering and synthetic biology of extremophilic bacteria for tailored PHA production

Metabolic engineering (ME) and synthetic biology (SB) have generated numerous useful tools and engineering approaches of which some have revolutionized microbial biotechnology and the use of microorganisms, their metabolic pathways, and enzymes for biosynthesis of valuable chemicals (including PHA) or biodegradation of environmental pollutants (Choi et al., 2019; Dvořák et al., 2017). Initially, these powerful tools were restricted for modifications of microbial models *E. coli*, *Bacillus subtilis*, or *Saccharomyces cerevisiae*. The advent of ME in the 1990s enabled interspecies transplantation of whole metabolic pathways and gave rise, beside others, to recombinant PHA-producing *E. coli* strains with exogenous *phaCAB* operon (Choi et al., 1998). Some of the derivatives of these strains are now used for PHA manufacturing on an industrial scale (Tan et al., 2021). Research communities that work with non-canonical microbial hosts including paradigmatic PHA producers *C. necator* or *Pseudomonas putida* also understood the potential of ME and SB and quickly adopted these disciplines in their work. This step enabled the construction of upgraded bacterial factories capable of enhanced accumulation of PHA with altered properties from a variety of low-cost substrates (Budde et al., 2011; Dvořák et al., 2020b; Park et al., 2013; Salvachúa et al., 2020). Extremophilic PHA producers - especially halophiles and thermophiles - have drawn the attention of biotechnologists more recently, which means that a less colorful palette of ME and SB methods and gadgets is available for them. The unsatisfactory situation is nonetheless rapidly changing for PHA-forming halophiles, namely for certain strains from genus *Halomonas* or the haloarchaeal genus *Haloferax* (Haque et al., 2020; Zhang et al., 2020). In the last couple of years, *Halomonas bluephagenesis* TD01 became a testbed for ME and SB tools in halophiles and, as mentioned previously in this article, its engineered derivatives are being used in several Chinese companies that develop NGIB based on this bacterium (Tan et al., 2021). To the best of our knowledge, no study reporting the use of genetically enhanced thermophilic microorganisms for PHA production has been published to date. But the wealth of inspiration for such an endeavor can be taken from the increasing number of publications that unveil the potential of moderately and extremely thermophilic bacteria and archaea for the biosynthesis of biofuels from lignocellulosic residues (Crosby et al., 2019; Jiang et al., 2017b).

In this chapter, we will discuss selected examples of studies from the last decade in which ME and SB played a key role in improving the yield, properties, and downstream processing of PHA in halophilic bacteria, or in broadening the substrate scope of these organisms. We will also map the history of adoption and development of genetic engineering, ME and SB tools and techniques for PHA-forming halophiles (Fig. 3). The prerequisites for successful engineering interventions in thermophilic PHA producers will be highlighted in the last part of the chapter.

4.1. Engineering halophiles for high-yield production of PHA with tailored properties

High production cost and limited biopolymer diversity are the two major issues that hinder wider commercialization of PHA produced by conventional microbial strains of *E. coli* or *C. necator* as well as by the next generation bacterial platforms derived from genus *Halomonas* (Tan et al., 2021; Zheng et al., 2020). Metabolic engineering of suitable natural PHA producers can lead to substantially increased PHA content in the cell, higher PHA yield, titer, and substrate-to-product conversion, or diversified arrangements of polymer structures with new functionalities and properties that fit market needs. Professor Chen's group work on *H. bluephagenesis* TD01 is an outstanding showcase to demonstrate the step-by-step adoption and development of ME and SB tools for a non-canonical extremophilic host and pushing the limits of the cell factory towards economically feasible bioprocess (Chen et al., 2017).

In 2011, Tan and co-workers (Tan et al., 2011) reported unsterile continuous cultivation of the wild-type strain TD01 in seawater with glucose as a sole carbon source during which PHB content in CDM reached up to 80 wt% and substrate-to-biopolymer conversion ranged from 20 to 50%. This engaging study drew attention to *H. bluephagenesis* as a potential candidate for the NGIB but also pointed to the lack of reliable ME tools for this organism. Basic molecular biology toolkit including a restricted number of cloning and expression shuttle plasmid vectors (such as pHS15, pEE5), selection markers (e.g., kanamycin, gentamycin, trimethoprim) and reporters (GFP), conjugation and electroporation protocols, some regulatory sequences for gene expression, or transposon Tn*1732* and suicide plasmid pKS18*mobsac* for random gene knockouts in halomonads was already available by that time (Argandoña et al., 2012). However, a true ME Swiss knife was urgently needed for advancing *H. bluephagenesis* from microbial Cinderella to a

biotech star.

4.1.1. Development of fundamental tools for genetic manipulations of Halomonas

Following studies of Chen's group therefore initially focused on the design of genetic tools for TD01 and their utilization for increasing content, productivity, and applicability of PHB, the biopolyester typically synthesized by *H. bluephagenesis*. Fu and co-workers (Fu et al., 2014) developed a scarless gene knockout and integration system based on the I-SceI endonuclease-mediated homologous recombination technique established for *E. coli* (Posfai et al., 1999). Constructed suicide vector pRE112-6I-SceI with R6kγ replicon and mobilizable broad-host-range plasmid pBBR1MCS1-I-SceI with constitutively expressed I-SceI gene were transferred to TD01 via conjugation with *E. coli* S17-1 (electroporation and chemical transformation did not work for *H. bluephagenesis*). The system was used to delete *prpC* gene encoding 2-methylcitrate synthase and to improve the otherwise poor production of PHBHV in mineral medium with glucose and propionate. The deletion prevented degradation of 3HV precursor propionyl-CoA in methylcitric cycle. As a result, 3HV fraction in PHBHV increased substantially and PHBHV content in CDM reached 70 wt% both in shake flask and 500 L fermentor experiments.

In another study from the same year, the conjugation efficiency and stability of exogenous plasmid DNA in *H. bluephagenesis* was improved by partial inactivation of the restriction-modification system via deletion of 8.4-kb *hsdRMS* gene cluster and *re1* and *re2* genes (Tan et al., 2014). The authors also looked for the optimal plasmid backbone that would secure the stable expression of heterologous genes in *Halomonas*. High-copy broad-host-range plasmid pSEVA341 (with chloramphenicol resistance) from Standard European Vector Architecture collection (Silva-Rocha et al., 2013) surpassed previously prepared plasmids (e.g., pRE112-pMB1) in its stability and conjugation efficiency. The expression vector was prepared by cloning the strong hybrid $LacI^{q}$-P_{trc} promoter inducible with IPTG together with *Halomonas* porin gene RBS (ribosome binding site) into pSEVA341 polylinker. The new vector was used for the overexpression of the threonine synthesis pathway (*thrACB*) and threonine dehydrogenase gene (*ilvA*) which enabled PHBHV copolymer production solely from glucose or other unrelated carbon sources (glycerol, sucrose, maltose, fructose) without the need for co-feeding with costly and toxic propionic acid.

Subsequently, also the chromosomal expression of exogenous genes was tested and *Halomonas* operon outer membrane porin gene expressed from strong constitutive P_{porin} promoter was used as a hot spot. Yin *et al.* employed the I-SceI-based system for the insertion of *phaC* gene from *C. necator* downstream the porin gene in the *H. bluephagenesis* Δ*phaC* mutant (Yin et al., 2014). They demonstrated that $phaC_{Cn}$ expression

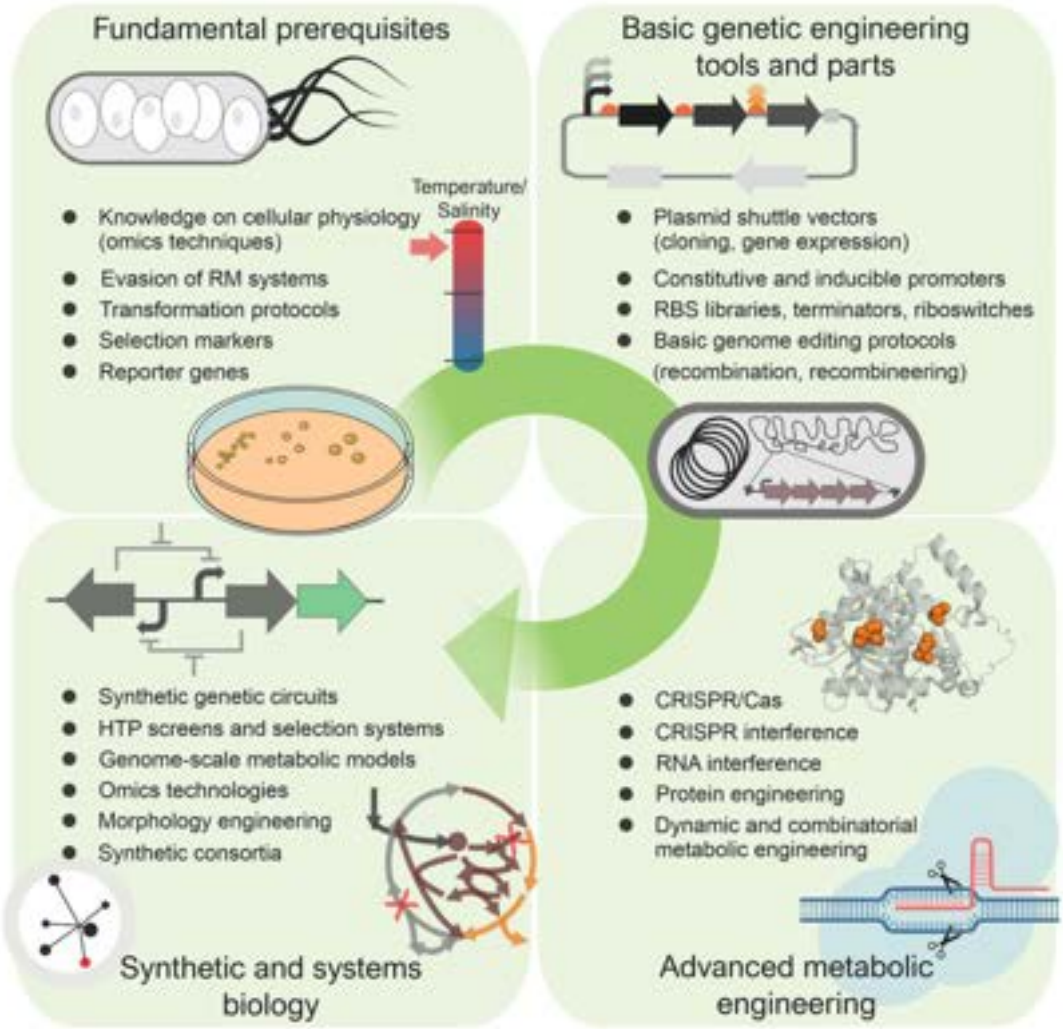


Fig. 3. Tools and approaches of modern microbial bioengineering disciplines that can be adopted to domesticate and upgrade environmental halophilic and thermophilic PHA producers. The figure outlines the roadmap towards the preparation of next-generation bacterial catalysts for biotechnological manufacturing of PHA. Abbreviations: RM systems, restriction-modification systems; RBS, ribosome binding site; CRISPR, clustered regularly interspaced short palindromic repeats; Cas, CRISPR-associated protein; HTP screens, high-throughput screens.

from this spot, in contrast to expression from several different chromosomal loci, fully compensated deletion of native *phaC* and enabled levels of PHB in *H. bluephagenesis* mutant very similar to the wild-type strain. The verified P_{porin} was later characterized and randomized using degenerate oligos to give rise to the promoter library with over 300-fold variation in transcriptional activity (Li et al., 2016b). In general, constitutive promoters can be instrumental for stable robust expression of target genes especially in long-term large-scale cultivations under non-sterile conditions in which the addition of costly, sometimes even harmful chemical inducers is not practical (Dvorak et al., 2015). This was demonstrated by adopting broad-host-range hybrid promoter P_{tac} for chromosomal expression of the *orfZ* gene of 4-hydroxybutyrate-CoA transferase from *Clostridium kluyveri* in TD01 strain. The mutant strain could for the first time synthesize the attractive elastic poly(3-hydroxybutyrate-co-4-hydroxybutyrate) [poly(3HB-co-4HB)] copolymer with a 4HB content of up to 16 mol.-% from glucose and γ-butyrolactone in a 48 h-lasting fed-batch cultivation in a 1000-L pilot fermentor under non-sterile conditions (Chen et al., 2017). The suboptimal productivity of the mutant (1.04 g/(L•h)) was surpassed by 53% in the follow-up work in which the authors placed the *orfZ* gene downstream the mutant variant of P_{porin} promoter integrated into the genome of recombinant *H. bluephagenesis*; here a volumetric productivity for poly(3HB-co-11%-4HB) of 1.59 g/(L•h) was obtained (Shen et al., 2018).

Tightly regulated inducible promoters, on the other hand, are required especially for advanced ME and SB applications in the design of synthetic genetic circuits or for the precise control of expression of key enzymes in a metabolic network. Li *et al.* adopted the *E. coli lac* repressor operator system to construct an inducible expression machinery suitable for *H. bluephagenesis* (Li et al., 2016b). The *lacI* repressor gene was inserted downstream the porin ORF, while the *lacO* operator sequence was placed downstream the −10 element of P_{porin} cloned into pBBR1-MCS1 or pSEVA341 bearing $phaCAB_{Cn}$ operon or only $phaC_{Cn}$ gene from *C. necator*, respectively (Yin et al., 2014; Li et al., 2016b). The controllability of the expression from the resulting plasmids was demonstrated by the PHB content in *Halomonas* cells, which could be tuned by the varying concentration of synthetic IPTG inducer in the culture medium.

The spectrum of inducible expression systems available for *Halomonas* spp. was substantially expanded by Zhao *et al.* in 2017. As the attempts to adopt the paradigmatic T7 system in *H. bluephagenesis* failed, the authors decided to mine novel T7-like expression machineries using BLAST searches in NCBI genome database and PHIRE (PHage In silico REgulatory elements) software package (Lavigne et al., 2004). Three new T7-like RNA polymerase-promoter pairs named MmP1, VP4, and K1F were described and studied in more detail. RNA polymerase modules were integrated into the chromosome, while the respective promoters were cloned into the pSEVA321 plasmid backbone bearing the GFP reporter gene. All three systems displayed broad-host-range functionality (proven in *E. coli*, *Halomonas* TD01, and *Pseudomonas entomophila*), orthogonality, and tight regulation (>3000-fold induction range depending on IPTG concentration). *Halomonas* TD01 with *phaCAB* operon expressed from MmP1 system embedded in chromosome produced by 38% more PHB than the wild-type control (69 g/L vs. 50 g/L) and broke the actual record of PHB content in dry *Halomonas* cells (92 wt.-%).

Besides transcription, the translation level is vital for balancing gene expression in native and synthetic operons encoding the key metabolic pathways for PHA formation. In combinatorial ME, these two levels are often varied simultaneously by combining pathway modules expressed from two or more plasmids, promoters, and RBS sites with different characteristics and searching for the optimum in the possible solution space (Ajikumar et al., 2010; Jeschek et al., 2017; Kurumbang et al., 2014). Ren et al. (2018) optimized the expression of $phaCAB_{Cn}$ operon in *H. bluephagenesis* by combining two mutated variants of P_{porin} promoter with two synthetic RBS sites from the previously prepared library (Li et al., 2016a). Synthetic RBS sites were used for *phaA* and *phaB* genes, while *phaC* retained the native RBS. The combination of individual elements gave rise to four *phaCAB* operon variants that were expressed from pSEVA331 plasmid backbone. The PHB content determined in the cells with the best operon variant was 12.3% higher than in the parental strain without the optimized *phaCAB* cassette.

4.1.2. Adoption of CRISPR/Cas technology in ME of halophiles

The new era of extremophile engineering started with the onset of CRISPR/Cas genome editing technology, which was readily adopted by the *Halomonas* community. Qin et al. (2018) designed and tested for the first time an editing system based on (i) the *Streptococcus pyogenes cas9* gene placed on low-copy plasmid pSEVA321 and (ii) sgRNA (single-guide RNA) cloned together with donor DNA for homology-directed repair into the high-copy pSEVA241 with CAS constitutive promoter of *cas9* gene. This setup enabled up to 80% efficiency of *phaC* gene disruption in the genome of TD01 strain. The slightly modified system was adopted also for genomic insertions of more than 4 kb-long ORFs. The authors used the new technology for the study of glucose metabolism in *Halomonas* and identified the Entner-Doudoroff pathway as a major route for glucose catabolism in strain TD01. The efficiency of the prepared deletions of glycolysis genes ranged from 12.5 to 100%. Importantly, developed CRISPR/Cas9 system with minor modifications was shown functional also for *H. campaniensis* LS21. As a real quantum leap, this work reduced the time needed for genome editing of *Halomonas* from months to weeks. Ling and others (2018) took the advantage of the new rapid genome-editing technique and intended to enhance PHB production in *H. bluephagenesis* by engineering its redox metabolism (Ling et al., 2018a). They identified that *H. bluephagenesis* naturally generates more biomass and PHA under oxygen-limiting conditions and revealed that acetoacetyl-CoA reductase PhaB is NADH-dependent. Extraordinary high $NADH/NAD^+$ ratio (1.5) and surplus NADH generated in metabolism of glucose-grown TD01 strain was beneficial for PhaB function but could inhibit pyruvate dehydrogenase in parallel. The authors used acetate in cell culture as a redox regulator to remove the inhibition and provide more acetyl-CoA for PHB synthesis under oxygen limitation. Additional disruption of the NADH utilization pathway – the NADH-dependent respiratory chain – by deleting the b subunit of electron transfer flavoprotein ETF using CRISPR/Cas9 system enabled a remarkable accumulation of PHB in mutant cells (94 wt.-%) on 30 g/L glucose and 3 g/L acetate without affecting growth. In another study, Chen and colleagues (Chen et al., 2019) achieved to obtain 65% PHBHV in CDM of glucose and gluconate-grown *H. bluephagenesis*. The PHBHV contained 25 mol.-% 3HV, which was the highest reported 3HV content in *Halomonas* cells with chromosomally encoded PHBHV pathway. Complex changes in the best production strain were conducted using CRISPR/Cas9. The genetic modifications together with added gluconate co-substrate enhanced the activity of TCA cycle and enabled a high 3 HV ratio in PHBHV copolyester. Moreover, the previously prepared P_{porin} promoter library (Shen et al., 2018) found its use in this study for finetuning the expression of an exogenous operon from *E. coli*. A new genome-editing technique was also utilized for the construction of a *H. bluephagenesis* strain capable of producing functional scl- and mcl-PHA co-polymers containing unsaturated bonds (Yu et al., 2020). A heterologous biosynthetic pathway from *Aeromonas hydrophila* consisting of PHA synthase PhaC and enoyl-CoA-hydratase PhaJ for polymerization of 3HB and the mcl-PHA monomers 3-hydroxyhexanoate (3HHx; saturated monomer) or 3-hydroxy-5-hexenoate (3HHxE; unsaturated) into repsective co-polymers was introduced into *H. bluephagenesis* Δ*phaC* with overexpressed endogenous acyl-CoA synthetase (FadD). Optimization of promoter and RBS in introduced expression cassette, together with the finetuning of cell redox state and fed-batch cultivation strategy, resulted in the synthesis of functional PHA with unprecedented attractive properties that have never been produced by *Halomonas* before.

In contrast to null mutations, CRISPR interference (CRISPRi) enables

tunable downregulation of target genes on the translation level. The CRISPRi plasmid for halophilic PHA-producing bacteria was constructed by Tao and co-workers (Tao et al., 2017). The pli-dCas9-sgRNA bearing the dead Cas9 gene and the sgRNAs cassette expressed from IPTG-inducible Ptrc promoter was prepared using the previously published CRISPRi plasmid plv-dCas9-sgRNA for *E. coli* (Lv et al., 2015) and pSEVA321 backbone. The sgRNA cassette was designed using the recognition sites for isocaudomers and Biobricks standards, thus, several sgRNAs could be integrated into a single plasmid and target multiple genes in parallel. The system functioning in *H. bluephagenesis* was demonstrated on the partial repression of *prpC* and *gltA* (citrate synthase) genes which allowed flexible regulation of 3HV fraction in PHBHV copolymer and re-routing acetyl-CoA from TCA cycle to PHB synthesis, respectively.

4.1.3. *Advanced metabolic flux fine-tuning in halophiles for higher product yields*

Modern ME is based on much more than just knockouts and knockins of genes in a target host organism. Development of high-performance strains for NGIB requires knowledge-driven tuning of intracellular carbon fluxes, which should prevent system perturbations and suboptimal product yields caused by disorganized expression of exogenous operons and possible accumulation of toxic pathway intermediates. At this point, synthetic and systems biology come to the scene. Synthetic genetic circuits for tightly regulated gene expression, multi-omics analyses, high-throughput screens, and mathematical modeling of metabolic networks form an inseparable part of state-of-the-art ME approaches (Choi et al., 2019; Dvořák et al., 2017). This trend must be, of course, reflected also in the engineering of extremophilic PHA producers and although there is some debt in adopting synthetic and systems biology, first attempts have indeed emerged in the last couple of years.

In this context, Ma *et al.* aimed to manipulate *H. bluephagenesis* to become a suitable chassis for the co-production of PHA with another valuable chemical ectoine – above-mentioned compatible solute for hypersalinity resistance which has potential applications in cosmetics and medicine (Ma et al., 2020).

They combined three metabolic modules comprised of *de novo* ectoine pathway (endogenous *etcABC* gene cluster) with endogenous L-aspartate-semialdehyde-dehydrogenase (Asd) and aspartokinase (LysC) from *Corynebacterium glutamicum* that were overexpressed from the chromosome of *H. bluephagenesis* lacking competing pathways to channel more flux to ectoine synthesis. Importantly, two orthogonal inducible types of machinery based on LuxR/AHL (N-acyl homoserine lactone) and T7-like MmP1 systems, embedded in *Halomonas* chromosome and in low-copy pSEVA321 plasmid, were designed and used for GFP reporter-mediated transcriptional fine-tuning (Ye et al., 2020) of the three ectoine synthesis modules. Once the optimal transcriptional level for each of the three modules was identified with the help of the plasmid-based system, the P_{porin} promoter library was searched to provide similar optimal ectoine titer with the expression of *etcABC*, *asd*, and *lysC* from the chromosome. The resulting *H. bluephagenesis* strain produced 32 g/L CDM with 75 wt.-% PHB and 8 g/L ectoine after 44 h cultivation on glucose and urea (low-cost nitrogen source for ectoine synthesis) under open unsterile conditions in a 7 L fermentor. The recent work of Jiang and colleagues (Jiang et al., 2021) is a notable example of a study in which omics techniques play an important role. The authors focused on the production of industrially valuable platform chemical 3-hydroxypropionic acid (3HP) and its copolymer with 3HB, poly(3-hydroxybutyrate-*co*-3-hydroxypropionate), in *H. bluephagenesis* from glucose and 1,3-propanediol. Extraordinary high titer of 3HP in final fed-batch culture (154 g/L) would not be possible without utilizing transcriptomics for the identification of a competitive 3HP-degradation pathway and endogenous 1,3-propanediol dehydrogenase, which enabled the efficient conversion of 1,3-propanediol to 3HP. The complete 3HP biosynthesis route comprising three metabolic modules with genes from four microorganisms was optimized using combinatorial ME and promoter, RBS, and gene order adjustments.

The development of mathematical modeling towards optimization of PHA production, in general, has been reviewed by Novak *et al.* in 2015. They suggest that a compromise solution can be achieved with hybrid modeling approaches that combine existing high-level (low-structured) kinetic models with fluid dynamics and neural networks. To understand and synthesize optimal PHA pathways dynamics, low-level (highly structured) models have to be developed (Novak et al., 2015). In the context of halophiles, Cui *et al.* have proposed a high-level differential model targeting PHA production in *H. mediterranei* under different temperatures suggesting that high temperatures may provide a good strategy for improving the PHA productivity (Cui et al., 2017b). To bring the utility of such models to fine-tuning of PHA production in halophiles, a complete genome-scale reconstruction of the organism's metabolic network yet needs to be built to form a fundamental base for kinetic optimizations.

4.2. *Expanding the substrate scope of halophiles for cost-effective PHA production*

A significant portion (dozens of %) of the PHA production cost in both conventional and non-conventional hosts can be attributed to the substrate price (Braunegg et al., 2004; Dietrich et al., 2019; Lettner et al., 2017). PHA-related carbon sources (pure fatty acids, oils) or pure glucose have been frequently used in research studies coping with microbial PHA production. However, sustainable NGIB require cheap and abundant next-generation substrates. Glycerol from biodiesel production, waste organic polymers (lignocellulosic residues, kitchen waste, waste petroplastics), and C1 substrates (syngas, CO_2, CH_4) represent extremely attractive carbon sources for future biotechnological production of bulk chemicals including PHA and other biopolymers (Dietrich et al., 2019; Tiso et al., 2021; Weiss et al., 2017). Xylose and arabinose-rich hemicellulose fraction and lignin aromatics from lignocellulose waste processing have become the targets of intense research in the last years (Dietrich et al., 2019; Salvachúa et al., 2020). Some extremophilic bacteria including *H. halophila* or *S. thermodepolymerans* have the natural ability to grow and produce PHA on xylose or arabinose, but especially the former species is quite sensitive to the inhibitory effects of aromatic chemicals in lignocellulosic hydrolysates (Kucera et al., 2018; Kourilova et al., 2020). Metabolic engineers should therefore aim at improving the robustness of these organisms and at enhancing PHA yields and productivities on pentoses, but, to the best of our knowledge, such studies are not yet available.

The first attempt to convert organic polymeric feedstock directly to PHA using an engineered extremophilic bacterium has just recently been reported. Lin et al. (2021) modified *H. bluephagenesis* for growth and PHB production on starch. Codon-optimized α-amylase gene from *Bacillus licheniformis* and endogenous amyloglucosidase gene with a pre-selected Sec-secretion signal sequence were inserted into the chromosome of *Halomonas* and in a plasmid, respectively, and expressed from optimal constitutive P_{porin} and inducible P_{Mrp1} promoter variants. Well secreted enzymes hydrolyzed commercial insoluble corn starch to monomeric glucose. The thus designed recoenbinant *Halomonas* grew on 30 g/L starch in shake flasks to 9.5 g/L CDM containing 51.5 wt.-% PHB. These values are lower than those achieved on pure glucose, but future advances in engineering recombinant protein secretion in halophiles can push these limits and maybe even direct the authors towards second-generation polymeric feedstocks such as lignocellulosic residues.

All previously discussed halophilic and thermophilic heterotrophs cannot naturally utilize C1 substrates such as CO_2 or CH_4 for PHA synthesis. The engineering of efficient CO_2 fixation in heterotrophs is very challenging and is currently restricted to mesophilic model microorganisms (Gassler et al., 2020; Gleizer et al., 2019). This barrier can be nonetheless bypassed by a smart combination of PHA-producing extremophile with a suitable autotroph in a rationally designed synthetic consortium. An interesting example of such an approach is the work of

Weiss and co-workers (2017) in which an engineered cyanobacterium *Synechococcus elongatus* with implanted sucrose permease was co-cultured with PHB-forming *H. boliviensis*. The cyanobacterium autotrophically fixed CO_2 and generated sucrose, the major portion of it (up to 85%) was secreted into the bulk medium where it served as a heterpotrophic carbon source for *Halomonas*. *H. boliviensis* was selected as a partner for *Synechococcus* in synthetic co-culture because it has low nutritional demands and grows well in the cyanobacterial minimal medium. *Synechococcus* cells were physically separated from *Halomonas* by encapsulation into alginate beads. This step improved co-culture stability, facilitated the selective recovery of PHB-producing *Halomonas* cells, and, quite unexpectedly, doubled sucrose secretion by the cyanobacterium. The co-culture was maintained stable and productive without any selection agents for remarkable five months. The PHB content in *Halomonas* biomass reached 37 wt.-%. It is not very surprising that the PHB productivity was low (peaked at 28.3 mg/(L•d)). However, it should be stressed that this value is equal or even higher than productivities achieved in monocultures of cyanobacteria engineered for autotrophic PHA production from CO_2 (Wang et al., 2013). This consortium approach thus represents an interesting alternative to single strain engineering in future research.

4.3. Engineering Halomonas for facilitated downstream processing of PHA

Another costly part of industrial PHA production includes downstream processes such as separation and disruption of cells and product purification (Wang et al., 2019). Conventional physical, physicochemical, or biological techniques used for this purpose are often expensive (centrifugation), time-consuming (gravity sedimentation of cells, enzymatic treatments), or harsh to the environment (acidic or alkaline treatments, application of solvents for product recovery) and their efficiency in terms of recovery yield and product purity can be limited. Genetic engineering and SB provide complementary or entirely alternative solutions for downstream processing of intracellular products including PHA. Some of them have already been tried in *Halomonas*. For instance, morphology engineering has the potential to kill two birds with one stone and produce more PHA in enlarged bacteria that settle faster in the medium than standard-shaped cells. The size of the *H. bluephagenesis* cells was expanded by blocking the cell division through inhibition of Z ring formation. This was achieved by inhibiting the polymerization of tubulin-like protein FtsZ (Tan et al., 2014; Zhao et al., 2017), which is necessary for binary fission of bacteria, or by the direct deletion of *ftsZ* gene (Jiang et al., 2017a). Proteins MinC and MinD belong among inhibitors of FtsZ polymerization. IPTG-induced expression of *minCD* from pSEVA341 plasmid resulted in 1.4-fold longer *H. bluephagenesis* cells (1.84 μm on average) when compared with uninduced control, and 19% higher PHB content in cells (Tan et al., 2014). Interestingly, expression of *minCD* genes from novel T7-like expression systems MmP1 in *H. bluephagenesis* chromosome gave rise to cells with a much higher average length of 102 μm (Zhao et al., 2017). Unfortunately, the effect of morphology engineering on PHB content and yield was not quantified in the latter study.

The drawback of the interventions in bacterial morphology is reduced cell growth and consequently lower PHA titer. Jiang and co-workers (2017a) developed a plasmid-based system for temperature-inducible morphology changes in *H. campaniensis* LS21. The strain with *ftsZ* deletion and temperature-sensitive plasmid pTKmf bearing constitutively expressed *ftsZ* gene was initially grown as usual at 30 °C for 12 h and then the morphology change was induced by removing the plasmid with the temperature raised to 37 °C for the rest of the culture. Both biomass and PHB content in cells was increased more than 1.3-fold when compared with the wild-type strain to 16 g/L CDM and 78 wt.-% PHB, respectively. The inhibition of FtsZ described previously (Zhao et al., 2017) also improved cell biomass separation by gravity sedimentation, but the reported 12 h time interval required to fully separate the cells would probably not meet the needs of the industrial process. Therefore, it might be beneficial to promote self-flocculation of engineered PHA-producing cells which would accelerate their sedimentation. This was achieved for *H. campaniensis* LS21 (not yet for *H. bluephagenesis*) by increasing the hydrophobicity of cells through the deletion of electron transferring flavoprotein (*etf* operon) (Ling et al., 2018b). The mutant cells prepared by homologous recombination-based technique had reduced surface charge and could sediment rapidly within less than 1 min after stopping the agitation in the bioreactor.

Genetic engineering-based solutions are being developed also for facilitated PHA recovery from collected biomass of halophilic bacteria. The size of PHA granules in bacterial cells is typically around hundreds of nm in diameter (Anderson and Dawes, 1990). PHA recovery processes can benefit from bigger granules that would enable faster separation from aqueous suspensions obtained after chemical or biological treatment of collected cells. Shen and co-workers engineered a *H. bluephagenesis* strain which generated PHB granules of the unique and unprecedented size of up to 10 μm (Shen et al., 2019). The enlarged granules were observed only in filamentous recombinants that combined the deletion of phasin gene *phaP1* with overexpressed *minCD*. Deletion of *phaP1* alone did not result in enlarged PHB particles. The study thus identified that the size of PHB granules in *H. bluephagenesis* is primarily controlled by the cell size. Larger PHB granules were observed also in outer membrane-defective *H. bluephagenesis* mutants (Wang et al., 2021) prepared by deleting the genes *waaC* and *lpxL* whose products take part in lipopolysaccharide synthesis. The *lpxL* gene lacking mutant also showed improved secretion of the low molecular compound ectoine. Secretion of bigger molecules such as recombinant proteins was regrettably not tested. Recently, improved secretion and surface attachment of recombinant proteins including artificial adhesins or cellulosomal binding domains was reported for surface-shaved *P. putida* KT2440 mutant with removed lipopolysaccharide layer (Dvořák et al., 2020a; Fraile et al., 2021). One can envision that enhanced secretion of proteins in more permeable engineered PHA producers, including *Halomonas* spp. and other extremophilic candidates for NGIB, will further support the attempts to develop cost-effective downstream processes via facilitated self-flocculation or selective adhesion of cells (Fraile et al., 2021; Ling et al., 2018b), controlled cellular autolysis for intracellular product release (Borrero-de Acuña et al., 2017), or direct secretion of PHA granules out of the bacterium (Rahman et al., 2013).

4.4. The perspective of the development of genetic engineering tools for thermophilic PHA producers

In contrast to the world of halophiles, the bioengineering of PHA-producing thermophilic microorganisms is practically a non-existent field at the moment. Nevertheless, considering the great potential of this group of extremophiles, the emergence of ME and SB tools for it is just a matter of time. Researchers can find inspiration in the *Halomonas* story described above as well as in the toolkit already available for numerous biotechnologically relevant thermophilic microorganisms that do not form PHA granules but can synthesize other attractive chemicals at elevated temperatures. These include extreme thermophiles from the genera *Thermus*, *Thermococcus*, *Pyrococcus*, or *Caldicellulosiruptor* and moderate thermophiles such as *Geobacillus* spp., certain species from the *Bacillaceae* family, or *Clostridium thermocellum* (Drejer et al., 2018; Kananavičiūtė and Čitavičius, 2015; Lee et al., 2020; Mazzoli and Olson, 2020; Zeldes et al., 2015). The latter Gram-positive cellulolytic bacterium with a temperature optimum of around 60 °C is attracting lots of attention of bioengineers for its use in consolidated bioprocessing and its example could be instrumental for future engineering of thermophilic PHA producers (Mazzoli and Olson, 2020). It is, however, necessary to realize the specifics of individual species as well as the specifics of thermophilic bacteria in general. However, the devil is in the details, and these can make the engineering of particular species extremely challenging. Fortunately, massive advancements in omics

technologies and data analysis achieved during the last two decades substantially reduced the time needed to understand the fundamental physiology of newly isolated strains and made the selection of the best candidates for the process of microbial domestication the matter of rational knowledge-based choice rather than a fully empirical exercise (Fig. 3). Genomic sequences (available for most bacteria listed in Table 2), single-molecule real-time sequencing, and methylome analyses can help to reveal the host's restriction-modification (RM) systems. RM systems digest foreign DNA with a methylation pattern that differs from the host's one. Hence, their removal or evasion is a key prerequisite for genetic engineering attempts (Riley and Guss, 2021; Riley et al., 2019). They were identified in almost 90% of all known prokaryotic genomes including genomes of thermophiles (Vasu and Nagaraja, 2013). The RM systems of the target host can be either deleted (if functional genome editing machinery for a given host is already available) or evaded by the use of plasmid propagated in engineered *E. coli* strain with implanted restriction-associated methyltransferases that secure a methylation pattern acceptable for the target host (Riley et al., 2019).

Mitigation of unwanted restriction enables the next step - recombinant DNA transfer into the host cell. Both native and exogenous plasmids can be employed for genetic engineering purposes. Unfortunately, very little is known about endogenous plasmids in thermophilic PHA producers and, to the best of our knowledge, no molecules utilizable in ME have been identified yet. There is plenty of space for further research in this area. Meanwhile, popular broad-host-range plasmids such as pBC1 (medium-to-high copy replicon for G+ hosts), pBBR1 (medium-to-high copy replicon G- hosts) and pRK2 (low-copy replicon G- hosts), or SEVA vectors bearing pBBR1, RK2, or pBC1 origins of replication and suitable selection markers can be tested as backbones for heterologous gene expression in PHA-forming thermophiles (De Rossi et al., 1991; Drejer et al., 2018; Silva-Rocha et al., 2013). Available plasmids for thermophilic *Bacilli* and *Geobacilli* have been recently summarized in reviews of Drejer et al. (2018) and Kananavičiūtė and Čitavičius (2015) respectively. These shuttle vectors that are convenient for work in *E. coli* and target thermophiles often contain two replicons and two markers because equal functionality of the same component in two hosts with different temperature optima is rare. Thermophiles certainly have specific demands for selection markers. Only a few antibiotics (e.g., kanamycin, bleomycin, hygromycin, chloramphenicol, or simvastatin) are applicable for longer time intervals at temperatures above 50 °C (Zeldes et al., 2015). Also, the stability of products of resistance-conferring genes must be considered. Mutant variants of kanamycin resistance marker with improved thermal stability at temperatures as high as 70 °C were prepared by directed evolution (Hoseki et al., 1999). Nutritional selection (such as uracil or tryptophan prototrophy) is a possible alternative to the use of antibiotics in thermophile cultures, but this approach requires construction or selection of auxotrophic strain deficient in an essential nutrient gene (Tripathi et al., 2010; Tripathi et al., 2010). Concerning DNA transfer techniques, electroporation of circular plasmids or simple mixing of naturally competent cells with target linear or circular DNA are favorable for thermophilic bacteria (Olson and Lynd, 2012). Protoplast transformation and conjugation with *E. coli* were reported for thermophilic *Geobacilli* (Kananavičiūtė and Čitavičius, 2015).

Advanced genetic modifications in thermophilic PHA producers will require the establishment of a reliable gene expression toolbox including standardized constitutive and inducible promoters, terminators, libraries of RBS, and thermostable reporters. RBS sites, terminators, and constitutive promoters of essential genes can be relatively easily mined from thermophile genomes with the help of diverse available software packages or using RNAseq of cells grown under specific conditions (Drejer et al., 2018; Olson et al., 2015). Adopting well-characterized constitutive and inducible promoters from models in non-model microorganisms can be a challenge (Zeldes et al., 2015; Zhao et al., 2017). Instead, researchers may develop inducible expression types of machinery that would respond to the drop of temperature or to (non)native substrate molecules (Mearls et al., 2015; Williams-Rhaesa et al., 2018; Zheng et al., 2019). Attractive alternatives for tunable gene expression in thermophiles are ligand-dependent mRNA leader sequences known as „riboswitches". Some natural and synthetic thermostable riboswitches have been recently proven well functional for gene up- and down-regulation in *Geobacillus thermoglucosidasius* and *C. thermocellum* at 55 °C (Marcano-Velazquez et al., 2019). Reliable molecular reporters for thermophiles are rare but they can be prepared by stabilization of mesophilic molecules via mutagenesis. New superfolder GFP variants with almost 900-fold enhanced fluorescence at temperatures as high as 60 °C were recently prepared by random mutagenesis coupled to fluorescence-activated cell sorting (Frenzel et al., 2018).

Reported genome editing techniques for thermophiles are mostly based on homologous recombination (Drejer et al., 2018; Zeldes et al., 2015). One of the first gene deletion protocols for *C. thermocellum* took advantage of (i) prepared Δ*pyrF* (orotidine 5′-phosphate decarboxylase) auxotrophic strain, (ii) a replicating plasmid with a sequence homologous to upstream and downstream regions (usually 500–1000 bp long) of the target gene in the chromosome, and (iii) *pyrF* gene used as a dual selection marker (Tripathi et al., 2010). Recombination-mediated genetic engineering methods („recombineering") can increase the efficiency of homologous recombination even with homologous sequences as short as 30 bp (Muyrers, 1999). However, the popular Red complex from bacteriophage l (including Exo, Beta, and Gam recombination proteins) which is used for recombineering in *E. coli* often does not function in non-model bacteria. Thermostable homologs to l Red proteins Exo/Beta were recently isolated from *Acidithiobacillus caldus* (Walker et al., 2020). These new recombinases helped to improve the performance of CRISPR/Cas genome editing systems for *C. thermocellum* firstly described in the same study. Endogenous Type I–B CRISPR system from *C. thermocellum* and exogenous Type II CRISPR system from *Geobacillus stearothermophilus* combined with thermophilic Cas9 variant from *G. stearothermophilus* and stable recombinases enabled 70% and 94% genome editing efficiency, respectively. The two-step CRISPR/Cas-recombineering method reduced the time needed for genome editing in *C. thermocellum* from four to two weeks. *Geobacillus thermodenitrificans* T12 became a source of another thermophilic Cas9 nuclease whose engineered inactive variant was employed for CRISPRi (CRISPR interference) gene silencing in *Bacillus smithii* and *C. thermocellum* (Ganguly et al., 2019; Mougiakos et al., 2017). These and other successful examples of efficient genome editing in thermophilic bacteria hold promise for accelerated development of new advanced ME and SB tools for thermophiles in near future. The vast experience collected during the last two decades for organisms such as *C. thermocellum* will help to reduce the time needed for genetic domestication of thermophilic PHA producers in the following years.

5. Conclusions/outlook

The development of sustainable and feasible biotechnological production of PHA is a holy grail of industrial biotechnology. Reaching this goal would decrease our dependency on non-renewable resources and reduce the amount of polymer-based resistant solid waste, which is currently generated in colossal amounts by modern civilization. The extremophile-based concept of Next-Generation Industrial Biotechnology holds a promise to provide the foundation for such processes since it principally enhances robustness and reduces costs of the biotechnological process. This work demonstrated that there are numerous promising PHA producers among halophiles and thermophiles. The biotechnological production of PHA employing *H. bluephagenesis* has already reached initial industrial scale, other microorganisms such as *Haloferax mediterranei*, *H. halophila* or *S. thermodepolymerans* can be also considered being strong candidates for industrial production of PHA; nevertheless, processes that utilize these microorganisms have not reached high technology readiness levels so far. Therefore, further research and development are needed in this field. Table 3 summarizes the most important advantages associated

Table 3
Comparison of the most important features associated with PHA production employing halophiles and thermophiles as compared to mesophiles.

	Mesophiles	Halophiles	Thermophiles
Range of utilizable substrates	High	High	Moderate
Robustness of the process against contamination	Low	High	High
Influence of cultivation conditions on the solubility of substrates	None	None	Positive
Influence of cultivation conditions on the solubility of oxygen	None	Negative	Negative
Rate of the metabolic processes	Normal - depending upon the employed culture	Normal - depending upon the employed culture	High
Energy demands related to the sterility of the process	High	Low	Low
Energy demands related to cultivation	Standard	Standard	Questionable
Down-stream processing	Standard	Hypotonic lysis of the cells can be employed	Standard
Availability of the tools for synthetic biology and metabolic engineering for improved PHA production	High	Medium	Low

with the employment of halophiles and thermophiles for PHA production as compared to mesophilic bacteria.

Furthermore, extremophiles can be considered not only as biotechnological chassis for direct PHA production but also very interesting sources of genes and enzymes for both *in-vivo* and *in-vitro* PHA synthesis as was demonstrated for PHA synthases from thermophilic bacteria revealing superior stability and activity (Tajima et al., 2016). Further, PHA synthases present especially in thermophilic bacteria reveal extraordinary substrate specificity resulting in the formation of co-polymers with unique monomer compositions and, therefore, also interesting mechanical and technological properties (Choonut et al., 2020b; Pantazaki et al., 2003; Sedlacek et al., 2020; Xi et al., 2016).

The biotechnological potential of extremophiles for commercial PHA production can be further enhanced by employing approaches of metabolic engineering. Despite substantial progress during the last decade, metabolic engineering and synthetic biology of halophilic PHA producers is still in its adolescence, or childhood when other species than *H. bluephagenesis* are considered. Bioengineering of PHA-forming thermophiles was not yet born, but this review clearly demonstrates that the new baby is awaited with great expectations. We summarized here the milestones that the genetic engineering-driven domestication of attractive thermophilic PHA producers must pass to get at least on the level achieved with halophiles: starting from the identification and evasion of restriction-modification systems, through the development or collection of basic genetic engineering tools and components, to the adoption of advanced metabolic engineering techniques and protocols (Fig. 3). The most recent trends in the engineering of industrially relevant microorganisms highlight interdisciplinary approaches that combine protein engineering and metabolic engineering with synthetic and systems biology for the knowledge-driven design and construction of high-performance strains (Choi et al., 2019; Dvořák et al., 2017). If the field of the bioengineering of extremophilic PHA producers aims to reach its full maturity, it should soon go in the same direction. Mutant variants of halophilic and thermophilic PhaC and other key enzymes in PHA biosynthesis machinery should be prepared and tested for altered activity, specificity, stability, or selectivity that can have an immense effect on the quantity and quality of the product (Zheng et al., 2020). Available genomic sequences and other omics data should be applied for the preparation of high-resolution genome-scale metabolic models that will enable theoretical and experimental flux analyses and targeted re-direction of carbon fluxes in the cells. The obtained strains can be further finetuned by adaptive laboratory evolution combined with high-throughput screening or selection protocols based on the synthetic genetic circuits implanted in the evolved cells.

One more prerequisite and challenge for the bloom of bioengineering of extremophilic PHA producers is the popularization of bioplastics made in genetically modified organisms (GMOs). Some companies, especially in Europe, still refuse to release on the market biopolymers made in genetically modified organisms due to the alleged negative public perception of GMOs. It is quite a paradoxical position in the world in which, for instance, hundreds of millions of people with *diabetes mellitus* are cured with recombinant insulin produced in *E. coli* and *Saccharomyces cerevisiae* cell factories (Baeshen et al., 2014). We believe that the pros of the above-mentioned genetic engineering-based technologies for the biomanufacturing of PHA by far outweigh the potential risks linked to the use of GMOs in closed fermentation systems. It is thus important that the interested researchers dedicate part of their time to the activities that improve public awareness of GMOs and their benefits for NGIB.

Funding

This work was supported by the projects GA19-20697S and GA22-12505S of the Czech Science Foundation (GACR).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Ajikumar, P., Xiao, W., Tyo, K., Wang, Y., Simeon, F., Leonard, E., Mucha, O., Phon, T., Pfeifer, B., Stephanopoulos, G., 2010. Isoprenoid pathway optimization for taxol precursor overproduction in Escherichia coli. Science 330, 70–74. https://doi.org/10.1126/science.1191652.

Ali, I., Jamil, N., 2016. Polyhydroxyalkanoates: current applications in the medical field. Front. Biol. 11, 19–27. https://doi.org/10.1007/s11515-016-1389-z.

Alsafadi, D., Al-Mashaqbeh, O., 2017. A one-stage cultivation process for the production of poly-3-(hydroxybutyrate-co-hydroxyvalerate) from olive mill wastewater by Haloferax mediterranei. New Biotechnol. 34, 47–53. https://doi.org/10.1016/j.nbt.2016.05.003.

Alves, L., Plucani do Amaral, F., Kim, D., Todo Bom, M., Piñero Gavidia, M., Silvano Teixeira, C., Holthman, F., de Oliveira Pedrosa, F., Maltempi de Souza, E., Chubatsu, L., Müller-Santos, M., Stacey, G., Müller, V., 2019. Importance of poly-3-hydroxybutyrate metabolism to the ability of *Herbaspirillum seropedicae* to promote plant growth. Appl. Environ. Microbiol. 85 https://doi.org/10.1128/AEM.02586-18.

Alves, L., Santana-Filho, A., Sassaki, G., de Oliveira Pedrosa, F., Maltempi de Souza, E., Chubatsu, L., Müller-Santos, M., Stabb, E., 2020. 3-hydroxybutyrate derived from poly-3-hydroxybutyrate mobilization alleviates protein aggregation in heat-stressed *Herbaspirillum seropedicae* SmR1. Appl. Environ. Microbiol. 86 https://doi.org/10.1128/AEM.01265-20.

Anastatz, V., Zinn, M., 2020. Syngas as a sustainable carbon source for PHA production. In: The Handbook of Polyhydroxyalkanoates. CRC Press, Boca Raton, pp. 377–416.

Anderson, A., Dawes, E., 1990. Occurrence, metabolism, metabolic role, and industrial uses of bacterial polyhydroxyalkanoates. Microbiol. Rev. 54, 450–472. https://doi.org/10.1128/mr.54.4.450-472.1990.

Argundoña, M., Vargas, C., Reina-Bueno, M., Rodriguez-Moya, J., Salvador, M., Nieto, J., 2012. An extended suite of genetic tools for use in bacteria of the halomonadaceae: an overview. In: Lorence, A. (Ed.), Recombinant Gene Expression, Methods in Molecular Biology. Humana Press, Totowa, NJ, pp. 167–201. https://doi.org/10.1007/978-1-61779-433-9_9.

Baeshen, N., Baeshen, M., Sheikh, A., Bora, R., Ahmed, M., Ramadan, H., Saini, K., Redwan, E., 2014. Cell factories for insulin production. Microbial Cell Factor. 13 https://doi.org/10.1186/s12934-014-0141-0.

Baliga, N.S., Bonneau, R., Facciotti, M.T., Pan, M., Glusman, G., Deutsch, E.W., Shannon, P., Chiu, Y., Weng, R.S., Gan, R.R., Hung, P., Date, S.V., Marcotte, E.,

Ng, W.V., 2004. Genome sequence of Haloarcula marismortui: a halophilic archaeon from the Dead Sea. [published correction appears in Genome Res. 2004 Dec;14(12): 2510]. Genome Res. 14, 2221–2234. https://doi.org/10.1101/gr.2700304.

Batista, M., Teixeira, C., Sfeir, M., Alves, L., Valdameri, G., Pedrosa, F., Sassaki, G., Steffens, M., de Souza, E., Dixon, R., Müller-Santos, M., 2018. PHB Biosynthesis Counteracts Redox Stress in Herbaspirillum seropedicae. Front. Microbiol. 9 https://doi.org/10.3389/fmicb.2018.00472.

Bhattacharyya, A., Saha, J., Haldar, S., Bhowmic, A., Mukhopadhyay, U., Mukherjee, J., 2014. Production of poly-3-(hydroxybutyrate-co-hydroxyvalerate) by Haloferax mediterranei using rice-based ethanol stillage with simultaneous recovery and re-use of medium salts. Extremophiles 18, 463–470. https://doi.org/10.1007/s00792-013-0622-9.

Bioplastics market data, 2020. EuropeanBioplastics. https://www.european-bioplastics.org/market/ (accessed 8.20.2021).

Bolhuis, H., Palm, P., Wende, A., Falb, M., Rampp, M., Rodriguez-Valera, F., Pfeiffer, F., Oesterhelt, D., 2006. The genome of the square archaeon Haloquadratum walsbyi: life at the limits of water activity. BMC Genomics 7, 1–12. https://doi.org/10.1186/1471-2164-7-169.

Bonete, M., Pire, C., Llorca, F., Camacho, M., 1996. Glucose dehydrogenase from the halophilic Archaeon Haloferax mediterranei: enzyme purification, characterisation and N-terminal sequence. FEBS Lett. 383, 227–229. https://doi.org/10.1016/0014-5793(96)00235-9.

Borrero-de Acuña, J., Hidalgo-Dumont, C., Pacheco, N., Cabrera, A., Poblete-Castro, I., 2017. A novel programmable lysozyme-based lysis system in Pseudomonas putida for biopolymer production. Sci. Rep. 7, 4373. https://doi.org/10.1038/s41598-017-04741-2.

Braunegg, G., Bona, R., Koller, M., 2004. Sustainable polymer production. Polym.-Plast. Technol. Eng. 43, 1779–1793. https://doi.org/10.1081/PPT-200040130.

Britton, K., Baker, P., Fisher, M., Ruzheinikov, S., Gilmour, D., Bonete, M., Ferrer, J., Pire, C., Esclapez, J., Rice, D., 2006. Analysis of protein solvent interactions in glucose dehydrogenase from the extreme halophile Haloferax mediterranei. Proc. Natl. Acad. Sci. 103, 4846–4851. https://doi.org/10.1073/pnas.0508854103.

Budde, C., Riedel, S., Willis, L., Rha, C., Sinskey, A., 2011. Production of poly(3-hydroxybutyrate- co −3-hydroxyhexanoate) from plant oil by engineered ralstonia eutropha strains. Appl. Environ. Microbiol. 77, 2847–2854. https://doi.org/10.1128/AEM.02429-10.

Castro, J., Nouioui, I., Asenjo, J., Andrews, B., Bull, A., Goodfellow, M., 2019. New genus-specific primers for PCR identification of Rubrobacter strains. Antonie Van Leeuwenhoek 112, 1863–1874. https://doi.org/10.1007/s10482-019-01314-3.

Chen, G.-Q., 2009. A microbial polyhydroxyalkanoates (PHA) based bio- and materials industry. Chem. Soc. Rev. 38, 2434–2446. https://doi.org/10.1039/b812677c.

Chen, G.-Q., 2010. Introduction of bacterial plastics PHA, PLA, PBS, PE, PTT, and PPP. In: Plastics From Bacteria: Natural Functions And Applications. Springer, Berlin, Heidelberg, pp. 1–16.

Chen, G., Jiang, X., 2018. Next generation industrial biotechnology based on extremophilic bacteria. Curr. Opin. Biotechnol. 50, 94–100. https://doi.org/10.1016/j.copbio.2017.11.016.

Chen, C., Don, T., Yen, H., 2006a. Enzymatic extruded starch as a carbon source for the production of poly(3-hydroxybutyrate-co-3-hydroxyvalerate) by Haloferax mediterranei. Process Biochem. 41, 2289–2296. https://doi.org/10.1016/j.procbio.2006.05.026.

Chen, T., Chou, Y., Chen, W., Arun, B., Young, C., 2006b. Tepidimonas taiwanensis sp. nov., a novel alkaline-protease-producing bacterium isolated from a hot spring. Extremophiles 10, 35–40. https://doi.org/10.1007/s00792-005-0469-9.

Chen, C., Hsu, S., Lin, M., Hsu, Y., 2015. Mass production of C50 carotenoids by Haloferax mediterranei in using extruded rice bran and starch under optimal conductivity of brined medium. Bioprocess Biosyst. Eng. 38, 2361–2367. https://doi.org/10.1007/s00449-015-1471-y.

Chen, X., Yin, J., Ye, J., Zhang, H., Che, X., Ma, Y., Li, M., Wu, L., Chen, G., 2017. Engineering Halomonas bluephagensis TD01 for non-sterile production of poly(3-hydroxybutyrate-co-4-hydroxybutyrate). Bioresour. Technol. 244, 534–541. https://doi.org/10.1016/j.biortech.2017.07.149.

Chen, Y., Chen, X., Du, H., Zhang, X., Ma, Y., Chen, J., Ye, J., Jiang, X., Chen, G., 2019. Chromosome engineering of the TCA cycle in Halomonas bluephagensis for production of copolymers of 3-hydroxybutyrate and 3-hydroxyvalerate (PHBHV). Metab. Eng. 54, 69–82. https://doi.org/10.1016/j.ymben.2019.03.006.

Choi, J., Lee, S., Han, K., 1998. Cloning of the Alcaligenes latus polyhydroxyalkanoate biosynthesis genes and use of these genes for enhanced production of poly(3-hydroxybutyrate) in Escherichia coli. Appl. Environ. Microbiol. 64, 4897–4903. https://doi.org/10.1128/AEM.64.12.4897-4903.1998.

Choi, K., Jang, W., Yang, D., Cho, J., Park, D., Lee, S., 2019. Systems metabolic engineering strategies: integrating systems and synthetic biology with metabolic engineering. Trends Biotechnol. 37, 817–837. https://doi.org/10.1016/j.tibtech.2019.01.003.

Choonut, A., Prasertsan, P., Klomklao, S., Sangkharak, K., 2020a. Bacillus thermoamylovorans-related strain isolated from high temperature sites as potential producers of medium-chain-length polyhydroxyalkanoate (mcl-PHA). Curr. Microbiol. 77, 3044–3056. https://doi.org/10.1007/s00284-020-02118-9.

Choonut, A., Prasertsan, P., Klomklao, S., Sangkharak, K., 2020b. Study on mcl-PHA production by novel thermotolerant gram-positive isolate. J. Polym. Environ. 28, 2410–2421. https://doi.org/10.1007/s10924-020-01779-8.

Chou, Y., Shen, S., Sheu, D., Wang, J., Chen, W., 2006. Schlegelella aquatica sp. nov., a novel thermophilic bacterium isolated from a hot spring. Int. J. Syst. Evol. Microbiol. 56, 2793–2797. https://doi.org/10.1099/ijs.0.64446-0.

Cristea, A., Fiaure, C., Borodi, G., Leopold, N., Kasco, I., Cadar, O., Bancis, H., 2017. Physical and chemical detection of polyhydroxybutyrate from Halomonas Elongata strain 2FFF. New front. Chem 26. S4_OP3.

Crosby, J., Laemthong, T., Lewis, A., Straub, C., Adams, M., Kelly, R., 2019. Extreme thermophiles as emerging metabolic engineering platforms. Curr. Opin. Biotechnol. 59, 55–64. https://doi.org/10.1016/j.copbio.2019.02.006.

Cui, B., Huang, S., Xu, F., Zhang, R., Zhang, Y., 2015. Improved productivity of poly (3-hydroxybutyrate) (PHB) in thermophilic Chelatococcus daeguensis TAD1 using glycerol as the growth substrate in a fed-batch culture. Appl. Microbiol. Biotechnol. 99, 6009–6019. https://doi.org/10.1007/s00253-015-6489-1.

Cui, Y.W., Gong, X., Shi, Y., Wang, Z., 2017a. Salinity effect on production of PHA and EPS by Haloferax mediterranei. RSC Adv. 7, 53587–53595. https://doi.org/10.1039/C7RA09652F.

Cui, Y.-W., Zhang, H.-Y., Ji, S.-Y., Wang, Z.-W., 2017b. Kinetic analysis of the temperature effect on polyhydroxyalkanoate production by haloferax mediterranei in synthetic molasses wastewater. J. Polym. Environ. 25, 277–285. https://doi.org/10.1007/s10924-016-0807-2.

Cummings, S., Gilmour, D., 1995. The effect of NaCl on the growth of a Halomonas species: accumulation and utilization of compatible solutes. Microbiology 141, 1413–1418. https://doi.org/10.1099/13500872-141-6-1413.

Danson, M., Hough, D., 1998. Structure, function and stability of enzymes from the Archaea. Trends Microbiol. 6, 307–314. https://doi.org/10.1016/S0966-842X(98)01316-X.

Das, P., Behera, B., Meena, D., Azmi, S., Chatterjee, S., Meena, K., Sharma, A., 2015. Salt stress tolerant genes in halophilic and halotolerant bacteria: paradigm for salt stress adaptation and osmoprotection. Int. J. Curr. Microbiol. App. Sci. 4, 642–658.

de Almeida, A., Catone, M., Rhodius, V., Gross, C., Pettinari, M., 2011. Unexpected stress-reducing effect of PhaP, a Poly(3-Hydroxybutyrate) granule-associated protein, in Escherichia coli. Appl. Environ. Microbiol. 77, 6622–6629. https://doi.org/10.1128/AEM.05469-11.

De Rossi, E., Brigidi, P., Rossi, M., Matteuzzi, D., Riccardi, G., 1991. Characterization of Gram-positive broad host-range plasmids carrying a thermophilic replicon. Res. Microbiol. 142, 389–396. https://doi.org/10.1016/0923-2508(91)90108-M.

Derner, E., Kolari, M., Hoornstra, D., Tsitko, I., Kämpfer, P., Busse, H., Salkinoja-Salonen, M., 2006. Rubellimicrobium thermophilum gen. nov., sp. nov., a red-pigmented, moderately thermophilic bacterium isolated from coloured slime deposits in paper machines. Int. J. Syst. Evol. Microbiol. 56, 1355–1362. https://doi.org/10.1099/ijs.0.63751-0.

Deole, R., Heff, W., 2020. A potassium chloride to glycine betaine osmoprotectant switch in the extreme halophile Halorhodospira halophila. Sci. Rep. 10, 1–10. https://doi.org/10.1038/s41598-020-59231-9.

Di Caprio, F., 2021. Cultivation processes to select microorganisms with high accumulation ability. Biotechnol. Adv. 49 https://doi.org/10.1016/j.biotechadv.2021.107740.

Di Donato, P., Fiorentino, G., Anzelmo, G., Tommonaro, G., Nicolaus, B., Poli, A., 2011. Re-use of vegetable wastes as cheap substrates for extremophile biomass production. Waste Biomass Valorization 2, 103–111. https://doi.org/10.1007/s12649-011-9062-x.

Dietrich, K., Dumont, M., Del Rio, L., Orsat, V., 2019. Sustainable PHA production in integrated lignocellulose biorefineries. New Biotechnol. 49, 161–168. https://doi.org/10.1016/j.nbt.2018.11.004.

Dinnbier, U., Limpinsel, E., Schmid, R., Bakker, E., 1988. Transient accumulation of potassium glutamate and its replacement by trehalose during adaptation of growing cells of Escherichia coli K-12 to elevated sodium chloride concentrations. Arch. Microbiol. 150, 348–357. https://doi.org/10.1007/BF00408306.

Drejer, E., Hakvåg, S., Irla, M., Brautaset, T., 2018. Genetic tools and techniques for recombinant expression in Thermophilic Bacillaceae. Microorganisms 6, 42. https://doi.org/10.3390/microorganisms6020042.

D'Souza, S., Altekar, W., D'Souza, S., 1997. Adaptive response of *Haloferax mediterranei* to low concentrations of NaCl (20%) in the growth medium. Arch. Microbiol. 168, 68–71. https://doi.org/10.1007/s002030050471.

Dvorak, P., Chrast, L., Nikel, P., Fedr, R., Soucek, K., Sedlackova, M., Chaloupkova, R., de Lorenzo, V., Prokop, Z., Damborsky, J., 2015. Exacerbation of substrate toxicity by IPTG in Escherichia coli BL21(DE3) carrying a synthetic metabolic pathway. Microb. Cell Factories 14. https://doi.org/10.1186/s12934-015-0393-3.

Dvořák, P., Nikel, P., Damborský, J., de Lorenzo, V., 2017. Bioremediation 3. 0: Engineering pollutant-removing bacteria in the times of systemic biology. Biotechnol. Adv. 35, 845–866. https://doi.org/10.1016/j.biotechadv.2017.08.001.

Dvořák, P., Bayer, E., de Lorenzo, V., 2020a. Surface display of designer protein scaffolds on genome-reduced strains of Pseudomonas putida. ACS Synth. Biol. 9, 2749–2764. https://doi.org/10.1021/acssynbio.0c00276.

Dvořák, P., Kovář, J., Lorenzo, V., 2020b. Biotransformation of d -xylose to d -xylonate coupled to medium-chain-length polyhydroxyalkanoate production in cellobiose-grown Pseudomonas putida EM42. Microb. Biotechnol. 13, 1273–1283. https://doi.org/10.1111/1751-7915.13574.

Elbanna, K., Lütke-Eversloh, T., Van Trappen, S., Mergaert, J., Swings, J., Steinbüchel, A., 2003. Schlegelella thermodepolymerans gen. nov., sp. nov., a novel thermophilic bacterium that degrades poly(3-hydroxybutyrate-co-3-mercaptopropionate). Int. J. Syst. Evol. Microbiol. 53, 1165–1168. https://doi.org/10.1099/ijs.0.02562-0.

Elbanna, K., Lütke-Eversloh, T., Jendrossek, D., Luftmann, H., Steinbüchel, A., 2004. Studies on the biodegradability of polythioester copolymers and homopolymers by polyhydroxyalkanoate (PHA)-degrading bacteria and PHA depolymerases. Arch. Microbiol. 182, 212–225. https://doi.org/10.1007/s00203-004-0715-z.

El-malek, F., Farag, A., Omar, S., Khairy, H., 2020. Polyhydroxyalkanoates (PHA) from Halomonas pacifica ASL10 and Halomonas salifodiane ASL11 isolated from Mariout

salt lakes. Int. J. Biol. Macromol. 161, 1318–1328. https://doi.org/10.1016/j.ijbiomac.2020.07.258.

Fang, C., Ku, K., Lee, M., Su, N., 2010. Influence of nutritive factors on C50 carotenoids production by Haloferax mediterranei ATCC 33500 with two-stage cultivation. Bioresour. Technol. 101, 6487–6493. https://doi.org/10.1016/j.biortech.2010.03.044.

Fathima, S.N., Krishnaswamy, V., Shen, N., 2017. Isolation and screening of phb producing halotolerant bacterial strains from a saline environment. J. Biotechnol. Biomed. Sci. 1, 34–45. https://doi.org/10.14302/issn.2576-6694.jbbs-17-1783.

Fraile, S., Briones, M., Revenga-Parra, M., de Lorenzo, V., Lorenzo, E., Martínez-García, E., 2021. Engineering tropism of pseudomonas putida toward target surfaces through ectopic display of recombinant nanobodies. ACS Synth. Biol. 10, 2049–2059. https://doi.org/10.1021/acssynbio.1c00227.

Frenzel, E., Legebeke, J., van Stralen, A., van Kranenburg, R., Kuipers, O., 2018. In vivo selection of sfGFP variants with improved and reliable functionality in industrially important thermophilic bacteria. Biotechnol. Biofuels 11. https://doi.org/10.1186/s13068-017-1008-5.

Fu, X., Tan, D., Aibaidula, G., Wu, Q., Chen, J., Chen, G., 2014. Development of Halomonas TD01 as a host for open production of chemicals. Metab. Eng. 23, 78–91. https://doi.org/10.1016/j.ymben.2014.02.006.

Ganguly, J., Martin-Pascual, M., Kranenburg, R., 2019. CRISPR interference (CRISPRi) as transcriptional repression tool for Hungateiclostridium thermocellum DSM 1313. Microb. Biotechnol. 13, 339–349. https://doi.org/10.1111/1751-7915.13516.

Gassler, T., Sauer, M., Gasser, B., Egermeier, M., Troyer, C., Causon, T., Hann, S., Mattanovich, D., Steiger, M., 2020. The industrial yeast Pichia pastoris is converted from a heterotroph into an autotroph capable of growth on CO2. Nat. Biotechnol. 38, 210–216. https://doi.org/10.1038/s41587-019-0363-0.

Gedikli, S., Çelik, P., Demirbilek, M., Mutlu, M., Denkbaş, E., Çabuk, A., 2019. Experimental exploration of thermostable poly (β-Hydroxybutyrates) by Geobacillus kaustophilus using Box-Behnken design. J. Polym. Environ. 27, 245–255. https://doi.org/10.1007/s10924-018-1335-z.

Giani, M., Martínez-Espinosa, R., 2020. Carotenoids as a protection mechanism against oxidative stress in haloferax mediterranei. Antioxidants 9, 1060. https://doi.org/10.3390/antiox9111060.

Giani, M., Garbayo, I., Vílchez, C., Martínez-Espinosa, R., 2019. Haloarchaeal carotenoids: healthy novel compounds from extreme environments. Marine Drugs 17, 524. https://doi.org/10.3390/md17090524.

Ginzburg, M., Sachs, L., Ginzburg, B., 1970. Ion metabolism in a Halobacterium. J. Gen. Physiol. 55, 187–207. https://doi.org/10.1085/jgp.55.2.187.

Gleizer, S., Ben-Nissan, R., Bar-On, Y., Antonovsky, N., Noor, E., Zohar, Y., Jona, G., Krieger, E., Shamshoum, M., Bar-Even, A., Milo, R., 2019. Conversion of Escherichia coli to generate all biomass carbon from CO2. Cell 179. https://doi.org/10.1016/j.cell.2019.11.009, 1255-1263.e12.

Gonçalves, P., Hume, C., Ferreira, A., Tsui, S., Brocchi, M., Wren, B., Araujo, W., 2019. Environmental interactions are regulated by temperature in Burkholderia seminalis TC3.4.2R3. Sci. Rep. 9 https://doi.org/10.1038/s41598-019-41778-x.

Grillo, R., do Pereira, A.E.S., de Melo, N.F.S., Porto, R.M., Feitosa, L.O., Tonello, P.S., Filho, N.L.D., Rosa, A.H., Lima, R., Fraceto, L.F., 2011. Controlled release system for ametryn using polymer microspheres: preparation, characterization and release kinetics in water. J. Hazard. Mater. 186, 1645–1651. https://doi.org/10.1016/j.jhazmat.2010.12.044.

Guzmán, H., Van-Thuoc, D., Martín, J., Hatti-Kaul, R., Quillaguamán, J., 2009. A process for the production of ectoine and poly(3-hydroxybutyrate) by Halomonas boliviensis. Appl. Microbiol. Biotechnol. 84, 1069–1077. https://doi.org/10.1007/s00253-009-2036-2.

Han, J., Hou, J., Zhang, F., Ai, G., Li, M., Cai, S., Lin, H., Wang, L., Wang, Z., Zhang, S., Cai, L., Zhao, D., Zhou, J., Xiang, H., 2013. Multiple propionyl coenzyme A-supplying pathways for production of the bioplastic poly(3-hydroxybutyrate- co −3-hydroxyvalerate) in Haloferax mediterranei. Appl. Environ. Microbiol. 79, 2922–2931. https://doi.org/10.1128/AEM.03915-12.

Haque, R., Paradisi, F., Allers, T., 2020. Haloferax volcanii for biotechnology applications: challenges, current state and perspectives. Appl. Microbiol. Biotechnol. 104, 1371–1382. https://doi.org/10.1007/s00253-019-10314-2.

Hermann-Krauss, C., Koller, M., Muhr, A., Fasl, H., Stelzer, F., Braunegg, G., 2013. Archaeal production of polyhydroxyalkanoate (PHA) co- and terpolyesters from biodiesel industry-derived by-products. Archaea 2013, 1–10. https://doi.org/10.1155/2013/129268.

Henayen, F., Relm, B., Eberhardt, R., Steinbüchel, A., 2000. Polymer production by two newly isolated extremely halophilic archaea: application of a novel corrosion-resistant bioreactor. Appl. Microbiol. Biotechnol. 54, 319–325. https://doi.org/10.1007/s002530000394.

Hickey, D., Singer, G., 2004. Genomic and proteomic adaptations to growth at high temperature. Genome Biol. 5, 117. https://doi.org/10.1186/gb-2004-5-10-117.

Hong, J., Song, H., Moon, Y., Hong, Y., Bhatia, S., Jung, H., Choi, T., Yang, S., Park, H., Choi, Y., Yang, Y., 2019. Polyhydroxybutyrate production in halophilic marine bacteria Vibrio proteolyticus isolated from the Korean peninsula. Bioprocess Biosyst. Eng. 42, 603–610. https://doi.org/10.1007/s00449-018-02066-6.

Hoseki, J., Yano, T., Koyama, Y., Kuramitsu, S., Kagamiyama, H., 1999. Directed evolution of thermostable kanamycin-resistance gene: a convenient selection marker for Thermus thermophilus. J. Biochem. 126, 951–956. https://doi.org/10.1093/oxfordjournals.jbchem.a022539.

Hsiao, L., Lin, J., Sankatumvong, P., Wu, T., Li, S., 2016. The feasibility of thermophilic caldimonas manganoxidans as a platform for efficient PHB production. Appl. Biochem. Biotechnol. 180, 852–871. https://doi.org/10.1007/s12010-016-2138-0.

Hsiao, L., Lee, M., Chuang, P., Kuo, Y., Lin, J., Wu, T., Li, S., 2018. The production of poly (3-hydroxybutyrate) by thermophilic Caldimonas manganoxidans from glycerol. J. Polym. Res. 25, 85. https://doi.org/10.1007/s10965-018-1486-6.

Hussein, A., Lisowska, B., Leak, D., 2015. The genus geobacillus and their biotechnological potential. Adv. Appl. Microbiol. 94, 1–48. https://doi.org/10.1016/bs.aambs.2015.03.001.

Ibrahim, M., Steinbüchel, A., 2010. High-cell-density cyclic fed-batch fermentation of a poly(3-hydroxybutyrate)-accumulating thermophile, Chelatococcus sp. Strain MW10. Appl. Environ. Microbiol. 76, 7890–7895. https://doi.org/10.1128/AEM.01488-10.

Ibrahim, M., Willems, A., Steinbüchel, A., 2010. Isolation and characterization of new poly(3HB)-accumulating star-shaped cell-aggregates-forming thermophilic bacteria. J. Appl. Microbiol. 5 https://doi.org/10.1111/j.1365-2672.2010.04786.x no-no.

Ibrahim, M., Lebbe, L., Willems, A., Steinbüchel, A., 2016. Chelatococcus thermostellatus sp. nov., a new thermophile for bioplastic synthesis: comparative phylogenetic and physiological study. AMB Express 6, 39. https://doi.org/10.1186/s13568-016-0209-9.

Israni, N., Shivakumar, S., 2019. Polyhydroxyalkanoates in Packaging, in: Biotechnological Applications Of Polyhydroxyalkanoates. Springer Singapore, Singapore, pp. 363–388. https://doi.org/10.1007/978-981-13-3759-8_14.

Jau, M., Yew, S., Toh, P., Chong, A., Chu, W., Phang, S., Najimudin, N., Sudesh, K., 2005. Biosynthesis and mobilization of poly(3-hydroxybutyrate) [P(3HB)] by Spirulina platensis. Int. J. Biol. Macromol. 36, 144–151. https://doi.org/10.1016/j.ijbiomac.2005.05.002.

Jendrossek, D., 2009. Polyhydroxyalkanoate granules are complex subcellular organelles (carbonosomes). J. Bacteriol. 191, 3195–3202. https://doi.org/10.1128/JB.01723-08.

Jeschek, M., Gerngross, D., Panke, S., 2017. Combinatorial pathway optimization for streamlined metabolic engineering. Curr. Opin. Biotechnol. 47, 142–151. https://doi.org/10.1016/j.copbio.2017.06.014.

Jiang, X., Yao, Z., Chen, G., 2017a. Controlling cell volume for efficient PHB production by Halomonas. Metab. Eng. 44, 30–37. https://doi.org/10.1016/j.ymben.2017.09.004.

Jiang, Y., Xin, F., Lu, J., Dong, W., Zhang, W., Zhang, M., Wu, H., Ma, J., Jiang, M., 2017b. State of the art review of biofuels production from lignocellulose by thermophilic bacteria. Bioresour. Technol. 245, 1498–1506. https://doi.org/10.1016/j.biortech.2017.05.142.

Jiang, X., Yan, X., Yu, L., Liu, X., Chen, G., 2021. Hyperproduction of 3-hydroxypropionate by Halomonas bluephagenesis. Nat. Commun. 12 https://doi.org/10.1038/s41467-021-21632-3.

Kananavičiūtė, R., Čitavičius, D., 2015. Genetic engineering of Geobacillus spp. J. Microbiol. Methods 111, 31–39. https://doi.org/10.1016/j.mimet.2015.02.002.

Kaur, R., Tiwari, S., 2021. Purification and characterization of a new halocin HA4 from Haloferax larsenii HA4 isolated from a Salt Lake. Probiot. Antimicrob. Proteins 1-9. https://doi.org/10.1007/s12602-021-09823-2.

Kennedy, S., Ng, W., Salzberg, S., Hood, L., DasSarma, S., 2001. Understanding the adaptation of halobacterium species NRC-1 to its extreme environment through computational analysis of its genome sequence. Genome Res. 11, 1641–1650. https://doi.org/10.1101/gr.190201.

Khan, I., Habib, N., Asem, M., Salam, N., Xiao, M., Zhou, E., Zhi, X., Li, W., 2019. Aquabacterium tepidiphilum sp. nov., a moderately thermophilic bacterium isolated from a hot spring. Int. J. Syst. Evol. Microbiol. 69, 337–342. https://doi.org/10.1099/ijsem.0.003103.

Kim, J., Won, Y., Nikoh, N., Nakayama, H., Han, S., Kikuchi, Y., Rhee, Y., Park, H., Kwon, J., Kurokawa, K., Dohmae, N., Fukatsu, T., Lee, B., 2013. Polyester synthesis genes associated with stress resistance are involved in an insect-bacterium symbiosis. Proc. Natl. Acad. Sci. 110, E2381–E2389. https://doi.org/10.1073/pnas.1303228110.

Koller, M., 2015. Recycling of waste streams of the biotechnological poly (hydroxyalkanoate) production by Haloferax mediterranei on Whey. Int. J. Polym. Sci. 2015, 1–8. https://doi.org/10.1155/2015/370164.

Koller, M., 2017. Production of polyhydroxyalkanoate (PHA) biopolyesters by extremophiles. MOJ Polym. Sci. 1 https://doi.org/10.15406/mojps.2017.01.00011.

Koller, M., 2019a. Linking food industry to "green plastics" – polyhydroxyalkanoate (PHA) biopolyesters from agro-industrial by-products for securing food safety. Appl. Food Biotechnol. 6, 1–6. https://doi.org/10.22037/afb.v6i1.17979.

Koller, M., 2019b. Polyhydroxyalkanoate biosynthesis at the edge of water activity-haloarchaea as biopolyester factories. Bioengineering 6, 34. https://doi.org/10.3390/bioengineering6020034.

Koller, M., 2020. Established and advanced approaches for recovery of microbial polyhydroxyalkanoate (PHA) biopolyesters from surrounding microbial biomass. EuroBiotech J. 4, 113–126. https://doi.org/10.2478/ebtj-2020-0013.

Koller, M., Hesse, P., Bona, R., Kutschera, C., Atlić, A., Braunegg, G., 2007a. Biosynthesis of high quality polyhydroxyalkanoate co- and terpolyesters for potential medical application by the ArchaeonHaloferax mediterranei. Macromol. Symp. 253, 33–39. https://doi.org/10.1002/masy.200750704.

Koller, M., Hesse, P., Bona, R., Kutschera, C., Atlić, A., Braunegg, G., 2007b. Potential of various archae- and eubacterial strains as industrial polyhydroxyalkanoate producers from whey. Macromol. Biosci. 7, 218–226. https://doi.org/10.1002/mabi.200600211.

Koller, M., Sandholzer, D., Salerno, A., Braunegg, G., Narodoslawsky, M., 2013. Biopolymer from industrial residues: Life cycle assessment of poly (hydroxyalkanoates) from whey. Resour. Conserv. Recycl. 73, 64–71. https://doi.org/10.1016/j.resconrec.2013.01.017.

Koller, M., Chiellini, E., Braunegg, G., 2015. Study on the production and re-use of poly (3-hydroxybutyrate-co-3- hydroxyvalerate) and extracellular polysaccharide by the

archaeon haloferax mediterranei strain DSM 1411. Chem. Biochem. Eng. Q. 29, 87–98. https://doi.org/10.15255/CABEQ.2014.2058.

Koller, M., Maršálek, L., de Sousa Dias, M., Braunegg, G., 2017. Producing microbial polyhydroxyalkanoate (PHA) biopolyesters in a sustainable manner. New Biotechnol. 37, 24–38. https://doi.org/10.1016/j.nbt.2016.05.001.

Koller, M., Braunegg, Obruca, S., 2020. Linking salinity to microbial biopolyesters biosynthesis: polyhydroxyalkanoate production by Haloarchaea and Halophilic Eubacteria. In: Koller, M. (Ed.), The Handbook of Polyhydroxyalkanoates. Vol. 2: Kinetics, Bioengineering, and Industrial Aspects. CRC Press, Boca Raton, pp. 133–212.

Koskimäki, J., Kajula, M., Hokkanen, J., Ihantola, E., Kim, J., Hautajärvi, H., Hankala, E., Suokas, M., Pohjanen, J., Podolich, O., Kozyrovska, N., Turpeinen, A., Pääkkönen, M., Mattila, S., Campbell, B., Pirttilä, A., 2016. Methyl-esterified 3-hydroxybutyrate oligomers protect bacteria from hydroxyl radicals. Nat. Chem. Biol. 12, 332–338. https://doi.org/10.1038/nchembio.2043.

Kourilova, X., Pernicova, I., Sedlar, K., Musilova, J., Sedlacek, P., Kalina, M., Koller, M., Obruca, S., 2020. Production of polyhydroxyalkanoates (PHA) by a thermophilic strain of Schlegelella thermodepolymerans from xylose rich substrates. Bioresour. Technol. 315, 123885. https://doi.org/10.1016/j.biortech.2020.123885.

Kourilova, X., Novackova, I., Koller, M., Obruca, S., 2021a. Evaluation of mesophilic Burkholderia sacchari, thermophilic Schlegelella thermodepolymerans and halophilic Halomonas halophila for polyhydroxyalkanoates production on model media mimicking lignocellulose hydrolysates. Bioresour. Technol. 325, 124704. https://doi.org/10.1016/j.biortech.2021.124704.

Kourilova, X., Schwarzerova, J., Pernicova, I., Sedlar, K., Mrazova, K., Krzyzanek, V., Nebesarova, J., Obruca, S., 2021b. The first insight into polyhydroxyalkanoates accumulation in multi-extremophilic Rubrobacter xylanophilus and Rubrobacter spartanus. Microorganisms 9, 909. https://doi.org/10.3390/microorganisms9050909.

Kourilova, X., Pernicova, I., Vidlakova, M., Krejcirik, R., Mrazova, K., Hrubanova, K., Krzyzanek, V., Nebesarova, J., Obruca, S., 2021c. Biotechnological conversion of grape pomace to poly(3-hydroxybutyrate) by moderately thermophilic bacterium Tepidimonas ignavus. Bioengineering 8, 141. https://doi.org/10.3390/bioengineering8100141.

Kourmentza, C., Plácido, J., Venetsaneas, N., Burniol-Figols, A., Varrone, C., Gavala, H., Reis, M., 2017. Recent advances and challenges towards sustainable polyhydroxyalkanoate (PHA) production. Bioengineering 4. https://doi.org/10.3390/bioengineering4020055.

Kourmentza, K., Kachrimanidou, V., Pseki, O., Pateraki, C., Ladakis, D., Koutinas, A., 2020. Competitive advantage and market introduction of PHA polymers and potential use of PHA monomers. In: The Handbook Of Polyhydroxyalkanoates. CRC Press, Boca Raton, pp. 167–202.

Krüger, A., Schäfers, C., Schröder, C., Antranikian, G., 2018. Towards a sustainable biobased industry – highlighting the impact of extremophiles. New Biotechnol. 40, 144–153. https://doi.org/10.1016/j.nbt.2017.05.002.

Kucera, D., Pernicová, I., Kovalcik, A., Koller, M., Mullerova, L., Sedlacek, P., Mravec, F., Nebesarova, J., Kalina, M., Marova, I., Krzyzanek, V., Obruca, S., 2018. Characterization of the promising poly(3-hydroxybutyrate) producing halophilic bacterium Halomonas halophila. Bioresour. Technol. 256, 552–556. https://doi.org/10.1016/j.biortech.2018.02.062.

Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K., Battistuzzi, F., 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol. Biol. Evol. 35, 1547–1549. https://doi.org/10.1093/molbev/msy096.

Kumar, V., Kumar, S., Singh, D., 2020. Microbial polyhydroxyalkanoates from extreme niches: Bioprospection status, opportunities and challenges. Int. J. Biol. Macromol. 147, 1255–1267. https://doi.org/10.1016/j.ijbiomac.2019.09.253.

Kurumbang, N., Dvorak, P., Bendl, J., Brezovsky, J., Prokop, Z., Damborsky, J., 2014. Computer-assisted engineering of the synthetic pathway for biodegradation of a toxic persistent pollutant. ACS Synth. Biol. 3, 172–181. https://doi.org/10.1021/sb400147n.

Lakshmanan, M., Foong, C.P., Abe, H., Sudesh, K., 2019. Biosynthesis and characterization of co and ter-polyesters of polyhydroxyalkanoates containing high monomeric fractions of 4-hydroxybutyrate and 5-hydroxyvalerate via a novel PHA synthase. Polym. Degrad. Stab. 163, 122–135. https://doi.org/10.1016/j.polymdegradstab.2019.03.005.

Lavigne, R., Sun, W., Volckaert, G., 2004. PHIRE, a deterministic approach to reveal regulatory elements in bacteriophage genomes. Bioinformatics 20, 629–635. https://doi.org/10.1093/bioinformatics/btg456.

Lee, L., Crosby, J., Rubinstein, G., Laemthong, T., Bing, R., Straub, C., Adams, M., Kelly, R., 2020. The biology and biotechnology of the genus Caldicellulosiruptor: recent developments in 'Caldi World'. Extremophiles 24, 1–15. https://doi.org/10.1007/s00792-019-01116-5.

Lemoigne, M., 1926. Produits de déshydration et de polymérisation de l'acide β-oxobutyrique. BULLETIN DE LA SOCIETÉ DE CHIMIE BIOLOGIQUE 8, 770–782.

Lettner, M., Schöggl, J., Stern, T., 2017. Factors influencing the market diffusion of bio-based plastics: Results of four comparative scenario analyses. J. Clean. Prod. 157, 289–298. https://doi.org/10.1016/j.jclepro.2017.04.077.

Li, T., Ye, J., Shen, R., Zong, Y., Zhao, X., Lou, C., Chen, G., 2016a. Semirational approach for ultrahigh poly(3-hydroxybutyrate) accumulation in Escherichia coli by combining one-step library construction and high-throughput screening. ACS Synth. Biol. 5, 1308–1317. https://doi.org/10.1021/acssynbio.6b00083.

Li, T., Li, T., Ji, W., Wang, Q., Zhang, H., Chen, G., Lou, C., Ouyang, Q., 2016b. Engineering of core promoter regions enables the construction of constitutive and inducible promoters in Halomonas sp. Biotechnol. J. 11, 219–227. https://doi.org/10.1002/biot.201400828.

Lin, Y., Guan, Y., Dong, X., Ma, Y., Wang, X., Leng, Y., Wu, F., Ye, J., Chen, G., 2021. Engineering Halomonas bluephagenesis as a chassis for bioproduction from starch. Metab. Eng. 64, 134–145. https://doi.org/10.1016/j.ymben.2021.01.014.

Ling, C., Qiao, G., Shuai, B., Olavarria, K., Yin, J., Xiang, R., Song, K., Shen, Y., Guo, Y., Chen, G., 2018a. Engineering NADH/NAD+ ratio in Halomonas bluephagenesis for enhanced production of polyhydroxyalkanoates (PHA). Metab. Eng. 49, 275–286. https://doi.org/10.1016/j.ymben.2018.09.007.

Ling, C., Qiao, G., Shuai, B., Song, K., Yao, W., Jiang, X., Chen, G., 2018b. Engineering self-flocculating halomonas campaniensis for wastewaterless open and continuous fermentation. Biotechnol. Bioeng. https://doi.org/10.1002/bit.26897.

Lippert, K., Galinski, E., Trüper, H., 1993. Biosynthesis and function of trehalose inEctothiorhodospira halochloris. Antonie Van Leeuwenhoek 63, 85–91. https://doi.org/10.1007/BF00871735.

Liu, C., Wang, X., Yang, H., Liu, C., Zhang, Z., Chen, G., 2021. Biodegradable polyhydroxyalkanoates production from wheat straw by recombinant Halomonas elongata A1. Int. J. Biol. Macromol. 187, 675–682. https://doi.org/10.1016/j.ijbiomac.2021.07.137.

Lusk, B., 2019. Thermophiles; or, the modern prometheus: the importance of extreme microorganisms for understanding and applying extracellular electron transfer. Front. Microbiol. 10, 818. https://doi.org/10.3389/fmicb.2019.00818.

Lv, L., Ren, Y., Chen, J., Wu, Q., Chen, G., 2015. Application of CRISPRi for prokaryotic metabolic engineering involving multiple genes, a case study: Controllable P(3HB-co-4HB) biosynthesis. Metab. Eng. 29, 160–168. https://doi.org/10.1016/j.ymben.2015.03.013.

Ma, H., Zhao, Y., Huang, W., Zhang, L., Wu, F., Ye, J., Chen, G., 2020. Rational flux-tuning of Halomonas bluephagenesis for co-production of bioplastic PHB and ectoine. Nat. Commun. 11, 1–12. https://doi.org/10.1038/s41467-020-17223-3.

Mahansaria, R., Dhara, A., Saha, A., Haldar, S., Mukherjee, J., 2018. Production enhancement and characterization of the polyhydroxyalkanoate produced by Natrinema ajinwuensis (as synonym) = Natrinema altunense strain RM-G10. Int. J. Biol. Macromol. 107, 1480–1490. https://doi.org/10.1016/j.ijbiomac.2017.10.009.

Mahler, N., Tschirren, S., Pflügl, S., Herwig, C., 2018. Optimized bioreactor setup for scale-up studies of extreme halophilic cultures. Biochem. Eng. J. 130, 39–46. https://doi.org/10.1016/j.bej.2017.11.006.

Majara, M., O'Connor-Cox, E., Axcell, B., 2018. Trehalose—a stress protectant and stress indicator compound for yeast exposed to adverse conditions. J. Am. Soc. Brew. Chem. 54, 221–227. https://doi.org/10.1094/ASBCJ-54-0221.

Manaia, C., Nogales, B., Nunes, O., 2003. Tepidiphilus margaritifer gen. nov., sp. nov., isolated from a thermophilic aerobic digester. Int. J. Syst. Evol. Microbiol. 53, 1405–1410. https://doi.org/10.1099/ijs.0.02538-0.

Marcano-Velazquez, J., Lo, J., Nag, A., Maness, P., Chou, K., 2019. Developing riboswitch-mediated gene regulatory controls in thermophilic bacteria. ACS Synth. Biol. 8, 633–640. https://doi.org/10.1021/acssynbio.8b00487.

Masojídek, J., Torzillo, G., Kopecký, J., Koblížek, M., Nidiaci, L., Komenda, J., Lukavská, A., Sacchi, A., 2000. Changes in chlorophyll fluorescence quenching and pigment composition in the green alga Chlorococcum sp. grown under nitrogen deficiency and salinity stress. J. Appl. Phycol. 12, 417–426. https://doi.org/10.1023/A:1008165900780.

Mazzoli, R., Olson, D., 2020. Clostridium thermocellum: A microbial platform for high-value chemical production from lignocellulose. Adv. Appl. Microbiol., Adv. Appl. Microbiol. 113, 111–161. https://doi.org/10.1016/bs.aambs.2020.07.004.

Mearls, E., Olson, D., Herring, C., Lynd, L., 2015. Development of a regulatable plasmid-based gene expression system for Clostridium thermocellum. Appl. Microbiol. Biotechnol. 99, 7589–7599. https://doi.org/10.1007/s00253-015-6610-5.

Medeiros Garcia Alcântara, J., Distante, F., Storti, G., Moscatelli, D., Morbidelli, M., Sponchioni, M., 2020. Current trends in the production of biodegradable bioplastics: the case of polyhydroxyalkanoates. Biotechnol. Adv. 42 https://doi.org/10.1016/j.biotechadv.2020.107582.

Mezzina, M.P., Wettler, D.E., de Almeida, A., Dinjaski, N., Prieto, M.A., Pettinari, M.J., 2015. A phasin with extra talents: a polyhydroxyalkanoate granule-associated protein has chaperone activity. Environ. Microbiol. 17, 1765–1776. https://doi.org/10.1111/1462-2920.12636.

Mezzina, M.P., Pettinari, M.J., Müller, V., 2016. Phasins, multifaceted polyhydroxyalkanoate granule-associated proteins. Appl. Environ. Microbiol. 82, 5060–5067. https://doi.org/10.1128/AEM.01161-16.

Miyake, M., Erata, M., Asada, Y., 1996. A thermophilic cyanobacterium, Synechococcus sp. MA19, capable of accumulating poly-β-hydroxybutyrate. J. Ferment. Bioeng. 82, 512–514. https://doi.org/10.1016/S0922-338X(97)86995-4.

Montero-Lobato, Z., Ramos-Merchante, A., Fuentes, J., Sayago, A., Fernández-Recamales, Á., Martínez-Espinosa, R., Vega, J., Vílchez, C., Garbayo, I., 2018. Optimization of growth and carotenoid production by Haloferax mediterranei using response surface methodology. Marine Drugs 16, 372. https://doi.org/10.3390/md16100372.

Moreira, C., Rainey, F., Nobre, M., da Silva, M., da Costa, M., 2000. Tepidimonas ignava gen. nov., sp. nov., a new chemolithoheterotrophic and slightly thermophilic member of the beta-Proteobacteria. Int. J. Syst. Evol. Microbiol. 50, 735–742. https://doi.org/10.1099/00207713-50-2-735.

Mothes, G., Schubert, T., Harms, H., Maskow, T., 2008. biotechnological coproduction of compatible solutes and polyhydroxyalkanoates using the genus Halomonas. Eng. Life Sci. 8, 658–662. https://doi.org/10.1002/elsc.200800097.

Mougiakos, I., Mohanraju, P., Bosma, E., Vrouwe, V., Finger Bou, M., Naduthodi, M., Gussak, A., Brinkman, R., van Kranenburg, R., van der Oost, J., 2017. Characterizing a thermostable Cas9 for bacterial genome editing and silencing. Nat. Commun. 8, 1647. https://doi.org/10.1038/s41467-017-01591-4.

Mozejko-Ciesielska, J., Szacherska, K., Marciniak, P., 2019. Pseudomonas species as producers of eco-friendly polyhydroxyalkanoates. J. Polym. Environ. 27, 1151–1166. https://doi.org/10.1007/s10924-019-01422-1.

Mozumder, M., Garcia-Gonzalez, L., De Wever, H., Volcke, E., 2015. Effect of sodium accumulation on heterotrophic growth and polyhydroxybutyrate (PHB) production by Cupriavidus necator. Bioresour. Technol. 191, 213–218. https://doi.org/10.1016/j.biortech.2015.04.110.

Mravec, F., Obruca, S., Krzyzanek, V., Sedlacek, P., Hrubanova, K., Samek, O., Kucera, D., Benesova, P., Nebesarova, J., Steinbüchel, A., 2016. Accumulation of PHA granules in Cupriavidus necator as seen by confocal fluorescence microscopy. FEMS Microbiol. Lett. 363 https://doi.org/10.1093/femsle/fnw094.

Muhammadi, Shabina, Afzal, M., Hameed, S., 2015. Bacterial polyhydroxyalkanoates-eco-friendly next generation plastic: production, biocompatibility, biodegradation, physical properties and applications. Green Chem. Lett. Rev. 8, 56–77. https://doi.org/10.1080/17518253.2015.1109715.

Müller-Santos, M., Koskimäki, J., Alves, L., de Souza, E., Jendrossek, D., Pirttilä, A., 2020. The protective role of PHB and its degradation products against stress situations in bacteria. FEMS Microbiol. Rev. 45 https://doi.org/10.1093/femsre/fuaa058.

Musilova, J., Kourilova, X., Bezdicek, M., Lengerova, M., Obruca, S., Skutkova, H., Sedlar, K., Ochman, H., 2021. First complete genome of the thermophilic polyhydroxyalkanoates-producing bacterium schlegelella thermodepolymerans DSM 15344. Genome Biol. Evolut. 13 https://doi.org/10.1093/gbe/evab007 evab007.

Muyrers, J., 1999. Rapid modification of bacterial artificial chromosomes by ET-recombination. Nucleic Acids Res. 27, 1555–1557. https://doi.org/10.1093/nar/27.6.1555.

Ng, W., Kennedy, S., Mahairas, G., Berquist, B., Pan, M., Shukla, H., Lasky, S., Baliga, N., Thorsson, V., Sbrogna, J., Swartzell, S., Weir, D., Hall, J., Dahl, T., Welti, R., Goo, Y., Leithauser, B., Keller, K., Cruz, R., Danson, M., Hough, D., Maddocks, D., Jablonski, P., Krebs, M., Angevine, C., Dale, H., Isenbarger, T., Peck, R., Pohlschroder, M., Spudich, J., Jung, K., Alam, M., Freitas, T., Hou, S., Daniels, C., Dennis, P., Omer, A., Ebhardt, H., Lowe, T., Liang, P., Riley, M., Hood, L., DasSarma, S., 2000. Genome sequence of Halobacterium species NRC-1. Proc. Natl. Acad. Sci. 97, 12176–12181. https://doi.org/10.1073/pnas.190337797.

Nguyen, T., Ishizuna, F., Sato, Y., Arai, H., Ishii, M., 2019. Physiological characterization of poly-β-hydroxybutyrate accumulation in the moderately thermophilic hydrogen-oxidizing bacterium Hydrogenophilus thermoluteolus TH-1. J. Biosci. Bioeng. 127, 686–689. https://doi.org/10.1016/j.jbiosc.2018.11.011.

Nishioka, M., Nakai, K., Miyake, M., Asada, Y., Taya, M., 2001. Production of poly-β-hydroxybutyrate by thermophilic cyanobacterium, Synechococcus sp. MA19, under phosphate-limited conditions. Biotechnol. Lett. 23, 1095–1099. https://doi.org/10.1023/A:1010551614648.

Nonato, R., Mantelatto, P., Rossel, C., 2001. Integrated production of biodegradable plastic, sugar and ethanol. Appl. Microbiol. Biotechnol. 57, 1–5. https://doi.org/10.1007/s002530100732.

Novak, M., Koller, M., Braunegg, G., Horvat, P., 2015. Mathematical modelling as a tool for optimized PHA production. Chem. Biochem. Eng. Q. 29, 183–220. https://doi.org/10.15255/CABEQ.2014.2101.

Nowroth, V., Marquart, L., Jendrossek, D., Imperial, J., 2016. Low temperature-induced viable but not culturable state of *Ralstonia eutropha* and its relationship to accumulated polyhydroxybutyrate. FEMS Microbiology Letters vol. 363. https://doi.org/10.1093/femsle/fnw249.

Obruca, S., Benesova, P., Petrik, S., Oborna, J., Prikryl, R., Marova, I., 2014. Production of polyhydroxyalkanoates using hydrolysate of spent coffee grounds. Process Biochem. 49, 1409–1414. https://doi.org/10.1016/j.procbio.2014.05.013.

Obruca, S., Benesova, P., Marsalek, L., Marova, I., 2015. Use of lignocellulosic materials for PHA production. Chem. Biochem. Eng. Quart. 29, 135–144. https://doi.org/10.15255/CABEQ.2014.2253.

Obruca, S., Sedlacek, P., Krzyzanek, V., Mravec, F., Hrubanova, K., Samek, O., Kucera, D., Benesova, P., Marova, I., Chen, G., 2016a. Accumulation of poly(3-hydroxybutyrate) helps bacterial cells to survive freezing. PLOS ONE 11. https://doi.org/10.1371/journal.pone.0157778.

Obruca, S., Sedlacek, P., Mravec, F., Samek, O., Marova, I., 2016b. Evaluation of 3-hydroxybutyrate as an enzyme-protective agent against heating and oxidative damage and its potential role in stress response of poly(3-hydroxybutyrate) accumulating cells. Appl. Microbiol. Biotechnol. 100, 1365–1376. https://doi.org/10.1007/s00253-015-7162-4.

Obruca, S., Sedlacek, P., Mravec, F., Krzyzanek, V., Nebesarova, J., Samek, O., Kucera, D., Benesova, P., Hrubanova, K., Milerova, M., Marova, I., 2017. The presence of PHB granules in cytoplasm protects non-halophilic bacterial cells against the harmful impact of hypertonic environments. New Biotechnol. 39, 68–80. https://doi.org/10.1016/j.nbt.2017.07.008.

Obruca, S., Sedlacek, P., Koller, M., Kucera, D., Pernicova, I., 2018. Involvement of polyhydroxyalkanoates in stress resistance of microbial cells: Biotechnological consequences and applications. Biotechnol. Adv. 36, 856–870. https://doi.org/10.1016/j.biotechadv.2017.12.006.

Obruca, S., Sedlacek, P., Slaninova, E., Fritz, I., Daffert, C., Meixner, K., Sedrlova, Z., Koller, M., 2020. Novel unexpected functions of PHA granules. Appl. Microbiol. Biotechnol. 104, 4795–4810. https://doi.org/10.1007/s00253-020-10568-1.

Obruca, S., Sedlacek, P., Koller, M., 2021. The underexplored role of diverse stress factors in microbial biopolymer synthesis. Bioresource Technol. 326. https://doi.org/10.1016/j.biortech.2021.124767.

Obulisamy, P., Mehariya, S., 2021. Polyhydroxyalkanoates from extremophiles: a review. Bioresour. Technol. 325 https://doi.org/10.1016/j.biortech.2020.124653.

O'Leary, N., Wright, M., Brister, J., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., Astashyn, A., Badretdin, A., Bao, Y., Blinkova, O., Brover, V., Chetvernin, V., Choi, J., Cox, E., Ermolaeva, O., Farrell, C., Goldfarb, T., Gupta, T., Haft, D., Hatcher, E., Hlavina, W., Joardar, V., Kodali, V., Li, W., Maglott, D., Masterson, P., McGarvey, K., Murphy, M., O'Neill, K., Pujar, S., Rangwala, S., Rausch, D., Riddick, L., Schoch, C., Shkeda, A., Storz, S., Sun, H., Thibaud-Nissen, F., Tolstoy, I., Tully, R., Vatsan, A., Wallin, C., Webb, D., Wu, W., Landrum, M., Kimchi, A., Tatusova, T., DiCuccio, M., Kitts, P., Murphy, T., Pruitt, K., 2016. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. 44, D733–D745. https://doi.org/10.1093/nar/gkv1189.

Oliveira, C., Silva, C., Carvalho, G., Reis, M., 2017. Strategies for efficiently selecting PHA producing mixed microbial cultures using complex feedstocks: Feast and famine regime and uncoupled carbon and nitrogen availabilities. New Biotechnol. 37, 69–79. https://doi.org/10.1016/j.nbt.2016.10.008.

Olson, D., Lynd, L., 2012. Transformation of clostridium thermocellum by electroporation. In: Cellulases, Methods in Enzymology. Elsevier, pp. 317–330. https://doi.org/10.1016/B978-0-12-415931-0.00017-3.

Olson, D., Maloney, M., Lanahan, A., Hon, S., Hauser, L., Lynd, L., 2015. Identifying promoters for gene expression in Clostridium thermocellum. Metabol. Eng. Commun. 2, 23–29. https://doi.org/10.1016/j.meteno.2015.03.002.

Oren, A., 2002. Diversity of halophilic microorganisms: environments, phylogeny, physiology, and applications. J. Ind. Microbiol. Biotechnol. 28, 56–63. https://doi.org/10.1038/sj/jim/7000176.

Oren, A., Ginzburg, M., Ginzburg, B., Hochstein, L., Volcani, B., 1990. Haloarcula marismortui (Volcani) sp. nov., nom. rev., an extremely halophilic bacterium from the dead sea. Int. J. Syst. Bacteriol. 40, 209–210. https://doi.org/10.1099/00207713-40-2-209.

Pacholak, A., Gao, Z., Gong, X., Kaczorek, E., Cui, Y., 2021. The metabolic pathways of polyhydroxyalkanoates and exopolysaccharides synthesized by Haloferax mediterranei in response to elevated salinity. J. Proteome 232, 104065. https://doi.org/10.1016/j.jprot.2020.104065.

Pantazaki, A., Tambaka, M., Langlois, V., Guerin, P., Kyriakidis, D., 2003. Polyhydroxyalkanoate (PHA) biosynthesis in Thermus thermophilus: purification and biochemical properties of PHA synthase. Mol. Cell. Biochem. 254, 173–183. https://doi.org/10.1023/A:1027373100955.

Pantazaki, A., Papaneophytou, C., Pritsa, A., Liakopoulou-Kyriakides, M., Kyriakidis, D., 2009. Production of polyhydroxyalkanoates from whey by Thermus thermophilus HB8. Process Biochem. 44, 847–853. https://doi.org/10.1016/j.procbio.2009.04.002.

Panuschka, S., Deosg, B., Ellersdorfer, M., Meixner, K., Fritz, I., 2019. Photoautotrophic production of poly-hydroxybutyrate – First detailed cost estimations. Algal Res. 41 https://doi.org/10.1016/j.algal.2019.101558.

Park, J., Cho, B., Simpson, A., 2006. Halocafeteria seosinensis gen. et sp. nov. (Bicosoecida), a halophilic bacterivorous nanoflagellate isolated from a solar saltern. Extremophiles 10, 493–504. https://doi.org/10.1007/s00792-006-0001-x.

Park, S., Kim, T., Kim, M., Lee, S., Lim, S., 2012. Advanced bacterial polyhydroxyalkanoates: towards a versatile and sustainable platform for unnatural tailor-made polyesters. Biotechnol. Adv. 30, 1196–1206. https://doi.org/10.1016/j.biotechadv.2011.11.007.

Park, S., Jang, Y., Lee, H., Park, A., Yang, J., Shin, J., Oh, Y., Song, B., Jegal, J., Lee, S., Lee, S., 2013. Metabolic engineering of Ralstonia eutropha for the biosynthesis of 2-hydroxyacid-containing polyhydroxyalkanoates. Metab. Eng. 20, 20–28. https://doi.org/10.1016/j.ymben.2013.08.002.

Park, Y., Bhatia, S., Gurav, R., Choi, T., Kim, H., Song, H., Park, J., Han, Y., Lee, S., Park, S., Lee, H., Kim, Y., Yang, Y., 2020. Fructose based hyper production of poly-3-hydroxybutyrate from Halomonas sp. YLGW01 and impact of carbon sources on bacteria morphologies. Int. J. Biol. Macromol. 154, 929–936. https://doi.org/10.1016/j.ijbiomac.2020.03.129.

Peleg, M., Normand, M., Corradini, M., 2012. The arrhenius equation revisited. Crit. Rev. Food Sci. Nutr. 52, 830–851. https://doi.org/10.1080/10408398.2012.667460.

Pernicova, I., Kucera, D., Nebesarova, J., Kalina, M., Novackova, I., Koller, M., Obruca, S., 2019. Production of polyhydroxyalkanoates on waste frying oil employing selected Halomonas strains. Bioresour. Technol. 292, 122028. https://doi.org/10.1016/j.biortech.2019.122028.

Pernicova, I., Novackova, I., Sedlacek, P., Kourilova, X., Kalina, M., Kovalcik, A., Koller, M., Nebesarova, J., Krzyzanek, V., Hrubanova, K., Masilko, J., Slaninova, E., Obruca, S., 2020a. Introducing the newly isolated bacterium aneurinibacillus sp. H1 as an auspicious thermophilic producer of various polyhydroxyalkanoates (PHA) copolymers–1. Isolation and characterization of the bacterium. Polymers 12, 1235. https://doi.org/10.3390/polym12061235.

Pernicova, I., Novackova, I., Sedlacek, P., Kourilova, X., Koller, M., Obruca, S., 2020b. Application of osmotic challenge for enrichment of microbial consortia in polyhydroxyalkanoates producing thermophilic and thermotolerant bacteria and their subsequent isolation. Int. J. Biol. Macromol. 144, 698–704. https://doi.org/10.1016/j.ijbiomac.2019.12.128.

Pfeifer, K., Ergal, İ., Koller, M., Basen, M., Schuster, B., Rittmann, S.K.-M.R., 2021. Archaea biotechnology. Biotechnol. Adv. 47, 107668. https://doi.org/10.1016/j.biotechadv.2020.107668.

Posfai, G., Kolisnychenko, V., Bereczki, Z., Blattner, F., 1999. Markerless gene replacement in Escherichia coli stimulated by a double-strand break in the chromosome. Nucleic Acids Res. 27, 4409–4415. https://doi.org/10.1093/nar/27.22.4409.

Prieto, A., Escapa, I., Martínez, V., Dinjaski, N., Herencias, C., de la Peña, F., Tarazona, N., Revelles, O., 2016. A holistic view of polyhydroxyalkanoate metabolism in Pseudomonas putida. Environ. Microbiol. 18, 341–357. https://doi.org/10.1111/1462-2920.12760.

Qin, Q., Ling, C., Zhao, Y., Yang, T., Yin, J., Guo, Y., Chen, G., 2018. CRISPR/Cas9 editing genome of extremophile Halomonas spp. Metab. Eng. 47, 219–229. https://doi.org/10.1016/j.ymben.2018.03.018.

Quillaguamán, J., Hashim, S., Bento, F., Mattiasson, B., Hatti-Kaul, R., 2005. Poly (β -hydroxybutyrate) production by a moderate halophile, Halomonas boliviensis LC1 using starch hydrolysate as substrate. J. Appl. Microbiol. 99, 151–157. https://doi.org/10.1111/j.1365-2672.2005.02589.x.

Rahman, A., Linton, E., Hatch, A., Sims, R., Miller, C., 2013. Secretion of polyhydroxybutyrate in Escherichia coli using a synthetic biological engineering approach. J. Bioleg. Eng. 7 https://doi.org/10.1186/1754-1611-7-24.

Ren, Y., Ling, C., Hajnal, I., Wu, Q., Chen, G., 2018. Construction of Halomonas bluephagensis capable of high cell density growth for efficient PHA production. Appl. Microbiol. Biotechnol. 102, 4499–4510. https://doi.org/10.1007/s00253-018-8931-7.

Riley, L., Guss, A., 2021. Approaches to genetic tool development for rapid domestication of non-model microorganisms. Biotechnol. Biofuels 14. https://doi.org/10.1186/s13068-020-01872-x.

Riley, L., Ji, L., Schmitz, R., Westpheling, J., Guss, A., 2019. Rational development of transformation in Clostridium thermocellum ATCC 27405 via complete methylome analysis and evasion of native restriction–modification systems. J. Ind. Microbiol. Biotechnol. 46, 1435–1443. https://doi.org/10.1007/s10295-019-02218-x.

Rodríguez, Y., Pérez, V., López, J., Bordel, S., Firmino, P., Lebrero, R., Muñoz, R., 2020. Coupling biogas with PHA biosynthesis. In: The Handbook of Polyhydroxyalkanoates. CRC Press, Boca Raton, pp. 357–376.

Rodriguez-Contreras, A., Koller, M., Brunnegg, G., Marqués-Calvo, M., 2016. Poly[(R)-3-hydroxybutyrate] production under different salinity conditions by a novel Bacillus megaterium strain. New Biotechnol. 33, 73–77. https://doi.org/10.1016/j.nbt.2015.08.006.

Ruth, K., Grubelnik, A., Hartmann, R., Egli, T., Zinn, M., Ren, Q., 2007. Efficient production of (R)-3-hydroxycarboxylic acids by biotechnological conversion of polyhydroxyalkanoates and their purification. Biomacromolecules 8, 279–286. https://doi.org/10.1021/bm060585a.

Sadykov, M., Ahn, J., Widhelm, T., Eckrich, V., Endres, J., Driks, A., Rutkowski, G., Wingerd, K., Bayles, K., 2017. Poly(3-hydroxybutyrate) fuels the tricarboxylic acid cycle and de novo lipid biosynthesis during Bacillus anthracis sporulation. Mol. Microbiol. 104, 793–803. https://doi.org/10.1111/mmi.13665.

Salgaonkar, B., Bragança, J., 2015. Biosynthesis of poly(3-hydroxybutyrate- co −3-hydroxyvalerate) by Halogeometricum borinquense strain E3. Int. J. Biol. Macromol. 78, 339–346. https://doi.org/10.1016/j.ijbiomac.2015.04.016.

Salgaonkar, B., Bragança, J., 2017. Utilization of Sugarcane Bagasse by Halogeometricum borinquense Strain E3 for Biosynthesis of Poly(3-hydroxybutyrate-co-3-hydroxyvalerate). Bioengineering 4, 50. https://doi.org/10.3390/bioengineering4020050.

Salgaonkar, B., Mani, K., Bragança, J., 2019. Sustainable bioconversion of cassava waste to poly(3-hydroxybutyrate-co-3-hydroxyvalerate) by Halogeometricum borinquense strain E3. J. Polym. Environ. 27, 299–308. https://doi.org/10.1007/s10924-018-1346-9.

Salvachúa, D., Rydzak, T., Auwae, R., De Capite, A., Black, B., Bouvier, J., Cleveland, N., Elmore, J., Furches, A., Huenemann, J., Katahira, R., Michener, W., Peterson, D., Rohrer, H., Vardon, D., Beckham, G., Guss, A., 2020. Metabolic engineering of Pseudomonas putida for increased polyhydroxyalkanoate production from lignin. Microb. Biotechnol. 13, 290–298. https://doi.org/10.1111/1751-7915.13481.

Sangkharak, K., Khaithongkaeo, P., Chuaikhunpakarn, T., Choonut, A., Prasertsan, P., 2020. The production of polyhydroxyalkanoate from waste cooking oil and its application in biofuel production. Biomass Conversion Biorefinery. https://doi.org/10.1007/s13399-020-00657-6.

Satoh, Y., Tajima, K., Nakamoto, S., Xuerong, H., Matsushima, T., Obshima, T., Kawano, S., Erata, T., Dairi, T., Munekata, M., 2011. Isolation of a thermotolerant bacterium producing medium-chain-length polyhydroxyalkanoate. J. Appl. Microbiol. 111, 811–817. https://doi.org/10.1111/j.1365-2672.2011.05093.x.

Sauer, T., Galinski, E., 1998. Bacterial milking: a novel bioprocess for production of compatible solutes. Biotechnol. Bioeng. 59, 128. https://doi.org/10.1002/(SICI)1097-0290(19980705)59:1128::AID-BIT173.0.CO;2-E.

Sayers, E., Cavanaugh, M., Clark, K., Ostell, J., Pruitt, K., Karsch-Mizrachi, I., 2019. GenBank. Nucleic Acids Res. D1, D84–D86. https://doi.org/10.1093/nar/gkz956.

Schmid, M., Song, H., Raschbauer, M., Emerstoefer, F., Omann, M., Stelzer, F., Neureiter, M., 2019. Utilization of desugarized sugar beet molasses for the production of poly(3-hydroxybutyrate) by halophilic Bacillus megaterium uyuni S29. Process Biochem. 86, 9–15. https://doi.org/10.1016/j.procbio.2019.08.001.

Schmid, M., Raschbauer, M., Song, H., Bauer, C., Neureiter, M., 2021a. Effects of nutrient and oxygen limitation, salinity and type of salt on the accumulation of poly(3-hydroxybutyrate) in Bacillus megaterium uyuni S29 with sucrose as a carbon source. New Biotechnol. 61, 137–144. https://doi.org/10.1016/j.nbt.2020.11.012.

Schmid, M., Sykacek, E., O'Connor, K., Omann, M., Mundigler, N., Neureiter, M., 2021b. Pilot scale production and evaluation of mechanical and thermal properties of P (3HB) from Bacillus megaterium cultivated on desugarized sugar beet molasses. J. Appl. Polym. Sci. 51503 https://doi.org/10.1002/app.51503.

Sedlacek, P., Slaninova, E., Koller, M., Nebesarova, J., Marova, I., Krzyzanek, V., Obruca, S., 2019. PHA granules help bacterial cells to preserve cell integrity when exposed to sudden osmotic imbalances. New Biotechnol. 49, 129–136. https://doi.org/10.1016/j.nbt.2018.10.005.

Sedlacek, P., Pernicova, I., Novackova, I., Kourilova, X., Kalina, M., Kovalcik, A., Koller, M., Nebesarova, J., Krzyzanek, V., Hrubanova, K., Masilko, J., Slaninova, E., Trudicova, M., Obruca, S., 2020. Introducing the newly isolated bacterium aneurinibacillus sp. H1 as an auspicious thermophilic producer of various polyhydroxyalkanoates (PHA) copolymers–2. material study on the produced copolymers. Polymers 12, 1298. https://doi.org/10.3390/polym12061298.

Shahid, S., Mosrati, R., Ledauphin, J., Amiel, C., Fontaine, P., Gaillard, J., Corroler, D., 2013. Impact of carbon source and variable nitrogen conditions on bacterial biosynthesis of polyhydroxyalkanoates: evidence of an atypical metabolism in Bacillus megaterium DSM 509. J. Biosci. Bioeng. 116, 302–308. https://doi.org/10.1016/j.jbiosc.2013.02.017.

Shen, R., Yin, J., Ye, J., Xiang, R., Ning, Z., Huang, W., Chen, G., 2018. Promoter engineering for enhanced P(3HB-co-4HB) production by Halomonas bluephagensis. ACS Synth. Biol. 7, 1897–1906. https://doi.org/10.1021/acssynbio.8b00102.

Shen, R., Ning, Z., Lan, Y., Chen, J., Chen, G., 2019. Manipulation of polyhydroxyalkanoate granular sizes in Halomonas bluephagensis. Metab. Eng. 54, 117–126. https://doi.org/10.1016/j.ymben.2019.03.011.

Sheu, D., Chen, W., Yang, J., Chang, R., 2009. Thermophilic bacterium Caldimonas taiwanensis produces poly(3-hydroxybutyrate-co-3-hydroxyvalerate) from starch and valerate as carbon sources. Enzym. Microb. Technol. 44, 289–294. https://doi.org/10.1016/j.enzmictec.2009.01.004.

Sheu, D., Chen, W., Lai, Y., Chang, R., 2012. Mutations derived from the thermophilic polyhydroxyalkanoate synthase PhaC enhance the thermostability and activity of PhaC from cupriavidus necator H16. J. Bacteriol. 194, 2620–2629. https://doi.org/10.1128/JB.06543-11.

Shivaraud, P., Mugeraya, G., 2011. Halophilic bacteria and their compatible solutes – osmoregulation and potential applications. Curr. Sci. 100, 1516–1521.

Silva-Rocha, R., Martínez-García, E., Calles, B., Chavarría, M., Arce-Rodríguez, A., de las Heras, A., Páez-Espino, A., Durante-Rodríguez, G., Kim, J., Nikel, P., Platero, R., de Lorenzo, V., 2013. The Standard European Vector Architecture (SEVA): a coherent platform for the analysis and deployment of complex prokaryotic phenotypes. Nucleic Acids Res. 41, D666–D675. https://doi.org/10.1093/nar/gks1119.

Simon-Colin, C., Raguénès, G., Cozien, J., Guezennec, J., 2008. Halomonas profundus sp. nov., a new PHA-producing bacterium isolated from a deep-sea hydrothermal vent shrimp. J. Appl. Microbiol. 104, 1425–1432. https://doi.org/10.1111/j.1365-2672.2007.03667.x.

Singh, A., Srivastava, J., Chandel, A., Sharma, L., Mallick, N., Singh, S., 2019. Biomedical applications of microbially engineered polyhydroxyalkanoates: an insight into recent advances, bottlenecks, and solutions. Appl. Microbiol. Biotechnol. 103, 2007–2032. https://doi.org/10.1007/s00253-018-09604-y.

Slaninova, E., Sedlacek, P., Mravec, F., Mullerova, L., Samek, O., Koller, M., Hesko, O., Kucera, D., Marova, I., Obruca, S., 2018. Light scattering on PHA granules protects bacterial cells against the harmful effects of UV radiation. Appl. Microbiol. Biotechnol. 102, 1923–1931. https://doi.org/10.1007/s00253-018-8760-8.

Soto, G., Setten, L., Lisi, C., Maurelis, C., Mozzicafreddo, M., Cuccioloni, M., Angeletti, M., Ayub, N., 2012. Hydroxybutyrate prevents protein aggregation in the halotolerant bacterium Pseudomonas sp. CT13 under abiotic stress. Extremophiles 16, 455–462. https://doi.org/10.1007/s00792-012-0445-0.

Stanley, A., Punil Kumar, H., Matturi, S., Vijayendra, S., 2018. Fed-batch strategies for production of PHA using a native isolate of halomonas venusta KT832796 Strain. Appl. Biochem. Biotechnol. 184, 935–952. https://doi.org/10.1007/s12010-017-2601-6.

Stanley, A., Matturi, S., Vijayendra, S., 2021. Halophilic microorganisms as potential producers of polyhydroxyalkanoates. In: Kuddus, M., Roohi (Eds.), Bioplastics for Sustainable Development. Springer Singapore, Singapore, pp. 277–294. https://doi.org/10.1007/978-981-16-1623-9_10.

Steinbüchel, A., 1995. Diversity of bacterial polyhydroxyalkanoic acids. FEMS Microbiol. Lett. 128, 219–228. https://doi.org/10.1016/0378-1097(95)00125-O.

Tajima, K., Han, X., Hashimoto, Y., Satoh, Y., Satoh, T., Taguchi, S., 2016. In vitro synthesis of polyhydroxyalkanoates using thermostable acetyl-CoA synthetase, CoA transferase, and PHA synthase from thermotolerant bacteria. J. Biosci. Bioeng. 122, 660–665. https://doi.org/10.1016/j.jbiosc.2016.06.001.

Takeda, M., Kamagata, Y., Ghiorse, W., Hanada, S., Koizumi, J., 2002. Caldimonas manganoxidans gen. nov., sp. nov., a poly(3-hydroxybutyrate)-degrading, manganese-oxidizing thermophile. Int. J. Syst. Evol. Microbiol. 52, 895–900. https://doi.org/10.1099/00207713-52-3-895.

Tan, D., Xue, Y., Aibaidula, G., Chen, G., 2011. Unsterile and continuous production of polyhydroxybutyrate by Halomonas TD01. Bioresour. Technol. 102, 8130–8136. https://doi.org/10.1016/j.biortech.2011.05.068.

Tan, D., Wu, Q., Chen, J., Chen, G., 2014. Engineering Halomonas TD01 for the low-cost production of polyhydroxyalkanoates. Metab. Eng. 26, 34–47. https://doi.org/10.1016/j.ymben.2014.09.001.

Tan, I.K.P., Foong, C.P., Tan, H.T., Lim, H., Zain, N.A.A., Tan, Y.C., Hoh, C.C., Sudesh, K., 2020. Polyhydroxyalkanoate (PHA) synthase genes and PHA-associated gene clusters in Pseudomonas spp. and Janthinobacterium spp. isolated from Antarctica. J. Biotechnol. 313, 18–28. https://doi.org/10.1016/j.jbiotec.2020.03.006.

Tan, D., Wang, Y., Tong, Y., Chen, G., 2021. Grand challenges for industrializing polyhydroxyalkanoates (PHAs). Trends Biotechnol. 39, 953–963. https://doi.org/10.1016/j.tibtech.2020.11.010.

Tao, W., Lv, L., Chen, G., 2017. Engineering halomonas species TD01 for enhanced polyhydroxyalkanoates synthesis via CRISPRi. Microb. Cell Factories 16. https://doi.org/10.1186/s12934-017-0655-3.

Tao, G., Tan, B., Li, Z., 2021. Production of polyhydroxyalkanoates by a moderately halophilic bacterium of Salinivibrio sp. TGB10. Int. J. Biol. Macromol. 186, 574–579. https://doi.org/10.1016/j.ijbiomac.2021.07.038.

Tiso, T., Narancic, T., Wei, R., Pollet, E., Beagan, N., Schröder, K., Honak, A., Jiang, M., Kenny, S., Wierckx, N., Perrin, R., Avérous, L., Zimmermann, W., O'Connor, K., Blank, L., 2021. Towards bio-upcycling of polyethylene terephthalate. Metab. Eng. 66, 167–178. https://doi.org/10.1016/j.ymben.2021.03.011.

Tribelli, P., Pezzoni, M., Brito, M., Montesinos, N., Costa, C., López, N., 2020. Response to lethal UVA radiation in the Antarctic bacterium Pseudomonas extremaustralis: polyhydroxybutyrate and cold adaptation as protective factors. Extremophiles 24, 265–275. https://doi.org/10.1007/s00792-019-01152-1.

Tripathi, S., Olson, D., Argyros, D., Miller, B., Barrett, T., Murphy, D., McCool, J., Warner, A., Rajgarhia, V., Lynd, L., Hogsett, D., Caiazza, N., 2010. Development of pyrF- based genetic system for targeted gene deletion in clostridium thermocellum and creation of a pta mutant. Appl. Environ. Microbiol. 76, 6591–6599. https://doi.org/10.1128/AEM.01484-10.

Urbieta, M.S., Donati, E.R., Chan, K.-G., Shahar, S., Sin, L.L., Goh, K.M., 2015. Thermophiles in the genomic era: biodiversity, science, and applications. Biotechnol. Adv. 33, 633–647. https://doi.org/10.1016/j.biotechadv.2015.04.007.

Van Thuoc, D., My, D., Loan, T., Sudesh, K., 2019. Utilization of waste fish oil and glycerol as carbon sources for polyhydroxyalkanoate production by Salinivibrio sp. M318. Int. J. Biol. Macromol. 141, 885–892. https://doi.org/10.1016/j.ijbiomac.2019.09.063.

Vandewijngaarden, J., Wauters, R., Murariu, M., Dubois, P., Carleer, R., Yperman, J., D'Haen, J., Ruttens, B., Schreurs, S., Lepot, N., Peeters, R., Buntinx, M., 2016. Poly (3-hydroxybutyrate-co-3-hydroxyhexanoate)/organomodified montmorillonite nanocomposites for potential food packaging applications. J. Polym. Environ. 24, 104–118. https://doi.org/10.1007/s10924-016-0751-1.

Van-Thuoc, D., Huu-Phong, T., Minh-Khoong, D., Hatti-Kaul, R., 2015. Poly(3-hydroxybutyrate-co-3-Hydroxyvalerate) production by a moderate Halophile Yangia sp. ND199 using glycerol as a carbon source. Appl. Biochem. Biotechnol. 175, 3120–3132. https://doi.org/10.1007/s12010-015-1479-4.

Vasu, K., Nagaraja, V., 2013. Diverse functions of restriction-modification systems in addition to cellular defense. Microbiol. Mol. Biol. Rev. 77, 53–72. https://doi.org/10.1128/MMBR.00044-12.

Walker, J., Lanahan, A., Zheng, T., Tormo, C., Lynd, L., Cameron, J., Olson, D., Eckert, C., 2020. Development of both type I-B and type II CRISPR/Cas genome editing systems in the cellulolytic bacterium Clostridium thermocellum. Metabol. Eng. Commun. 10, e00116 https://doi.org/10.1016/j.mec.2019.e00116.

Wang, B., Pugh, S., Nielsen, D., Zhang, W., Meldrum, D., 2013. Engineering cyanobacteria for photosynthetic production of 3-hydroxybutyrate directly from CO2. Metab. Eng. 16, 68–77. https://doi.org/10.1016/j.ymben.2013.01.001.

Wang, Y., Ling, C., Chen, Y., Jiang, X., Chen, G., 2019. Microbial engineering for easy downstream processing. Biotechnol. Adv. 37, 107365. https://doi.org/10.1016/j.biotechadv.2019.03.004.

Wang, Z., Qin, Q., Zheng, Y., Li, F., Zhao, Y., Chen, G., 2021. Engineering the permeability of Halomonas bluephagenesis enhanced its chassis properties. Metab. Eng. 67, 53–66. https://doi.org/10.1016/j.ymben.2021.05.010.

Weiss, T., Young, E., Ducat, D., 2017. A synthetic, light-driven consortium of cyanobacteria and heterotrophic bacteria enables stable polyhydroxybutyrate production. Metab. Eng. 44, 236–245. https://doi.org/10.1016/j.ymben.2017.10.009.

Whelan, S., Goldman, N., 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Mol. Biol. Evol. 18, 691–699. https://doi.org/10.1093/oxfordjournals.molbev.a003851.

Williams-Rhaesa, A., Awuku, N., Lipscomb, G., Poole, F., Rubinstein, G., Conway, J., Kelly, R., Adams, M., 2018. Native xylose-inducible promoter expands the genetic tools for the biomass-degrading, extremely thermophilic bacterium Caldicellulosiruptor bescii. Extremophiles 22, 629–638. https://doi.org/10.1007/s00792-018-1023-x.

Wood, J., Bremer, E., Csonka, L., Kraemer, R., Poolman, B., van der Heide, T., Smith, L., 2001. Osmosensing and osmoregulatory compatible solute accumulation by bacteria. Comp. Biochem. Physiol. A Mol. Integr. Physiol. 130, 437–460. https://doi.org/10.1016/S1095-6433(01)00442-1.

Xi, L., Zhang, Z., Qiao, N., Zhang, Y., Li, J., Zhao, J., Xiao, Z., 2016. Complete genome sequence of the novel thermophilic polyhydroxyalkanoates producer Aneurinibacillus sp. XH2 isolated from Gudao oilfield in China. J. Biotechnol. 227, 54–55. https://doi.org/10.1016/j.jbiotec.2016.03.056.

Xiao, Z., Zhang, Y., Xi, L., Huo, F., Zhao, J., Li, J., 2015. Thermophilic production of polyhydroxyalkanoates by a novel Aneurinibacillus strain isolated from Gudao oilfield, China. J. Basic Microbiol. 55, 1125–1133. https://doi.org/10.1002/jobm.201400843.

Xu, F., Huang, S., Liu, Y., Zhang, Y., Chen, S., 2014. Comparative study on the production of poly(3-hydroxybutyrate) by thermophilic Chelatococcus daeguensis TAD1: a good candidate for large-scale production. Appl. Microbiol. Biotechnol. 98, 3965–3974. https://doi.org/10.1007/s00253-014-5524-y.

Ye, J., Huang, W., Wang, D., Chen, F., Yin, J., Li, T., Zhang, H., Chen, G., 2018. Pilot Scale-up of Poly(3-hydroxybutyrate- co −4-hydroxybutyrate) Production by Halomonas bluephagenesis via Cell Growth Adapted Optimization Process. Biotechnol. J. 13, 1800074. https://doi.org/10.1002/biot.201800074.

Ye, J., Hu, D., Yin, J., Huang, W., Xiang, R., Zhang, L., Wang, X., Han, J., Chen, G., 2020. Stimulus response-based fine-tuning of polyhydroxyalkanoate pathway in Halomonas. Metab. Eng. 57, 85–95. https://doi.org/10.1016/j.ymben.2019.10.007.

Yin, J., Fu, X., Wu, Q., Chen, J., Chen, G., 2014. Development of an enhanced chromosomal expression system based on porin synthesis operon for halophile Halomonas sp. Appl. Microbiol. Biotechnol. 98, 8987–8997. https://doi.org/10.1007/s00253-014-5959-1.

Yin, J., Chen, J., Wu, Q., Chen, G., 2015. Halophiles, coming stars for industrial biotechnology. Biotechnol. Adv. 33, 1433–1442. https://doi.org/10.1016/j.biotechadv.2014.10.008.

You, J., Li, Y., Hong, S., Wang, J., Yu, J., Mu, B., Ma, X., Xue, Y., 2019. Tepidicella baoligensis sp. nov., a novel member of betaproteobacterium isolated from an oil reservoir. Curr. Microbiol. 76, 410–414. https://doi.org/10.1007/s00284-018-1604-z.

Yu, L., Wu, F., Chen, G., 2019. Next-generation industrial biotechnology-transforming the current industrial biotechnology into competitive processes. Biotechnol. J. 14 https://doi.org/10.1002/biot.201800437.

Yu, L., Yan, X., Zhang, X., Chen, X., Wu, Q., Jiang, X., Chen, G., 2020. Biosynthesis of functional polyhydroxyalkanoates by engineered Halomonas bluephagenesis. Metab. Eng. 59, 119–130. https://doi.org/10.1016/j.ymben.2020.02.005.

Zelder, B., Keller, M., Loder, A., Straub, C., Adams, M., Kelly, R., 2015. Extremely thermophilic microorganisms as metabolic engineering platforms for production of fuels and industrial chemicals. Front. Microbiol. 6, 1209. https://doi.org/10.3389/fmicb.2015.01209.

Zhang, M., Thomas, N.L., 2011. Blending polylactic acid with polyhydroxybutyrate: the effect on thermal, mechanical, and biodegradation properties. Adv. Polym. Technol. 30, 67–79. https://doi.org/10.1002/adv.20235.

Zhang, X., Lin, Y., Wu, Q., Wang, Y., Chen, G., 2020. Synthetic biology and genome-editing tools for improving PHA metabolic engineering. Trends Biotechnol. 38, 689–700. https://doi.org/10.1016/j.tibtech.2019.10.006.

Zhao, Y., Li, H., Qin, L., Wang, H., Chen, G., 2007. Disruption of the polyhydroxyalkanoate synthase gene in Aeromonas hydrophila reduces its survival ability under stress conditions. FEMS Microbiol. Lett. 276, 34–41. https://doi.org/10.1111/j.1574-6968.2007.00904.x.

Zhao, Y., Rao, Z., Xue, Y., Gong, P., Ji, Y., Ma, Y., 2015. Poly(3-hydroxybutyrate-co-3-hydroxyvalerate) production by Haloarchaeon Halogranum amylolyticum. Appl. Microbiol. Biotechnol. 99, 7639–7649. https://doi.org/10.1007/s00253-015-6609-y.

Zhao, H., Zhang, H., Chen, X., Li, T., Wu, Q., Ouyang, Q., Chen, G., 2017. Novel T7-like expression systems used for Halomonas. Metab. Eng. 39, 128–140. https://doi.org/10.1016/j.ymben.2016.11.007.

Zheng, Y., Meng, F., Zhu, Z., Wei, W., Sun, Z., Chen, J., Yu, B., Lou, C., Chen, G., 2019. A tight cold-inducible switch built by coupling thermosensitive transcriptional and proteolytic regulatory parts. Nucleic Acids Res. 47, e137. https://doi.org/10.1093/nar/gkz785.

Zheng, Y., Chen, J., Ma, Y., Chen, G., 2020. Engineering biosynthesis of polyhydroxyalkanoates (PHA) for diversity and cost reduction. Metab. Eng. 58, 82–93. https://doi.org/10.1016/j.ymben.2019.07.004.

Zou, H., Shi, M., Zhang, T., Li, L., Li, L., Xian, M., 2017. Natural and engineered polyhydroxyalkanoate (PHA) synthase: key enzyme in biopolyester production. Appl. Microbiol. Biotechnol. 101, 7417–7426. https://doi.org/10.1007/s00253-017-8485-0.

5.16 Article XVI

MUSILOVA, Jana, KOURILOVA, Xenie, PERNICOVA, Iva, BEZDICEK, Matej, LENGEROVA, Martina, OBRUCA, Stanislav, and **SEDLAR, Karel**. Novel thermophilic polyhydroxyalkanoates producing strain Aneurinibacillus thermoaerophilus CCM 8960. *Applied Microbiology and Biotechnology*. 2022. 106(12), p. 4669–4681. (2022 IF not announced yet, 2021 IF = 5.560, Q1 in BIOTECHNOLOGY & APPLIED MICROBIOLOGY)

DOI: https://doi.org/10.1007/s00253-022-12039-1

Applied Microbiology and Biotechnology (2022) 106:4669–4681
https://doi.org/10.1007/s00253-022-12039-1

GENOMICS, TRANSCRIPTOMICS, PROTEOMICS

Novel thermophilic polyhydroxyalkanoates producing strain *Aneurinibacillus thermoaerophilus* CCM 8960

Jana Musilova[1,2] · Xenie Kourilova[3] · Iva Pernicova[3] · Matej Bezdicek[4] · Martina Lengerova[4] · Stanislav Obruca[3] · Karel Sedlar[1,5]

Received: 14 March 2022 / Revised: 16 June 2022 / Accepted: 18 June 2022 / Published online: 27 June 2022

Abstract

Aneurinibacillus thermoaerophilus CCM 8960 is a thermophilic bacterium isolated from compost in Brno. The bacterium accumulates polyhydroxyalkanoates (PHAs), a biodegradable and renewable alternative to petrochemical polymers. The bacterium reveals several features that make it a very interesting candidate for the industrial production of PHA. At first, due to its thermophilic character, the bacterium can be utilized in agreement with the concept of next-generation industrial biotechnology (NGIB), which relies on extremophiles. Second, the bacterium is capable of producing PHA copolymers containing a very high portion of 4-hydroxybutyrate (4HB). Such materials possess unique properties and can be advantageously used in multiple applications, including but not limited to medicine and healthcare. Therefore, this work focuses on the in-depth characterization of *A. thermoaerophilus* CCM 8960. In particular, we sequenced and assembled the genome of the bacterium and identified its most important genetic features, such as the presence of plasmids, prophages, CRISPR arrays, antibiotic-resistant genes, and restriction-modification (R-M) systems, which might be crucial for the development of genome editing tools. Furthermore, we focused on genes directly involved in PHA metabolism. We also experimentally studied the kinetics of glycerol and 1,4-butanediol (1,4BD) utilization as well as biomass growth and PHA production during cultivation. Based on these data, we constructed a metabolic model to reveal metabolic fluxes and nodes of glycerol and 1,4BD concerning their incorporation into the poly(3-hydroxybutyrate-*co*-4-hydroxybutyrate (P(3HB-*co*-4HB)) structure.

Key points

- *Aneurinibacillus sp. H1 was identified as Aneurinibacillus thermoaerophilus.*
- *PHA metabolism pathway with associated genes was presented.*
- *Unique monomer composition of produced PHAs was reported.*

Keywords De novo assembly · PHA · *Aneurinibacillus* species H1 · Plasmid pAT1 · 4-hydroxybutyrate · Next-generation industrial biotechnology

✉ Karel Sedlar
sedlar@vut.cz

[1] Department of Biomedical Engineering, Faculty of Electrical Engineering and Communication, Brno University of Technology, Brno, Czech Republic

[2] Department of Biochemistry, University of Nebraska-Lincoln, Lincoln, NE, USA

[3] Department of Food Chemistry and Biotechnology, Faculty of Chemistry, Brno University of Technology, Brno, Czech Republic

[4] Department of Internal Medicine – Hematology and Oncology, University Hospital Brno, Brno, Czech Republic

[5] Department of Informatics, Institute of Bioinformatics, Ludwig-Maximilians-Universität München, Munich, Germany

Introduction

The pollution of the earth by non-biodegradable materials, such as plastics, is one of the main issues of our time directly associated with human activity. Generally, there is a strong consensus that this unwise contamination of the environment should be avoided. The pollution may be significantly reduced by partially replacing these harmful materials with environmentally friendly ones. Eco-friendly, biodegradable polyesters, polyhydroxyalkanoates (PHAs) are accumulated by numerous prokaryotes that, in their material properties, are similar to industrially processed synthetic polymers (Koller et al. 2017).

Extremophilic organisms live and thrive in extreme conditions, such as high or low temperatures, alkaline or acidic pH, and high salt concentrations, or they can be resistant to various toxic elements (Martin and McMinn 2018). The role of extremophiles in the biotechnology industry has come to the fore in recent years; due to their interesting capability to survive and prosper under extreme living conditions, their cultivation can be carried out in a semi-sterile or even unsterile mode, which substantially reduces the cost of biotechnological processes (Coker 2016). However, they are also very interesting producers of various extreme enzymes; they are able to utilize interesting waste substrates which further leads to increased competitiveness. The use of extremophiles in the biotechnology industry is often referred to as the next-generation industrial biotechnology (NGIB) (Chen and Jiang 2018).

The ability to accumulate PHAs is also widespread among extremophiles (Obulisamy and Mehariya 2021; Obruča et al. 2022). PHAs not only serve the bacteria as a source of carbon and energy but also help protect them against various stress factors (Obruca et al. 2018). Production among halophilic bacteria is already fairly well described (Simon-Colin et al. 2008; Alsafadi and Al-Mashaqbeh 2017; Obruca et al. 2018; Pernicova et al. 2019), whereas production among thermophilic bacteria is not as well documented (Obruca et al. 2018).

We have developed a novel method enabling the isolation of PHA-producing bacteria from mixed microbial consortia. This technique called "osmoselection" takes advantage of the protective function of PHAs against fluctuations in osmotic pressure (Pernicova et al. 2020b). This protocol yielded several promising thermophilic PHA producers, and the bacterium classified as *Aneurinibacillus thermoaerophilus* CCM 8960 (previously *Aneurinibacillus* sp. H1) was the most promising. *A. thermoaerophilus* CCM 8960 is a gram-positive bacterium first isolated in the central urban composting plant in the city of Brno, Czech Republic (Centrální kompostárna Brno operated by SUEZ CZ a.s.) and deposited in the Czech Collection of Microorganisms as patent culture CCM 8960 under the name *Aneurinibacillus* sp. H1 (Pernicova et al. 2020a; Sedlacek et al. 2020). It has been discovered that the *A. thermoaerophilus* CCM 8960 belongs to thermophiles, i.e., a subgroup of extremophiles that optimally grow in temperatures above 45 °C (Zeldes et al. 2015). This fact has significant advantages for the biotechnological use of the bacterium such as reduced cooling costs, sterilization demands and costs, risk of microbial contamination, and improved solubility of substrates. Moreover, the bacterium is biotechnologically promising due to its capability to accumulate a high amount of PHAs under given conditions (Pernicova et al. 2020a).

This bacterium is able to use glycerol as a carbon source, which leads to a reduction in production cost because glycerol can be cheaply obtained as a by-product from biodiesel production. In addition, *A. thermoaerophilus* CCM 8960 was also found to be able to utilize 1,4-butanediol (1,4BD) or gamma-butyrolactone (GBL) as the sole carbon sources (Pernicova et al. 2020a). Furthermore, when using these substrates as a carbon source, it is capable of forming the copolymer poly(3-hydroxybutyrate-*co*-4-hydroxybutyrate) [P(3HB-*co*-4HB)], which has unique properties. Homopolymer of 3-hydroxybutyrate (3HB), poly(3-hydroxybutyrate) [P(3HB)], the most abundant member of the PHAs family, is a crystalline material with poor flexibility and thermal stability. However, when monomer 4HB is incorporated into the chain, its crystallinity decreases, the material becomes more elastic and its melting temperature substantially increases, which simplifies the processing via melting. The use of PHA, especially copolymer P(3HB-*co*-4HB), has great potential. This material can be used in the cosmetic industry, in medicine as surgical threads, in scaffolds, tissue-engineering applications, or even in agriculture. The production of the copolymer by *A. thermoaerophilus* CCM 8960 is also unique since the bacterium is capable of incorporating very high fractions of 4HB (up to 90 mol. %) into the copolymer; moreover, the composition of the copolymer can be controlled simply by manipulating the composition of the cultivation media, in particular with concentration and ratios of glycerol as the main carbon substrate and precursor of 3HB and 1,4BD as a precursor of 4HB (Pernicova et al. 2020a).

Studying an organism at the genomic level allows us to gain a deeper understanding of both its structure and functional properties. With the availability of whole-genome sequencing methods as well as the computationally powerful machines necessary for the assembly, the availability of genomic information that was previously hidden from us has greatly increased (Stephens et al. 2015). Although the published information in databases is currently mostly based on model organisms (MOs) (Ho et al. 2010; Agarwala et al. 2016; Huerta-Cepas et al. 2019; Karp et al.

2019; Safran et al. 2021; Chen et al. 2021; Kanehisa et al. 2021; Lee et al. 2022), this knowledge can be exploited to better understand the properties of newly discovered or previously unexplored non-model organisms (NMOs), which undoubtedly includes *A. thermoaerophilus* CCM 8960. The revealing of homologies with MOs allows estimation of the properties of the organism under study, such as the function of individual genes, or even the proposal of a metabolic model, which provides essential information in the context of subsequent biotechnological applications.

Therefore, in this work, we decided to investigate the metabolism of *A. thermoaerophilus* CCM 8960 with respect to PHA metabolism. In particular, we sequenced and assembled the genome of the bacterium and identified the most important genetic features such as the presence of plasmids, prophages, CRISPR arrays, antibiotic-resistant genes, and restriction-modification (R-M) systems, and we have also reconsidered the taxonomic classification of the bacterium. Furthermore, we focused on genes directly involved in PHA metabolism. We also experimentally studied the kinetics of glycerol and 1,4BD utilization as well as biomass growth and PHA production during cultivation. Based on these data, we constructed a metabolic model to reveal metabolic fluxes of glycerol and 1,4BD with respect to their incorporation into the P(3HB-*co*-4HB) structure.

Materials and methods

Growth conditions

Nutrient Broth complex medium (10 g/L beef extract, 10 g/L peptone, 5 g/L NaCl) for the growth of *Aneurinibacillus* sp. H1 was used at 45 °C and stirred at 190 rpm. The complex medium was used as an inoculum, in which cultivation lasted for 24 h. For PHA production, a mineral medium (MSM) was used consisting of: Na_2HPO_4•12 H_2O, 9.0 g/L; KH_2PO_4, 1.5 g/L; $MgSO_4$•7 H_2O, 0.2 g/L; NH_4NO_3 1.0 g/L; $CaCl_2$•2 H_2O, 0.02 g/L; $Fe^{III}NH_4$ citrate, 0.0012 g/L; tryptone, 0.5 g/L with 1 mL/L trace element solution (TES). The TES contained g per liter of water: EDTA, 50.0; $FeCl_3$•6 H_2O, 13.8; $ZnCl_2$, 0.84; $CuCl_2$•2 H_2O, 0.13; $CoCl_2$•6 H_2O, 0.1; $MnCl_2$•6 H_2O, 0.016; H_3BO_3, 0.1. PHA production was carried out for 0 to 72 h at 45 °C and 190 rpm. The carbon source was 1,4-butanediol at a concentration of 4 g/L, glycerol at 4 g/L, or a mixture of 1,4BD and glycerol, each at a concentration of 4 g/L. All culture conditions were carried out in duplicate. At 0, 12, 24, 36, 48, and 72 h, the culture was terminated. The bacterial cells were centrifuged (6000 rpm, 5 min). Biomass was determined gravimetrically.

DNA extraction and sequencing

The genomic DNA was extracted and purified using the GenElute Bacterial Genomic DNA Kit (Sigma-Aldrich, St. Louis, MI, USA) according to the manufacturer's protocols. NanoDrop spectrophotometer (Thermo Scientific, Wilmington, DE, USA) was used to measure the purity and Qubit 3.0 (Thermo Scientific, Wilmington, DE, USA) to measure the concentration of extracted genomic DNA. DNA library construction was performed using the KAPA HyperPlus kit according to the standard protocol. The sequencing was carried out using the Miseq Reagent Kit v2 (500 cycles) and the Illumina MiSeq platform (Illumina, San Diego, CA, USA).

For long-read sequencing, genomic high molecular weight DNA was extracted using the MagAttract HMW DNAKit (Qiagene, Venlo. NL). The extracted DNA purity was measured using the NanoDrop (Thermo Fisher Scientific, Waltham, MA, USA), the concentration was measured using the Qubit 3.0 (Thermo Scientific, Wilmington, DE, USA), and proper length was checked using Agilent 4200 TapeStation (Agilent Technologies, Santa Clara, CA, USA), respectively. Library preparation for Oxford Nanopore sequencing was performed using the Ligation sequencing 1D Kit (Oxford Nanopore Technologies, Oxford, UK). The library was sequenced using the R9.4.1 flowcell and the MinION platform (Oxford Nanopore Technologies).

Genome assembly

The whole assembly process was done in two parts: Nanopore reads were assembled into initial sequence and, subsequently, Illumina reads were mapped as a hybrid assembly. The initial sequence assembly included several steps. First, Guppy v3.4.4 was used to basecall raw Nanopore reads. Next, reads were assembled using Flye v2.8.1. The assembly was polished with Racon v1.4.13 (Vaser et al. 2017) and Medaka. Quality check during the initial assembly was done with MinIONQC (Lanfear et al. 2019), and auxiliary PAF files were generated using minimap2 (Li 2018). In the next step, trimmed Illumina paired-reads were mapped to the initial sequence; the adapter and the quality trimming was done using Trimomatic v1.36 (Bolger et al. 2014). The reads were mapped using BWA v07.17 (Li 2013) and polished with Pilon v1.23 (Walker et al. 2014). The quality of Illumina reads was checked with FastQC v0.11.5 and MultiQC v1.7 (Ewels et al. 2016) tools; auxiliary BAM files were generated using SAMtools (Li et al. 2009). In the last step, the final sequence was rearranged according to the origin of replication (oriC) to *DnaA* be the first gene using the Ori-finder (Luo et al. 2019) tool.

Genome annotation and analysis

Both chromosome and plasmid were annotated using NCBI Prokaryotic Genome Annotation Pipeline (PGAP) (Tatusova et al. 2016). Operons were predicted with PathwayTools (Karp et al. 2015). In the next step, protein-coding genes were classified into clusters of orthologous groups (COGs) using the eggNOG database and eggNOG-mapper (Huerta-Cepas et al. 2019). Subsequently, circular chromosomal maps of both chromosome and plasmid were generated with DNAPlotter (Carver et al. 2009) integrated in the Artemis (Rutherford et al. 2000) tool. CRISPR arrays were identified using CRISPRDetect tool (Biswas et al. 2016). Antibiotic-resistant genes were detected with Resistance Gene Identifier (RGI) 5.2.0, using the Comprehensive Antibiotic Resistance Database (CARD) 3.1.4 (Alcock et al. 2020). Methylation bases were inferred using deepsignal2 (Ni et al. 2019); motifs were further detected with STREME Command-Line Version (Bailey 2021). The detection of R-M systems was completed using the internal tools of The Restriction Enzyme Database (Roberts et al. 2015).

Plasmid identification was done in two ways. First, phages were searched using experimental laboratory methods and, subsequently, bioinformatics tools were used. Furthermore, online tools Prophage Hunter (Song et al. 2019) and PHASTER (Arndt et al. 2016) were used to identify prophage DNA. Moreover, similarities with other available phages were searched using NCBI's BLAST tool (Altschul et al. 1990).

Species identification

Digital DNA to DNA hybridization (dDDH) values were calculated using the type strain genome server (TYGS) (Meier-Kolthoff and Göker 2019). A phylogenomic tree was generated using the bootstrapping method with PhyloPhlAn 3.0.58 (Segata et al. 2013) and its internal database of circa 400 genes conserved across the bacterial domain.

Construction of a genome scale model

Organic-specific database development and metabolic pathways were identified using Pathway Tools 24.5 (Karp et al. 2015); an organism-specific database was built using annotated genomic data and the PathoLogic integrated tool. Genes involved in PHA metabolism were primarily searched using the PHA Depolymerase Engineering Database (Knoll et al. 2009) and NCBI BLAST tool (Altschul et al. 1990) in addition to a literature search.

Phenotype characterization

The quality and quantity of PHAs were determined by gas chromatography with a flame ionization detector (GC-FID) as previously described (Obruca et al. 2014). Residual 1,4BD and glycerol were determined by HPLC using a Shimadzu LC-10AD with a refractive index detector. Separation was performed on a Phenomenex Rezex ROA-Organic Acid column. A sulfuric acid solution of 0.005 M concentration was used as the mobile phase and the flow rate was set at 1 ml/min. The column temperature was set at 50 °C and the detector temperature was set at 40 °C.

Results

Genome assembly and properties

Aneurinibacillus thermoaerophilus CCM 8960 genome was assembled using both long reads and short reads as a hybrid approach with initial coverage of 870×. First, the initial sequence was assembled from 372,378 Oxford Nanopore Technologies (ONT) reads with a median read length of 3.23 kbp. Next, 1.1 million Illumina read pairs were mapped to the initial sequences (1.2 million paired reads in total—98% of all Illumina reads) with an average Phred score Q≈36. As a result, two circular sequences with coverage of 850×, corresponding to a circular chromosome and a circular plasmid, were retrieved and deposited at the DDBJ/EMBL/GenBank under accession numbers CP080764.1 and CP080765.1.

The genome length is 3,731,915 bp and contains 3840 open reading frames (ORFs), some of them organized into 668 polycistronic operons. While 3604 ORFs present protein-coding sequences (CDSs), 82 genes have corrupted ORFs and formed pseudogenes. The remaining loci correspond to RNA coding genes. The chromosomal GC content is 44.8%. The sequence features are summarized in Table 1.

Although sequencing data proved the plasmid pAT1 to be circular, no gene-coding replication-initiator protein

Table 1 Chromosomal and plasmidic features of *Aneurinibacillus thermoaerophilus* CCM 8960

Feature	Chromosome	Plasmid pAT1
Length (bp)	3,662,904	69,011
GC content (%)	44.8	41.9
ORF	3,745	95
Polycistronic operons	668	0
CDS	3,511	93
rRNA (5S, 16S, 23S)	11, 11, 12	0, 0, 0
tRNA	114	1
ncRNA	5	0

Rep was found. However, prophage prediction tools predicted several incomplete and putative active prophages covering almost all genes predicted for the pAT1 plasmid (see Table S1). By manual inspection, even the remaining two prophage unassigned loci, K3F53_18730 and K3F53_18735, were identified as prophage-associated genes coding small- and large-phage terminase subunits. Since the whole plasmid pAT1 carried the complete gene machinery of a prophage, it corresponded to the novel, hitherto unreported, phage. The start of the pAT1 plasmid sequence in a linear form was set according to the small-terminase subunit gene. A subset of the same prophage genes was also found in a draft genome of the type strain *Aneurinibacillus thermoaerophilus* DSM 10154 ^{T} (= L 420-91 ^{T}), GenBank accession no. FNDE00000000.1. Its whole contig 45 sequence containing prophage genes mapped to the pAT1 sequence.

Functional annotation

The functional annotation classified genes according to clusters of orthologous groups (COGs) into 20 categories. In total, 3592 chromosomal genes and 94 plasmidic genes were assigned to a COG category. Unfortunately, 35.77% of chromosomal genes and 80.85% of plasmidic genes were assigned to group S—function unknown. The second- and third-most prevalent COGs of chromosomal genes formed groups E (amino acid transport and metabolism) and L (replication, recombination, and repair) with 8.55% and 7.29% of all assigned genes, respectively. In contrast, plasmid contained groups K (transcription) and L, with 8.51% and 7.45% of all assigned genes in the second- and third-most prevalent COGs. Individual chromosomal and plasmidic features are shown in Fig. 1. COGs and RNAs are highlighted with different colors; RNAs are divided into tRNA, rRNA, and ncRNA categories and displayed in the fourth outermost circle.

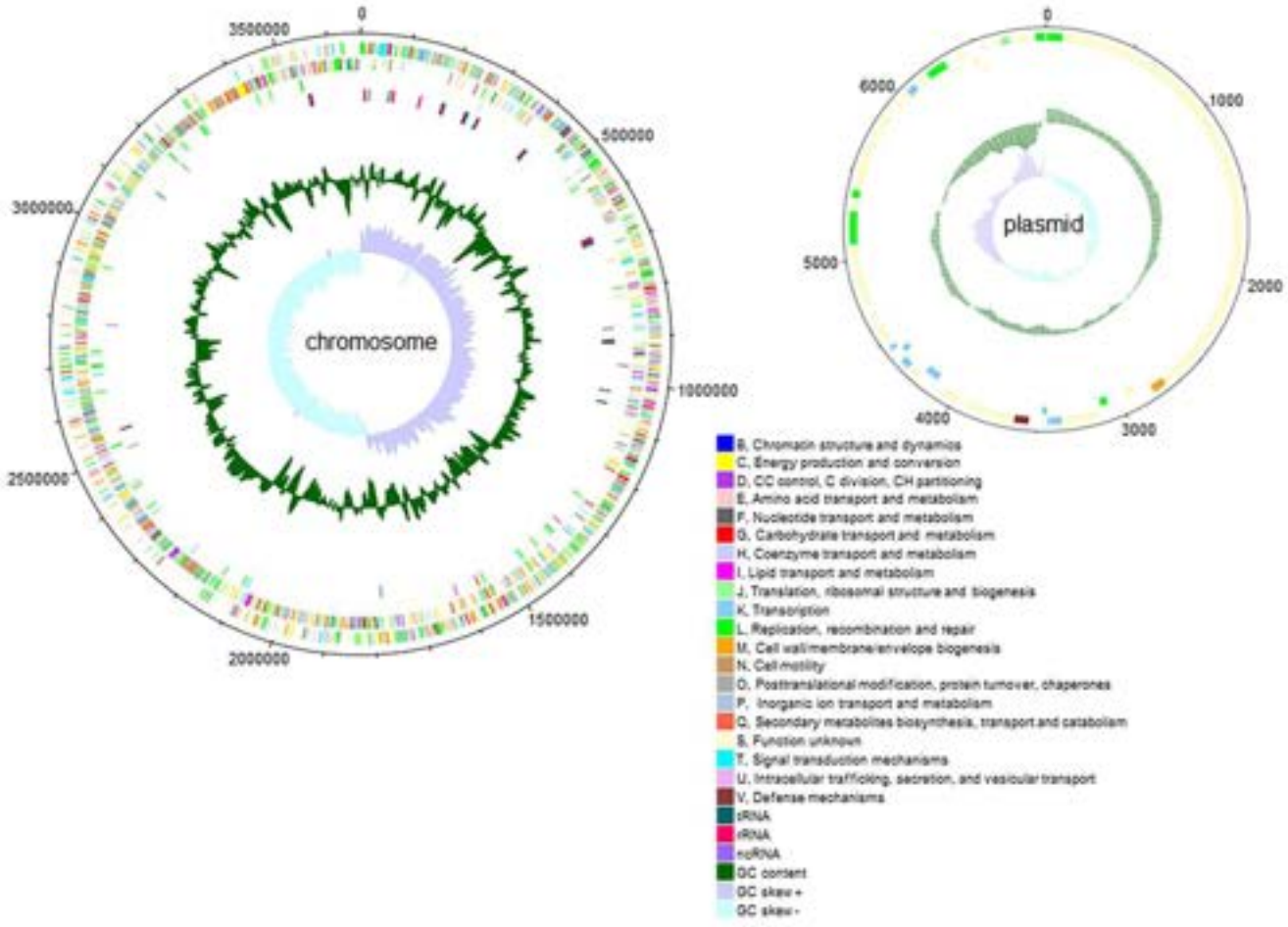


Fig. 1 Chromosomal maps of *Aneurinibacillus thermoaerophilus* CCM 8960 chromosome and plasmid. The first, second, and third outermost circles represent CDSs on the forward and backward strands, and pseudogenes, respectively. Classification of COGs is represented by colors. Next, RNA genes, distinguishing among tRNA, rRNA, and ncRNA, are represented in the fourth outermost circle. The inner area represents the GC content and GC skew

Analysis of CRISPR revealed five arrays in total, four of them belonging to chromosomal genome part (see Table S2), with arrays ranging from 228 to 4424 bp. Although cas-like genes have not been identified directly in the arrays, *cas6* (K3F53_02890), *cas7i* (K3F53_02900), *cas5b* (K3F53_02905), *cas4* (K3F53_02915), *cas1b* (K3F53_02920), and *cas2* (K3F53_02925) genes are located between the first and second chromosomal CRISPR arrays.

Detection of antibiotic-resistant genes against the Comprehensive Antibiotic Resistance Database (CARD) did not find perfect hits, but eight genes with sequence similarity of > 55% and length 100% ± 5% of known antibiotic-resistant genes were identified (see Table S3). Moreover, four different mechanisms of resistance (antibiotic target alteration, antibiotic target protection, antibiotic target replacement, and antibiotic efflux) were found, with the most prevalent group, antibiotic target alternation, having four genes.

In the subsequent annotation, we studied the presence of restriction-modification (R-M) systems based on methylation states of DNA using the Oxford Nanopore sequencing data. The study resulted in the detection of 438,889 5-methylcytosine (5-mC) methylations. Methylated motifs were further searched using the 17-mer sequences, including the detected modified base. In total, eight motifs were discovered; for five of them, the *p*-value of 0.05 was not exceeded. All motifs discovered are listed in the supplementary Table S4. Next, we detected R-M systems through the genome using The Restriction Enzyme Database (REBASE) (Roberts et al. 2015). As a result, ten systems were detected. Nine of them were located in the chromosome including types I, II, and IV; one type II system was uncovered in the plasmid. Details can be found in the supplementary Table S5. Subsequently, REBASE match the methylated motifs with R-M systems. Unfortunately, none of the discovered motifs were assigned to the system.

Species identification

Typing of the *A. thermoaerophilus* CCM 8960 genome using digital DNA-DNA hybridization (dDDH) resulted in a very high similarity of 98.2% to the type strain *Aneurinibacillus thermoaerophilus* DSM 10154 ^{T} (= L 420-91 ^{T}). Thus, we reidentified the strain formerly known as *Aneurinibacillus* sp. H1 as *A. thermoaerophilus* CCM 8960. The genome similarity to other type strains did not exceed the value of 36%. However, an even higher similarity of 98.4% was reached in comparison with the genome of *Aneurinibacillus* sp. XH2.

Phylogenomic analysis of both gram-positive and gram-negative PHA-producing bacteria classified *A. thermoaerophilus* CCM 8960 among G + species with the type strain *A. thermoaerophilus* DSM 10154 ^{T} and an unclassified strain *Aneurinibacillus sp.* XH2 being the closest relatives, as shown in Fig. 2. This outcome supported the previous dDDH results.

Phenotypic analysis

The phenotypic analysis of *Aneurinibacillus thermoaerophilus* CCM 8960 revealed 14 genes involved in PHA metabolism (see Table 2). Moreover, several of these genes were identified as being gathered in the operons. The first operon was created by genes *phaC*, *phbB*, and *phaR*; the second one covered *phaQ*, *phaP*, and *phaJ* genes. The third operon was formed by genes *fadN*, *phaA*, and *fadE*. *sucC* and *sucD* genes belong to the fourth identified operon.

The bacterium is capable of utilizing 1,4BD and glycerol. In Fig. 3, the carbon-substrate utilization can be observed where 1,4BD was used as the sole carbon source at a concentration of 4 g/L, glycerol was used as the sole carbon source at a concentration of 4 g/L, and a mixture of 1,4-butanediol and glycerol each at a concentration of 4 g/L was also used. The utilization of both substrates was analyzed by HPLC during the cultivation of the bacterium in the flasks. It is evident that glycerol was consumed more rapidly by the bacteria, both as a sole source and in the mixture with 1,4BD, being depleted within 36 h. In contrast, the progression of 1,4BD was much more gradual and was not completely depleted even after 72 h of cultivation. In the mixture, it can be observed that its rapid utilization occurred only after the glycerol was depleted from the medium which indicates the preference of glycerol prior to 1,4BD by the bacterium.

The course of carbon source utilization also affects the course of cellular growth and total PHA production, which is shown in Fig. 4. Depletion of glycerol in as the sole carbon source as early as 36 h results in a decrease in accumulated PHAs. For the mixture of substrates, the onset of copolymer formation can be observed only after glycerol depletion. Once 1,4BD is used as the sole carbon source (without glycerol), incorporation of 4HB into the PHAs chain occurs from the beginning of the cultivation, and 4HB reaches a very high content of 83 mol. %. However, when 1,4BD is used in combination with glycerol, the synthesis of the copolymer is achieved in the later stage of the cultivation and the portion of 4HB in copolymer reaches rather low values.

Thus, the fact that the bacterium is capable of growing in the presence of 1,4BD as a sole substrate proves the fact that 1,4BD serves not only as a precursor of 4HB but also as a carbon source that can fuel cell growth and metabolism. Based on experimental data and gene detection, a metabolic pathway that the bacterium is likely to use is proposed (see Fig. 5). Moreover, the database development and subsequent metabolic pathways reconstruction using Pathway Tools 24.5 (Karp et al. 2015) revealed a model containing four pathways with PHA-associated genes (see Table 2) and, thus, were associated with PHA metabolism. The pathways

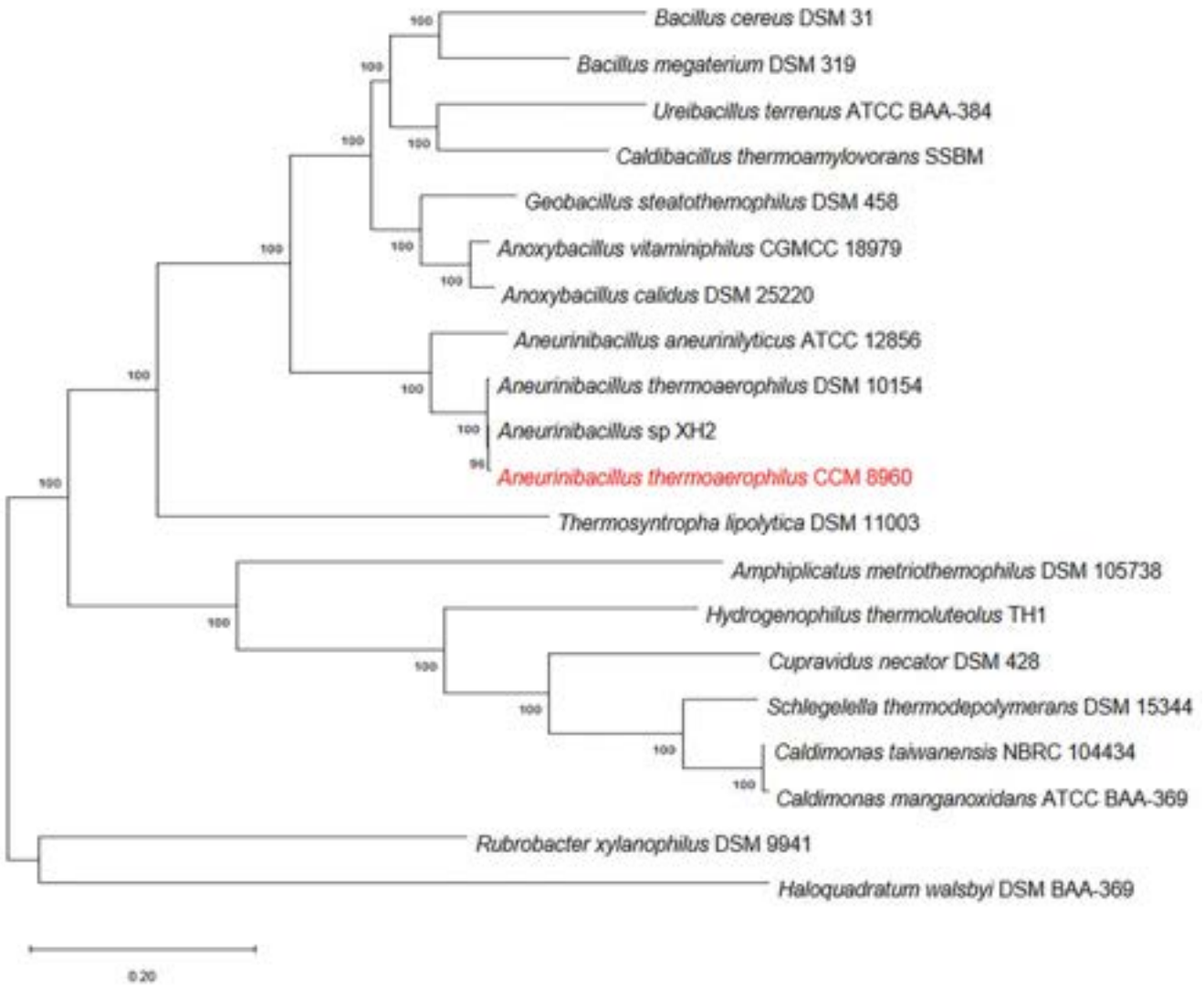


Fig. 2 Phylogenomic tree of PHA-producing bacteria and the placement of *Aneurinibacillus thermoaerophilus* CCM 8960 (highlighted in red). The tree was constructed using Phylophlan 3.0.58 together with its internal database of circa 400 genes conserved across the bacterial domain. The values represent the bootstrap support based on 100 replicates

include PHA biosynthesis, which is incorporated into PHA metabolism, displayed in Fig. 5. Besides the PHA metabolism, the model contains 209 pathways and 1098 enzymatic reactions. Unfortunately, 183 pathway holes occur in the 639 total reactions as well. The model was uploaded to the FAIRDOMHub (Wolstencroft et al. 2017) under the SEEK ID: fairdomhub.org/projects/269 and can be downloaded and imported into the Pathway Tools software.

Discussion

Genomic and phenotypic characterization of the polyhydroxyalkanoates producing strain *Aneurinibacillus thermoaerophilus* CCM 8960 may provide important insights toward the replacement of synthetic polymers with environmentally friendly polyesters. The genome assembly using the pipeline combining long Oxford Nanopore reads and short Illumina reads revealed chromosomal and plasmidic circular contigs. Although the presence of the circular plasmid was evident from the initial assembly of long reads by Flye tool, a series of high-quality short reads served as a necessary step of polishing the final assembly. Almost all short reads were unambiguously mapped to the final genome, confirming the correctness of the assembly. Moreover, origin of replication (oriC) prediction was unambiguous in the chromosome. Although the chromosomal GC content is slightly above the average for gram-positive bacteria (Li and Du 2014), as shown in Table 1, the value corresponds to the GC content of the type strain

Table 2 List of genes involved into P(3HB-*co*-4HB) metabolism of *Aneurinibacillus thermoaerophilus* CCM 8960

Locus tag	Gene name	Function
K3F53_05230	*phaZ*	Intracellular nPHAmcl depolymerase
K3F53_05285	*phaC*	PHA synthase
K3F53_05290	*phbB*	Acetoacetyl-CoA reductase
K3F53_05295	*phaR*	PHA synthase
K3F53_05300	*phaQ*	Transcriptional regulator
K3F53_05305	*phaP*	Phasin (PHA-granule associated protein) gene
K3F53_05310	*phaJ*	Specific enoyl-CoA hydratase (R-hydratase)
K3F53_08065	*fadN*	Acetyl-3-hydroxyacyl-CoA dehydrogenase
K3F53_08070	*phaA*	Acetyl-CoA C-acetyltransferase
K3F53_08075	*fadE*	Acyl-CoA dehydrogenase
K3F53_10675	*gabD*	Succinate semialdehyde dehydrogenase
K3F53_11680	*sucD*	Succinate–CoA ligase subunit alpha
K3F53_11685	*sucC*	ADP-forming succinate–CoA ligase subunit beta
K3F53_17725	*hbd*	3-Hydroxybutyryl-CoA dehydrogenase

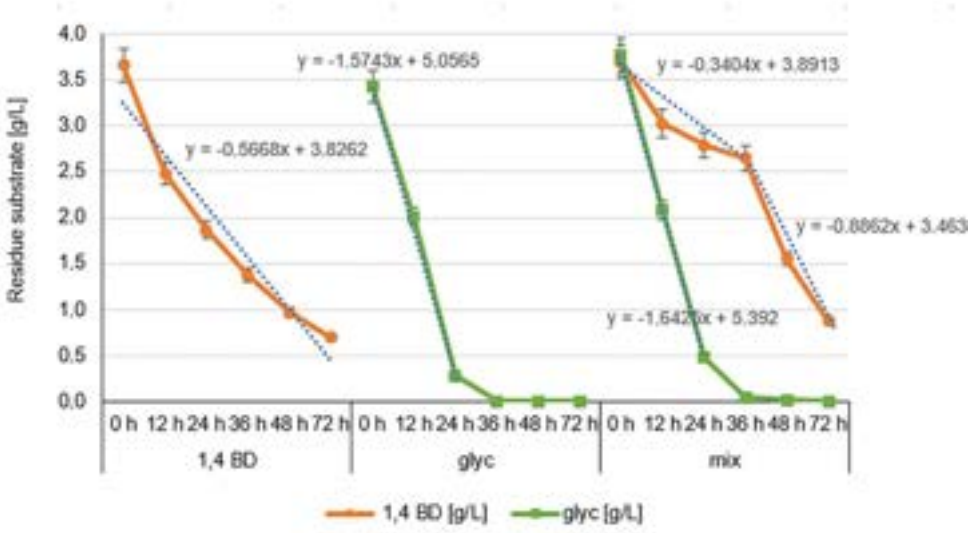


Fig. 3 The course of utilization of carbon sources. 1,4BD — 1,4-butanediol at 4 g/L; glyc — glycerol at 4 g/L; mix — mixture of 1,4-butanediol (4 g/L) and glycerol (4 g/L). The cultivation was performed in duplicate and each sample was analyzed twice as well. The error bars represent the standard deviation

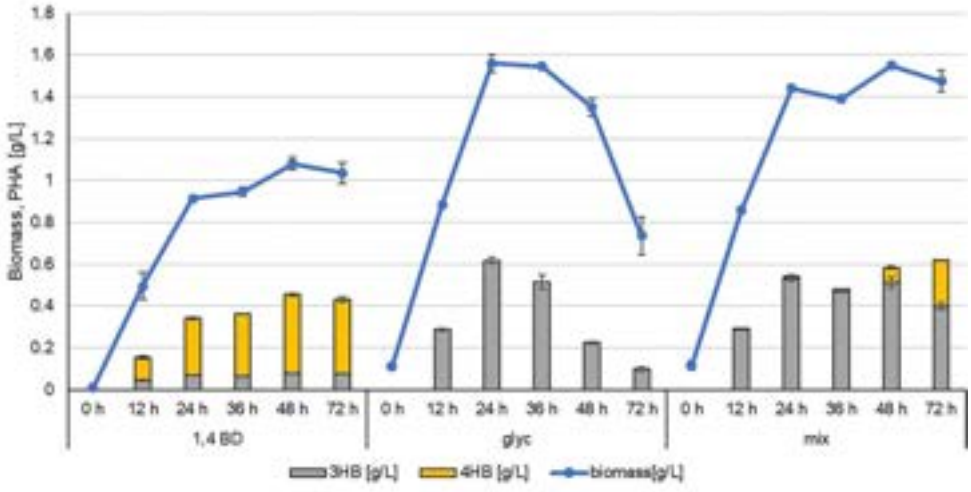


Fig. 4 Dependence of biomass and PHA concentration on cultivation time. 1,4BD — 1,4-butanediol at 4 g/L; glyc — glycerol at 4 g/L; mix — mixture of 1,4-butanediol (4 g/L) and glycerol (4 g/L). The cultivation was performed in duplicate and each sample was analyzed twice as well. The error bars represent the standard deviation

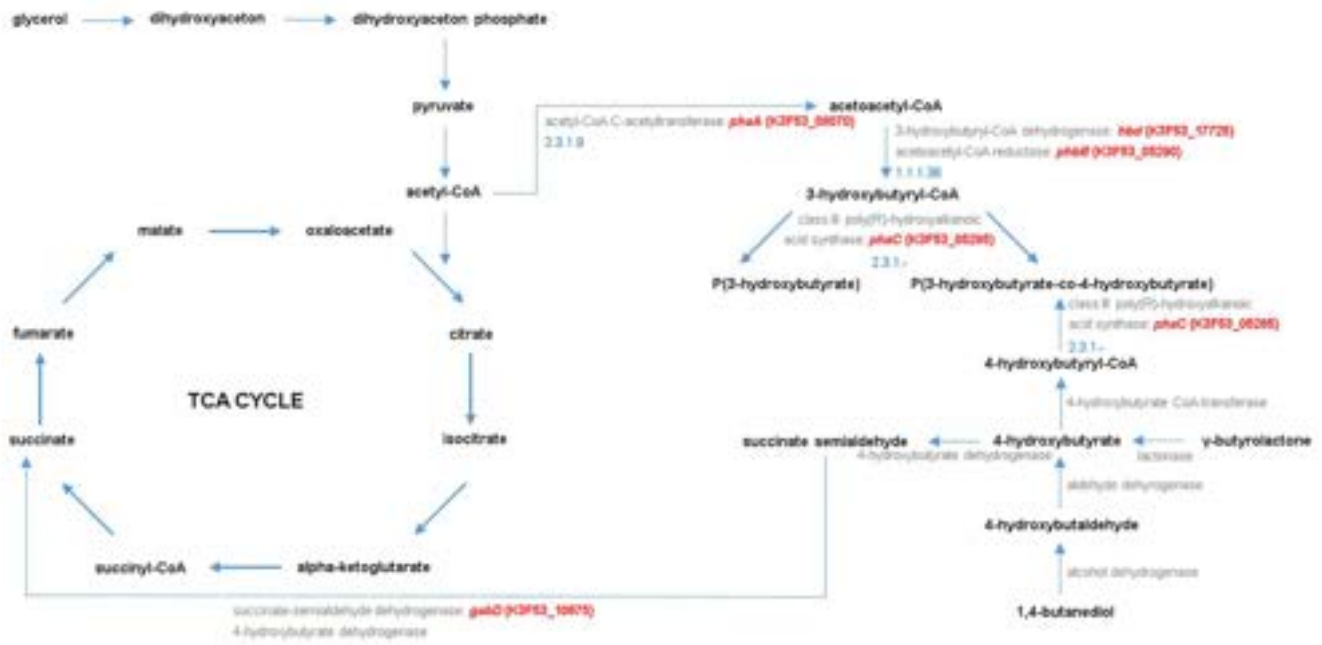


Fig. 5 PHA metabolism pathway of *Aneurinibacillus thermoaerophilus* CCM 8960. Genes involved in PHA metabolism and identified in the genome are highlighted in red. If associated, EC numbers are displayed under the list of the involved genes in blue

Aneurinibacillus thermoaerophilus DSM 10,194. Chromosome length 3.66 Mbp and a number of ORFs in a whole genome correspond to the type strain as well.

Plasmid prophages are very common in *Firmicutes*, as was recently demonstrated by Piligrimova et al. (2021). Although their identification is cumbersome, the quality of our assembly gives strong evidence of the existence of the plasmid prophage pAT1 in the *A. thermoaerophilus* CCM 8960 genome. Uniform coverage along the whole sequence and utilization of long reads, some of them even covering almost the whole plasmid, proved circularity of the sequence. The fact that plasmid prediction tools predicted more overlapping prophages on the plasmid was caused by the novelty of the found prophage and fully corresponded to the conclusions of the study by Piligrimova et al. (2021). Moreover, the prophage might not be unique for the studied strain, as the same genes were found in the assembly of the type strain. Unfortunately, due to the draft nature of the type strain's assembly, the presence of the prophage could not be confirmed for the type strain.

Functional annotation revealed key insights for understanding the organism, such as assigning function to 64% of chromosomal genes and 20% of plasmidic genes. However, *A. thermoaerophilus* CCM 8960 is a bacterium whose potential functional capacity is currently difficult to predict in detail because there are not enough similar model organisms described. Nevertheless, CRISPR and resistome identification detected a total of five CRISPR arrays and eight antibiotic-resistant genes against the Comprehensive Antibiotic Resistance Database (CARD).

The study of restriction-modification (R-M) led to the discovery of ten systems along the genome, one of which is located in the plasmid. In detail, three type I systems, considered the most complex type containing restriction, modification, and specificity enzymes, are located in the chromosome; two complete and one incomplete. Additionally, five occurrences of the type II systems including one located in the plasmid were revealed; however, only one located in the chromosome seems to be active as the remaining predicted methylases lacked coupled restriction endonucleases. Conversely, both observed type IV systems might be active as the type is formed only by a restriction enzyme that was identified in the chromosome. In addition, eight methylated motifs were discovered. However, the systems contained enzymes recognizing unknown motifs and, therefore, due to the number of possible combinations, we were not able to match observed motifs to predicted systems. Given the finding that the bacterium disposes of a strong system protecting it from foreign DNA, we can conclude that the genetic manipulation of this strain might be challenging (Kolek et al. 2016; Fu et al. 2017). These observations might be of crucial importance, with respect to further genome editing of the bacterium, to improve its PHA production potential.

Genome typing of the strain *A. thermoaerophilus* CCM 8960, formerly known as *Aneurinibacillus* sp. H1, revealed clear evidence of its belonging to the genus *A.*

thermoaerophilus as calculated dDDH value highly exceeded the 70% threshold. We used the d_4 formula, robust against the use of incomplete draft genomes, which is the case of the type strain *Aneurinibacillus thermoaerophilus* DSM 10154^T (=L 420-91^T) genome. Our results also showed that the thermophilic PHA-producing strain *Aneurinibacillus* sp. XH2, which is currently not completely taxonomically classified, belongs to the *A. thermoaerophilus* species, too. Low genome similarities to other type strains further support the uniqueness of the genome that was already showed by preceding COGs analysis (see Fig. 1). Phylogenomic tree construction further supports the bacterium's classification into the *A. thermoaerophilus* species, as the *A. thermoaerophilus* CCM 8960 formed the cluster together with other *A. thermoaerophilus* strains, namely XH2 and the type strain DSM 10,154 (see Fig. 2). Moreover, the bacterium belongs to the clade containing well-described PHA producers. The outcome corresponded to the published findings (Obruča et al. 2022). In the study, analysis of several hundred genes placed the strain in a taxonomic hierarchy of thermophilic bacteria, in which PHA production has been described or otherwise documented. The study's result clearly confirms that the bacterium clusters into G + bacteria together with other representatives of the genus *Aneurinibacillus*, and especially with the most similar strains of the *A. thermoaerophilus* species.

The phenotypic analysis provided key insights into PHA metabolism. A total of 14 genes associated with PHA metabolism were identified (see Table 2), including genes responsible for class IV PHA synthase, in particular its subunits encoded by *phaC* and *phaR* genes. In addition, PHA depolymerase genes were identified based on the PHA Depolymerase Engineering Database (Knoll et al. 2009). The presence of a *Bacillus megaterium*-type PHAs cluster described in the study by Kihara et al. (2017) was further verified. The cluster contains two operons formed of PHAs *phaC*, *phbB*, and *phaR*; *phaQ*, *phaP*, and *phaJ* genes which were revealed in the *A. thermoaerophilus* CCM 8960 as well. Moreover, four metabolic pathways were associated with PHA metabolism out of a total of 209 metabolic pathways discovered. The presence of 28.6% enzymatic reactions, for which no corresponding enzyme has been identified in the genome, further confirms the uniqueness of the newly described genome and the need for follow-up studies. Our results are, moreover, proven in the published work, which states that a new isolate in this work, classified as *A. thermoaerophilus* CCM 8960, is a promising PHAs producer. It is capable of producing a P(3HB-*co*-4HB) copolymer with a very high content of 4HB monomer, which provides the material with unique properties and expands its application potential (Pernicova et al. 2020a; Sedlacek et al. 2020). The ability to produce PHA was demonstrated both on the phenotypic and genotypic level, where genes involved in PHA metabolism, and especially the key enzyme PHA synthase, were identified. The class IV PHA synthase characteristic of the genus *Bacillus* and its related genera is composed of two subunits encoded by the *phaC* and *phaR* genes, both of which were found in the bacterial genome, as can be seen in the published work (Chen et al. 2015; Chek et al. 2017). PHA class IV synthase exhibits alcoholysis activity, resulting in a lower molecular weight of the resulting polymer (Tsuge et al. 2015).

The bacterium is also capable of utilizing glycerol, which it prefers to glucose. It is also capable of utilizing 4HB precursors as the sole carbon source and, in particular, 1,4BD (Figs. 3 and 4). Our experimental data indicate that the bacterium is capable of using both glycerol and 1,4BD to synthesize PHAs as well as sources of carbon and energy for other metabolic processes (Utsunomia et al. 2020). The data show that glycerol is the preferred substrate; it is utilized at a higher rate and is also used preferentially when both substrates are provided in the mixture. Moreover, glycerol can be converted to a 3HB monomer unit of PHA. Conversely, 1,4BD serves as a precursor of 4HB, and its introduction into cultivation media results in the formation of the valuable copolymer P(3HB-*co*-4HB). Our metabolic model in Fig. 5 describes some of the important metabolic nodes concerning fluxes of both glycerol and 1,4BD in *A. thermoaerophilus* CCM 8960. This knowledge might be of crucial importance for the improvement of the bacterium using genome editing tools in order to further improve the process of PHA production.

When there is enough glycerol, the bacterium uses the substrate for growth but also for the production of a storage product—PHA. When glycerol is depleted from the medium, the amount of PHA, specifically P(3HB), in the biomass also decreases as it is used as a carbon and energy source (Raza et al. 2018). This also verified the presence and functionality of intracellular depolymerases, which were again found in the bacterial genome. When 1,4BD is the only carbon source, the formation of the copolymer P(3HB-*co*-4HB) occurs. According to the proposed pathway, this means that, on the one hand, the bacteria use the substrate as a carbon source and, on the other hand, that it is metabolized in the Krebs cycle and related pathways. However, this too can be partially converted to acetyl-CoA (via its conversion to succinate and utilization of the Krebs cycle), which leads to the formation of the 3HB monomer. In addition, a proportion is directly incorporated into the polymer chain as the 4HB monomer. During cultivation, the ratio of 3HB to 4HB monomer does not change much, and the 4HB monomer is more abundant, at about 80 mol. %.

The dynamics of substrate utilization were observed for both the individual substrates and the mixture of substrates used (see Fig. 3). The utilization rate of 1,4BD alone is relatively low, reaching only 0.57 g/h, which is also related to the gradual increase in biomass concentration. The glycerol utilization rate is similar for both glycerol alone (1.57 g/h)

and glycerol in the mixture (1.64 g/h). Glycerol would be depleted at this rate within 36 h of cultivation. The dynamics of 1,4BD utilization in the mixture are different. If glycerol is present in the medium, the rate of 1,4BD utilization is only 0.34 g/h. After glycerol is depleted from the medium, the rate increases to 0.89 g/h. This is indicative of the preference of the carbonaceous substrate for the bacterium and is also related to the formation of the PHA copolymer.

In conclusion, the *Aneurinibacillus thermoaerophilus* CCM 8960 is a promising producer of eco-friendly, biodegradable polyesters demonstrated by genomic and phenotypic characterization. As mentioned, the bacterium is characterized by the unique properties of the genome that were described in this study for the first time. Therefore, further research is necessary to unravel the detailed characteristics of the organism, with the aim of biotechnological PHA production as some of its properties still remain hidden.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00253-022-12039-1.

Author contribution KS, SO, and ML conceived and designed the research. XK, IP, and MB conducted experiments. JM and KS analyzed data. JM, XK, IP, SO, and KS wrote the manuscript. All authors read and approved the manuscript.

Funding This study was supported by Brno University of Technology intra-university junior project FCH/FEKT-J-21–7331. PhD study of Jana Musilova is supported by the Brno Ph.D. Talent Scholarship – Funded by the Brno City Municipality.

Data availability The whole-genome sequence and plasmid sequence have been deposited in the DDBJ/ENA/GenBank under the accession numbers CP080764.1 and CP080765.1, respectively. The NCBI BioProject and BioSample IDs are PRJNA753004 and SAMN20669237. The raw reads have been deposited in the NCBI SRA database under the accession numbers SRR17854630 (paired-end Illumina) and SRR17854629 (Oxford Nanopore Technologies). The metabolic model is stored in the FAIRDOMHub database under the SEEK ID fairdomhub.org/projects/269.

Declarations

Ethical statement This article does not contain any studies with human participants or animals performed by any of the authors.

Conflict of interest The authors declare no competing interests.

References

Agarwala R, Barrett T, Beck J, Benson DA, Bollin C, Bolton E, Bourexis D, Brister JR, Bryant SH, Canese K, Charowhas C, Clark K, Dicuccio M, Dondoshansky I, Federhen S, Feolo M, Funk K, Geer LY, Gorelenkov V, Hoeppner M, Holmes B, Johnson M, Khotomlianski V, Kimchi A, Kimelman M, Kitts P, Klimke W, Krasnov S, Kuznetsov A, Landrum MJ, Landsman D, Lee JM, Lipman DJ, Lu Z, Madden TL, Madej T, Marchler-Bauer A, Karsch-Mizrachi I, Murphy T, Orris R, Ostell J, O'sullivan C, Panchenko A, Phan L, Preuss D, Pruitt KD, Rodarmer K, Rubinstein W, Sayers E, Schneider V, Schuler GD, Sherry ST, Sirotkin K, Siyan K, Slotta D, Soboleva A, Soussov V, Starchenko G, Tatusova TA, Todorov K, Trawick BW, Vakatov D, Wang Y, Ward M, Wilbur WJ, Yaschenko E, Zbicz K (2016) Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 44:D7. https://doi.org/10.1093/NAR/GKV1290

Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Bouchard M, Edalatmand A, Huynh W, Nguyen ALV, Cheng AA, Liu S, Min SY, Miroshnichenko A, Tran HK, Werfalli RE, Nasir JA, Oloni M, Speicher DJ, Florescu A, Singh B, Faltyn M, Hernandez-Koutoucheva A, Sharma AN, Bordeleau E, Pawlowski AC, Zubyk HL, Dooley D, Griffiths E, Maguire F, Winsor GL, Beiko RG, Brinkman FSL, Hsiao WWL, Domselaar GV, McArthur AG (2020) CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. Nucleic Acids Res 48:D517–D525. https://doi.org/10.1093/nar/gkz935

Alsafadi D, Al-Mashaqbeh O (2017) A one-stage cultivation process for the production of poly-3-(hydroxybutyrate-co-hydroxyvalerate) from olive mill wastewater by *Haloferax mediterranei*. N Biotechnol 34:47–53. https://doi.org/10.1016/J.NBT.2016.05.003

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, Wishart DS (2016) PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res 44:W16–W21. https://doi.org/10.1093/nar/gkw387

Bailey TL (2021) STREME: accurate and versatile sequence motif discovery. Bioinformatics 37:2834–2840. https://doi.org/10.1093/BIOINFORMATICS/BTAB203

Biswas A, Staals RHJ, Morales SE, Fineran PC, Brown CM (2016) CRISPRDetect: a flexible algorithm to define CRISPR arrays. BMC Genomics 17:356. https://doi.org/10.1186/s12864-016-2627-0

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Carver T, Thomson N, Bleasby A, Berriman M, Parkhill J (2009) DNAPlotter: circular and linear interactive genome visualization. Bioinformatics 25:119–120. https://doi.org/10.1093/bioinformatics/btn578

Chek MF, Kim SY, Mori T, Arsad H, Samian MR, Sudesh K (2017) Hakoshima T (2017) Structure of polyhydroxyalkanoate (PHA) synthase PhaC from Chromobacterium sp USM2, producing biodegradable plastics. Sci Reports 71(7):1–15. https://doi.org/10.1038/s41598-017-05509-4

Chen GQ, Hajnal I, Wu H, Lv L, Ye J (2015) Engineering biosynthesis mechanisms for diversifying polyhydroxyalkanoates. Trends Biotechnol 33:565–574. https://doi.org/10.1016/J.TIBTECH.2015.07.007

Chen GQ, Jiang XR (2018) Next generation industrial biotechnology based on extremophilic bacteria. Curr Opin Biotechnol 50:94–100. https://doi.org/10.1016/J.COPBIO.2017.11.016

Chen IMA, Chu K, Palaniappan K, Ratner A, Huang J, Huntemann M, Hajek P, Ritter S, Varghese N, Seshadri R, Roux S, Woyke T, Eloe-Fadrosh EA, Ivanova NN, Kyrpides NC (2021) The IMG/M data management and analysis system vol 6.0: new tools and advanced capabilities. Nucleic Acids Res 49:D751–D763. https://doi.org/10.1093/NAR/GKAA939

Coker JA (2016) Extremophiles and biotechnology current uses and prospects. F1000 Research 5:396

Ewels P, Magnusson M, Lundin S, Käller M (2016) MultiQC: summarize analysis results for multiple tools and samples in a single

report. Bioinformatics 32:3047–3048. https://doi.org/10.1093/bioinformatics/btw354

Fu P, Ge Y, Wu Y, Zhao N, Yuan Z, Hu X (2017) The LspC3–411 restriction-modification system is the major determinant for genetic manipulations of *Lysinibacillus sphaericus* C3–41. BMC Microbiol 17:116. https://doi.org/10.1186/s12866-017-1014-6

Ho MR, Chen C houh, Lin W chang (2010) Gene-oriented ortholog database: a functional comparison platform for orthologous loci. Database J Biol Databases Curation 2010. https://doi.org/10.1093/DATABASE/BAQ002

Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, Mende DR, Letunic I, Rattei T, Jensen LJ, Von Mering C, Bork P (2019) EggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res 47:D309–D314. https://doi.org/10.1093/nar/gky1085

Kanehisa M, Furumichi M, Sato Y, Ishiguro-Watanabe M, Tanabe M (2021) KEGG: integrating viruses and cellular organisms. Nucleic Acids Res 49:D545–D551. https://doi.org/10.1093/NAR/GKAA970

Karp PD, Billington R, Caspi R, Fulcher CA, Latendresse M, Kothari A, Keseler IM, Krummenacker M, Midford PE, Ong Q, Ong WK, Paley SM, Subhraveti P (2019) The BioCyc collection of microbial genomes and metabolic pathways. Brief Bioinform 20:1085–1093. https://doi.org/10.1093/BIB/BBX085

Karp PD, Paley SM, Midford PE, Krummenacker M, Billington R, Kothari A, Ong WK, Subhraveti P, Keseler IM, Caspi R (2015) Pathway Tools version 24.0: integrated software for pathway/genome informatics and systems biology. Brief Bioinform 11:40–79

Kihara T, Hiroe A, Ishii-Hyakutake M, Mizuno K, Tsuge T (2017) *Bacillus cereus*-type polyhydroxyalkanoate biosynthetic gene cluster contains R-specific enoyl-CoA hydratase gene. OUP 81:1627–1635. https://doi.org/10.1080/09168451.2017.1325314

Knoll M, Hamm TM, Wagner F, Martinez V, Pleiss J (2009) The PHA Depolymerase Engineering Database: a systematic analysis tool for the diverse family of polyhydroxyalkanoate (PHA) depolymerases. BMC Bioinforma 101(10):1–8. https://doi.org/10.1186/1471-2105-10-89

Kolek J, Sedlar K, Provaznik I, Patakova P (2016) Dam and Dcm methylations prevent gene transfer into *Clostridium pasteurianum* NRRL B-598: development of methods for electrotransformation, conjugation, and sonoporation. Biotechnol Biofuels 9:14. https://doi.org/10.1186/s13068-016-0436-y

Koller M, Maršálek L, de Sousa Dias MM, Braunegg G (2017) Producing microbial polyhydroxyalkanoate (PHA) biopolyesters in a sustainable manner. N Biotechnol 37:24–38. https://doi.org/10.1016/J.NBT.2016.05.001

Lanfear R, Schalamun M, Kainer D, Wang W, Schwessinger B (2019) MinIONQC: fast and simple quality control for MinION sequencing data. Bioinformatics 35:523–525. https://doi.org/10.1093/bioinformatics/bty654

Lee BT, Barber GP, Benet-Pagès A, Casper J, Clawson H, Diekhans M, Fischer C, Gonzalez JN, Hinrichs AS, Lee CM, Muthuraman P, Nassar LR, Nguy B, Pereira T, Perez G, Raney BJ, Rosenbloom KR, Schmelter D, Speir ML, Wick BD, Zweig AS, Haussler D, Kuhn RM, Haeussler M, Kent WJ (2022) The UCSC Genome Browser database: 2022 update. Nucleic Acids Res 50:D1115–D1122. https://doi.org/10.1093/NAR/GKAB959

Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997v2 [q-bio.GN]. https://doi.org/10.48550/arXiv.1303.3997

Li H (2018) Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34:3094–3100. https://doi.org/10.1093/bioinformatics/bty191

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li X-Q, Du D (2014) Variation, evolution, and correlation analysis of C+G content and genome or chromosome size in different kingdoms and phyla. PLoS ONE 9:e88339. https://doi.org/10.1371/journal.pone.0088339

Luo H, Quan C-L, Peng C, Gao F (2019) Recent development of Ori-Finder system and DoriC database for microbial replication origins. Brief Bioinform 20:1114–1124. https://doi.org/10.1093/bib/bbx174

Martin A, McMinn A (2018) Sea ice, extremophiles and life on extra-terrestrial ocean worlds. Int J Astrobiol 17:1–16. https://doi.org/10.1017/S1473550416000483

Meier-Kolthoff JP, Göker M (2019) TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. Nat Commun 101(10):1–10. https://doi.org/10.1038/s41467-019-10210-3

Ni P, Huang N, Zhang Z, Wang DP, Liang F, Miao Y, Le XC, Luo F, Wang J (2019) DeepSignal: detecting DNA methylation state from Nanopore sequencing reads using deep-learning. Bioinformatics 35:4586–4595. https://doi.org/10.1093/BIOINFORMATICS/BTZ276

Obruca S, Benesova P, Oborna J, Marova I (2014) Application of protease-hydrolyzed whey as a complex nitrogen source to increase poly(3-hydroxybutyrate) production from oils by *Cupriavidus necator*. Biotechnol Lett 36:775–781. https://doi.org/10.1007/S10529-013-1407-Z

Obruča S, Dvořák P, Sedláček P, Koller M, Sedlář K, Pernicová I, Šafránek D (2022) Polyhydroxyalkanoates synthesis by halophiles and thermophiles: towards sustainable production of microbial bioplastics. Biotechnol Adv 107906.https://doi.org/10.1016/J.BIOTECHADV.2022.107906

Obruca S, Sedlacek P, Koller M, Kucera D, Pernicova I (2018) Involvement of polyhydroxyalkanoates in stress resistance of microbial cells: biotechnological consequences and applications. Biotechnol Adv 36:856–870

Obulisamy PK, Mehariya S (2021) Polyhydroxyalkanoates from extremophiles: a review. Bioresour Technol 325.https://doi.org/10.1016/J.BIORTECH.2020.124653

Pernicova I, Kucera D, Nebesarova J, Kalina M, Novackova I, Koller M, Obruca S (2019) Production of polyhydroxyalkanoates on waste frying oil employing selected *Halomonas* strains. Bioresour Technol 292:122028. https://doi.org/10.1016/J.BIORTECH.2019.122028

Pernicova I, Novackova I, Sedlacek P, Kourilova X, Kalina M, Kovalcik A, Koller M, Nebesarova J, Krzyzanek V, Hrubanova K, Masilko J, Slaninova E, Obruca S (2020) Introducing the newly isolated bacterium Aneurinibacillus sp H1 as an auspicious thermophilic producer of various polyhydroxyalkanoates (PHA) copolymers-1 isolation and characterization of the bacterium. Polymers (Basel) 12:1235. https://doi.org/10.3390/POLYM12061235

Pernicova I, Novackova I, Sedlacek P, Kourilova X, Koller M, Obruca S (2020b) Application of osmotic challenge for enrichment of microbial consortia in polyhydroxyalkanoates producing thermophilic and thermotolerant bacteria and their subsequent isolation. Int J Biol Macromol 144:698–704. https://doi.org/10.1016/J.IJBIOMAC.2019.12.128

Piligrimova EG, Kazantseva OA, Kazantsev AN, Nikulin NA, Skorynina AV, Koposova ON (2021) Shadrin AM (2021) Putative plasmid prophages of *Bacillus cereus* sensu lato may hold the key to undiscovered phage diversity. Sci Reports 111(11):1–16. https://doi.org/10.1038/s41598-021-87111-3

Raza ZA, Abid S, Banat IM (2018) Polyhydroxyalkanoates: characteristics, production, recent developments and applications. Int

Biodeterior Biodegradation 126:45–56. https://doi.org/10.1016/J.IBIOD.2017.10.001

Roberts RJ, Vincze T, Posfai J, Macelis D (2015) REBASE—a database for DNA restriction and modification: enzymes, genes and genomes. Nucleic Acids Res 43:D298–D299. https://doi.org/10.1093/nar/gku1046

Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream M-A, Barrell B (2000) Artemis: sequence visualization and annotation. Bioinformatics 16:944–945. https://doi.org/10.1093/bioinformatics/16.10.944

Safran M, Rosen N, Twik M, BarShir R, Stein TI, Dahary D, Fishilevich S, Lancet D (2021) The GeneCards Suite. Pract Guid to Life Sci Databases 27–56.https://doi.org/10.1007/978-981-16-5812-9_2

Sedlacek P, Pernicova I, Novackova I, Kourilova X, Kalina M, Kovalcik A, Koller M, Nebesarova J, Krzyzanek V, Hrubanova K, Masilko J, Slaninova E, Trudicova M, Obruca S (2020) Introducing the newly isolated bacterium Aneurinibacillus sp H1 as an auspicious thermophilic producer of various polyhydroxyalkanoates (PHA) copolymers-2 Material study on the produced copolymers. Polymers (Basel) 12:1298. https://doi.org/10.3390/POLYM12061298

Segata N, Börnigen D, Morgan XC, Huttenhower C (2013) PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. Nat Commun 4:2304. https://doi.org/10.1038/ncomms3304

Simon-Colin C, Raguénès G, Cozien J, Guezennec JG (2008) *Halomonas profundus* sp. nov., a new PHA-producing bacterium isolated from a deep-sea hydrothermal vent shrimp. J Appl Microbiol 104:1425–1432. https://doi.org/10.1111/J.1365-2672.2007.03667.X

Song W, Sun H-X, Zhang C, Cheng L, Peng Y, Deng Z, Wang D, Wang Y, Hu M, Liu W, Yang H, Shen Y, Li J, You L, Xiao M (2019) Prophage Hunter: an integrative hunting tool for active prophages. Nucleic Acids Res 47:W74–W80. https://doi.org/10.1093/nar/gkz380

Stephens ZD, Lee SY, Faghri F, Campbell RH, Zhai C, Efron MJ, Iyer R, Schatz MC, Sinha S, Robinson GE (2015) Big data: astronomical or genomical? PLoS Biol 13:e1002195. https://doi.org/10.1371/journal.pbio.1002195

Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J (2016) NCBI prokaryotic genome annotation pipeline. Nucleic Acids Res 44:6614–6624. https://doi.org/10.1093/nar/gkw569

Tsuge T, Hyakutake M, Mizuno K (2015) Class IV polyhydroxyalkanoate (PHA) synthases and PHA-producing *Bacillus*. Appl Microbiol Biotechnol 99:6231–6240. https://doi.org/10.1007/S00253-015-6777-9/FIGURES/5

Utsunomia C, Ren Q, Zinn M (2020) Poly(4-hydroxybutyrate): current state and perspectives. Front Bioeng Biotechnol 8:257. https://doi.org/10.3389/FBIOE.2020.00257/BIBTEX

Vaser R, Sović I, Nagarajan N, Šikić M (2017) Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res 27:737–746. https://doi.org/10.1101/gr.214270.116

Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM (2014) Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS ONE 9:e112963. https://doi.org/10.1371/journal.pone.0112963

Wolstencroft K, Krebs O, Snoep JL, Stanford NJ, Bacall F, Golebiewski M, Kuzyakiv R, Nguyen Q, Owen S, Soiland-Reyes S, Straszewski J, Van Niekerk DD, Williams AR, Malmström L, Rinn B, Müller W, Goble C (2017) FAIRDOMHub: a repository and collaboration environment for sharing systems biology research. Nucleic Acids Res 45:D404–D407. https://doi.org/10.1093/NAR/GKW1032

Zeldes BM, Keller MW, Loder AJ, Straub CT, Adams MWW, Kelly RM (2015) Extremely thermophilic microorganisms as metabolic engineering platforms for production of fuels and industrial chemicals. Front Microbiol 6. https://doi.org/10.3389/fmicb.2015.01209