Research article

Genomic and phenotypic comparison of polyhydroxyalkanoates producing strains of genus *Caldimonas/Schlegelella*

Jana Musilova [a], Xenie Kourilova [b], Kristyna Hermankova [a], Matej Bezdicek [c], Anastasiia Ieremenko [d], Pavel Dvorak [d], Stanislav Obruca [b], Karel Sedlar [a,e,*]

[a] *Department of Biomedical Engineering, Faculty of Electrical Engineering and Communication, Brno University of Technology, Brno, Czech Republic*
[b] *Department of Food Chemistry and Biotechnology, Faculty of Chemistry, Brno University of Technology, Brno, Czech Republic*
[c] *Department of Internal Medicine – Hematology and Oncology, University Hospital Brno, Brno, Czech Republic*
[d] *Department of Experimental Biology (Section of Microbiology, Microbial Bioengineering Laboratory), Faculty of Science, Masaryk University, Brno, Czech Republic*
[e] *Institute of Bioinformatics, Department of Informatics, Ludwig-Maximilians-Universität München, Munich, Germany*

ARTICLE INFO

ABSTRACT

Polyhydroxyalkanoates (PHAs) have emerged as an environmentally friendly alternative to conventional polyesters. In this study, we present a comprehensive analysis of the genomic and phenotypic characteristics of three non-model thermophilic bacteria known for their ability to produce PHAs: *Schlegelella aquatica* LMG 23380[T], *Caldimonas thermodepolymerans* DSM 15264, and *C. thermodepolymerans* LMG 21645 and the results were compared with the type strain *C. thermodepolymerans* DSM 15344[T]. We have assembled the first complete genomes of these three bacteria and performed the structural and functional annotation. This analysis has provided valuable insights into the biosynthesis of PHAs and has allowed us to propose a comprehensive scheme of carbohydrate metabolism in the studied bacteria. Through phylogenomic analysis, we have confirmed the synonymity between *Caldimonas* and *Schlegelella* genera, and further demonstrated that *S. aquatica* and *S. koreensis*, currently classified as orphan species, belong to the *Caldimonas* genus.

1. Introduction

Polyhydroxyalkanoates (PHAs) are polyesters accumulated by numerous prokaryotes in the form of intracellular granules to serve as carbon and energy storage materials and to enhance stress robustness of bacterial cells [40]. Moreover, PHAs have emerged as environmentally friendly substitutes for petroleum-based polymers, offering a sustainable, renewable, biodegradable, compostable and also biocompatible alternative [23,48]. Although the production of bioplastics is seen as the way of the future and an integral part of the circular economy, less than 1% of total plastics production comes from the bioplastics industry [50].

Extremophiles are organisms able to survive and even prosper in extreme conditions such as acidic or basic pH levels, the presence of toxic elements, and immoderate temperatures [47,34]. Since the cultivation of extremophiles can be processed in a semi-sterile or even un-sterile mode, which significantly lowers the cost of biotechnological procedures, these organisms have gained a key role in the so-called "Next-generation Industrial Biotechnology" (NGIB) concept [9] in

recent years [11]. Further benefits come from the employment of thermophiles, bacteria that thrive at temperatures above 45 °C. Although biotechnological processes operate at high temperatures, they can be energetically feasible because the metabolic heat and energy released during mixing can be used to heat the bioreactor. In addition, cooling costs are modest because ambient air can be used to cool the process [20, 42].

Several potent thermophilic PHAs producers can be found in the recently proposed family *Sphaerotilaceae,* which, after the taxonomic revisions based on phylogenomic comparisons, contains several genera, including the genus *Caldimonas* [31]. The most promising group of organisms within this genus is formed by species belonging originally to the genus *Schlegelella* [39] with the type species *Schlegelella thermodepolymerans*, which was initially studied for its ability to degrade PHAs such as 3-hydroxybutyrate and 3-mercaptopropionate copolymers [14]. Moreover, an extraordinary ability of the type strain *S. thermodepolymerans* DSM 15344[T] to utilize xylose and synthesize 3-hydroxybutyrate and 3-hydroxyvalerate copolymers was recently

reported by our group [26] and the bacterium was identified as a promising candidate for industrial production of PHAs from various xylose rich lignocellulose-based resources [25]. Subsequently, we provided the first complete genome assembly and the functional annotation of *S. thermodepolymerans* DSM 15344[T] [37]. With the availability of complete genome sequences, *Schlegelella thermodepolymerans* was found to be a homotypic synonym of *Caldimonas thermodepolymerans* [31]. Since the genomes of the two other *Schlegelella* species were not available and 16S rRNA gene phylogeny was found to be insufficient to infer evolutionary relationships in the family *Sphaerotilaceae*, the two orphaned species, *Schlegelella aquatica* and *Schlegelella koreensis*, remained in the genus *Schlegelella*. However, additional research is needed to gain a comprehensive understanding of the evolutionary relationships within the genus, as well as to uncover the full genomic, metabolic, and biotechnological potential of these non-model bacteria.

Here, we compared the newly assembled genome of the type strain of *S. aquatica* LMG 23380[T] with the recently published genome of the type strain of *S. koreensis* ID0723[T] and other genomes from the genus *Caldimonas* including two newly assembled genomes of non-type *Caldimonas thermodepolymerans*, formerly *Schlegelella thermodepolymerans*, species. Besides further revisiting their taxonomy, we aimed at their genotypic and phenotypic comparison because all *Caldimonas/Schlegelella* species remain underexplored, which prevents their possible use in industrial biotechnology. These comparisons included examining the substrate range and antibiotic susceptibility, or identifying genes for the key enzymes in core carbohydrate and PHA metabolism.

2. Material and methods

2.1. Growth conditions and phenotypic analysis

In the experiments, four different strains of the genus *Caldimonas/Schlegelella* were used. Two of them were purchased from Leibnitz Institute DSMZ German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany: *Caldimonas thermodepolymerans* DSM 15344[T] and *Caldimonas thermodepolymerans* DSM 15264. The other two were obtained in BCCM/LMG Bacteria Collection in the Laboratory of Microbiology, Department of Biochemistry and Microbiology, Faculty of Sciences of Ghent University: *Schlegelella aquatica* LMG 23380[T] and *Caldimonas thermodepolymerans* LMG 21645. The bacteria were cultivated in two steps. In the first part of the cultivation, a nutritionally rich medium (Nutrient broth) was used (100 mL Erlenmeyer flasks, filling 50%). This phase entailed 20 h at 50 °C and constant shaking at 180 rpm. Subsequently, for the production phase, mineral medium was used (9.0 g/L $Na_2HPO_4 \cdot 12\ H_2O$, 1.5 g/L KH_2PO_4, 1.0 g/L NH_4Cl, 0.2 g/L $MgSO_4 \cdot 7\ H_2O$, 0.02 g/L $CaCl_2 \cdot 2\ H_2O$, 0.0012 g/L $Fe^{(III)}NH_4$citrate, 0.5 g/L yeast extract) with 20 g/L carbon sources (cellobiose, glucose, xylose) and trace element solution (50.0 g/L EDTA, 13.8 g/L $FeCl_3 \cdot 6\ H_2O$, 0.84 g/L $ZnCl_2$, 0.13 g/L $CuCl_2 \cdot 2\ H_2O$, 0.1 g/L $CoCl_2 \cdot 6\ H_2O$, 0.016 g/L $MnCl_2 \cdot 6\ H_2O$, 0.1 g/L H_3BO_3, dissolved in distilled water) and 5% v/v of culture from complex medium (250 mL Erlenmeyer flasks, filling 40%). This part of the cultivation took 72 h, also at 50 °C and constant shaking at 180 rpm.

Duplicates of biomass samples (2 × 10 mL) were determined gravimetrically (dried at 80 °C until constant weight was obtained). The biomass was centrifuged (6 000 × *g*, 5 min), then washed with 10 mL distilled water, and centrifuged again. The supernatant was discarded and the pellet of biomass was dried to constant weight. PHA content in dry biomass was determined by gas chromatography with a flame ionization detector as described previously [41].

2.2. Antibiotic susceptibility testing

The sensitivity of studied *Caldimonas/Schlegelella* strains to kanamycin (Serva, Germany), streptomycin, (Serva, Germany), gentamycin (Roth, Germany), tetracycline (Roth, Germany) and chloramphenicol (Serva, Germany) was examined to cover all antibiotics commonly used in genetic and metabolic engineering. Cell cultures inoculated from overnight cultures (these were prepared in 20 mL NB medium in 100 mL Erlen flasks inoculated directly from glycerol stocks) to the starting OD_{600} of 0.05 were grown in 3 mL of NB medium (Himedia, India) in 15 mL plastic tube at 50 °C with shaking (200 rpm, Biosan ES-20/60) for 72 h. With all of the antibiotics except streptomycin, ten concentrations were tested – 0, 2.5, 5, 7, 12.5, 25, 50, 100, 150, and 200 μg/mL. Three additional concentrations of 300, 400, and 500 μg/mL were checked with regards to streptomycin. All the experiments were performed in two biological replicates. Raw OD_{600} data from all of the measurements is provided in Supplementary File 2.

2.3. DNA extraction and sequencing

High molecular weight genomic DNA for long-read sequencing was extracted using MagAttract HMW DNA kit (Qiagen, NL) in accordance with the manufacturer's instructions. The DNA purity was checked using NanoDrop (Thermo Fisher Scientific, USA), the concentration was measured using Qubit 4.0 Fluorometer (Thermo Fisher Scientific, USA), and the length was measured using Agilent 4200 TapeStation (Agilent Technologies, USA). The Ligation Sequencing 1D Kit (Oxford Nanopore Technologies, UK) was used for libraries preparation and the sequencing was performed using R9.4.1 Flow Cell on the Oxford Nanopore Technologies (ONT) MinION platform. Genomic DNA for the high-throughput short-read sequencing was extracted using GenElut Bacterial Genomic DNA Kit (Sigma-Aldrich, USA) in accordance with the manufacturer's instructions. Sequencing libraries were prepared using KAPA HyperPlus kit and sequencing was carried out using Miseq Reagent kit v2 (500 cycles) on the MiSeq platform (Illumina, USA).

2.4. Genome assembly

The assemblies were completed in two steps. First, ONT reads were assembled into the initial sequence, and afterward, Illumina reads were mapped onto the initial sequence to produce hybrid assemblies. In the first step, the ONT reads were initially basecalled using Guppy v3.4.4 to translate picoamper signals, so called squiggles, into character form, and the resulting reads were assembled using Flye v2.8.1 [24]. Quality check was done using MinIONQC [27] to exclude possible low-quality reads and contaminations. The initial assembly was polished with Racon v1.4.20 [57] and Medaka (available from https://github.com/nanoporetech/medaka) to reduce the influence of sequencing errors typical for nanopore sequencing and improve the following mapping of high-quality short reads; PAF files were generated using minimap2 v2.24 [28]. In the second step, Illumina paired end (PE) reads were firstly quality-checked using FastQC v0.11.5 and MultiQC v1.7 [15] and, subsequently, adapter and the quality trimming of the Illumina reads were performed using Trimomatic v0.36 [6] to exclude detected adapters and low-quality reads. The trimmed reads were mapped on the Nanopore initial assembly using BWA v07.17 [18] and BAM files, needed for further polishing were generated using SAMtools v1.14 [29]. The final polishing was done using Pilon v1.24 [58] to produce high-quality complete genome sequences. Finally, the sequences were rearranged according to the origin of replication (oriC) in order for *DnaA* to be the first gene on the sense strand, which is a common representation for circular sequences that needs to be represented in a linear form.

2.5. Genome annotation

The genome structural annotation was completed through the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) [54] to predict various coding and non-coding elements. Operons' prediction was performed using Operon-mapper [53], and the results were further processed to obtain polycistronic operons, i.e. transcripts of two and more coding

genes. The genes involved in carbohydrate metabolism were primarily identified through the search for homologous sequences using the NCBI BLAST tool [2], along with a comprehensive literature search. For the genes associated with PHA metabolism, the PHA Depolymerase Engineering Database [22] was employed in addition to the aforementioned methods.

Functional annotation of protein-coding genes was performed with eggNOG-mapper [7], Operon-mapper [53], and batch CD-Search [33] by classifying them into clusters of orthologous groups (COGs) and assigning them particular functional categories. Outputs from the abovementioned tools were combined into the consensus result using COGtools (available from https://github.com/xpolak37/COGtools) and visualized in barplots for comparison of particular genomes and to provide informative visual representation of potential metabolic capacity of particular strains and species. Methylated bases that could prevent future genome engineering of the strains were detected from ONT raw data by nanopolish [51]; methylation motifs were inferred using STREME Command-Line version [3]. The detection of Restriction-Modification (R-M) systems was done using the Restriction Enzyme Databaseś (REBASE) tools [46] and KEGG Database [21]. Antibiotic-resistant genes search was performed using Resistance Gene Identifier (RGI) 6.0.0 included in the Comprehensive Antibiotic Resistance Database (CARD) 3.2.5 [1]. Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) arrays identification was carried out via the CRISPRDetect tool [5].

Digital DNA to DNA hybridization (dDDH) values were calculated using the Type (Strain) Genome Server (TYGS) [35] in order to propose correct taxonomic placement of analyzed strains and species. A phylogenomic tree was constructed with a bootstrapping method using PhyloPhlAn 3.0.60 [49] and its internal database of approximately 400 genes conserved across the bacterial domain to produce well reproducible phylogeny among closely related analyzed bacteria.

3. Results

3.1. Genome assembly & structural annotation

Schlegelella aquatica LMG 23880[T] and *Caldimonas thermodepolymerans* (formerly *Schlegelella thermodepolymerans*) DSM 15264 and LMG 21645 genomes were assembled using a hybrid approach with initial coverage of $261\times$, $1440\times$, and $160\times$, respectively. Over 44 thousand Oxford Nanopore Technologies (ONT) reads with a median read length of 3.44 kb were used to generate an initial assembly of *S. aquatica* LMG 23880[T]. Over 1.6 million high-quality (average Phred score Q $\approx$ 34) Illumina reads (97% of all Illumina reads) were subsequently mapped to the initial assembly. *C. thermodepolymerans* DSM 15264 was reconstructed from over 901 thousand ONT reads with a median read length of 2.51 kb and more than 2.67 million high-quality (average Phred score Q $\approx$ 36) Illumina reads (86% of all Illumina reads) mapped to the initial assembly. *C. thermodepolymerans* LMG 21645 initial assembly was reconstructed from over 14 thousand ONT reads with a median read length of 2.7 kb. Subsequently, more than 1.8 million high-quality (average Phred score Q $\approx$ 36) Illumina reads (89% of all Illumina reads) were mapped to the initial assembly. The whole process resulted in circular sequences with coverage of $194\times$, $1324\times$ and $85\times$, corresponding to *S. aquatica* LMG 23880[T], *C. thermodepolymerans* DSM 15264, and *C. thermodepolymerans* LMG 21645, respectively. The final assemblies have been deposited at the DDBJ/EMBL/GenBank under accession numbers CP110257, CP110416, and CP110415.

The genome lengths range from 3.3 to 4.0 Mbp and contain from 3069 to 3856 open reading frames (ORFs). The majority of all ORFs consist of protein-coding sequences (CDSs), but several pseudogenes were also identified. Table 1 summarizes the main features of the newly assembled genomes along with the genome of the type strain *C. thermodepolymerans* DSM 15344[T] [37], available in the GenBank NCBI database under the accession CP064338.

Table 1

Chromosomal features of *Schlegelella aquatica* (*Sa*) LMG 23880[T], *Caldimonas thermodepolymerans* (*Ct*) DSM 15264 and LMG 21645, and type strain DSM 15344[T].

Feature	*Sa* LMG 23880[T]	*Ct* DSM 15264	*Ct* LMG 21645	*Ct* DSM 15344[T]
Length (bp)	3348,564	4030,985	3940,444	3858,501
GC content (%)	69.34	70.22	70.20	70.28
Polycistronic operons	637	774	753	745
ORFs	3069	3856	3724	3670
CDSs	2976	3771	3627	3576
Pseudogenes	37	25	37	33
rRNA (5S, 16S, 23S)	2, 2, 2	2, 2, 2	2, 2, 2	2, 2, 2
tRNA	46	50	50	51
ncRNA	4	4	4	4

3.2. Functional annotation

Genes in the annotated genomes were further clustered into 26 categories according to clusters of orthologous groups (COGs). The relative abundance of COG categories in *Caldimonas/Schlegelella* species is shown in Fig. 1. Although the "Function unknown" (S) category is still the most prevalent group in each of the studied bacteria, the improvement in COGs annotation with COGtools reduced the number of genes in this category, and the representation of the other categories increased. All of the studied *Caldimonas/Schlegelella* species show a high similarity of genome content, although minor differences between genomes are present, especially in terms of *C. thermodepolymerans* DSM 15264 and *S. aquatica* LMG 23880[T]. The difference exists mainly due to the already mentioned category S, where *C. thermodepolymerans* DSM 15264 is the only genome with a relative abundance of this category exceeding 10%. Further, the group "Amino acid transport and metabolism" (E) constitutes more than 7.5% of studied *Caldimonas/Schlegelella* genomes except *C. thermodepolymerans* DSM 15264 with nearly 7.3%. In addition, the genome of *C. thermodepolymerans* DSM 15344[T] comprises the highest percentage of COG unknown category compared to others, even though it is a type strain bacterium. The further variations in the other COGs generate noticeable differences between *Caldimonas/Schlegelella* species (see Fig. 1, supplementary Table S1). The main dissimilarity between *S. aquatica* LMG 23880[T] and *C. thermodepolymerans* strains is in the relative abundance of the categories "Signal transduction mechanism" (T) and "Cell, wall/membrane/envelope biogenesis" (M), with the two comprising over 15% of *S. aquatica* LMG 23880[T] genome. Even though the S group still makes up the prevalent group of the genome, the difference between group S and M, or T, is not as marked as in *C. thermodepolymerans* strains. Detailed distribution of relative abundances in the individual categories across the studied *Caldimonas/Schlegelella* species is provided in supplementary Table S1.

Long ONT reads were used to detect the methylation state of cytosine nucleotides (5-mC), where a log-like ratio assigned to the target methylated site served as a measure of methylation support. The number of methylated sites in a genome differs across the studied bacteria. The highest number of methylated sites was detected in *C. thermodepolymerans* DSM 15344[T], whereas the genome of *C. thermodepolymerans* LMG 21645 carries the lowest content of 5-mC. Subsequently, methylation motifs in bacterial genomes were derived from the previously detected methylated sites. The highest number of methylated sites in *C. thermodepolymerans* DSM 15344[T] is consistent with its highest quantity of motifs, where six motifs satisfied the significance criteria set as 0.05 E-value. A complete overview of the methylated motifs can be found in supplementary Table S2.

Restriction-modification (R-M) analysis revealed type I, II, and III systems in the studied bacteria, based on the comparison with the REBASE and additionally confirmed with the KEGG database. No complete system was found in the *S. aquatica* LMG 23880[T]; only a partial
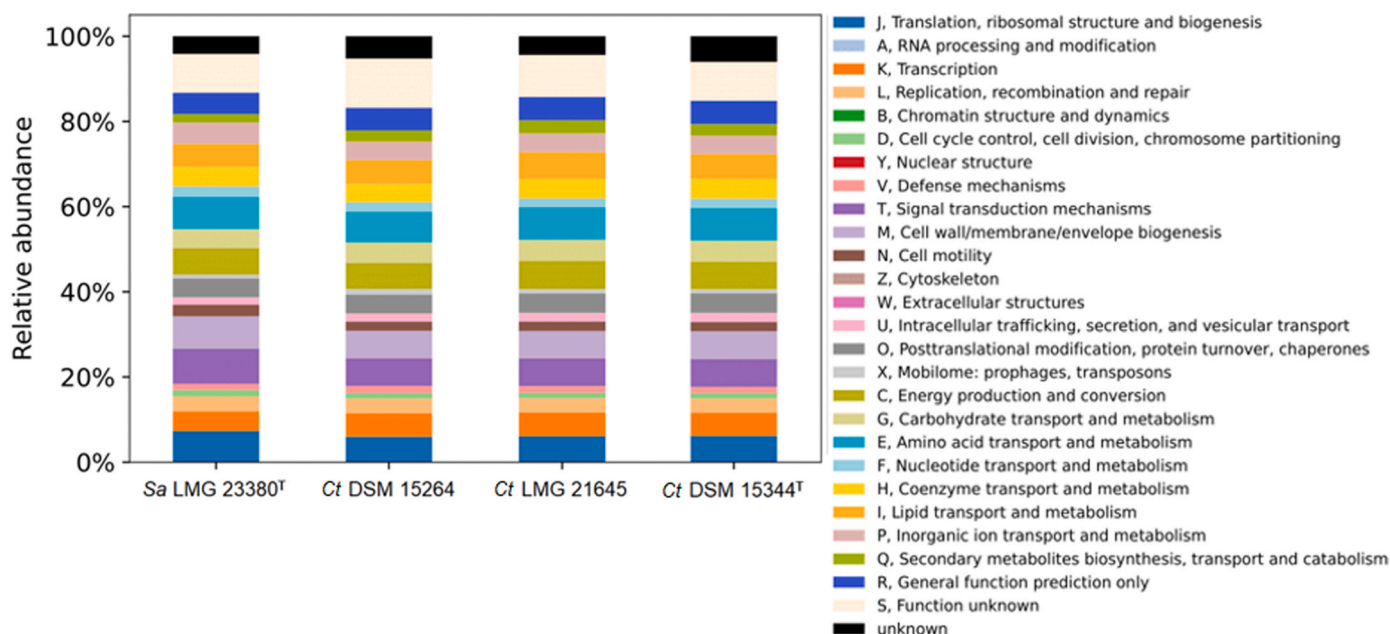
Fig. 1. Relative abundances of genes in particular clusters of orthologous groups (COGs) categories in the *Schlegelella aquatica* (*Sa*) LMG 23380[T], *Caldimonas thermodepolymerans* (*Ct*) DSM 15264 and LMG 21645, and reference type strain DSM 15344[T].

type I system was identified due to the absence of the R subunit of the system. *C. thermodepolymerans* DSM 15264 genome contains most protein-coding genes forming R-M systems, mainly genes encoding type II methylases. Furthermore, *C. thermodepolymerans* LMG 21645 possesses two complete type III R-M systems and the standalone type II methylase. The type strain of *C. thermodepolymerans* DSM 15344[T] has three complete type I and II R-M systems. Detailed information about R-M systems of *Caldimonas/Schlegelella* species is provided in the supplementary Table S3.

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) analysis revealed one array in each bacterium. The arrays are located near the end of each sequence, and the length of the discovered arrays ranges from 3.2kbp to 5.3kbp, except the type strain *C. thermodepolymerans* DSM 15344[T], whose array's length is only 163 bp. In addition, one more array with the length of 1.5kbp was found in the first third of the *C. thermodepolymerans* DSM 15264 sequence. The further information is provided in the Supplementary Table S4. Cas-like genes have not been directly identified in the arrays, however, several such genes are located in their close proximity (see supplementary Table S5). Putative *cas9* genes were identified in the genome of *C. thermodepolymerans* DSM 15264.

The antibiotic-resistance genes search was carried out through comparison of the sequences against the Comprehensive Antibiotic Resistance Database (CARD) and revealed three genes: *adeF*, *ANT(3'')-IIa*, and *aadA6* originally coding for a membrane fusion protein of the multidrug efflux complex AdeFGH (*adeF*) and aminoglycoside nucleotidyltransferases (*ANT(3'')-IIa*, *aadA6*). The first hit, *adeF* gene, was detected in all of the *C. thermodepolymerans* strains with 71% identity of matching region and length corresponding to the criterion 100% ± 5%. In contrast, the second gene *ANT(3'')-IIa* was discovered only in the *C. thermodepolymerans* LMG 21645 strain. The *aadA6* gene was identified only in the *C. thermodepolymerans* type strain DSM 15344[T]. The comparison of *S. aquatica* LMG 23380[T] genome did not reveal significant hits with at least 50% identity of the matching region. Details are provided in the supplementary Table S6.

The subsequent experimental verification of susceptibility to five selected antibiotics commonly used in genetic and metabolic engineering praxis revealed that all of the studied bacteria are sensitive to kanamycin, gentamycin, tetracycline, and chloramphenicol. The

minimal inhibitory concentrations of these antibiotics determined by our probe are shown in Table 2 (the raw data can be found in the Supplementary File 2). In contrast, all three *C. thermodepolymerans* (*Ct*) strains grew even when the highest concentration of streptomycin (500 μg/mL) was used. Streptomycin was also the least thermally stable antibiotic in the tested set. The only exception was *S. aquatica* (*Sa*), whose growth was fully inhibited with 400 μg/mL streptomycin. Interestingly, *aadA6* and *ANT(3'')-IIa* genes, which can confer resistance to streptomycin, were found only in the genomes of DSM 15344[T] and LMG 21645 strains. On the other hand, *S. aquatica* showed higher resistance to kanamycin and gentamycin than all of the *Ct* strains.

3.3. Genome-based phylogeny and taxonomic descriptions

Digital DNA-DNA hybridization (dDDH) confirmed the classification of *C. thermodepolymerans* DSM 15264 and *C. thermodepolymerans* LMG 21645 within the *Caldimonas thermodepolymerans* species. The former exhibited 94% and the latter 95% similarity to the type strain. In contrast, *S. aquatica* LMG 23380[T] exhibited a similarity of less than 30% to the *Caldimonas thermodepolymerans* species.

Phylogenomic analysis of *Caldimonas/Schlegelella* species and their comparison to the other genera of the family *Sphaerotilaceae* and additional Gram-positive and Gram-negative PHA-producing bacteria showed a well-distinguished cluster of *Caldimonas/Schlegelella* genomes confirming that these two genera are synonymous. Moreover, both *Schlegelella aquatica* and *Schlegelella koreensis* represent separate species as their genomes dDDH values compared to other *Caldimonas/Schlegelella* genomes are below the threshold for species dealineation. Therefore, we propose that *Caldimonas aquatica* comb. nov. and *Caldimonas koreensis* comb. nov. be the novel combinations for *Schlegelella aquatica* and *Schlegelella koreensis*, respectively.

Taxonomic description of new combinations for species:

Description of *Caldimonas aquatica* comb. nov.

Caldimonas aquatica (a.qua'ti.ca. L. fem. adj. *aquatica*, living in water).

Basonym: *Schlegelella aquatica* [10].

The description is the same as that of *S. aquatica* [10]. Genomic, phylogenetic, and phenotypic evidence strongly support the placement of this species in the genus *Caldimonas*. The type strain is LMG 23380[T]

Table 2

Minimal inhibitory concentrations of antibiotics in NB medium. The susceptibility to the tested antibiotics is demonstrated by minimal inhibitory concentrations (MIC, the lowest concentration of antibiotic, in μg/mL, necessary to inhibit visible bacterial growth).

Species	Kanamycin	Streptomycin	Gentamycin	Tetracycline	Chloramphenicol
Sa LMG 23880[T]	12.5	400.0	25.0	5.0	5.0
Ct DSM 15264	2.5	resistant	2.5	12.5	5.0
Ct LMG 21645	7.0	resistant	7.0	5.0	5.0
Ct DSM 15344[T]	2.5	resistant	2.5	5.0	2.5

(= wcf1[T] = BCRC 17557[T]). The genomic DNA G + C content is 69.34%. The GenBank accession numbers of the 16S rRNA gene and genome for the type strain are DQ417336.1 and CP110257.1, respectively.

Description of *Caldimonas koreensis* comb. nov.

Caldimonas koreensis (ko.re.en'sis. N.L. masc./fem. adj. *koreensis*, of Korea, from where the novel organisms were isolated).

Basonym: *Schlegelella koreensis* Chaudhary et al., 2022.

The description is the same as that of *S. koreensis* [8]. Genomic, phylogenetic, and phenotypic evidence strongly support the placement of this species in the genus *Caldimonas*. The type strain is ID0723[T] (= KCTC 72731[T] = NBRC 114611[T]). The genomic DNA G + C content is 69.87%. The GenBank accession numbers of the 16S rRNA gene and genome for the type strain are KP326334.1 and JABWMJ01, respectively.

3.4. PHA production on various sugars

The phenotypic investigation of various strains within the *Caldimonas/Schlegelella* genus focused on their growth and PHA production using xylose, glucose, and cellobiose as substrates. These sugars are commonly associated with lignocellulose-based resources that could potentially facilitate economically viable PHA production [12]. Notably,

the type strain *C. thermodepolymerans* DSM 15344[T] exhibited a significant preference for xylose over glucose, which was identified as its crucial characteristic [26]. Consequently, this attribute was assessed in other strains of *C. thermodepolymerans* as well as the closely related strain *S. aquatica*.

The results are presented in Table 3, indicating that *C. thermodepolymerans* achieves significantly higher PHA titers than *S. aquatica* when utilizing the specific substrates. Furthermore, all of the investigated strains of *C. thermodepolymerans* exhibited a substrate preference pattern similar to the type strain *Ct* DSM 15344[T], with a notable preference for xylose over glucose. These findings suggest that this substrate preference is a distinct characteristic of this bacterial species. In contrast, *S. aquatica* displayed the opposite trend. Although glucose as a carbon substrate led to the highest yields in biomass as well as poly(3-hydroxybutyrate) (PHB) titer in this bacterium, its low PHA yields render it unsuitable for industrial-scale production.

An intriguing observation across all of the tested strains of *C. thermodepolymerans* was their ability to utilize cellobiose, which resulted in higher biomass growth compared to glucose or even xylose. This is surprising given that cellobiose consists of two glucose units linked by a β-1,4 glycosidic bond. It implies that the relatively low efficiency of glucose utilization in *C. thermodepolymerans* strains is more
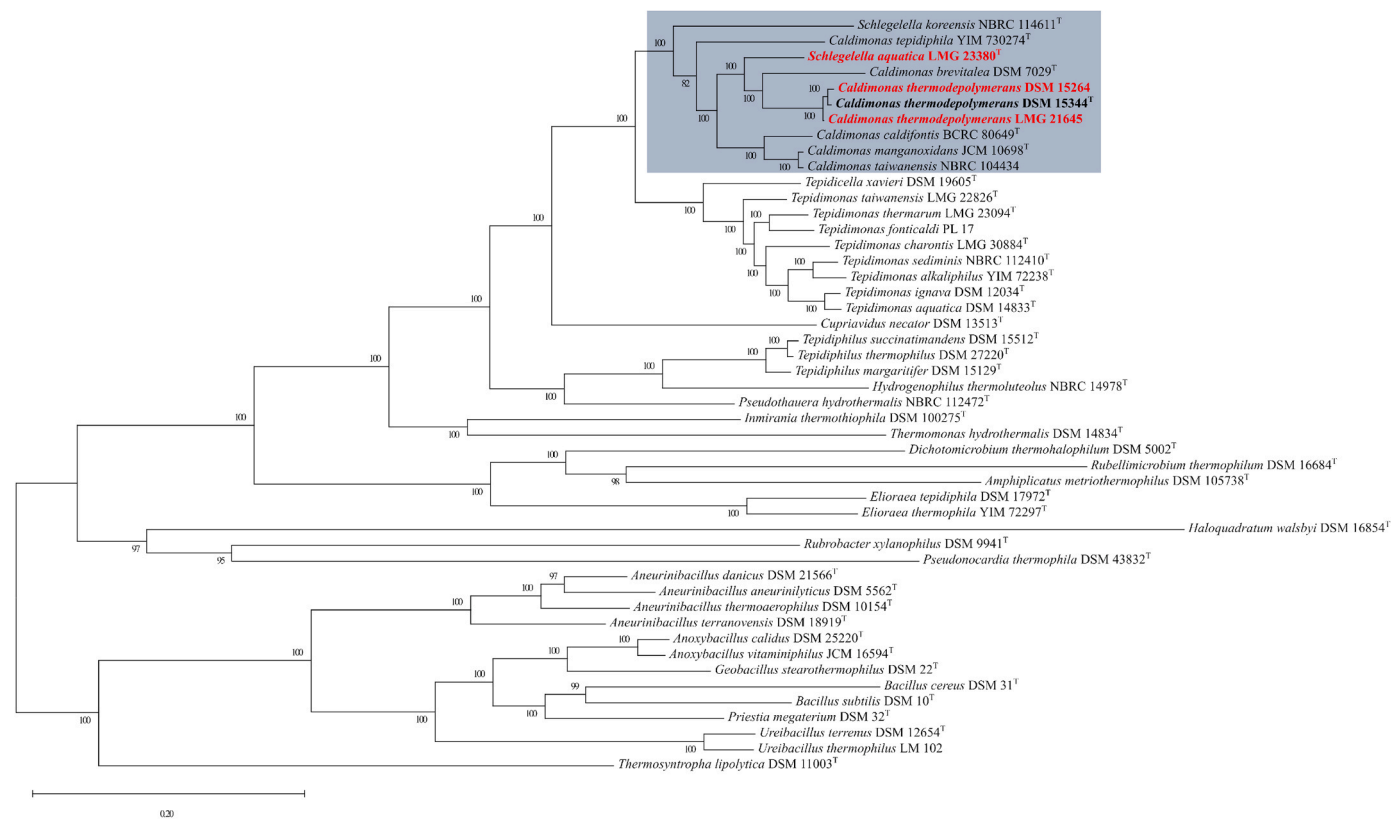

Fig. 2. Phylogenomic tree representing placement of *Schlegelella aquatica* LMG 23880[T] and *Caldimonas thermodepolymerans* DSM 15264 and LMG 21645 (highlighted in red). Bacteria placed in the *Caldimonas* genus are highlighted with the rectangle. The tree was constructed using the Phylophlan internal database of circa 400 genes conserved across bacterial domain. The values represent the bootstrap support based on 100 replicates.

Table 3

Comparison of PHB production of *Caldimonas/Schlegelella* strains utilizing different carbon substrates. Cultivation conditions – 72 h, 50 °C, 180 rpm, 20 g/L; carbon source – cellobiose, glucose and xylose.

Substrate	*Sa* LMG 23880[T]			*Ct* DSM 15264			*Ct* LMG 21645			*Ct* DSM 15344[T]		
	CDW [g/L]	PHB [g/L]	PHB [%]	CDW [g/L]	PHB [g/L]	PHB [%]	CDW [g/L]	PHB [g/L]	PHB [%]	CDW [g/L]	PHB [g/L]	PHB [%]
Xylose	0.51	0.03	5.88	6.22	4.21	67.64	2.78	1.66	59.60	5.50	3.63	66.02
	± 0.01	± 0.01	± 0.02	± 0.11	± 0.02	± 0.08	± 0.01	± 0.11	± 3.78	± 0.07	± 0.01	± 0.01
Glucose	1.68	0.53	31.55	0.10	*n.d.*	*n.d.*	0.09	*n.d.*	*n.d.*	2.63	1.35	51.33
	± 0.26	± 0.02	± 0.57	± 0.02			± 0.01			± 0.14	± 0.01	± 0.26
Cellobiose	0.19	*n.d.*	*n.d.*	8.34	3.03	36.39	4.63	3.41	73.79	6.70	5.27	78.66
	± 0.03			± 0.30	± 0.07	± 0.79	± 0.08	± 0.06	± 1.28	± 0.11	± 0.21	± 3.14

n.d. – not detected

likely due to the inefficient glucose transport into bacterial cells rather than a deficiency in glucose metabolism.

Analysis of genes linked to PHA synthesis revealed a total of eight genes involved in PHA metabolism summarized in Table 4. Each gene is presented only in a single copy in analyzed genomes. These genes encode for enzymes involved in PHA synthesis (*phaA, phaB, phaC*) and degradation. All the tested strains harbor genes encoding for intracellular PHA depolymerase (*phaZi*) but *S. aquatica* does not contain genes encoding for extracellular PHA depolymerase (*phaZe*) indicating that, unlike *C. thermodepolymerans*, *S. aquatica* is not capable of decomposition of PHA materials in the environment. Further, the strains also carry *phaP* genes encoding for phasins – proteins covering the surface of PHA granules in the bacterial cells acting as an interface between water-rich cytoplasm and hydrophobic polymer [36]. In addition, the genomes of all of the tested strains also contain the *phaR* gene, which is involved in the regulation of PHA metabolism [32]. Finally, one additional gene (locus tag OMP39_11465) related to PHA metabolism was predicted solely in *S. aquatica* LMG 23880[T]; nevertheless, the metabolic function of the encoded protein is not known yet as no sufficient homology to any well-annotated PHA metabolism gene was found. Further, as is common in most PHA-producing Gram-negative bacteria, *phaC*, *phaA*, and *phaB* genes are clustered in the *phaCAB* operon in all of the examined *Caldimonas/Schlegelella* strains, which enables effective orchestration of PHA synthesis.

3.5. Carbohydrate metabolism

To investigate the variations in the utilization of lignocellulosic sugars among the studied strains, we next focused on identifying genes responsible for encoding enzymes and transporters involved the upper carbohydrate metabolism. We confirmed the presence of genes associated with the complete Embden-Meyerhof-Parnas (EMP) pathway, which converts glucose to pyruvate and acetyl-CoA, in all four compared strains (supplementary Table S7). Regarding glucose transport into the cell, genes of the PEP-dependent phosphotransferase system (PTS) were identified in *Sa* LMG 23380[T] and in all three *Caldimonas* strains, but with a fructose-specific EIICB component. Therefore, it is likely that glucose is transported by a non-PTS transporter such as the GtsABCD glucose/mannose ABC transporter that we identified and whose genes form a *gts*

operon in *Sa* LMG 23380[T] and in *Caldimonas* strains (Fig. 3). In *Ct* strains, the *gts* operon also includes a β-glucosidase gene and a carbohydrate porin gene. The presence of the former gene correlates with the ability of *Ct* strains to grow on cellobiose (Table 3).

The xylose isomerase pathway (formed by xylose isomerase XylA and xylulose kinase XylB) and the non-oxidative branch of the pentose phosphate pathway (PPP) are responsible for the conversion of xylose to the EMP pathway intermediates glyceraldehyde 3-phosphate and fructose 6-phosphate in *Ct* DSM 15264, LMG 21645, and DSM 15344[T] (supplementary Table S7). *S. aquatica* lacks *xylA* and *xylB* genes, which is also indicated by its very low efficiency in xylose utilization (see Table 3). Genes for the oxidative metabolism of xylose or glucose are not present in any of the studied strains. Interestingly, the *xylA* and *xylB* genes in *Ct* strains form a single *xyl* operon together with the genes encoding the high-affinity xylose ABC transporter XylFGH and a ROK-family transcriptional regulator, indicating that all *xyl* genes are regulated and expressed simultaneously (Fig. 3).

An intriguing feature of the carbohydrate metabolism of all of the four compared bacteria is the absence of the two genes for enzymes of the oxidative PPP, namely 6-phosphogluconolactonase (*pgl* gene) and 6-phosphogluconate dehydrogenase (*gnd* gene). In the metabolism of the model Gram-negative bacterium *Escherichia coli*, Pgl catalyses the hydrolysis of 6-phosphogluconolactone to 6-phosphogluconate, which is converted by Gnd to ribulose 5-phosphate, which enters the non-oxidative PPP. The absence of Gnd is probably compensated for by the Entner-Doudoroff (ED) pathway formed by Edd 6-phosphogluconate dehydratase and Eda 2-keto-3-deoxy-6-phosphogluconate aldolase, which was found in all of the four bacteria studied and may explain their ability to grow on the lignocellulosic sugars such as xylose, glucose, and cellobiose.

4. Discussion

Uncovering genomic and phenotypic traits with a focus on PHAs production by strains *Schlegelella aquatica* LMG 23380[T] and *Caldimonas thermodepolymerans* DSM 15264 and LMG 21645 may provide important insights into microbial production of sustainable environmentally friendly polyesters. Based on our previous studies, *C. thermodepolymerans* can be considered a promising candidate for PHA

Table 4

List of genes involved in polyhydroxyalkanoates metabolism in *Schlegelella aquatica* (*Sa*) LMG 23380[T] *Caldimonas thermodepolymerans* (*Ct*) DSM 15264 and LMG 21645, and the reference type strain DSM 15344[T], referred to by their respective locus tags. The genes *phaC*, *phaA*, and *phaB* form one operon.

| Gene name	Function	*Sa* LMG 23380[T]	*Ct* DSM 15264	*Ct* LMG 21645	*Ct* DSM 15344[T]
phaC	class I poly(R)-hydroxyalkanoic acid synthase	OM39_07225	ONS87_09975	ONZ46_09255	IS481_08630
phaA	acetyl-CoA C-acetyltransferase	OMP39_07230	ONS87_09980	ONZ46_09260	IS481_08635
phbB	acetoacetyl-CoA reductase	OMP39_07235	ONS87_09985	ONZ46_09265	IS481_08640
phaP	Phasin (PHA-granule associated protein)	OMP39_02230	ONS87_09010	ONZ46_07840	IS481_07490
phaR	polyhydroxyalkanoate synthesis repressor PhaR	OMP39_08200	ONS87_09895	ONZ46_08730	IS481_08360
phaZi	polyhydroxyalkanoate depolymerase - intracellular	OMP39_06240	ONS87_08655	ONZ46_07490	IS481_07130
phaZe	polyhydroxybutyrate depolymerase - extracellular	-	ONS87_16240	ONZ46_15580	IS481_15185
-	polyhydroxyalkanoic acid system family protein of unknown function	OMP39_11465	-	-	-
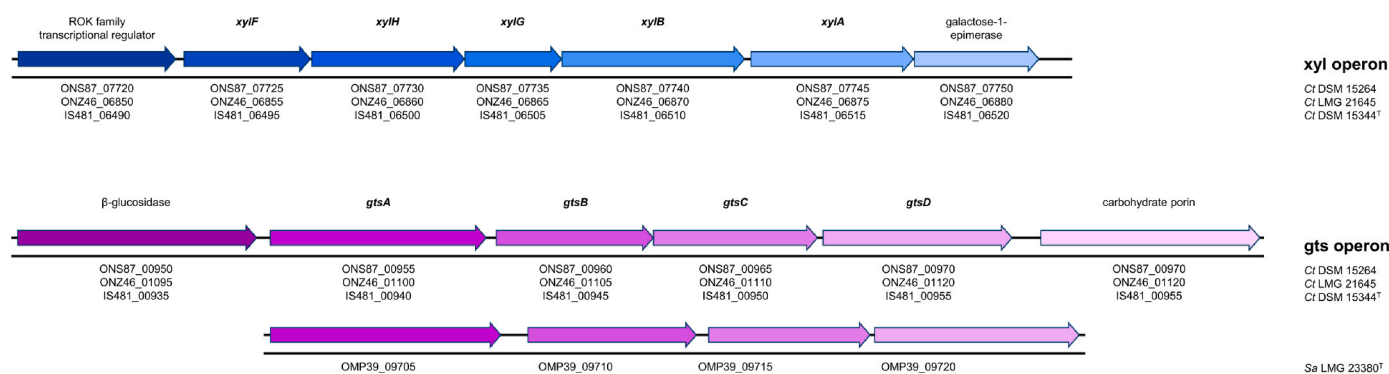
Fig. 3. Genes involved in carbohydrate metabolism forming xyl and gts operons across *Schlegelella aquatica* (*Sa*) LMG 23880[T], *Caldimonas thermodepolymerans* (*Ct*) DSM 15264 and LMG 21645, and the reference type strain DSM 15344[T].

production, especially from xylose-rich lignocellulose-based resources [25,26]. Since the genomic information was not previously available, genome assembly was a necessary first step for further studies. *De novo* genome assembly using ONT reads identified one circular contig in each bacterium, indicating the absence of plasmids and thus the presence of chromosomal DNA only. Furthermore, high-quality short Illumina reads were used to polish the assemblies. The mapping of nearly all of the short reads to the final genome, as well as the prediction of the replication origin oriC, were unambiguous, which confirms the genome had been assembled correctly. Although genome length and GC content are higher than the average for Gram-negative bacteria [30], the values are consistent with the assumption that GC content is positively correlated with genome length [60] and with growth temperature [19] as well.

Functional annotation, including the study of bacterial characteristics such as the function of individual genes, the ability to defend against foreign DNA, antibiotic resistance, and the requirements for genome editing, is the key prerequisite for understanding and manipulating the studied bacteria. Genes classified according to clusters of orthologous groups (COGs) revealed around 85% of gene functions in all of the studied *Caldimonas/Schlegelella*. However, the remaining 15% remains unknown, highlighting the uniqueness of these bacteria. Furthermore, the classification confirmed the genomic similarity among the *C. thermodepolymerans* strains and indicated a partial divergence *between S. aquatica* and *C. thermodepolymerans* based on different distributions across specific COGs within a narrow range. The prediction of Restriction-Modification (R-M) systems revealed the diversity of this cellular defense mechanism in the studied bacteria. The lack of active R-M systems was registered in *S. aquatica* LMG 23380[T] genome, with only a partial type I system being identified, suggesting its inactivity. In contrast, all of the studied *C. thermodepolymerans* possess at least one complete R-M system, indicating a strong system protecting *C. thermodepolymerans* strains from foreign DNA, which may pose challenges in genome editing [45]. Regarding further potential opportunities for future genome engineering, all of the studied bacteria were found to have the ability to receive foreign DNA as they contain CRISPR arrays. These arrays can be also adopted as a reservoir of components for genome editing of *Caldimonas* strains and potentially other thermophilic bacteria [17]. Native CRISPR-Cas systems in strains *S. aquatica* LMG 23880[T] and *C. thermodepolymerans* DSM 15264 and LGM 21645 will be further studied and can be repurposed for genome editing of these bacteria [59]. The observed susceptibility of all *Caldimonas* strains to four out of the five selected antibiotics will facilitate their future genetic engineering as various selection markers can be used. Identified antibiotic-resistance genes can be subsequently removed to prevent a potential outbreak of antibiotic resistance.

In the genomic era, a polyphasic approach combining examinations of genotypic and phenotypic traits is still necessary and has been practiced for more than a decade in prokaryotic taxonomy [55]. Here, the classification of the studied bacteria using a combination of digital

DNA-DNA hybridization (dDDH) and phylogenomic analysis confirmed that the newly analyzed non-type strains *C. thermodepolymerans* DSM 15264 and LMG 21645 belong to *C. thermodepolymerans* species and are very similar to the type strain *C. thermodepolymerans* DSM 15344[T]. Additional cultivation experiments also showed that these strains share unique phenotypic traits, particularly a preference for xylose over glucose, which seems to be a unique feature of *C. thermodepolymerans* that cannot be found in other species of the genus *Caldimonas*. This suggests, along with the fact that *C. thermodepolymerans*, formerly *Schlegellela thermodepolymerans*, was the type species of the genus *Schlegelella* covering four species, that the genus *Caldimonas/Schlegelella* is much more diverse than previously assumed. The previous absence of representative genomes of *S. aquatica* and *S. koreensis*, coupled with the limitations of the 16S rRNA gene in inferring the phylogeny within the family *Sphaerotilaceae* [31] prevented their accurate taxonomic placement. An analysis of the newly assembled *S. aquatica* and publicly available *S. koreensis* genomes revealed that *Schlegelella aquatica* and *Schlegelella koreensis* are homotypic synonyms for *Caldimonas aquatica* and *Caldimonas koreensis*, respectively. At the same time, genotypic analysis proved that both organisms represent separate species, as also shown by their different phenotypes. *S. aquatica*, in comparison to *C. thermodepolymerans*, exhibits slower growth and lower production of PHAs when utilizing various carbon sources and prefers glucose over xylose. Although we did not analyze the phenotype of *S. koreensis*, this species had been already shown to possess unique traits such as different temperature ranges for growth and the inability to utilize various saccharides while still producing PHA granules [8].

PHAs can be synthesized through four different pathways [52]. In the tested strains the most common metabolic a three-step pathway, encoded by the genes forming the *phaCAB* operon [26,56] was identified, this synthetic route utilizes predominantly acetyl-CoA as a primary substrate. This key metabolic intermediate is converted into acetoacetyl-CoA, followed by the conversion into 3-hydroxybutyryl-CoA [38,44]. Given that the *phaCAB* operon has been identified in each of the studied *Caldimonas* strains, exploitation of this pathway during the growth on the assessed lignocellulosic sugars (xylose, glucose, and cellobiose) can be assumed (Fig. 4). Additionally, PHA metabolism is closely linked with carbohydrate catabolism and fatty acids β-oxidation which provide precursors (especially acetyl-CoA) for PHA synthesis and fatty acid β-oxidation which can also supply suitable 3-hydroxyacids in various bacteria.

Genes involved in these three pathways have been identified in all the bacteria studied, which likely implies the ability to synthesize PHAs by all the pathways mentioned; however, further research is needed to confirm this hypothesis. In addition, several differences between *S. aquatica* (*Sa*) and *C. thermodepolymerans* (*Ct*) have been identified. Firstly, *S. aquatica* most likely does not possess the enzymatic apparatus for extracellular degradation of PHA materials present in the environment. On the other hand, *C. thermodepolymerans* was found to possess
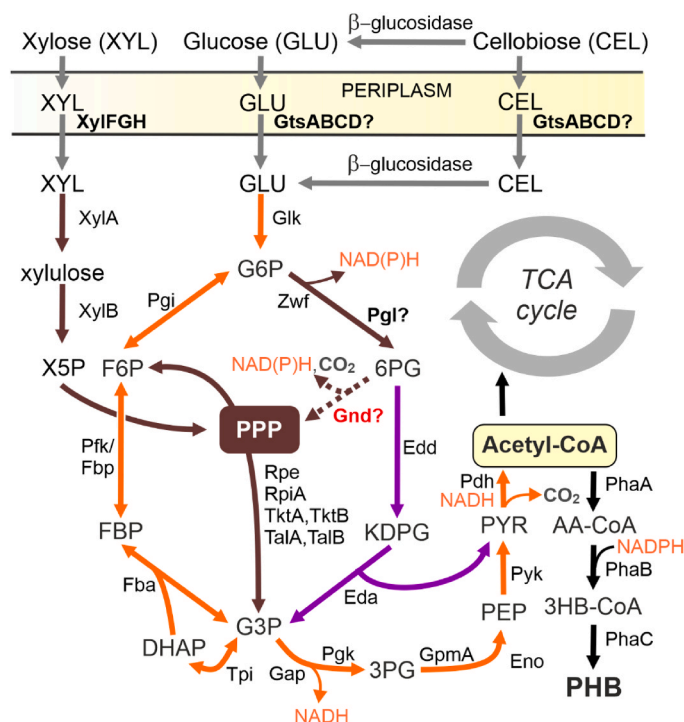
Fig. 4. Schematic illustration of proposed carbohydrate metabolism in studied *Caldimonas/Schlegelella* strains. Xylose isomerase pathway (formed by XylA xylose isomerase and XylB xylulokinase) and the pentose phosphate pathway are shown using brown arrows, the Embden-Meyerhof-Parnas pathway is shown with orange arrows, and the Entner-Doudoroff pathway with magenta arrows. Abbreviations (enzymes): Eda, 2-keto-3-deoxy-6-phosphogluconate aldolase; Edd, 6-phosphogluconate dehydratase; Eno, phosphopyruvate hydratase; Fba, fructose-1,6-biphosphate aldolase; Fbp, fructose-1,6-biphosphatase; Gap, glyceraldehyde-3-phosphate dehydrogenase; Gcd, glucose dehydrogenase; Gnd, 6-phosphogluconate dehydrogenase; Pdh, pyruvate dehydrogenase; Pgi, glucose-6-phosphate isomerase; Pgk, phosphoglycerate kinase; Pgl, 6-phosphogluconolactonase; Pgm, phosphoglycerate mutase; PhaA, acetyl-CoA acetyltransferase (3-ketothiolase); PhaB, acetoacetyl-CoA reductase; PhaC, poly(3-hydroxyalkanoate) polymerase; Pyk, pyruvate kinase; Rpe, ribulose-5-phosphate 3-epimerase; RpiA, ribose-5-phosphate isomerase; Tal, transaldolase; Tkt, transketolase; Tpi, triosephosphate isomerase; Zwf glucose-6-phosphate dehydrogenase. Abbreviations (metabolites): AA-CoA, acetoacetyl co-enzyme A; Acetyl-CoA, acetyl coenzyme A; DHPA, dihydroxyacetone phosphate; E4P, erythrose 4-phosphate; FBP, fructose 1,6-biphosphate; F6P, fructose 6-phosphate; GLL, glucono-δ-lactone; GLN, gluconate; G3P glyceraldehyde 3-phosphate; G6P, glucose 6-phosphate; 3HB-CoA, 3-hydroxybutyryl coenzyme A; KDPG, 2-keto-3-deoxy-6-phosphogluconate; 2KG, 2-ketogluconate; 2KG-6 P, 2-ketogluconate 6-phosphate; NADH, reduced nicotinamide adenine dinucleotide; NADPH, reduced nicotinamide adenine dinucleotide phosphate; PEP, phosphoenolpyruvate; 3PG, 3-phosphoglycerate; 6PG, 6-phosphogluconate; PHB, poly-3-hydroxybutyrate; R5P, ribose 5-phosphate; Ru5P, ribulose 5-phosphate; S7P, seduheptulose 7-phosphate; XLN, xylonate; X5P, xylulose 5-phosphate. GtsABCD and XylFGH stand for mannose/glucose and xylose ABC transporter, respectively. Note that xylose isomerase pathway and b-glucosidase are not present in *S. aquatica*.

highly effective extracellular PHA depolymerase. In fact, the great PHA degradation efficiency is one of the most notable features of *C. thermodepolymerans*, which is also reflected in its taxonomic name: thermodepolymerans – i.e. capable of degradation of PHA polymers under thermophilic conditions [14]. Further, a gene belonging to the polyhydroxyalkanoic acid group family with unknown function is exclusively present in the genome of *Sa*.

Genomic differences were also confirmed by the cultivation experiments. The ability to produce PHA was demonstrated in all of the investigated strains. However, a significantly lower ability to utilize carbohydrate substrates and produce PHA was observed in *S. aquatica* LMG 23380[T]. *C. thermodepolymerans* DSM 15264 showed the greatest similarity to the type strain *C. thermodepolymerans* DSM 15344[T]. A high increase in biomass was observed for both cellobiose and xylose substrates. The contrasting lack of growth and PHA formation in *S. aquatica* on xylose and cellobiose can be explained by the absence of the *xyl* operon and the β-glucosidase gene in its genome. Unlike *S. aquatica*, all *Ct* strains possess these genes, as reflected in their ability to utilize xylose and cellobiose. It is not clear now why all *Ct* strains grow much worse on glucose than on cellobiose. Since all the bacteria investigated in this study possess genes assumed to encode metabolic traits that ensure glucose utilization (Fig. 4), we hypothesize that the bottleneck limiting the utilization of this sugar may lie in its transport across the cyto-plasmic membrane. The identified GtsABCD glucose/mannose trans-porter might have a higher affinity for cellobiose than for monomeric

glucose, whose transport would be either slow (*Sa* LMG 23380[T], *Ct* DSM 15344[T]) or negligible (*Ct* DSM 15264 and *Ct* LMG 21645). Uptake of cellobiose and other cellooligosaccharides through the ABC-type glucose transporters is common in bacteria [13,43]. The transport of oligosac-charides is more economical with regards to ATP than glucose transport. Our hypothesis is further supported by the presence of the β-glucosidase gene in the *gts* operon of *Ct* strains and by the fact that residual glucose is often detected in culture supernatants of the strains grown on cellobiose (data not shown). The latter observation also implies that at least a part of the cellobiose substrate is cleaved to glucose monomers outside the cell. A signal sequence specific for the twin-arginine translocation (Tat) pathway was identified at the 5′ end of the β-glucosidase genes in the *Ct* strains. Since the Tat pathway is known to translocate already folded (and active) proteins across the cytoplasmic membrane [4], we argue that cellobiose hydrolysis by β-glucosidase in the *Ct* strains studied takes place both in the cytoplasm and outside the cell (Fig. 4).

The proposed scheme of the upper carbohydrate metabolism in the studied thermophilic PHA producers (Fig. 4) suggests that all three tested lignocellulosic sugars can be utilized, despite the lack of *pgl* and *gnd* homologs in the *Caldimonas* genomes as hydrolysis of 6-phosphoglu-conolactone can occur spontaneously [16]. A possible compensatory mechanism for the lack of Gnd is gluconate 6-phosphate undergoing the ED shunt to glyceraldehyde 3-phosphate and pyruvate. This way, 6-phosphogluconate would not become a dead-end product and at least 1 mol of NADPH per 1 mol of glucose would be produced in the

glucose-6-phosphate dehydrogenase (Zwf) reaction in the oxidative branch of the PPP. During the growth on glucose or cellobiose, at least a small reverse flux from glyceraldehyde 3-phosphate and fructose 6-phosphate to the non-oxidative PPP is required to ensure the formation of nucleotide precursors. Hence, the preference of xylose over glucose in *Caldimonas* strains might be explained by the fact that the pathway for securing important metabolic precursors in PPP such as ribulose 5-phosphate is shorter with pentose. However, more experimental data is necessary to investigate this phenomenon.

5. Conclusions

In conclusion, the genomic and phenotypic characterization of the non-model bacteria *Schlegelella aquatica* LMG 23380[T] and *Caldimonas thermodepolymerans* DSM 15264 and LMG 21645 provides valuable insights into microbial production of polyhydroxyalkanoates, sustainable and environmentally friendly polyesters. The genome assembly and functional annotation of these bacteria confirm their potential with regards to PHA production and reveal their individual characteristics. Particularly, the unique *xyl* operon present only in *C. thermodepolymerans* suggests their strong potential for biotechnological PHA production from xylose-rich resources. Furthermore, the fact that the strains of *C. thermodepolymerans* are also capable of utilizing cellobiose is of biotechnological interest since cellobiose is commonly present in lignocellulose-based media when enzymatic hydrolysis of cellulose is employed. Further, the fact that the studied strains of *C. thermodepolymerans* are capable of cellobiose utilization also indicates that the low efficiency of glucose metabolism might be linked to difficulties with the transport of glucose into the cells rather than a deficiency in its metabolization in the cells. In this case, glucose uptake and conversion into PHA might be potentially improved by introducing heterologous powerful glucose transporter using approaches of synthetic biology. Therefore, further investigation is required to fully understand the organism's properties and improve its potential for PHA production in agreement with the concept of NGIB.

CRediT authorship contribution statement

Jana Musilova: Methodology, Software, Formal analysis, Investigation, Data Curation, Writing - Original Draft, Writing - Original Draft, Visualization. **Xenie Kourilova**: Methodology, Validation, Formal analysis, Investigation, Writing - Original Draft, Writing - Review & Editing. **Kristyna Hermankova**: Software, Formal analysis, Data Curation, Writing - Review & Editing, Visualization. **Matej Bezdicek**: Investigation, Writing - Original Draft. **Anastasiia Ieremenko**: Investigation. **Pavel Dvorak**: Conceptualization, Formal analysis, Writing - Original Draft, Writing - Review & Editing, Visualization, Funding acquisition. **Stanislav Obruca**: Conceptualization, Writing - Original Draft, Writing - Review & Editing, Visualization, Funding acquisition. **Karel Sedlar**: Conceptualization, Formal analysis, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization, Funding acquisition.

Declaration of Competing Interest

None.

Data Availability

The whole-genome sequences of *S. aquatica* LMG 23380[T], *C. thermodepolymerans* DSM 15264 and *C. thermodepolymerans* LMG 21645 have been deposited in the DDBJ/ENA/GenBank under the accessions CP110257, CP110416, and CP110415, respectively. The raw paired-end Illumina reads have been deposited in the NCBI SRA database under the accession numbers SRX20788156, SRX20788022, and SRX20788024; Oxford Nanopore Technologies raw reads are accessible

under the numbers SRX20788157, SRX20788023, and SRX20788025.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.csbj.2023.10.051.

References

[1] 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. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. Nucleic Acids Res 2020;48: D517–25. https://doi.org/10.1093/nar/gkz935.
[2] Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol 1990;215:403–10. https://doi.org/10.1016/S0022-2836(05) 80360-2.
[3] Bailey TL. STREME: accurate and versatile sequence motif discovery. Bioinformatics 2021;37:2834–40. https://doi.org/10.1093/BIOINFORMATICS/ BTAB203.
[4] Berks BC, Palmer T, Sargent F. Tat Protein Translocat Pathw its role Microb Physiol 2003;187–254.
[5] Biswas A, Staals RHJ, Morales SE, Fineran PC, Brown CM. CRISPRDetect: a flexible algorithm to define CRISPR arrays. BMC Genom 2016;17:356. https://doi.org/ 10.1186/s12864-016-2627-0.
[6] Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 2014;30:2114–20. https://doi.org/10.1093/ bioinformatics/btu170.
[7] Cantalapiedra CP, Hernandez-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. Mol Biol Evol 2021;38:5825–9. https://doi.org/ 10.1093/MOLBEV/MSAB293.
[8] Chaudhary DK, Lee H, Dahal RH, Kim D-U. *Schlegelella koreensis* sp. nov., isolated from evaporator core of automobile air conditioning system. Arch Microbiol 2021; 203:2373–8. https://doi.org/10.1007/s00203-021-02206-9.
[9] Chen GQ, Jiang XR. Next generation industrial biotechnology based on extremophilic bacteria. Curr Opin Biotechnol 2018;50:94–100. https://doi.org/ 10.1016/j.copbio.2017.11.016.
[10] Chou Y-J, Sheu S-Y, Sheu D-S, Wang J-T, Chen W-M. *Schlegelella aquatica* sp. nov., a novel thermophilic bacterium isolated from a hot spring. Int J Syst Evol Microbiol 2006;56:2793–7. https://doi.org/10.1099/ijs.0.64446-0.
[11] Coker JA. Extremophiles and biotechnology: current uses and prospects. F1000Research 2016;5:396.
[12] Dietrich K, Dumont M-J, Del Rio LF, Orsat V. Sustainable PHA production in integrated lignocellulose biorefineries. N Biotechnol 2019;49:161–8. https://doi. org/10.1016/j.nbt.2018.11.004.
[13] Dvořák P, de Lorenzo V. Refactoring the upper sugar metabolism of Pseudomonas putida for co-utilization of cellobiose, xylose, and glucose. Metab Eng 2018;48: 94–108. https://doi.org/10.1016/j.ymben.2018.05.019.
[14] Elbanna K, Lütke-Eversloh T, Van Trappen S, Mergaert J, Swings J, Steinbüchel A. *Schlegelella thermodepolymerans* gen. nov., sp. nov., a novel thermophilic bacterium that degrades poly(3-hydroxybutyrate-co-3-mercaptopropionate). Int J Syst Evol Microbiol 2003;53:1165–8. https://doi.org/10.1099/ijs.0.02562-0.
[15] 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.
[16] Fatima T, Rani S, Fischer S, Efferth T, Kiani FA. The hydrolysis of 6-phosphoglu-conolactone in the second step of pentose phosphate pathway occurs via a two-water mechanism. Biophys Chem 2018;240:98–106. https://doi.org/10.1016/j. bpc.2018.06.002.
[17] Gaj T, Sirk SJ, Shui S, Liu J. Genome-editing technologies: principles and applications. Cold Spring Harb Perspect Biol 2016;8:a023754. https://doi.org/ 10.1101/cshperspect.a023754.
[18] Hi L. (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv: Genomics.
[19] Hu EZ, Lan XR, Liu ZL, Gao J, Niu DK. A positive correlation between GC content and growth temperature in prokaryotes. BMC Genom 2022;23:1–17. https://doi. org/10.1186/S12864-022-08353-7/FIGURES/4.
[20] Ibrahim MHA, Steinbüchel A. High-cell-density cyclic fed-batch fermentation of a Poly(3-Hydroxybutyrate)-accumulating thermophile, chelatococcus sp. strain

MW10. Appl Environ Microbiol 2010;76:7890–5. https://doi.org/10.1128/AEM.01488-10.

[21] Kanehisa M. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 2000;28:27–30. https://doi.org/10.1093/nar/28.1.27.

[22] Knoll M, Hamm TM, Wagner F, Martinez V, Pleiss J. The PHA depolymerase engineering database: a systematic analysis tool for the diverse family of polyhydroxyalkanoate (PHA) depolymerases. BMC Bioinforma 2009;10:89. https://doi.org/10.1186/1471-2105-10-89.

[23] Koller M, Maršálek L, de Sousa Dias MM, Braunegg G. Producing microbial polyhydroxyalkanoate (PHA) biopolyesters in a sustainable manner. N Biotechnol 2017;37:24–38. https://doi.org/10.1016/j.nbt.2016.05.001.

[24] Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol 2019 375 2019;37:540–6. https://doi.org/10.1038/s41587-019-0072-8.

[25] Kourilova X, Novackova I, Koller M, Obruca S. Evaluation of mesophilic *Burkholderia sacchari*, thermophilic *Schlegelella thermodepolymerans* and halophilic *Halomonas halophila* for polyhydroxyalkanoates production on model media mimicking lignocellulose hydrolysates. Bioresour Technol 2021;325:124704. https://doi.org/10.1016/j.biortech.2021.124704.

[26] Kourilova X, Pernicova I, Sedlar K, Musilova J, Sedlacek P, Kalina M, Koller M, Obruca S. Production of polyhydroxyalkanoates (PHA) by a thermophilic strain of *Schlegelella thermodepolymerans* from xylose rich substrates. Bioresour Technol 2020;315:123885. https://doi.org/10.1016/j.biortech.2020.123885.

[27] Lanfear R, Schalamun M, Kainer D, Wang W, Schwessinger B. MinIONQC: fast and simple quality control for MinION sequencing data. Bioinformatics 2019;35:523–5. https://doi.org/10.1093/bioinformatics/bty654.

[28] Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 2018;34:3094–100. https://doi.org/10.1093/bioinformatics/bty191.

[29] Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. The sequence alignment/map format and SAMtools. Bioinformatics 2009;25:2078–9. https://doi.org/10.1093/bioinformatics/btp352.

[30] Li X-Q, Du D. Variation, EVolution, and Correlation Analysis of C+G content and genome or chromosome size in different kingdoms and phyla. PLoS One 2014;9: 88339. https://doi.org/10.1371/journal.pone.0088339.

[31] Liu Y, Du J, Pei T, Du H, Feng GDa, Zhu H. Genome-based taxonomic classification of the closest-to-*Comamonadaceae* group supports a new family *Sphaerotilaceae* fam. nov. and taxonomic revisions. Syst Appl Microbiol 2022;45:126352. https://doi.org/10.1016/J.SYAPM.2022.126352.

[32] Maehara A, Taguchi S, Nishiyama T, Yamane T, Doi Y. A repressor protein, phar, regulates polyhydroxyalkanoate (PHA) synthesis via its direct interaction with PHA. J Bacteriol 2002;184:3992–4002. https://doi.org/10.1128/JB.184.14.3992-4002.2002.

[33] Marchler-Bauer A, Bryant SH. CD-Search: protein domain annotations on the fly. Nucleic Acids Res 2004;32. https://doi.org/10.1093/NAR/GKH454.

[34] Martin A, McMinn A. Sea ice, extremophiles and life on extra-terrestrial ocean worlds. Int J Astrobiol 2018;17:1–16. https://doi.org/10.1017/S1473550416000483.

[35] Meier-Kolthoff JP, Göker M. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. Nat Commun 2019;101(10):1–10. https://doi.org/10.1038/s41467-019-10210-3.

[36] Mezzina MP, Pettinari Phasina MJ. Multifaceted polyhydroxyalkanoate granule-associated proteins. Appl Environ Microbiol 2016;82:5060–7. https://doi.org/10.1128/AEM.01161-16.

[37] Musilova J, Kourilova X, Bezdicek M, Lengerova M, Obruca S, Skutkova H, Sedlar K. First complete genome of the thermophilic polyhydroxyalkanoates-producing bacterium *Schlegelella thermodepolymerans* DSM 15344. Genome Biol Evol 2021;13. https://doi.org/10.1093/GBE/EVAB007.

[38] Naik S, Venu Gopal SK, Somal P. Bioproduction of polyhydroxyalkanoates from bacteria: a metabolic approach. World J Microbiol Biotechnol 2008;24:2307–14. https://doi.org/10.1007/s11274-008-9745-z.

[39] Nunes OC, Manaia CM, Vaz-Moreira I. Schlegelella. Bergey's Man Syst Archaea Bact 2021:1–11. https://doi.org/10.1002/9781118960608.GBM01829.

[40] Obruca S, Sedlacek P, Koller M, Kucera D, Pernicova I. Involvement of polyhydroxyalkanoates in stress resistance of microbial cells: Biotechnological consequences and applications. Biotechnol Adv 2018;36:856–70. https://doi.org/10.1016/j.biotechadv.2017.12.006.

[41] Obruca S, Petrik S, Benesova P, Svoboda Z, Eremka L, Marova I. Utilization of oil extracted from spent coffee grounds for sustainable production of polyhydroxyalkanoates. Appl Microbiol Biotechnol 2014;98:5883–90. https://doi.org/10.1007/S00253-014-5653-3.

[42] Obruča S, Dvořák P, Sedláček P, Koller M, Sedlář K, Pernicová I, Šafránek D. Polyhydroxyalkanoates synthesis by halophiles and thermophiles: towards sustainable production of microbial bioplastics. Biotechnol Adv 2022;58:107906. https://doi.org/10.1016/j.biotechadv.2022.107906.

[43] Parisutham V, Chandran S-P, Mukhopadhyay A, Lee SK, Keasling JD. Intracellular cellobiose metabolism and its applications in lignocellulose-based biorefineries. Bioresour Technol 2017;239:496–506. https://doi.org/10.1016/j.biortech.2017.05.001.

[44] Pradhan S, Dikshit PK, Moholkar VS. Prod, Charact, Appl Biodegrad Polym: Polyhydroxyalkanoates 2020:51–94. https://doi.org/10.1007/978-981-15-1251-3_4.

[45] Riley LA, Guss AM. Approaches to genetic tool development for rapid domestication of non-model microorganisms. Biotechnol Biofuels 2021;14:30. https://doi.org/10.1186/s13068-020-01872-z.

[46] Roberts RJ, Vincze T, Posfai J, Macelis D. REBASE—a database for DNA restriction and modification: enzymes, genes and genomes. Nucleic Acids Res 2015;43: D298–9. https://doi.org/10.1093/nar/gku1046.

[47] Rothschild LJ, Mancinelli RL. Life in extreme environments. Nat 2001;4096823 (409):1092–101. https://doi.org/10.1038/35059215.

[48] Sabapathy PC, Devaraj S, Meixner K, Anburajan P, Kathirvel P, Ravikumar Y, Zabed HM, Qi X. Recent developments in Polyhydroxyalkanoates (PHAs) production – a review. Bioresour Technol 2020;306:123132.

[49] 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. https://doi.org/10.1038/ncomms3304.

[50] Shogren R, Wood D, Orts W, Glenn G. Plant-based materials and transitioning to a circular economy. Sustain Prod Consum 2019;19:194–215. https://doi.org/10.1016/j.spc.2019.04.007.

[51] Simpson JT, Workman RE, Zuzarte PC, David M, Dursi LJ, Timp W. Detecting DNA cytosine methylation using nanopore sequencing. Nat Methods 2017;14:407–10. https://doi.org/10.1038/nmeth.4184.

[52] Steinbüchel A. Polyhydroxyalkanoic acids. In: Byrom D, editor. Biomaterials. London: Palgrave Macmillan; 1991. p. 123–213.

[53] 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: 4118–20. https://doi.org/10.1093/bioinformatics/bty496.

[54] Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. NCBI prokaryotic genome annotation pipeline. Nucleic Acids Res 2016;44:6614–24. https://doi.org/10.1093/nar/gkw569.

[55] Tindall BJ, Rosselló-Móra R, Busse H-J, Ludwig W, Kämpfer P. Notes on the characterization of prokaryote strains for taxonomic purposes. Int J Syst Evol Microbiol 2010;60:249–66. https://doi.org/10.1099/ijs.0.016949-0.

[56] Umeda T, Kitano Y, Murakami Y, Yagi K, Miura Y, Mizoguchi T. Cloning and sequence analysis of the poly(3-hydroxyalkanoic acid)-synthesis genes of *Pseudomonas acidophila*. Appl Biochem Biotechnol 70–72 1998:341–52. https://doi.org/10.1007/BF02920150.

[57] Vaser R, Sović I, Nagarajan N, Šikić M. Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res 2017;27:737–46. https://doi.org/10.1101/gr.214270.116.

[58] Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 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.

[59] Walker JE, Lanahan AA, Zheng T, Toruno C, Lynd LR, Cameron JC, Olson DG, Eckert CA. Development of both type I–B and type II CRISPR/Cas genome editing systems in the cellulolytic bacterium *Clostridium thermocellum*. Metab Eng Commun 2020;10:e00116. https://doi.org/10.1016/j.mec.2019.e00116.

[60] Wu H, Zhang Z, Hu S, Yu J. On the molecular mechanism of GC content variation among eubacterial genomes. Biol Direct 2012;7:2. https://doi.org/10.1186/1745-6150-7-2.