Phenotypic and genomic analysis of isopropanol and 1,3-propanediol producer Clostridium diolis DSM 15410

SEDLÁŘ, K.; VASYLKIVSKA, M.; MUSILOVÁ, J.; BRANSKÁ, B.; PROVAZNÍK, I.; PATÁKOVÁ, P.

Genomics

2021, vol. 113, iss. 1, part 2, January 2021, pp. 1109-1119

ISSN: 0888-7543

DOI: https://doi.org/10.1016/j.ygeno.2020.11.007

Accepted manuscript

Research Article

Phenotypic and genomic analysis of isopropanol and 1,3-propanediol producer *Clostridium diolis* DSM 15410

Karel Sedlar[1,*], Maryna Vasylkivska[2], Jana Musilova[1], Barbora Branska[2], Ivo Provaznik[1], Petra Patakova[2]

[1] Department of Biomedical Engineering, Faculty of Electrical Engineering and Communication, Brno University of Technology, Technicka 12, Brno, Czechia

[2] Department of Biotechnology, University of Chemistry and Technology Prague, Technicka 5, 166 28 Prague, Czechia

*corresponding author, e-mail: sedlar@feec.vutbr.cz, tel: +420 54114 6670

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.

Keywords

Clostridium diolis; complete genome; solventogenic; IBE; 1,3-propanediol

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 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 1L 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.

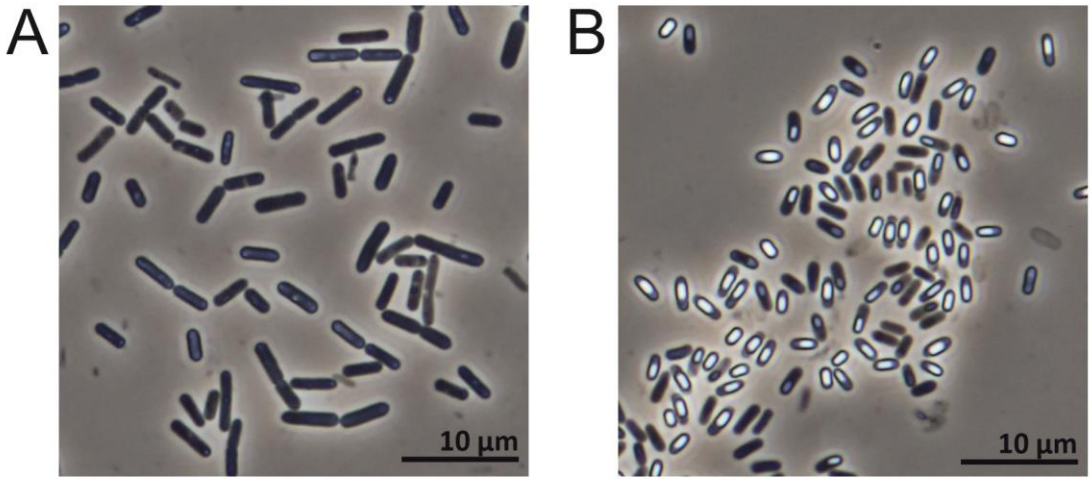


Fig. 1. Vegetative cells (A) and spores (B) of *C. diolis* DSM 15410 obtained in RCM medium after 24 and 48 h of cultivation, respectively.

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

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].

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].

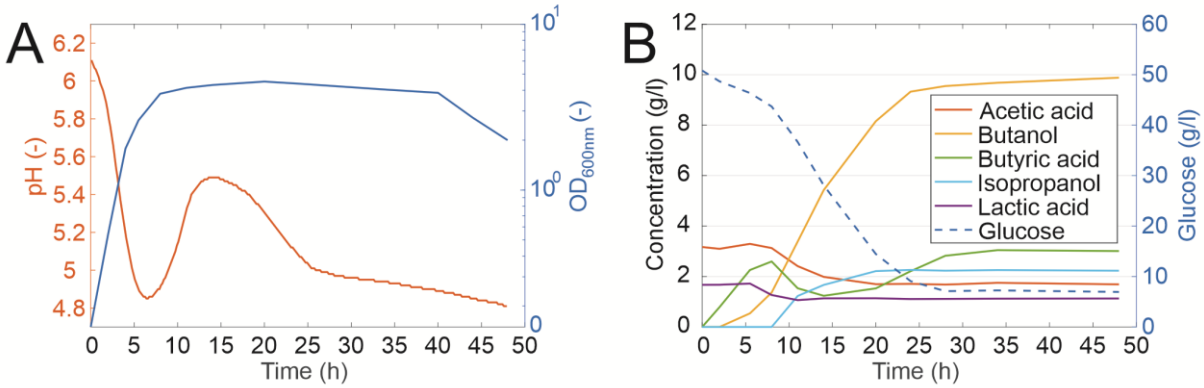


Fig. 2. The fermentation characteristics of *C. diolis* DSM 15410 cultivated in a bioreactor in an RCM medium containing glucose. (A) The growth curve represented by OD_{600nm} and pH course. (B) The glucose consumption and production of metabolites. The standard error of the measurements did not exceed 5 %.

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 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 chromosome 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 5,244 annotated open reading frames (ORFs) divided into 3,224 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 3,516 CDSs (out of all 5,102 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	5,244
Total number of operons	3,224
Protein coding genes	4,955
Pseudogenes	147
rRNA genes (5S, 16S, 23S)	15, 14, 14
tRNA	93
ncRNA	6

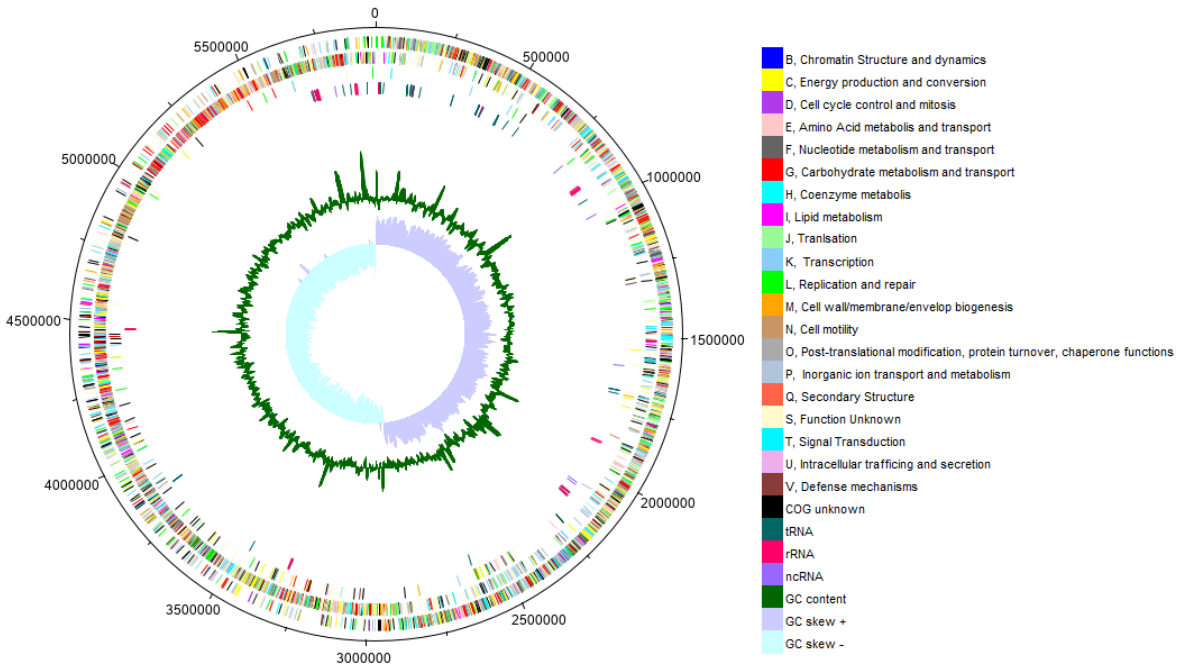


Fig. 3: A circular chromosome map of the *C. diolis* DSM 15410 genome. The outermost and second outermost circles represent CDSs on the forward and reverse strands, respectively. The third circle represents pseudogenes and the colors represent the COG functional classification. The fourth circle represents 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.

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 [58]). We used PacBio

sequencing data to study methylations on a genome wide scale. We detected 1,217 m6A and 2,023 m4C methylated positions and additional 40,787 modified bases in total. Roughly half of the detected methylations (622 m6A and 610 m5C) and 6,773 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. [59] 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. [59] meanwhile proved that *C. diolis* DSM 15410 can be transformed using DNA pre-methylated by methylatransferases (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 [60]. 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* [61]. 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, 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 [62]. 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 [62], 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.

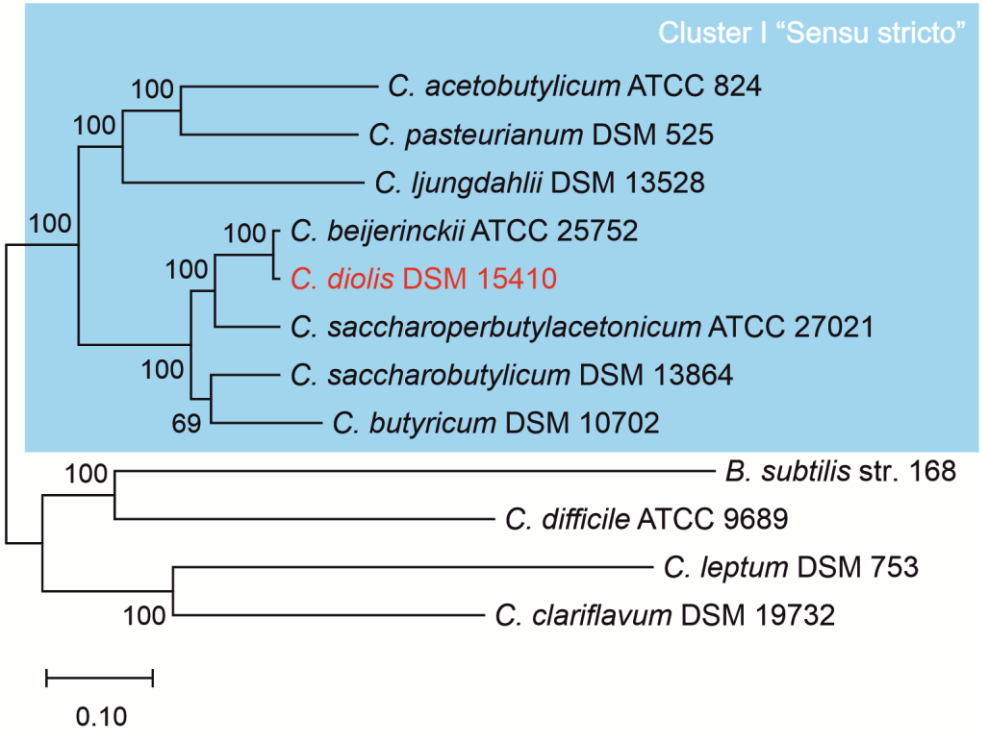


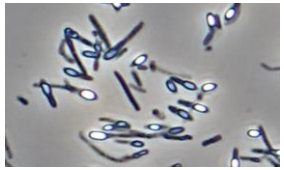
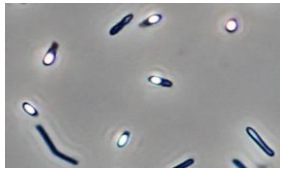
Fig 4. The phylogenetic placement of *C. diolis* DSM 15410. The blue rectangle contains species from the cluster I "Sensu stricto" of the genus *Clostridium*. The tree was constructed using Phylophlan 3.0, using its internal database of circa 400 genes conserved across bacterial domain. The values represent the boostrap support based on 100 replicates.

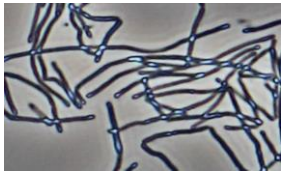
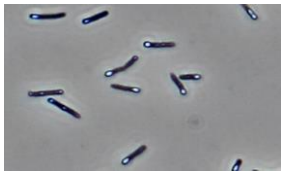
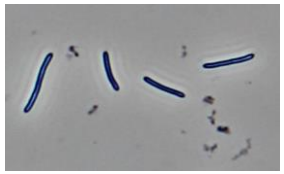
Digital DNA to DNA hybridization (dDDH) analysis using type strain genome server (TYGS) [63] and newly assembled complete genome confirmed that type strains *C. diolis* DSM 15410 and *C. beijerinckii* DSM 791 reached value of 85.2% confirming they are the same species. Nevertheless, using known complete 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.

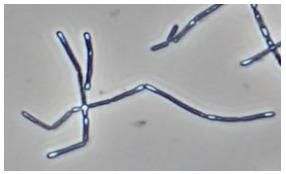
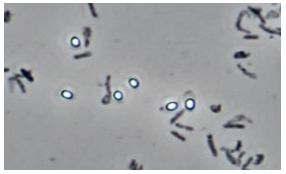
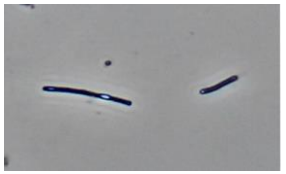
We performed comparative cultivation experiment of *C. diolis* DSM 15410 with two closely-related *C. beijerinckii* strains, *C. beijerinckii* NCIMB 8052 and *C. beijerinckii* NRRL B-598 (former *C. pasteurianum* NRRL B-598 [64]), see Table 3, which revealed some differences between them and *C. diolis* 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 [65] or isopropanol production by *C. beijerinckii* DSM 6423 or *C. beijerinckii* BGS1 [48,66]. 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.

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

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. 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)**								

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)**
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

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 [67] and *C. beijerinckii* NCIMB 8052 [15,68] 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* [69]. 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 [70]. 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) [71–73].

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 [74], 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.

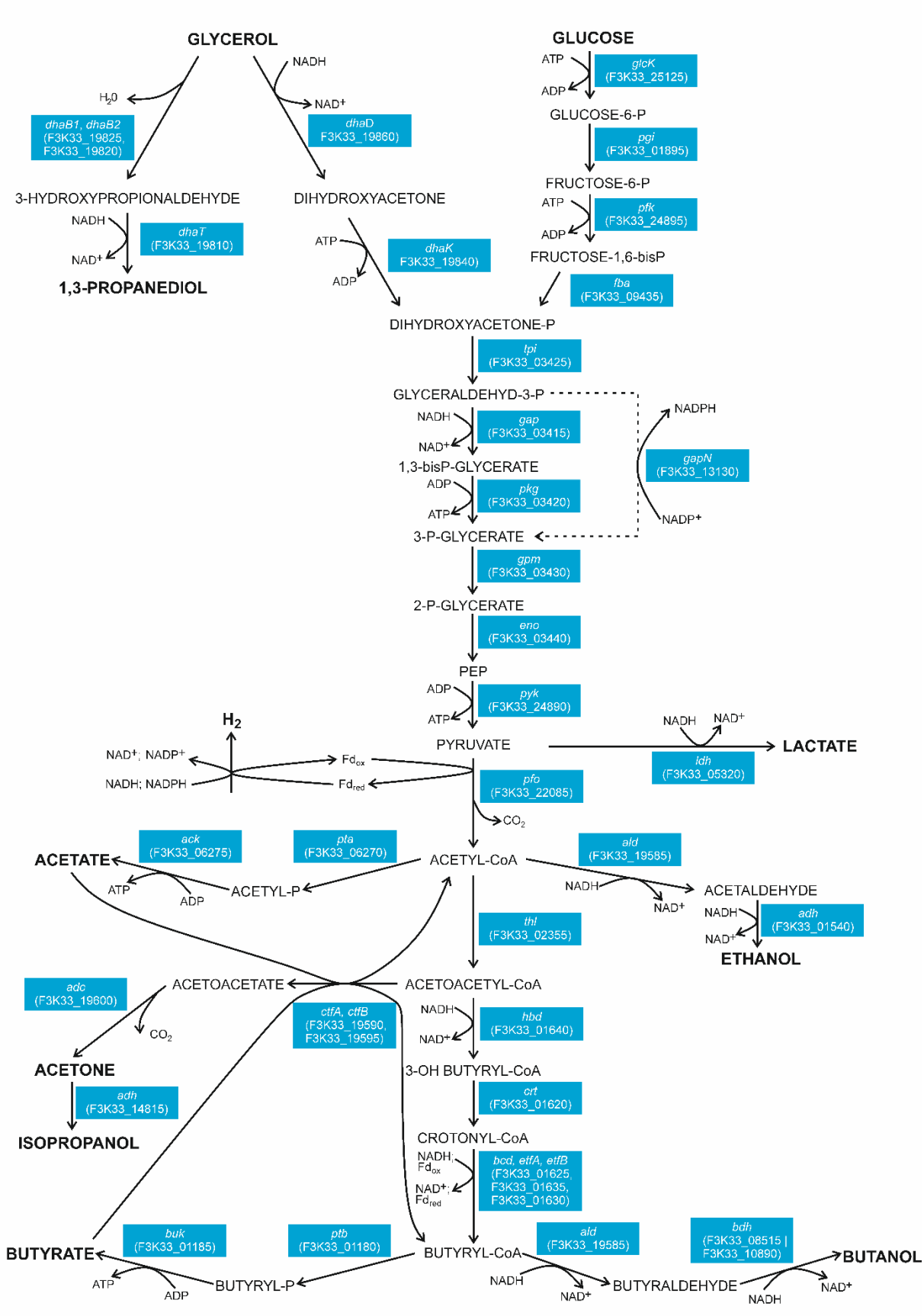


Fig. 5. Central metabolism of *Clostridium diolis* DSM 15410. Main substrates (glycerol, glucose) and main products (1,3-propanediol, H_2, lactate, acetate, acetone, ethanol, isopropanol,

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.

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 [59] 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 [75]. Expression of the pathway is more often studied in *E. coli*, where genes from the *Klebsiella pneumoniae* are usually expressed [76,77]. Nevertheless, also genes from *C. butyricum* were successfully used [78]. 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 [74]. 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*.

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 5,244 ORFs divided into 3,224 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. beijerickii* 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 17-00551S.

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.

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. doi:10.1016/j.renene.2020.01.151.

[2] B. Branska, L. Fortova, M. Dvorakova, 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. doi: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. doi: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. doi: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. doi: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. doi: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. doi: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. doi: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. doi: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. doi: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. doi:10.1099/ijsem.0.000824.

[12] K. Sedlar, J. Kolek, I. Provaznik, P. Patakova, Reclassification of non-type strain Clostridium pasteurianum NRRL B-598 as Clostridium beijerinckii NRRL B-598, J. Biotechnol. 244 (2017) 1–3. doi: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. doi: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) 4823–38. doi: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. doi:10.1186/1471-2164-12-479.

[16] K. Sedlar, J. Kolek, H. Skutkova, B. Branska, I. Provaznik, P. Patakova, Complete genome sequence of Clostridium pasteurianum NRRL B-598, a non-type strain producing butanol, J. Biotechnol. 214 (2015) 113–114. doi: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) e01591-14. doi: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. doi: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. doi: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. doi: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. doi: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. doi: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). doi: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. doi:10.1007/s00449-012-0700-x.

[25] P. Ewels, M. Magnusson, S. Lundin, M. Kaller, MultiQC: Summarize analysis results for multiple tools and samples in a single report, Bioinformatics. 32 (2016) 3047–3048. doi: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. doi: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. doi: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. doi:10.1093/bioinformatics/btv465.

[29] H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform., Bioinformatics. 25 (2009) 1754–1760. doi: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. doi: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. doi: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. doi: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. doi: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. doi: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. doi: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. doi: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. doi: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. doi: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. doi: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. doi: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. doi: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. doi: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. doi: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. doi:10.1186/s13068-018-1096-x.

[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. doi:10.1007/s00253-011-3322-3.

[46] H. Máté 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. doi:10.1186/s12864-018-4636-7.

[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. doi: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. doi: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. doi: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. doi: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. doi: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. doi: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. doi: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. doi: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. doi: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. doi: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. doi:10.1186/s13068-016-0436-y.

[58] K. Sedlar, J. Kolek, I. Provaznik, P. Patakova, Reclassification of non-type strain Clostridium pasteurianum NRRL B-598 as Clostridium beijerinckii NRRL B-598, J. Biotechnol. 244 (2017) 1–3. doi:10.1016/j.jbiotec.2017.01.003.

[59] 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. doi:10.1016/j.biortech.2020.123066.

[60] R. Sorek, C.M. Lawrence, B. Wiedenheft, CRISPR-Mediated Adaptive Immune Systems in Bacteria and Archaea, Annu. Rev. Biochem. 82 (2013) 237–266. doi:10.1146/annurev-biochem-072911-172315.

[61] 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. doi:10.1021/acssynbio.6b00060.

[62] P. Cruz-Morales, C.A. Orellana, G. Moutafis, G. Moonen, G. Rincon, L.K. Nielsen, E. Marcellin, E. Bapteste, Revisiting the Evolution and Taxonomy of Clostridia, a Phylogenomic Update, Genome Biol. Evol. 11 (2019) 2035–2044. doi:10.1093/gbe/evz096.

[63] 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. doi:10.1038/s41467-019-10210-3.

[64] K. Sedlar, J. Kolek, I. Provaznik, P. Patakova, Reclassification of non-type strain *Clostridium pasteurianum* NRRL B-598 as *Clostridium beijerinckii* NRRL B-598, J. Biotechnol. 244 (2017) 1–3. doi:10.1016/j.jbiotec.2017.01.003.

[65] 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. doi:10.1016/j.biortech.2016.04.020.

[66] 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 and Bioenergy. 128 (2019) 105331. doi:10.1016/j.biombioe.2019.105331.

[67] 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. doi:10.1038/s41598-018-37679-0.

[68] 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. doi:10.1186/1471-2164-13-102.

[69] 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. doi:10.1186/s13068-017-0742-z.

[70] 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) e02656-18. doi:10.1128/AEM.02656-18.

[71] 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. doi:10.1128/AEM.06382-11.

[72] 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. doi:10.1186/s13068-018-1024-0.

[73] 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. doi:10.1186/1754-6834-5-44.

[74] 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. doi:10.1073/pnas.0734105100.

[75] 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. doi:10.1016/j.ymben.2005.06.001.

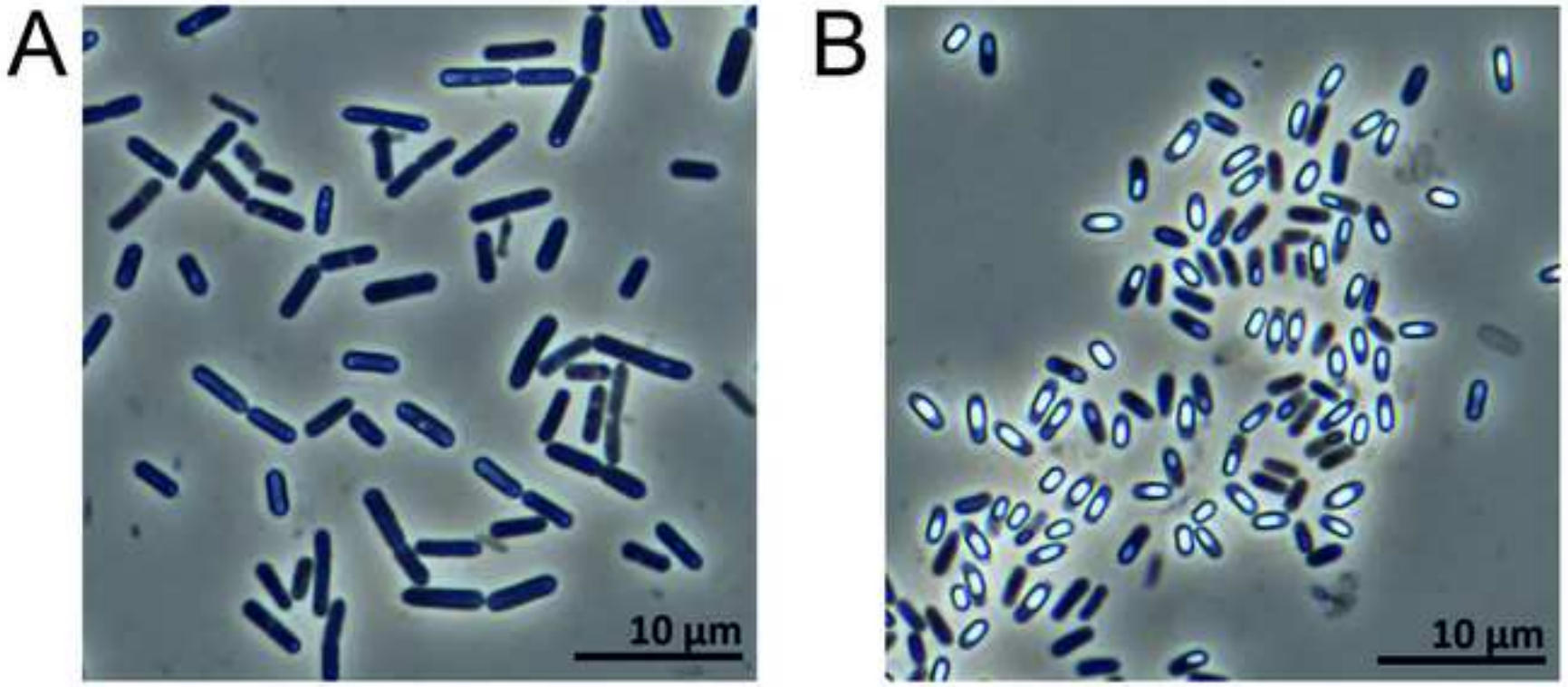
[76] I.T. Tong, H.H. Liao, D.C. Cameron, 1,3-Propanediol production by *Escherichia coli* expressing genes from the *Klebsiella pneumoniae dha* regulon, Appl. Environ. Microbiol. 57 (1991) 3541–3546.

[77] 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. doi:10.1007/s12257-018-0017-y.

[78] 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. doi:10.1128/AEM.02376-08.

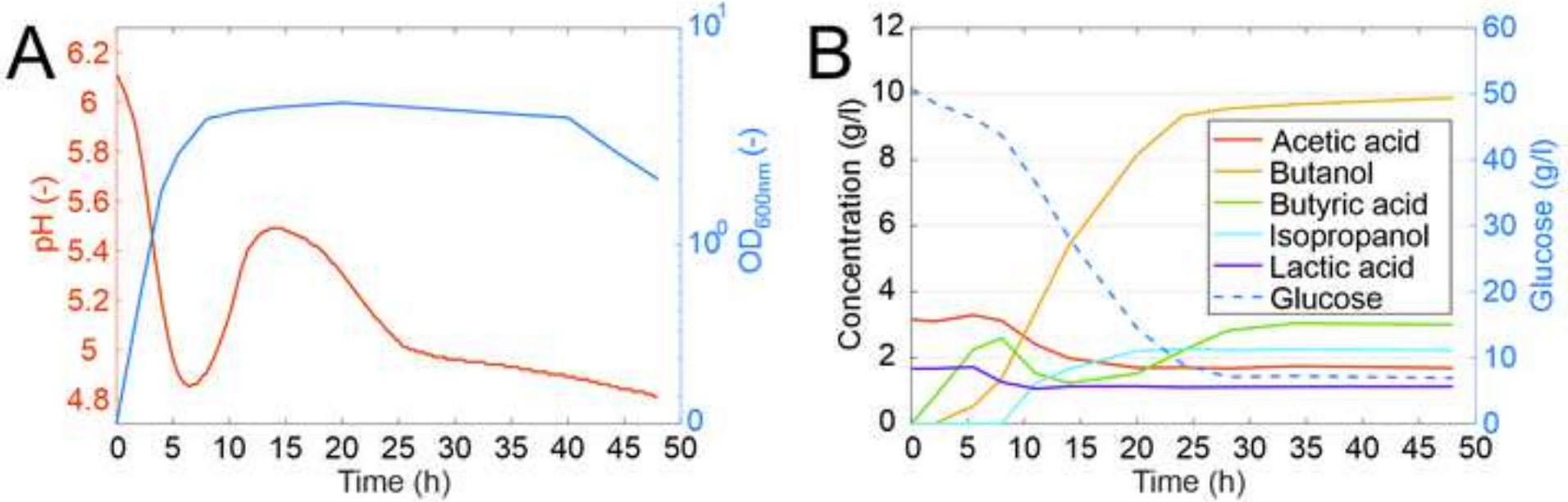
Figure

Click here to access/download;Figure;Fig1.tif



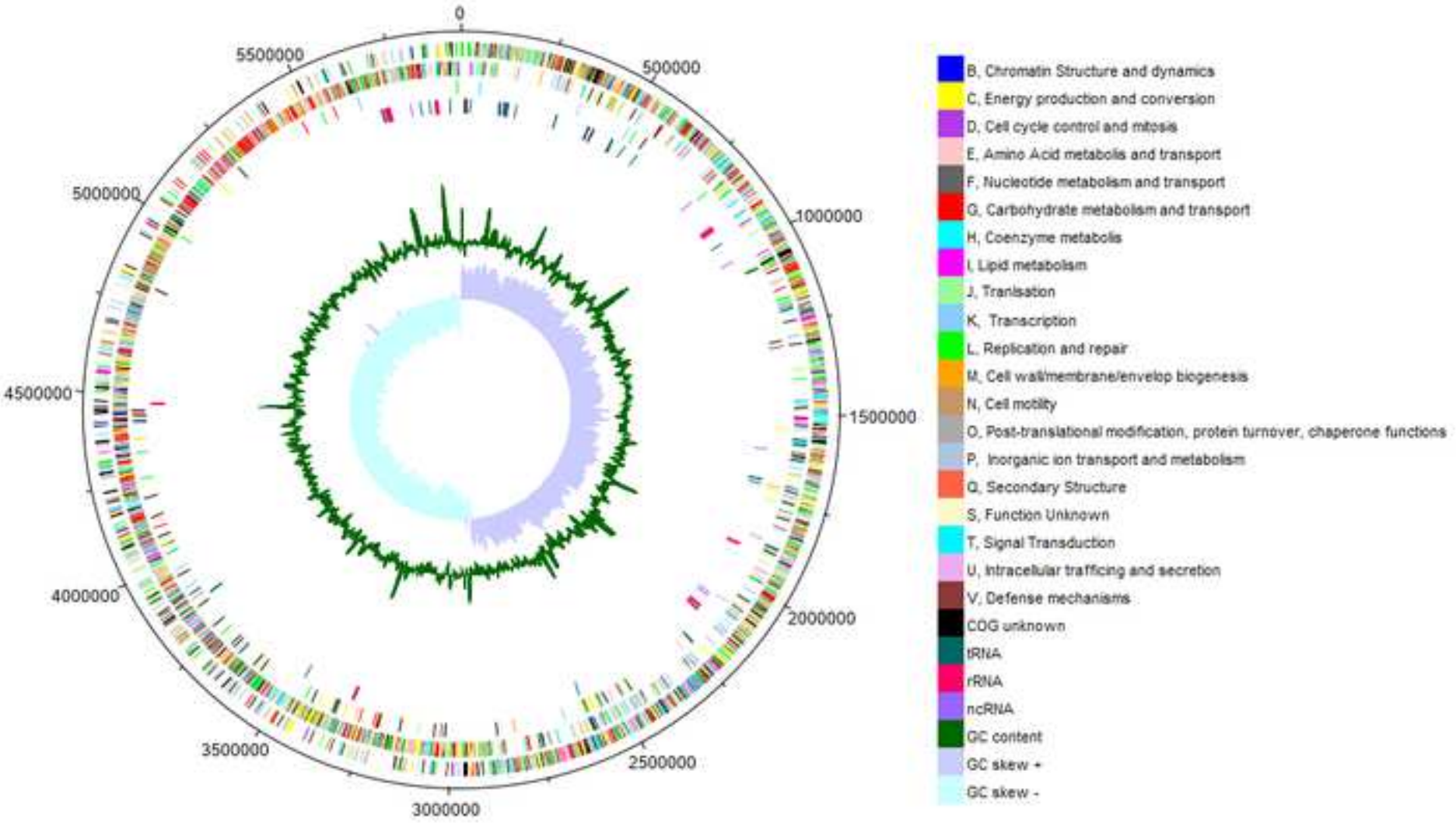
Figure

Click here to access/download;Figure;fig2.tif



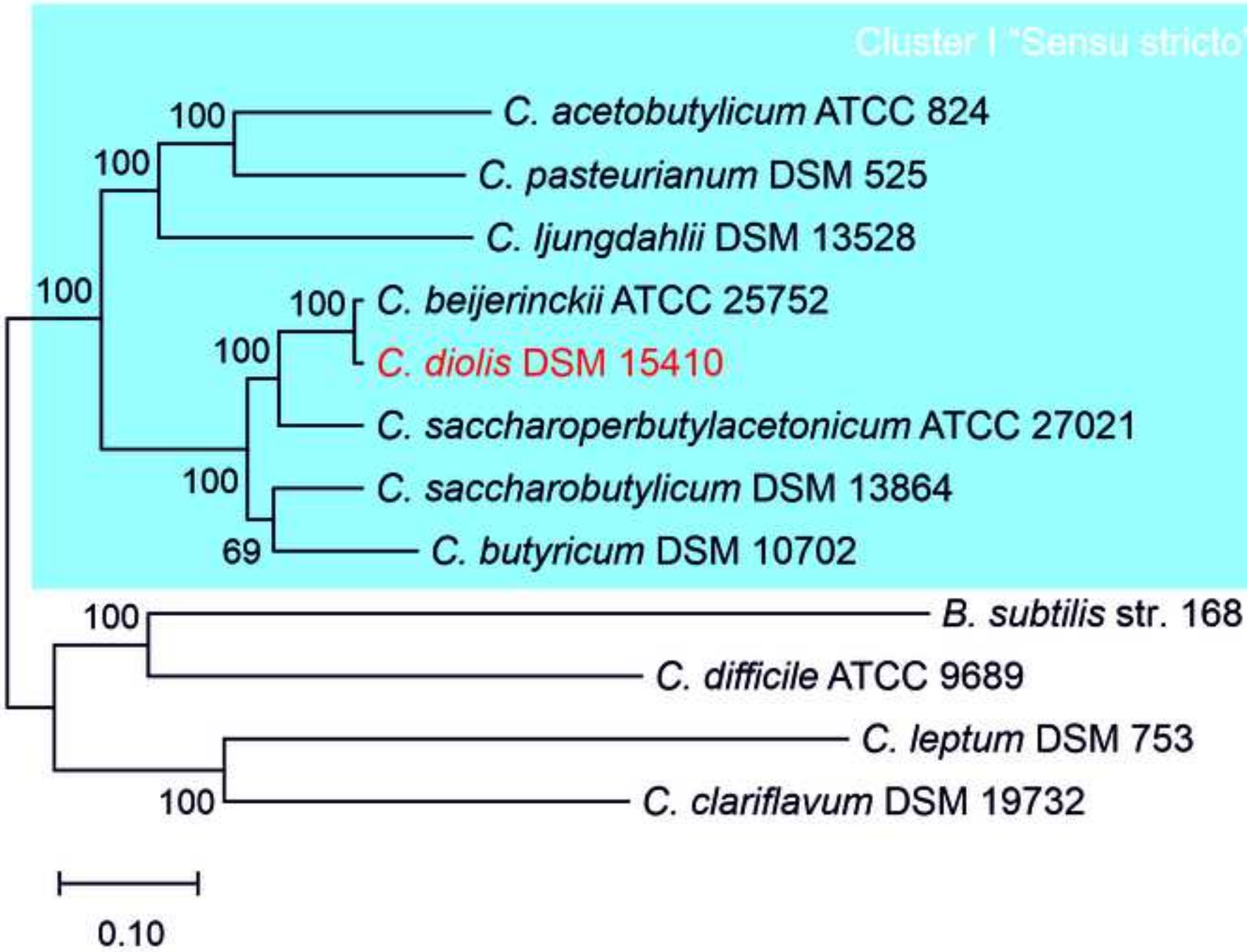
Figure

Click here to access/download;Figure;fig3.tif



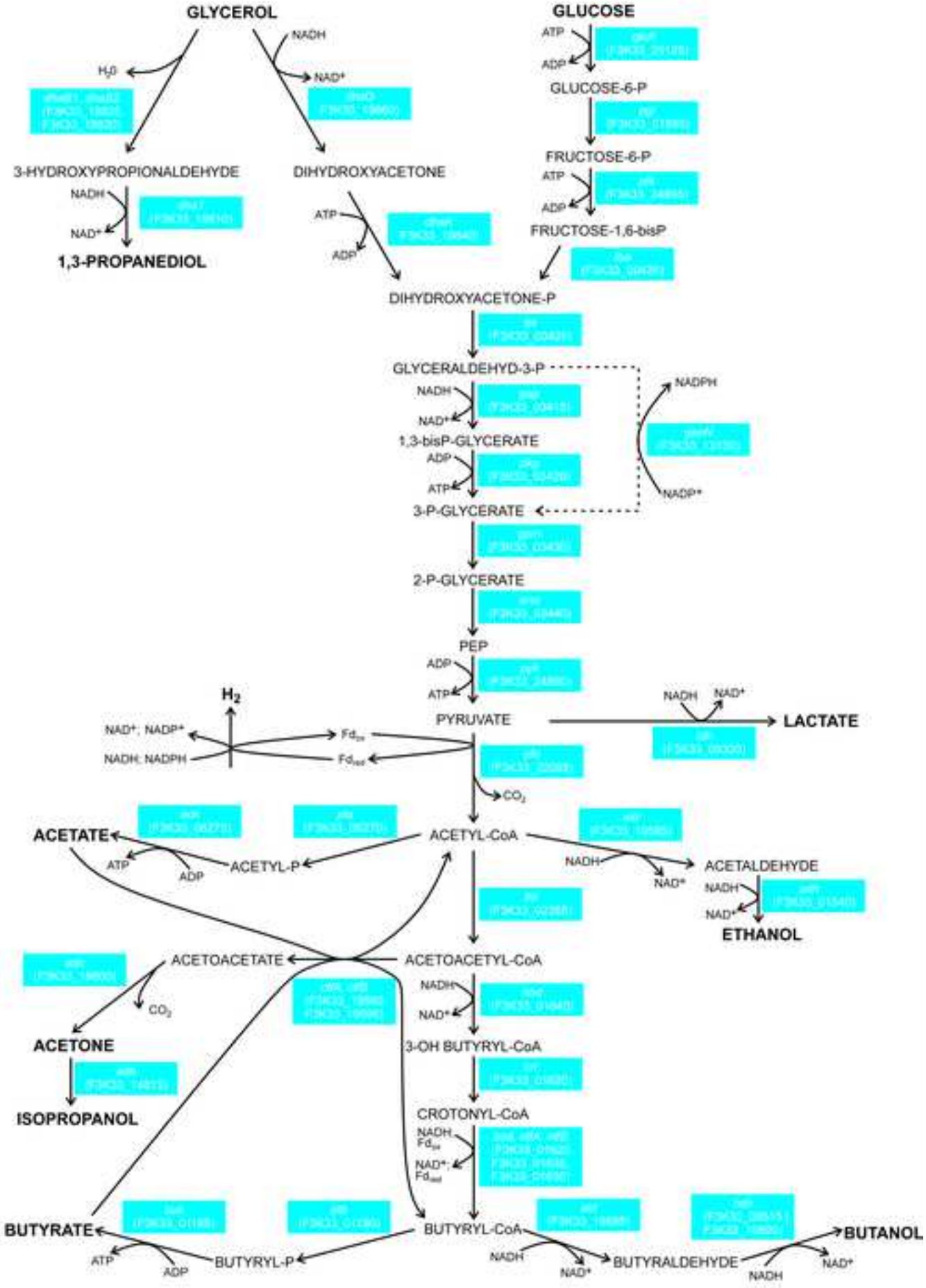
Figure

Click here to access/download;Figure;fig4.tif



Figure

Click here to access/download;Figure;fig5.tif



Karel Sedlar: Conceptualization, Methodology, Formal analysis, Data curation, Visualization, Writing- Original draft preparation, Writing- Reviewing and Editing. **Maryna Vasylkivska**: Validation, Investigation, Writing- Reviewing and Editing. **Jana Musilova**: Visualization, Formal analysis. ***Barbora Branska****:* Conceptualization, Investigation. **Ivo Provaznik**: Supervision.: **Petra Patakova:** Supervision, Conceptualization, Resources, Funding acquisition.