Microbial Pathogenesis 137 (2019) 103759

Contents lists available at ScienceDirect

Microbial Pathogenesis

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

# Complete genome sequencing of *Shigella* sp. PAMC 28760: Identification of CAZyme genes and analysis of their potential role in glycogen metabolism for cold survival adaptation

So-Ra Han[a,1], Do Wan Kim[b,1], Byeollee Kim[a], Young Min Chi[b], Seunghyun Kang[c], Hyun Park[c,d], Sang-Hee Jung[e], Jun Hyuck Lee[c,d,**], Tae-Jin Oh[a,f,g,*]

[a] Department of Life Science and Biochemical Engineering, Graduate School, SunMoon University, Asan, South Korea
[b] Division of Biotechnology, College of Life Sciences and Biotechnology, Korea University, Seoul, 02841, South Korea
[c] Unit of Polar Genomics, Korea Polar Research Institute, Incheon, South Korea
[d] Department of Polar Sciences, University of Science and Technology, Incheon, South Korea
[e] Department of Dental Hygiene, Gangneung Yeongdong University, Gangneung, South Korea
[f] Genome-based BioIT Convergence Institute, Asan, South Korea
[g] Department of Pharmaceutical Engineering and Biotechnology, SunMoon University, Asan, South Korea

ARTICLE INFO

*Keywords:*
Amylolytic characteristic
Antarctic lichen
CAZyme
Genome sequencing
*Shigella*

ABSTRACT

*Shigella* sp. PAMC 28760 (isolated from *Himantormia* sp. lichen in Antarctica) is a gram-negative, non-sporulating bacterium that has cellulolytic and amylolytic characteristics as well as glycogen metabolic pathways. In this study, we isolated *S.* sp. PAMC 28760 from Antarctic lichen, and present the complete genome sequence with annotations describing its unique features. The genome sequence has 58.85% GC content, 4,278 coding DNA sequences, 85 tRNAs, and 22 rRNA operons. 16S rRNA gene sequence analyses revealed strain PAMC 28760 as a potentially new species of genus *Shigella*, showing various differences from pathogenic bacteria reported previously. dbCAN2 analyses revealed 91 genes related to carbohydrate-metabolizing enzymes. *S.* sp. PAMC 28760 likely degrades polysaccharide starch to obtain glucose for energy conservation. This study provides a foundation for understanding *Shigella* survival adaptation mechanisms under extremely cold Antarctic conditions.

## 1. Introduction

Polysaccharides, the most abundant organic energy source found in nature, consist mainly of phenolic polymers such as cellulose, hemicellulose, and pectin [1]. Many microorganisms possess a series of synergistic enzymes capable of decomposing plant cell walls, which leads to the release of free glucose that fuels microbial metabolism for maintaining growth [2]. Starch is an excellent source of carbon and energy for many microbes that employ dedicated protein sets involved in extracellular hydrolysis of polysaccharides, in-cell absorption of short-chain fructose, and further decomposition into glucose [3,4]. In addition, strains with glycogen metabolism are responsible for various important physiological functions, including energy storage during glycogen synthesis and decomposition. These metabolic pathways act as carbon capacitors that regulate carbon flux [5]. Starch degradation in most microbes requires various amylolytic characteristics involving carbohydrate active enzymes (CAZymes). Several studies have reported cellulolytic or amylolytic bacteria that utilize CAZyme biodegradation [6,7]. Cellulose decomposition requires cellulase, a highly specific class of enzyme that can break down cellulose glycosidic bonds. These enzymes are closely related to those required for glycogen degradation and processing. Cellulolytic and amylolytic bacterial activities are induced by the presence of polysaccharides. However, the regulatory mechanisms of the synthesis and cohesion of cell surface complexes remain unclear. Several bacteria tend to utilize highly available carbohydrates, such as glucose, to suppress polysaccharide utilization systems [8].

*Shigella* sp., deposited as PAMC 28760 in the Polar and Alpine Microbial Collection (Korea Polar Research Institute, KOPRI, Incheon, Korea), was isolated from *Himantormia* sp., an Antarctic lichen collected

* Corresponding author. Department of Pharmaceutical Engineering and Biotechnology, Sunmoon University, Asan, 31460, South Korea. Tel.: +82 41 530 2677; fax.: +82 41 530 2279.
** Corresponding author. Unit of Polar Genomics, Korea Polar Research Institute, Incheon, 21990, South Korea.
*E-mail addresses:* junhyucklee@kopri.re.kr (J.H. Lee), tjoh3782@hanmail.net (T.-J. Oh).
[1] These two authors contributed equally to this work.

https://doi.org/10.1016/j.micpath.2019.103759
Received 6 June 2019; Received in revised form 18 September 2019; Accepted 20 September 2019
Available online 24 September 2019

from the Barton Peninsula, King George Island, Antarctica (62°13′S, 58°47′W). The continent of Antarctica is a land mass covered by ice up to 13,000 ft thick with areas of exposed rock; the major vegetation are small mosses and lichens [9]. As Antarctica is the highest, driest, windiest, and coldest continent in the world, only certain bacteria adapted to these harsh conditions can grow and survive. Microbes living in persistent cold environments, such as deep sea and polar habitats, have developed specialized characteristics to thrive in such environments [10]. Recent microbiological studies of specimens from Antarctica have led to discoveries of unusual microbial diversity, new species, and cold-adaptive enzymes. However, little is known about their polysaccharide utilization systems due to the difficulty in collecting samples [11,12].

*Shigella* causes endemic bacillary dysentery, which poses a significant public health threat [13]. There are four known *Shigella* species: *S. dysenteriae*, *S. boydii*, *S. sonnei* and *S. flexneri* (the most common of the species). These species have interesting discovery histories spanning ~100 years, and each species has unique pathogenic characteristics [14]. Several *Shigella* genomes have been sequenced, revealing dynamic and diverse features [15,16].

This study is the first to report the complete genome of a *Shigella* strain isolated from an Antarctic lichen. Furthermore, it is the first to describe the CAZyme gene characteristic in a *Shigella* strain. These results shed light on the microbial's use of available carbohydrates, such as glucose, to suppress polysaccharide utilization systems. This survival and adaptation strategy may be essential for subsisting in the extreme environments of Antarctica.

## 2. Materials and methods

### 2.1. Isolation of S. sp. PAMC 28760 and genomic DNA preparation

*S.* sp. PAMC 28760 was isolated from the Antarctic lichen *Himantormia* sp. using sterilized scissors and 0.85% NaCl solution. After vortexing for 1 min, the lichen solution was spread on R2A media (MB cell Ltd., Seoul, Korea) and incubated at 15 °C. A single colony was selected and stored at −70 °C in 20% glycerol. This strain was grown in 200 mL R2A broth at 15 °C for ~5–10 days before use in experiments. The genomic DNA of *S.* sp. PAMC 28760 was extracted using the QIAamp DNA Mini Kit (Qiagen Inc., Valencia, CA, USA) according to the manufacturer's protocol. DNA quality was verified by agarose gel electrophoresis, and DNA concentration and purity (A260/A280) were measured using a spectrophotometer (Biochrome, Libra S35PC, UK).

### 2.2. Genome sequencing and annotation

Genome sequencing was conducted by KOPRI, and the complete genome sequence was deposited in the GenBank database under the GenBank accession number CP014768.1 in March 2016. This sequence was then used to generate the comparative genomics data presented herein. DNA quantity and purity were determined using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). Genome sequencing was performed using PacBio RS II single-molecule real-time (SMRT) sequencing technology (Pacific Biosciences, Menlo Park, CA, USA). SMRTbell library inserts (20 kb) were sequenced using SMRT cells. Raw sequence data were generated from 77,075 reads and 821,081,934 bp that were assembled de novo using the hierarchical genome-assembly process (HGAP) protocol [17] and RS HGAP Assembly 2 in the SMRT analysis software (ver. 2.3; Pacific Biosciences, https://github.com/PacificBiosciences/SMRT-Analysis). The coding DNA sequences (CDSs) were predicted and annotation was performed by Rapid Annotation using Subsystem Technology (RAST) server [18]. The predicted gene sequences were translated and searched against the National Center for Biotechnology Information (NCBI) non-redundant database, the Clusters of Orthologous Group (COG) database, and the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. We present the genomic features as a circular map of the PAMC 28760 genome, created by the CGView comparison tool [19] (Fig. 1). CAZyme gene analyses were carried out by running HMMER3 [20] scans using hidden Markov model (HMM) profile downloaded from dbCAN2

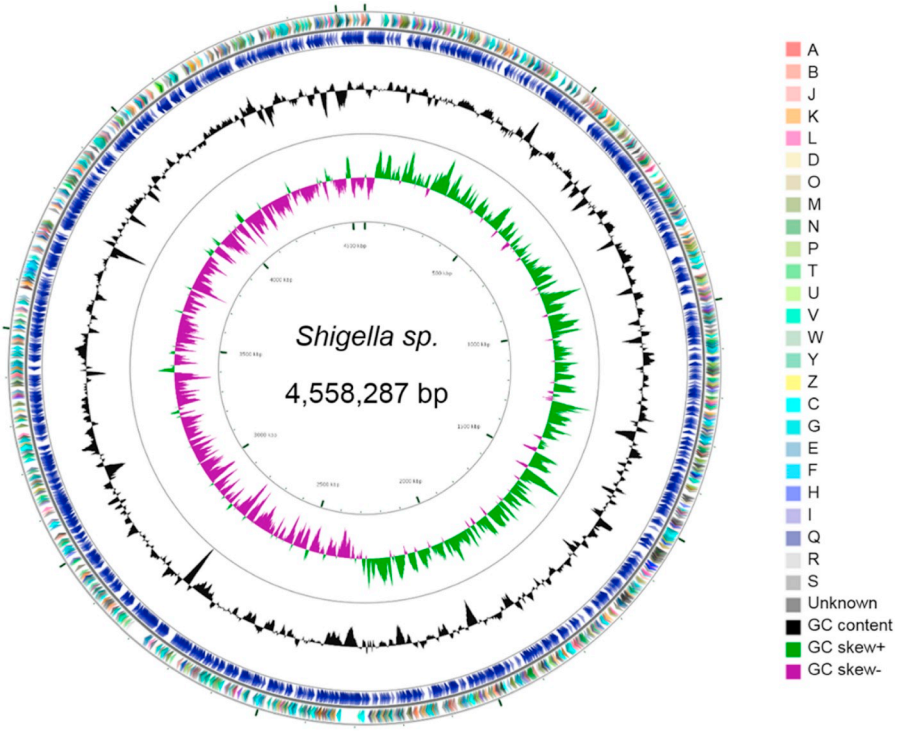


**Fig. 1. *Shigella* sp. PAMC 28760 genome: circular map generated using the CGView Comparison Tool.** The concentric circles (from outside to inside) represent the following sequence categories: 1, Clusters of Orthologous Group (COG) functional categories for forward coding sequences; 2, forward coding sequence features; 3, reverse coding sequence features; 4, COG functional categories for reverse coding sequences; 5, GC content; and 6, GC skew. The COG functional gene categories are as follows: A, RNA processing and modification; B, chromatin structure and dynamics; J, translation, ribosomal structure, and biogenesis; K, transcription; L, replication, recombination and repair; D, cell cycle control, cell division, and chromosome partitioning; O, posttranslational modification, protein turnover, and chaperones; M, cell wall/membrane/envelope biogenesis; N, cell motility; P, inorganic ion transport and metabolism; T, signal transduction mechanisms; U, intracellular trafficking, secretion, and vesicular transport; V, defense mechanisms; W, extracellular structures; Y, nuclear structure; Z, mobilome, prophages, and transposons; C, energy production and conversion; G, carbohydrate transport and metabolism; E, amino acid transport and metabolism; F, nucleotide transport and metabolism; H, coenzyme transport and metabolism; I, lipid transport and metabolism; Q, secondary metabolites biosynthesis, transport, and catabolism; R, general function prediction only; and S, function unknown.

HMMdb (version 7.0). The e-value cutoff was 1e-15 and the coverage cutoff was > 0.35. In addition, we used DIAMOND [21] (e-value < 1e-102) and Hotpep [22] (frequency > 2.6, hits > 6), to improve the prediction accuracy.

### 2.3. 16s rRNA phylogenetic tree, ANI analysis, and comparative genomics

We verified the 16s rRNA sequences for the selected bacteria. The sequences were confirmed by Basic Local Alignment Search Tool (BLAST) search of the NCBI database and analyzed using EzBio Cloud (www.ezbiocloud.net). The results confirmed that the genus of the strain was *Shigella*. The sequences were aligned using MUSCLE [23,24], and MEGA X [25] software, to reconstruct a neighbor-joining tree [26] with 1,000 bootstrap replications [27]. ANI analyses were performed to determine the relationship of *Shigella* species with other strains, as predicted by their complete genome sequences. The average nucleotide identity (ANI) values between the genomes were calculated using a method based on MUMmer alignment [28]. The genomes of many *Shigella* strains have previously been registered in GenBank. To compare the CAZymes of various registered *Shigella* strains, a list of strains was created using CAZy (http://www.cazy.org/) and compared to similar strains chosen based on ANI values; a total of 14 strains were selected in this manner. We downloaded the genome sequence for each of the 14 strains from the NCBI database and reannotated the CAZymes using the dbCAN2 server.

## 3. Results and discussion

### 3.1. 16s rRNA phylogenetic analysis, genome sequencing assembly, and annotation

As shown in Fig. 2, PAMC 28760 clusters with *Shigella* sp. in the phylogenetic tree. BLAST analyses revealed that the closely related strains include *S. flexneri* ATCC 29903(T) (99.86%), *S. sonnei* CECT 4887(T) (99.73%), *S. boydii* GTC 779(T) (99.39%), *Escherichia fergusonii* ATCC 35469(T) (99.80%), and *E. coli* ATCC 11775(T) (99.72%). These results confirmed that PAMC 28760 belongs to Gammaproteobacteria among the Proteobacteria. To determine whether PAMC 28760 lies between *S.* sp. and *E.* sp., we sequenced the genome to detect a possible subspecies between these two species. The complete genome of *S.* sp. PAMC 28760 contains a circular chromosome of 4,558,287 bp with 50.8% GC content. Our analyses predicted 4,278 CDSs, 85 tRNAs, and 22 rRNA operons in the *S.* sp. PAMC 28760 genome (Table 1). The genes were classified into 3,562 COG functional categories. Fig. 1 shows a circular map of the PAMC 28760 genome, which was generated using the CGView Comparison Tool [19]. We analyzed the following functional gene categories: G, carbohydrate transport and metabolism; E, amino acid transport and metabolism; K, transcription; C, energy production and conversion; and M, cell wall/membrane/envelope biogenesis. A comparison of the COG results among strains revealed that PAMC 28760 has large numbers of genes in categories G (375), E (350),

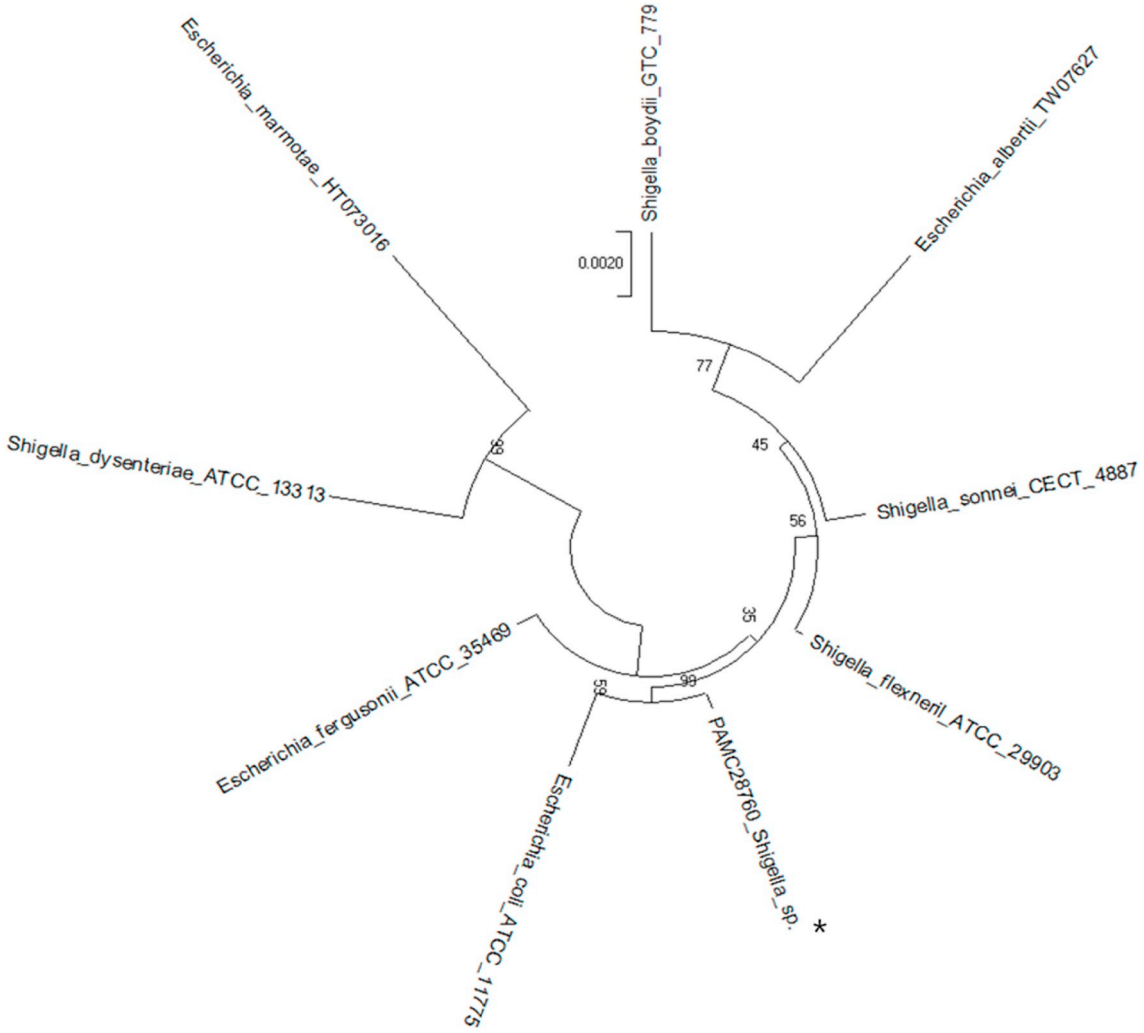


**Fig. 2. Phylogenetic tree of *S.* sp. PAMC 28760**. Phylogenetic tree showing the relationships of the four *Shigella* strains and four *Escherichia* strains and their phylogenetic position. It was prepared using MEGA X based on 16s rRNA sequences with neighbor-joining. * mean our strain.

**Table 1**
Genomic features of *Shigella* sp. PAMC 28760.

| Features | Value |
|---|---|
| Genome size (bp) | 4,558,287 |
| Contig | 1 |
| GC content (%) | 50.84 |
| Protein coding genes | 4,287 |
| rRNA genes | 22 |
| tRNA genes | 85 |

K (295), C (294), and M (278). We identified many genes related to carbohydrate transport and metabolism, suggesting that this strain utilizes CAZymes. Carbohydrates play a key role in many essential metabolic pathways [29]. Plant rely on photosynthesis to convert carbon dioxide and carbohydrates in water to produce stored energy [30]. Bacteria use cell respiration to degrade the stored carbohydrates in plants to obtain energy. The energy is stored temporarily in the form of high-energy molecules, such as ATP, for use in various cell processes [31]. Therefore, we predict that PAMC 28760 bacteria isolated from a polar lichen environment utilizes carbohydrate degradation for supplying energy to cells.

### 3.2. CAZymes

dbCAN2 [32], a database for predicting CAZyme coding genes, revealed the presence of CAZyme genes in the PAMC 28760 genome. In particular, dbCAN2 detected CAZyme genes encoding glycoside hydrolase (GH), glycosyl transferase (GT), carbohydrate esterase (CE), auxiliary activities (AA), and carbohydrate-binding modules (CBM). These results provide insight into the carbohydrate utilization mechanisms employed by PAMC 28760. Among the 91 CAZyme genes detected, Signal P analysis predicted that 27 contain signal peptides (Fig. 3). As shown in Table 2, we found 42 genes belonging to the GH family including GH1, GH2, GH3, GH4, GH8, GH13, GH23, GH24, GH25, GH31, GH37, GH38, GH63, GH65, GH73, GH77, GH102, GH103, GH104, GH127, and GH153. However, we found no genes associated with the PL family in the genome. Annotation of the GH genes revealed that the PAMC 28760 genome possesses genes encoding cellulase (endoglucanase, GH8), α-amylase (GH13_19), α-1,6-maltotetraose-hydrolase (GH13_11 + CBM48), and α-glucosidase (CBM34 + GH13_21, GH31). These genes may have potential for use in future biotechnological applications. In addition, this strain has four 6-phospho-β-glucosidase genes (three GH1 and one GH4) belonging to the glycosidases family, which hydrolyzes *O*- and *S*-glycosyl compounds. These genes are characteristic features of celluloytic bacterium, and they likely function as predicted.

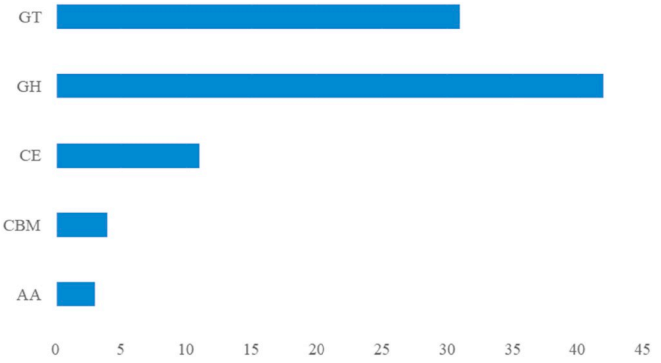


**Fig. 3. Numbers of carbohydrate active enzymes (CAZymes) in *S.* sp. PAMC 28760**. GT, glycosyl transferase; GH, glycoside hydrolase; CE, carbohydrate esterase; CBM, carbohydrate binding; and AA, auxiliary activities.

**Table 2**
Genomic information on CAZymes for 15 strains, including *S.* sp. PAMC 28760.

| | | PAMC 28760 | 1617 | E670/74 | Sd197 | ATCC 12039 | CFSAN 010956 | FORC_011 | CFSAN 030807 | LC 147718 | 2a str. 301 | G1663 | CDC 3083–94 | ATCC 9210 | Sb227 | ATCC 11775 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Genome size(Mb)** | | 4.56 | 4.48 | 5.18576 | 4.4802 | 5.06184 | 4.78297 | 5.13271 | 5.06295 | 4.90598 | 4.82882 | 4.81731 | 4.64652 | 4.57425 | 4.87466 | 5.03483 |
| **G + C content(%)** | | 50.84 | 50.92 | 50.7245 | 50.9501 | 50.8106 | 50.5998 | 50.7531 | 50.7423 | 50.9889 | 50.6659 | 50.669 | 51.0964 | 51.2 | 50.9975 | 50.6635 |
| **Genes** | | 4700 | 4834 | 5644 | 6503 | 5239 | 5012 | 5248 | 5612 | 5039 | 4788 | 5051 | 4616 | 4725 | 5336 | 5010 |
| **Proteins** | | 4335 | 4294 | 4672 | 6409 | 4226 | 4151 | 5130 | 4496 | 4119 | 4313 | 4578 | 4284 | 3908 | 4557 | 4723 |
| **plasmid** | | 0 | 1 | 1 | 2 | 1 | 1 | 3 | 8 | 3 | 1 | 1 | 5 | 0 | 1 | 1 |
| **CAZyme** | **AA** | 3 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 4 |
| | **CBM** | 4 | 2 | 3 | 2 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 2 | 2 | 3 |
| | **CE** | 11 | 7 | 9 | 7 | 9 | 11 | 10 | 10 | 10 | 9 | 9 | 9 | 7 | 7 | 13 |
| | **GH** | 42 | 37 | 38 | 36 | 45 | 34 | 43 | 43 | 43 | 41 | 41 | 35 | 35 | 35 | 43 |
| | **GT** | 31 | 29 | 31 | 29 | 34 | 34 | 28 | 28 | 28 | 30 | 30 | 29 | 28 | 28 | 34 |
| | **Total** | 91 | 76 | 83 | 75 | 92 | 83 | 85 | 85 | 85 | 84 | 84 | 76 | 72 | 72 | 97 |

PAMC 28760 = *Shigella* sp., 1617 = *S. dysenteriae*, E670/74 = *S. dysenteriae*, Sd197 = *S. dysenteriae*, ATCC 12039 = *S. dysenteriae*, CFSAN010956 = *S. dysenteriae*, FORC_011 = *S. sonnei*, CFSAN030807 = *S. sonnei*, LC1477_18 = *S. sonnei*, 2a str. 301 = *S. flexneri*, G1663 = *S. flexneri*, CDC 3083–94 = *S. boydii*, ATCC 9210 = *S. boydii*, Sb227 = *S. boydii*, and ATCC 11775 = *Escherichia coli*.

**Table 3**
ANI value (%) generated using the ANIm for 15 genomes, including *S.* sp. PAMC 28760.

| | PAMC 28760 | 1617 | E670/74 | Sd197 | ATCC 12039 | CFSAN 010956 | FORC _011 | CFSAN 030807 | LC 147718 | 2a str. 301 | G1663 | CDC 3083-94 | ATCC 9210 | Sb227 | ATCC 11775 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **PAMC 28760** | 100.00 | 97.81 | 97.84 | 97.84 | 97.22 | 98.42 | 98.39 | 98.34 | 98.41 | 98.14 | 98.15 | 98.25 | 98.38 | 98.30 | 96.95 |
| **1617** | 97.81 | 100.00 | 97.67 | 99.46 | 97.12 | 97.67 | 97.69 | 97.71 | 97.67 | 97.62 | 97.63 | 97.66 | 97.60 | 97.69 | 96.86 |
| **E670/74** | 97.84 | 97.67 | 100.00 | 97.76 | 97.24 | 98.25 | 98.29 | 98.32 | 98.29 | 98.16 | 98.25 | 98.14 | 98.11 | 98.13 | 96.92 |
| **Sd197** | 97.84 | 99.46 | 97.76 | 100.00 | 97.12 | 97.80 | 97.74 | 97.82 | 97.81 | 97.71 | 97.77 | 97.73 | 97.78 | 97.75 | 96.96 |
| **ATCC 12039** | 97.22 | 97.12 | 97.24 | 97.12 | 100.00 | 97.22 | 97.33 | 97.29 | 97.26 | 97.16 | 97.21 | 97.10 | 97.13 | 97.17 | 97.01 |
| **CFSAN 010956** | 98.42 | 97.67 | 98.25 | 97.80 | 97.22 | 100.00 | 98.65 | 98.63 | 98.70 | 98.38 | 98.36 | 98.69 | 98.71 | 98.77 | 96.96 |
| **FORC _011** | 98.39 | 97.69 | 98.29 | 97.74 | 97.33 | 98.65 | 100.00 | 99.96 | 99.97 | 98.40 | 98.44 | 98.60 | 98.72 | 98.74 | 96.84 |
| **CFSAN 030807** | 98.34 | 97.71 | 98.32 | 97.82 | 97.29 | 98.63 | 99.96 | 100.00 | 99.96 | 98.40 | 98.49 | 98.67 | 98.72 | 98.68 | 96.76 |
| **LC 147718** | 98.41 | 97.67 | 98.29 | 97.81 | 97.26 | 98.70 | 99.97 | 99.96 | 100.00 | 98.40 | 98.42 | 98.65 | 98.72 | 98.71 | 96.76 |
| **2a str. 301** | 98.14 | 97.62 | 98.16 | 97.71 | 97.16 | 98.38 | 98.40 | 98.40 | 98.40 | 100.00 | 99.91 | 98.24 | 98.31 | 98.36 | 96.85 |
| **G1663** | 98.15 | 97.63 | 98.25 | 97.77 | 97.21 | 98.36 | 98.44 | 98.49 | 98.42 | 99.91 | 100.00 | 98.24 | 98.32 | 98.45 | 96.94 |
| **CDC 3083–94** | 98.25 | 97.66 | 98.14 | 97.73 | 97.10 | 98.69 | 98.60 | 98.67 | 98.65 | 98.24 | 98.24 | 100.00 | 99.61 | 99.54 | 97.00 |
| **ATCC 9210** | 98.38 | 97.60 | 98.11 | 97.78 | 97.13 | 98.71 | 98.72 | 98.72 | 98.72 | 98.31 | 98.32 | 99.61 | 100.00 | 99.89 | 96.86 |
| **Sb227** | 98.30 | 97.69 | 98.13 | 97.75 | 97.17 | 98.77 | 98.74 | 98.68 | 98.71 | 98.36 | 98.45 | 99.54 | 99.89 | 100.00 | 96.86 |
| **ATCC 11775** | 96.95 | 96.86 | 96.92 | 96.96 | 97.01 | 96.96 | 96.84 | 96.76 | 96.76 | 96.85 | 96.94 | 97.00 | 96.86 | 96.86 | 100.00 |

PAMC 28760 = *Shigella* sp., 1617 = *S. dysenteriae*, E670/74 = *S. dysenteriae*, Sd197 = *S. dysenteriae*, ATCC 12039 = *S. dysenteriae*, CFSAN010956 = *S. dysenteriae*, FORC_011 = *S. sonnei*, CFSAN030807 = *S. sonnei*, LC1477_18 = *S. sonnei*, 2a str. 301 = *S. flexneri*, G1663 = *S. flexneri*, CDC 3083–94 = *S. boydii*, ATCC 9210 = *S. boydii*, Sb227 = *S. boydii*, and ATCC 11775 = *Escherichia coli.*

### 3.3. Comparative analysis of CAZyme genes

ANI values (Table 3) ranged from 98.42% (*S. dysenteriae* CFSAN010956) to 96.95% (*E. coli* ATCC 11775). dbCAN2 analysis of the CAZyme genes in the complete genomes of all strains (Table 2) revealed 72 CAZyme genes in *S. boydii* ATCC 9210 and Sb227, and as many as 97 genes in *E. coli* ATCC 11775; however, no PL family genes were detected among the complete genomes. Among the complete genome strains of *Shigella*, we found 15 GH-related genes in common, including GH1, GH2, GH3, GH4, GH8, GH13, GH23, GH24, GH25, GH31, GH37, GH73, GH77, GH102, and GH103 (Fig. 4).

Each CAZyme in strain *S.* sp. PAMC 28760 contained characteristic features involved in pathways for degrading cellulose, glycogen, or starch. Glycogen is the principle polysaccharide in living cells, and many microorganisms accumulate glycogen as energy reserves to cope with extreme environmental conditions [33]. Glycogen has a diameter of 20–50 nm, and consists of α-1,6 branches (8–12%) and α-1,4-glucosidic linkages. Glycogen deposits enable the survival of many microbes, including bacteria and yeast, when energy consumption is low [34,35]. However, most of these enzymes in bacteria are GT or glycosidase, which are not characteristic CAZymes. As summarized in Table 4, GH1, GH3, GH4, GH8, and GH13 families comprise enzymes that break down starch or cellulose. Through comparison of their complete genomes, we found that CAZymes are common in other *Shigella* strains. In addition, GH37 and GH65 (trehalose-related genes) are involved in energy storage. These glycogen degradation pathways have been found in *E. coli* from humans, as well as in *S.* sp., a similar strain. However, whereas various *S.* sp. specimens have been isolated from humans, our strain was isolated from lichen in an extreme environment. These genes show characteristic cellulolytic properties common among strains. In addition, our results revealed that they have a strong ability to store and release energy.

### 3.4. Glycogen metabolism and starch degradation in S. sp. PAMC 28760

Carbohydrate polymers and oligomers are stored by living cells for various purposes, such as to provide a stable source of energy and for stress resistance. Glycogen, one of the most prevalent carbohydrates, has been identified in archaea, bacteria, and eukaryotes [36]. To adapt and survive in a cold environment, organisms must have well-developed functional energy storage systems. The classical pathway (CP) of bacterial glycogen metabolism includes five essential enzymes: ADP-glucose pyrophosphorylase (GlgC), glycogen synthase (GlgA), glycogen branching enzyme (GlgB), glycogen phosphorylase (GlgP), and glycogen debranching enzyme (GlgX) [37]. Previous analyses have shown that most archaeal species lack genes for the CP, and *Picrophilus torridus* DSM

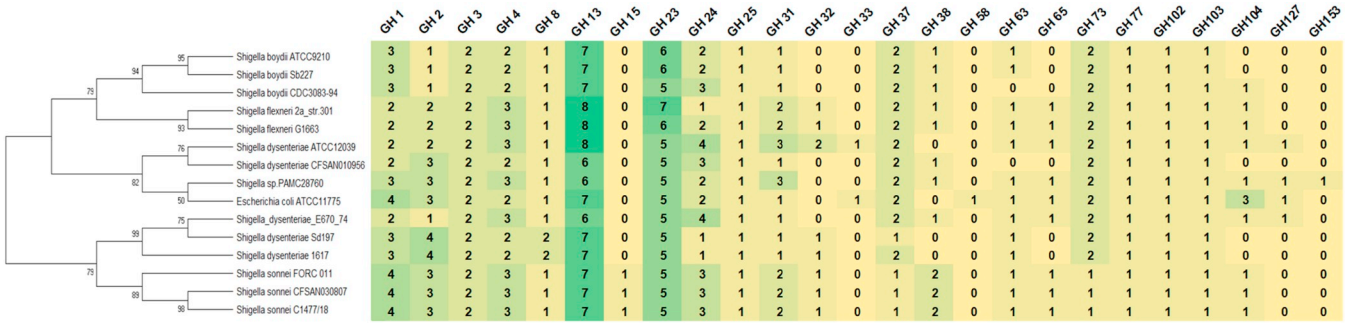


**Fig. 4.** The glycoside hydrolase (GH) enzyme-coding genes in the genomes of 15 strains, including *S.* sp. PAMC 28760.

**Table 4**
The GH families of CAZymes in *S.* sp. PAMC 28760.

| CAZyme family | Substrate | Enzyme activity | EC number | Number |
|---|---|---|---|---|
| GH1 | Cellulose | Alpha-glucosidase (1) | 3.2.1.86 | 3 |
| | Cellulose | 6-Phospho-beta-glucosidase (2) | 3.2.1.86 | |
| GH2 | Xylogulcan | Beta-galactosidase (2) | 3.2.1.23 | 3 |
| | Pectins | Beta-glucuronidase(1) | 3.2.1.31 | |
| GH3 | Peptidoglycan | Beta-N-acetylhexosaminidase | 3.2.1.52 | 2 |
| | Glucan | Beta-glucosidase | 3.2.1.21 | |
| GH4 | Starch | Maltose-6′-phosphate glucosidase | 3.2.1.122 | 3 |
| | Cellulose | 6-Phospho-beta-glucosidase | 3.2.1.86 | |
| | | Alpha-galactosidase | 3.2.1.22 | |
| GH8 | Cellulose | Endo-1,4-D-glucanase | 3.2.1.4 | 1 |
| GH13 | Amylose | Alpha-glucosidase (1) | 3.2.1.20 | 6 |
| | Starch | Limit dextrin alpha-1,6-maltotetraose-hydrolase (1) | 3.2.1.196 | |
| | Starch | 1,4-Alpha-glucan branching enzyme (1) | 2.4.1.18 | |
| | | Glucosylglycerate phosphorylase (1) | 2.4.1.352 | |
| | Starch | Alpha-amylase (1) | 3.2.1.1 | |
| | | Alpha-phosphotrehalase (1) | 3.2.1.93 | |
| GH23 | Peptidoglycan | Membrane-bound lytic murein transglycosylase (1) | 4.2.2.n1 | 5 |
| GH24 | Peptidoglycan | Lysozyme | 3.2.1.17 | 2 |
| GH25 | Peptidoglycan | Lysozyme | 3.2.1.17 | 1 |
| GH31 | Hemicellulose | Sulfoquinovosidase (1) | 3.2.1.199 | 3 |
| | | Alpha-glucosidase (1) | 3.2.1.20 | |
| | Xylogulcan | Alpha-D-xyloside xylohydrolase (1) | 3.2.1.177 | |
| GH37 | Trehalose | Alpha-trehalase | 3.2.1.28 | 2 |
| GH38 | | Mannosylglycerate hydrolase | 3.2.1.170 | 1 |
| GH63 | N-glycans | Glucosidase YgjK | 3.2.1.20 \| 3.2.1.84 | 1 |
| GH65 | Trehalose | Uncharacterized glycosyl hydrolase YcjT | 2.4.1.230 | 1 |
| GH73 | Peptidoglycan | Peptidoglycan hydrolase | 3.2.1.- | 2 |
| GH77 | | 4-Alpha-glucanotransferase | 2.4.1.25 | 1 |
| GH102 | Peptidoglycan | Membrane-bound lytic murein transglycosylase | 4.2.2.n1 | 1 |
| GH103 | Peptidoglycan | Membrane-bound lytic murein transglycosylase | 4.2.2.n1 | 1 |
| GH104 | Peptidoglycan | Lysozyme | 4.2.2.n2 | 1 |
| GH127 | Hemicellulose | Non-reducing end beta-L-arabinofuranosidase | 3.2.1.185 | 1 |
| GH153 | | Poly-beta-1,6-N-acetyl-D-glucosamine N-deacetylase | 3.5.1.- | 1 |

9790 was the only archaeal species identified that contained all four enzymes in new clinical pathway (NCP) I [38]. However, KEGG analysis confirmed that the pathways in *S.* sp. PAMC 28760 contained all four genes; in addition, we detected α-amylase. Thus, *S.* sp. PAMC 28760 contains the starch-degrading pathway, through which dextrin or maltose are obtained (Fig. 5). The roles of enzymes were confirmed using dbCAN2: we identified seven genes encoding enzymes involved in glycogen metabolism and starch degradation: GlgA (NZ_CP014768.1_3970, GT5), GlgB (NZ_CP014768.1_3973, CBM18 + GH13_9), GlgX (NZ_CP014768.1_3972, CBM48 + GH13_11), α-amylase (NZ_CP014768.1_4117, GH13_19), GlgP (NZ_CP014768.1_3969, GT35), and β-glucosidase (NZ_CP014768.1_2745, GH3). Most of the pathway genes identified involve enzymes in the GH13 family, which is known to be the most important group for starch degradation. The

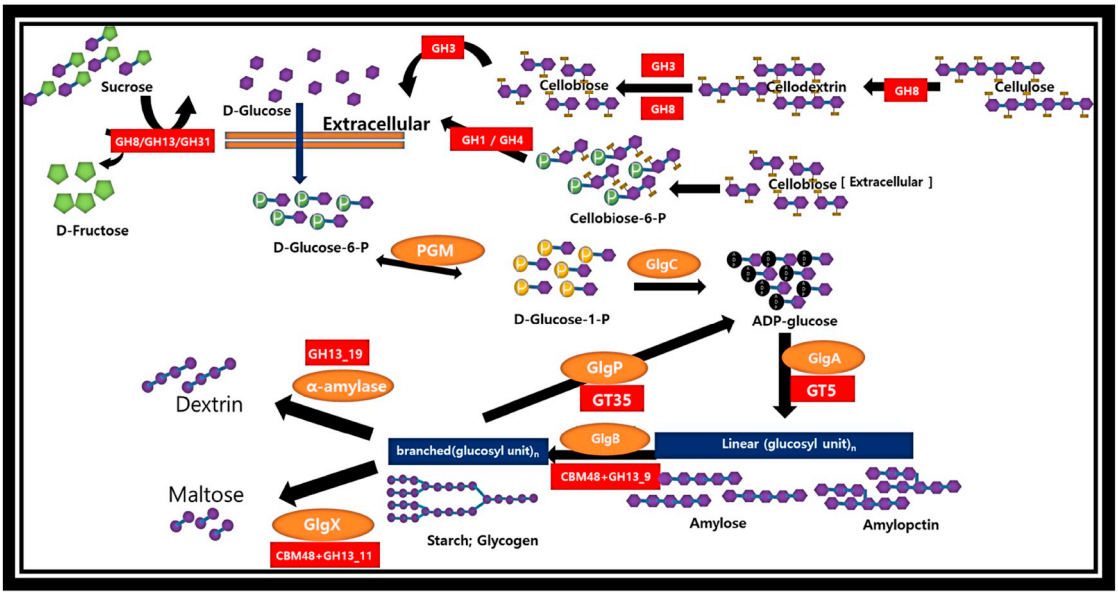


**Fig. 5. Predicted pathways for glycogen metabolism and starch degradation in *S.* sp. PAMC 28760.** Kyoto Encyclopedia of Genes and Genomes (KEGG)-predicted enzyme pathways (orange circle); dbCAN2-predicted CAZyme family (red square).

GH13 family comprises a major group of GH enzymes that act on substrates containing α-glucoside linkages. These enzymes have a variety of substrate preferences and products. For example, α-amylase prefers α-1,4-glucan polysaccharides, such as amylose and amylopectin, and may attack a supermolecular structure characterized by starch granules and glycogen particles. In addition, they show a slow polymerization transition into maltosaccharide [39,40]. Complete glycogen metabolic pathway genes are present only in species predominantly associated with mammalian hosts or natural environments [41]. As the PAMC 28760 strain was isolated from a lichen in Antarctica, symbiosis with the archaeal lichen likely plays a role in gene adaptation, by freeing up energy production and storage resources in the cold environment. *S.* sp. PAMC 28760 is considered a valuable resource that could be applied in a variety of industries because it has abundant GH13 amylolytic enzymes.

## 4. Conclusions

Communities of bacteria on the surface of the long-living lichen thallus surface remains almost constant with the season, despite exposure to significant cyclic factors in the habitat. It has already been reported that bacteria living on the surface of lichen thallus are well adapted to osmotic, oxidative stresses and other selective conditions barely survive these features are similar to the features of lichen in the habitat to withstand the environment. Resistance to biotic stress factors in symbiotic bacteria is interpreted as a pathogen defense mechanism, and without these mechanism bacteria are hard to survive. In fact, the metagenome analysis of lichen found the pathogens we knew (such as *E. coli*) [42]. We isolated several bacteria from lichen in the environment of lab and checked one of them as *Shigella* sp. Therefore we supposed that this process has allowed *Shigella* species to have a symbiotic process in lichen or become naturally separated because this strain are involved in the symbiotic system (that is pathogen defense). To summary up, we isolated *S.* sp. PAMC 28760 from Antarctic lichen; the complete genome sequence is 4.56 Mb in size with a GC content of 50.84%. Analysis of the complete genome suggested that it has cellulolytic and amylolytic characteristics, and diverse CAZyme genes. We confirmed that PAMC 28760 has 91 CAZyme genes: 3 in the AA family, 4 in the CBM family, 11 in the CE family, 42 in the GH family, and 31 in the GT family. In addition, comparative genome analyses of *Shigella* strains revealed that this species are a strong amylolytic bacterium, possibly because it is in the same family as *E. coli*. Genome sequence analysis provides valuable information regarding novel functional enzymes, which is useful for both biotechnological applications and fundamental research purposes. This study provides a foundation for understanding how this strain stores and produces energy in an extreme environment.

## Acknowledgment

This work was supported by the Korea Polar Research Institute (grant no. PE19210).

## References

[1] N. Aro, T. Pakula, M. Penttilä, Transcriptional regulation of plant cell wall degradation by filamentous fungi, FEMS Microbiol. Rev. 29 (2005) 719–739.
[2] W.R. de Souza, D.A. Chandel (Ed.), Microbial degradation of lignocellulosic Biomass, InTech, 2013, pp. 207–247.
[3] L. Avérous, P.J. Halley, Starch Polymers: from the Field to Industrial Products, Elsevier, Amsterdam, 2014, pp. 3–10.
[4] V. Valk, W. Eeuwema, F.D. Sarian, R.M. van der Kaaij, L. Dijkhuizen, Degradation of granular starch by the bacterium *Microbacterium aurum* strain B8. A involves a modular α-amylase enzyme system with FNIII and CBM25 domains, Appl. Environ. Microbiol. 81 (2015) 6610–6620.
[5] Y.J. Goh, T.R. Klaenhammer, Insights into glycogen metabolism in *Lactobacillus acidophilus*: impact on carbohydrate metabolism, stress tolerance and gut retention, Microb. Cell Factories 13 (2014) 94.
[6] S. Jobling, Improving starch for food and industrial applications, Curr. Opin. Plant Biol. 7 (2004) 210–218.
[7] F.W. Bai, W.A. Anderson, M. Moo-Young, Ethanol fermentation technologies from sugar and starch feedstocks, Biotechnol. Adv. 26 (2008) 89–105.
[8] K.L. Anderson, Degradation of Cellulose and Starch by Anaerobic Bacteria, Glycomicrobiology, Springer, Boston, 2002, pp. 359–386.
[9] R. Bargagli, Environmental contamination in Antarctic ecosystems, Sci. Total Environ. 400 (2008) 212–226.
[10] H. Takami, A. Inoue, F. Fjuk, K. Horikoshi, Microbial flora in the deepest sea mud of the Mariana Trench, FEMS Microbiol. Lett. 152 (1997) 279–285.
[11] G. Feller, Molecular adaptations to cold in psychrophilic enzymes, Cell. Mol. Life Sci. 60 (2003) 648–662.
[12] G. Feller, C. Gerday, Psychrophilic enzymes: hot topics in cold adaptation, Nat. Rev. Microbiol. 1 (2003) 200–208.
[13] G.T. Nhieu, P.J. Sansonetti, Mechanism of *Shigella* entry into epithelial cells, Curr. Opin. Microbiol. 2 (1999) 51–55.
[14] K.A. Lampel, S.B. Formal, A.T. Maurelli, A brief history of *Shigella*, EcoSal Plus 8 (2018), https://doi.org/10.1128/ecosalplus.ESP-0006-2017.
[15] J. Wei, M.B. Goldberg, V. Burland, M.M. Venkatesan, W. Deng, G. Fournier, et al., Complete genome sequence and comparative genomics of *Shigella flexneri* serotype 2a strain 2457T, Infect. Immun. 71 (2003) 2775–2786.
[16] H. Nie, F. Yang, X. Zhang, J. Yang, L. Chen, J. Wang, et al., Complete genome sequence of *Shigella flexneri* 5b and comparison with *Shigella flexneri* 2a, BMC Genomics 7 (2006) 173.
[17] C.S. Chin, D.H. Alexander, P. Marks, A.A. Klammer, J. Drake, C. Heiner, et al., Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data, Nat. Methods 10 (2013) 563–569.
[18] R.K. Aziz, D. Bartels, A.A. Best, M. DeJongh, T. Disz, R.A. Edwards, et al., The RAST Server: rapid annotations using subsystems technology, BMC Genomics 9 (2008) 75.
[19] J.R. Grant, A.S. Arantes, P. Stothard, Comparing thousands of circular genomes using the CGView Comparison Tool, BMC Genomics 13 (2012) 202.
[20] S.R. Eddy, Accelerated profile HMM searches, PLoS Comput. Biol. 7 (2011) e1002195.
[21] B. Buchfink, C. Xie, D.H. Huson, Fast and sensitive protein alignment using DIAMOND, Nat. Methods 12 (2015) 59–60.
[22] P.K. Busk, B. Pilgaard, M.J. Lezyk, A.S. Meyer, L. Lange, Homology to peptide pattern for annotation of carbohydrate-active enzymes and prediction of function, BMC Bioinf. 18 (2017) 214.
[23] R.C. Edgar, MUSCLE: a multiple sequence alignment method with reduced time and space complexity, BMC Bioinf. 5 (2004) 113.
[24] R.C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput, Nucleic Acids Res. 32 (2004) 1792–1797.
[25] S. Kumar, G. Stecher, M. Li, C. Knyaz, K. Tamura, X. MEGA, Molecular evolutionary genetics analysis across computing platforms, Mol. Biol. Evol. 35 (2018) 1547–1549.
[26] N. Saitou, M. Nei, The neighbor-joining method: a new method for reconstructing phylogenetic trees, Mol. Biol. Evol. 4 (1987) 406–425.
[27] J. Felsenstein, Confidence limits on phylogenies: an approach using the bootstrap, Evolution 39 (1985) 783–791.
[28] M. Richter, R. Rosselló-Móra, Shifting the genomic gold standard for the prokaryotic species definition, Proc. Natl. Acad. Sci. U.S.A. 106 (2009) 19126–19131.
[29] R. Maughan, Carbohydrate metabolism, Surgery (Oxford) 27 (2009) 6–10.
[30] D.L. Nelson, M.M. Cox, Lehninger Principles of Biochemistry, sixth ed., W. H. Freeman & Co., New York, 2013.
[31] L.M. Sanders, Carbohydrate: digestion, absorption and metabolism, Encycl. Food Health (2016) 643–650.
[32] H. Zhang, T. Yohe, L. Huang, S. Entwistle, P. Wu, Z. Yang, et al., dbCAN2: a meta server for automated carbohydrate-active enzyme annotation, Nucleic Acids Res. 46 (2018) W95–W101.
[33] A. McMeechan, M.A. Lovell, T.A. Cogan, K.L. Marston, T.J. Humphrey, P.A. Barrow, Glycogen production by different *Salmonella enterica* serotypes: contribution of functional *glgC* to virulence, intestinal colonization and environmental survival, Microbiology 151 (2005) 3969–3977.
[34] W.A. Wilson, P.J. Roach, M. Montero, E. Baroja-Fernández, F.J. Muñoz, G. Eydallin, et al., Regulation of glycogen metabolism in yeast and bacteria, FEMS Microbiol. Rev. 34 (2010) 952–985.
[35] J.T. Park, J.H. Shim, P.L. Tran, I.H. Hong, H.U. Yong, E.F. Oktavina, et al., Role of maltose enzymes in glycogen synthesis by *Escherichia coli*, J. Bacteriol. 193 (2011) 2517–2526.
[36] S. Ball, C. Colleoni, U. Cenci, J.N. Raj, C. Tirtiaux, The evolution of glycogen and starch metabolism in eukaryotes gives molecular clues to understand the establishment of plastid endosymbiosis, J. Exp. Bot. 62 (2011) 1775–1801.
[37] L. Wang, M.J. Wise, Glycogen with short average chain length enhances bacterial durability, Naturwissenschaften 98 (2011) 719–729.
[38] G. Chandra, K.F. Chater, S. Bornemann, Unexpected and widespread connections between bacterial glycogen and trehalose metabolism, Microbiology 157 (2011) 1565–1572.
[39] M.R. Stam, E.G. Danchin, C. Rancurel, P.M. Coutinho, B. Henrissat, Dividing the large glycoside hydrolase family 13 into subfamilies: towards improved functional annotations of alpha-amylase-related proteins, Protein Eng. Des. Sel. 19 (2006) 555–562.
[40] E. Hostinová, S. Janecek, J. Gasperík, Gene sequence, bioinformatics and enzymatic characterization of alpha-amylase from *Saccharomycopsis fibuligera* KZ, Protein J. 29 (2010) 355–364.
[41] T.A. Tompkins, G. Barreau, J.R. Broadbent, Complete genome sequence of *Lactobacillus helveticus* R0052, a commercial probiotic strain, J. Bacteriol. 194 (2012) 6349.
[42] M. Grube, T. Cemava, J. Soh, S. Fuchs, I. Aschenbrenner, C. Lassek, et al., Exploring functional contexts of symbiotic sustain within lichen-associated bacteria by comparative omics, ISME J. 9 (2015) 412–424.