frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 01 June 2023
DOI 10.3389/fmicb.2023.1112247

Check for updates

OPEN ACCESS

EDITED BY
Andreas Teske,
University of North Carolina at Chapel Hill, United States

REVIEWED BY
Jing Han,
Chinese Academy of Sciences (CAS), China
Shaoxing Chen,
Anhui Normal University, China

*CORRESPONDENCE
Alexander G. Elcheninov
✉ elcheninov.ag@gmail.com

†These authors have contributed equally to this work

RECEIVED 30 November 2022
ACCEPTED 15 May 2023
PUBLISHED 01 June 2023

CITATION
Elcheninov AG, Ugolkov YA, Elizarov IM, Klyukina AA, Kublanov IV and Sorokin DY (2023) Cellulose metabolism in halo(natrono)archaea: a comparative genomics study.
Front. Microbiol. 14:1112247.
doi: 10.3389/fmicb.2023.1112247

COPYRIGHT
© 2023 Elcheninov, Ugolkov, Elizarov, Klyukina, Kublanov and Sorokin. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Cellulose metabolism in halo(natrono)archaea: a comparative genomics study

Alexander G. Elcheninov[1]*, Yaroslav A. Ugolkov[1], Ivan M. Elizarov[1], Alexandra A. Klyukina[1], Ilya V. Kublanov[1†] and Dimitry Y. Sorokin[1,2†]

[1]Winogradsky Institute of Microbiology, Federal Research Centre of Biotechnology, Russian Academy of Sciences, Moscow, Russia, [2]Department of Biotechnology, Delft University of Technology, Delft, Netherlands

Extremely halophilic archaea are one of the principal microbial community components in hypersaline environments. The majority of cultivated haloarchaea are aerobic heterotrophs using peptides or simple sugars as carbon and energy sources. At the same time, a number of novel metabolic capacities of these extremophiles were discovered recently among which is a capability of growing on insoluble polysaccharides such as cellulose and chitin. Still, polysaccharidolytic strains are in minority among cultivated haloarchaea and their capacities of hydrolyzing recalcitrant polysaccharides are hardly investigated. This includes the mechanisms and enzymes involved in cellulose degradation, which are well studied for bacterial species, while almost unexplored in archaea and haloarchaea in particular. To fill this gap, a comparative genomic analysis of 155 cultivated representatives of halo(natrono)archaea, including seven cellulotrophic strains belonging to the genera *Natronobiforma, Natronolimnobius*, *Natrarchaeobius*, *Halosimplex*, *Halomicrobium* and *Halococcoides* was performed. The analysis revealed a number of cellulases, encoded in the genomes of cellulotrophic strains but also in several haloarchaea, for which the capacity to grow on cellulose was not shown. Surprisingly, the cellulases genes, especially of GH5, GH9 and GH12 families, were significantly overrepresented in the cellulotrophic haloarchaea genomes in comparison with other cellulotrophic archaea and even cellulotrophic bacteria. Besides cellulases, the genes for GH10 and GH51 families were also abundant in the genomes of cellulotrophic haloarchaea. These results allowed to propose the genomic patterns, determining the capability of haloarchaea to grow on cellulose. The patterns helped to predict cellulotrophic capacity for several halo(natrono)archaea, and for three of them it was experimentally confirmed. Further genomic search revealed that glucose and cellooligosaccharides import occurred by means of porters and ABC (ATP-binding cassette) transporters. Intracellular glucose oxidation occurred through glycolysis or the semi-phosphorylative Entner-Dudoroff pathway which occurrence was strain-specific. Comparative analysis of CAZymes toolbox and available cultivation-based information allowed proposing two possible strategies used by haloarchaea capable of growing on cellulose: so-called specialists are more effective in degradation of cellulose while generalists are more flexible in nutrient spectra. Besides CAZymes profiles the groups differed in genome sizes, as well as in variability of mechanisms of import and central metabolism of sugars.

KEYWORDS

haloarchaea, cellulotrophic, genomics, CAZymes, cellulose, polysaccharides degradation

Introduction

Extremely halophilic archaea, belonging to the class *Halobacteria* (*Euryarchaeota* phylum), are abundant in natural terrestrial and deep-see hypersaline lakes, men-made solar salterns, rock salt deposits and saline soils. The well-studied majority of cultivated haloarchaea are growing aerobically on rich media containing peptides or simple sugars. Recently, however, a number of novel metabolic capacities of haloarchaea were discovered, including capability to grow anaerobically by sulfur respiration (Sorokin et al., 2016, 2017, 2021) or to grow aerobically with insoluble polysaccharides as the sole substrate (Sorokin et al., 2015, 2018, 2019a,b). Still, the haloarchaea bearing novel metabolic features are in total minority among cultivated representatives of this class with their unique metabolic machinery practically unexplored on biochemical or genomic level. Another question is whether the numerous haloarchaea growing on simple substrates may have any of the mentioned above properties, more specifically would the saccharolytic haloarchaea be capable of growth on insoluble polysaccharides.

Polysaccharides are degraded under the action of different types of enzymes, belonging to so-called carbohydrate-active enzymes (CAZymes, Drula et al., 2022). CAZymes included glycosidases (GHs), polysaccharide lyases (PLs), carbohydrate esterases (CEs), glycosyl transferases (GTs, mostly involved in carbohydrate biosynthesis) as well as auxiliary proteins (AAs) and proteins with carbohydrate binding domains (CBMs). Hydrolysis of exogenous insoluble polysaccharides demand extracellular CAZymes (in total majority – GHs), responsible for initial degradation steps occurred outside the cell. Currently, four model of CAZymes export and the consequent mechanisms of polysaccharides degradation in bacteria are suggested (Gardner and Schreier, 2021): (a) CAZymes are exported *via* outer-membrane vesicles (Elhenawy et al., 2014), (b) CAZymes are exported from periplasm *via* type II secretion system (Gardner and Keating, 2010), (c) cellulosome – CAZymes and specific carbohydrate-binding proteins are attached to the scaffoldins anchored to the cytoplasmic membrane (Artzi et al., 2017), and (d) S-layer-bound CAZymes and tapirins (specific binding protein) are attached to S-layer glycoproteins and pili (Conway et al., 2016; Lee et al., 2019). In turn, the mechanisms and enzymes, involved in polysaccharides hydrolysis in archaea and specifically in halophilic archaea are almost unknown. In particular, this is true for one of the most abundant polysaccharide on Earth – cellulose. Cellulose is a recalcitrant structural homopolysaccharide consisted of beta-1,4-linked D-glucose residues. Despite no variation in primary structure, different forms of celluloses distinguished from each other by degree of crystallinity and ratio and layout of crystalline and amorphous domain – so-called allomorphs (Uusi-Tarkka et al., 2021) which defines the variability of the enzymes involved in cellulose hydrolysis.

Extracellular cellulose hydrolysis resulted in formation of cellooligosaccharides (maximal – C6, Zhong et al., 2020), cellobiose and glucose. The last is less preferable since it is more accessible for competitors and accounts for the preference of cellulotrophic microorganisms to carry out the final steps of cellooligosaccharides hydrolysis intracellularly. Although, a few studies on cellooligosaccharides and cellobiose import in hyperthermophilic archaea *Pyrococcus furiosus* (Koning et al., 2001) and *Sulfolobus solfataricus* (Elferink et al., 2001) revealed that high-affinity ATP-binding cassette (ABC) transporters to be involved in this process, but nothing is known about haloarchaea. The mechanisms of glucose import into the cells as well as proteins involved in this process also are very poorly studied in haloarchaea. It seems the ABC transporters play a key role (Williams et al., 2019) in glucose import in halophilic archaea. Glucose is a single final product of cellulose hydrolysis, and its oxidation in haloarchaea occurred by means of (a) semi-phosphorylative Entner-Doudoroff pathway (Johnsen et al., 2001; Pickl et al., 2014), (b) modified glycolysis involving ketohexokinase and 1-phosphofructokinase (Altekar and Rangaswamy, 1991; Pickl et al., 2012) or (c) canonical glycolysis with ADP-dependent phosphofructokinase.

Recent developments in sequence technologies as well as the overall interest to haloarchaea resulted in a high number of available haloarchaeal genome sequences, both cultivated and uncultured (metagenome assembled genomes - MAGs). This is a good background for a comprehensive comparative genomic study to reveal the mechanisms of cellulose utilization in haloarchaea. The aim of this work was to use comparative genomics for comprehensive annotation of haloarchaeal CAZymes, involved in cellulose hydrolysis to reveal their cellulolytic machinery in the strains capable of growing on cellulose and to be able predicting this possibility for strains, for which it was not verified experimentally. Finally, basing on the sets of GHs and other CAZymes and auxiliary proteins we attempted to predict haloarchaea's strategies of polysaccharides decomposing in hypersaline environments.

Materials and methods

Genome sequencing

Genomic DNA isolation, Illumina sequencing as well as genome assembly were performed as described earlier (Sorokin et al., 2018).

For strain AArcel5 additional sequencing using nanopore techonology (Oxford Nanopore Technology) was done. Genomic DNA of the strain was isolated using phenol-chloroform extraction (Gavrilov et al., 2016) and futher repurified using MagAttract HMW DNA Kit (Qiagen) according to manufacturer protocol. The DNA library was prepared with Rapid Barcoding Kit (SQK-RBK004, Oxford Nanopore Technologies). Sequencing was performed with FLO-MIN-106D flow cell (R9.4.1) and MinION device. Basecalling was performed using Guppy basecaller v.2.3.5 with flipflop model. In total of 94,910 reads were obtained by ONT sequencing (~152 Mbp). Assembly was performed as follows: Canu v.1.8 (Koren et al., 2017) was used to obtain *de novo* assembly using long ONT reads followed by Nanopolish v.0.11 (Loman et al., 2015) polishing with raw fast5 reads as well as several rounds of Pilon v.1.23 (Walker et al., 2014) polishing with Illumina reads.

Genome and phylogenetic analyses

High quality genomes of cultivated haloarchaea were downloaded from IMG/M system (Chen et al., 2019); genomes, which were *de novo* sequenced during this work, were also previously annotated using IMG/M system. To exclude almost identical genomes AAI matrix was constructed using aai_matrix.sh (Rodriguez-R and Konstantinidis, 2016). Completeness levels of assemblies with >95% AAI with each

other were estimated with CheckM v.1.1.5 (Parks et al., 2015): one assembly with better quality from each group will be selected for the further analysis.

For phylogenomic analysis based on the "ar122" set of conserved archaeal proteins, the sequences were identified and aligned in *in silico* proteomes of strains from AArcel and HArcel groups as well as described species within *Halobacteria* using the GTDB-tk v.1.2.0 with reference data v.89 (Chaumeil et al., 2019). The phylogenomic tree was constructed using RAxML v.8.2.12 (Stamatakis, 2014) with the PROTGAMMAILG model of amino acid substitution; local support values were 1,000 rapid bootstrap replications. Phylogenetic tree was visualized using iTOL v.6.5.2 (Letunic and Bork, 2019).

CAZymes genes were identified in the genomes using dbCAN v.2.0.11 (Zhang et al., 2018). Comparative analysis was performed with the complete set of revealed in each genome CAZymes as well as with families containing the enzymes with targeted (eg. endoglucanases) activities. Enzymes localization was predicted using SignalP v.6.0 (Teufel et al., 2022). Isoelectric points were estimated with IPC 2.0 (Kozlowski, 2021).

Putative carbohydrate-specific transporters as well as enzymes involved in central catabolism pathways were detected using blastp with characterized reference proteins, obtained from SwissProt (Boutet et al., 2016) and TCDB (Saier et al., 2014) databases, as queries and haloarchaeal genomes as subjects (e-value $<10^{-5}$). Positive hits were manually checked with blast against SwissProt database. For ABC transporters only the gene clusters encoding at least substrate-binding protein and permease subunits were taken into account (ATPase was not since it is relatively nonspecific component).

Clusters of Orthologous Groups (COGs) were identified with IMG Pipeline (Chen et al., 2019). NMDS ordination was performed with vegan package.[1]

Experimental support

Neutrophilic and alkaliphilic haloarchaea were cultivated on the medium prepared according to Sorokin et al. (2015). For the haloarchaea from salt lakes, a mineral base medium contained the following ($g\,l^{-1}$): 240 NaCl, 5 KCl, 0.25 NH_4Cl, 2.5 K_2HPO_4, pH 6.8. The medium was heat sterilized at 120°C for 30 min and after cooling supplemented with vitamin and trace metal mix (Pfennig and Lippert, 1966) (1 $mL\,l^{-1}$ each) and 2 mM $MgSO_4$. For alkaliphilic natronoarchaea from soda lakes, a sodium carbonate/bicarbonate buffered mineral base medium containing 4 M total Na^+ included ($g\,l^{-1}$): 190 Na_2CO_3, 30 $NaHCO_3$, 16 NaCl, 5 KCl and 1 K_2HPO_4 with a final pH 10 after heat sterilization was supplemented with the same additions as the neutral base medium, except that the amount of Mg was two times lower and that 4 mM NH_4Cl was added after sterilization. Finally, the ready to use alkaline base medium was mixed 1:3 with the neutral medium, resulting in the final pH of 9.6. Various forms of insoluble celluloses with different degrees of crystallinity were used as growth substrates at the final concentration of 1–2 $g\,l^{-1}$.

1 https://CRAN.R-project.org/package=vegan

Results and discussion

General genome properties and phylogenetic analysis

Genome properties of two cellulotrophic haloarchaeal groups, AArcel (alkaliphilic haloarchaea from soda lakes) and HArcel (neutrophilic haloarchaea from neutral salt lakes), were compared (Table 1). Two genome assemblies were obtained earlier [*Natronobiforma cellulositropha* AArcel2 (Sorokin et al., 2018) and *Halococcoides cellulosivorans* HArcel1 (Sorokin et al., 2019a)], one genome was resequenced and reassembled in the course of this work (*Natronobiforma cellulositropha* AArcel5^T, see *Material and Methods* section) and the others (*Natronolimnobius* sp. AArcel1, *Natrarchaeobius* sp. AArcel7, *Halosimplex* sp. HArcel2 and *Halomicrobium* sp. HArcel3) were sequenced *de novo* in the course of this work. The G + C content of all the genome assemblies laid within rather narrow boundaries: 58.85–68.31%. In turn the genome sizes greatly varied from 2.72 Mbp (HArcel1) to 5.12 Mbp (AArcel7) leading to fairly broad range of a number of a protein-coding genes: 2641–4,769.

Currently all haloarchaea are affiliated with the class *Halobacteria* within the *Euryarchaeaota* phylum. Phylogenomic analysis of cellulotrophic strains based on "ar122" set of conserved proteins showed that natronoarchaeal AArcel strains belong to the order *Natrialbales*, while neutrophilic HArcel strains – to the order *Halobacteriales* (Figure 1) thus confirming the results of 16S rRNA gene and RpoB protein sequence-based phylogenetic analyses (Sorokin et al., 2018, 2019a). The seven cellulotrophic haloarchaea were relatively equally distributed on the haloarchaeal tree indicating polyphyletic origin of cellulotrophy in this class. To estimate occurence of this capability among haloarchaea 155 genomes of cultivated representatives of *Halobacteria* class, including 7 cellulotrophic strains were analyzed in respect to the presence of cellulose-active CAZymes.

Cellulolytic capabilities of AArcel/HArcel strains compared with other haloarchaea

Currently there are 26 known CAZymes families (22 glycoside hydrolases and 4 polysaccharide monooxygenases) harboring the enzymes with confirmed cellulolytic activities *sensu lato* (including hydrolysis of cellooligosaccharides, e.g., beta-glucosidase or cellobiose phosphorylase, http://www.cazy.org; Supplementary Table S1): beta-glucosidase (GH1, GH2, GH3, GH30, GH39, GH116), endoglucanase (GH5, GH6, GH7, GH8, GH9, GH10, GH12, GH44, GH45, GH48, GH51, GH74, GH124, GH131, GH148), cellobiose/cellodextrin phosphorylase (GH94) and lytic cellulose monooxygenase (AA9, AA10, AA15, AA16). Among them, 13 families contain archaeal sequences and only 6 families contained biochemically characterized cellulases and related enzymes found in archaea (GH1, GH2, GH3, GH5, GH12 and GH116). Besides cellulases there is a number of auxiliary proteins responsible for binding and transportation of oligomers and glucose inside the cell.

To reveal the distribution of the cellulases *sensu lato* among the haloarchaeal genomes, the genes encoding selected GHs and AAs families members were searched in 155 high-quality genomes of representatives of *Halobacteria* class including 7 genomes of AArcel/

TABLE 1 General properties of genomes of haloarchaeal AArcel/HArcel strains.

Strain	IMG Genome ID	Genome size, bp	Gene count	Scaffold count	G+C, %	Reference
AArcel1	2,681,813,540	4,560,092	4,555	18	58.85	*de novo* sequencing
AArcel2	2,642,422,534	3,732,973	3,776	42	65.43	Sorokin et al., 2018
AArcel5	2,868,148,463	3,829,432	3,854	5	65.42	resequencing
AArcel7	2,808,606,451	5,121,137	4,820	22	62.82	*de novo* sequencing
HArcel1	2,681,813,541	2,723,120	2,757	1	65.74	Sorokin et al., 2019a
HArcel2	2,808,606,450	4,572,180	4,633	68	68.31	*de novo* sequencing
HArcel3	2,867,963,358	4,205,126	4,230	19	66.23	*de novo* sequencing

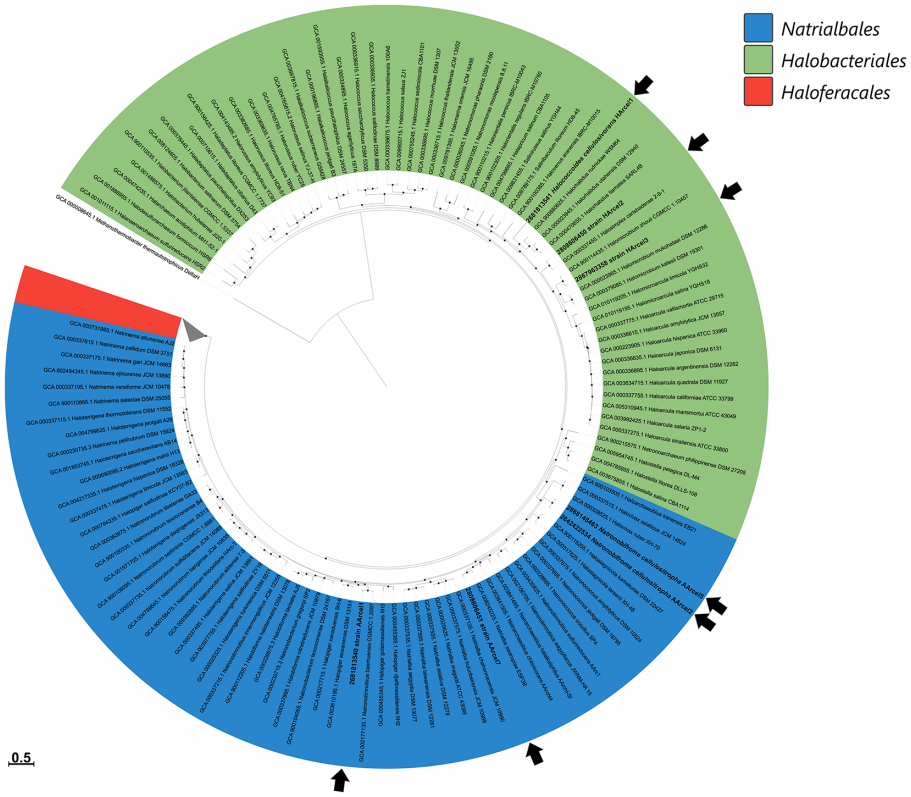

FIGURE 1
Maximum-likelihood phylogenetic tree of *Halobacteria* class based on 122 concatenated sequences of conservative archaeal proteins. Strains from AArcel/HArcel groups were marked by black arrows. The branch lengths correspond to the number of substitutions per site according to the corrections associated with the PROTGAMMAILG model in RAxML. The black circles at nodes indicate that the percentage of corresponding support values (1,000 rapid bootstrap replications) was above 50. *Methanothermobacter thermautotrophicus* DeltaH was used as an outgroup.

HArcel strains, playing a role of positive controls as they are known to be cellulose-utilizing organisms (Sorokin et al., 2015; Supplementary Figure S1). The search showed that 117 of 155 genomes possess at least one gene encoding protein from the abovementioned families (11 families were found).

The cellulases genes were unequally distributed within these 117 genomes and the AArcel/HArcel strains with the confirmed ability to utilize native insoluble forms of cellulose (Sorokin et al., 2015) were among the top in the number of such genes per genome. Among other haloarchaea for which this capacity is yet unknown there were examples with high number of cellulase genes per genome as well as genomes encoded single or few cellulases and the transition from the first to the latter variaty was seamless. With such distribution, it appeared impossible to distinguish genuine cellulotrophic representatives using this cellulases *sensu lato* dataset which is, most probably, related to the fact that besides cellulases these GH families contain enzymes which only indirectly involved in cellulose decomposition. In this regard, a set of query CAZymes was limited to CAZymes families containing endoglucanases – enzymes, playing crucial role in cellulose depolymerization (Mandeep et al., 2021) and which can be considered as signature enzymes for cellulotrophic organisms. The search with the endoglucanases set (Figure 2) resulted in selection of a much narrower group of haloarchaeal genomes with a high probability to be capable of degrading cellulose, not only its

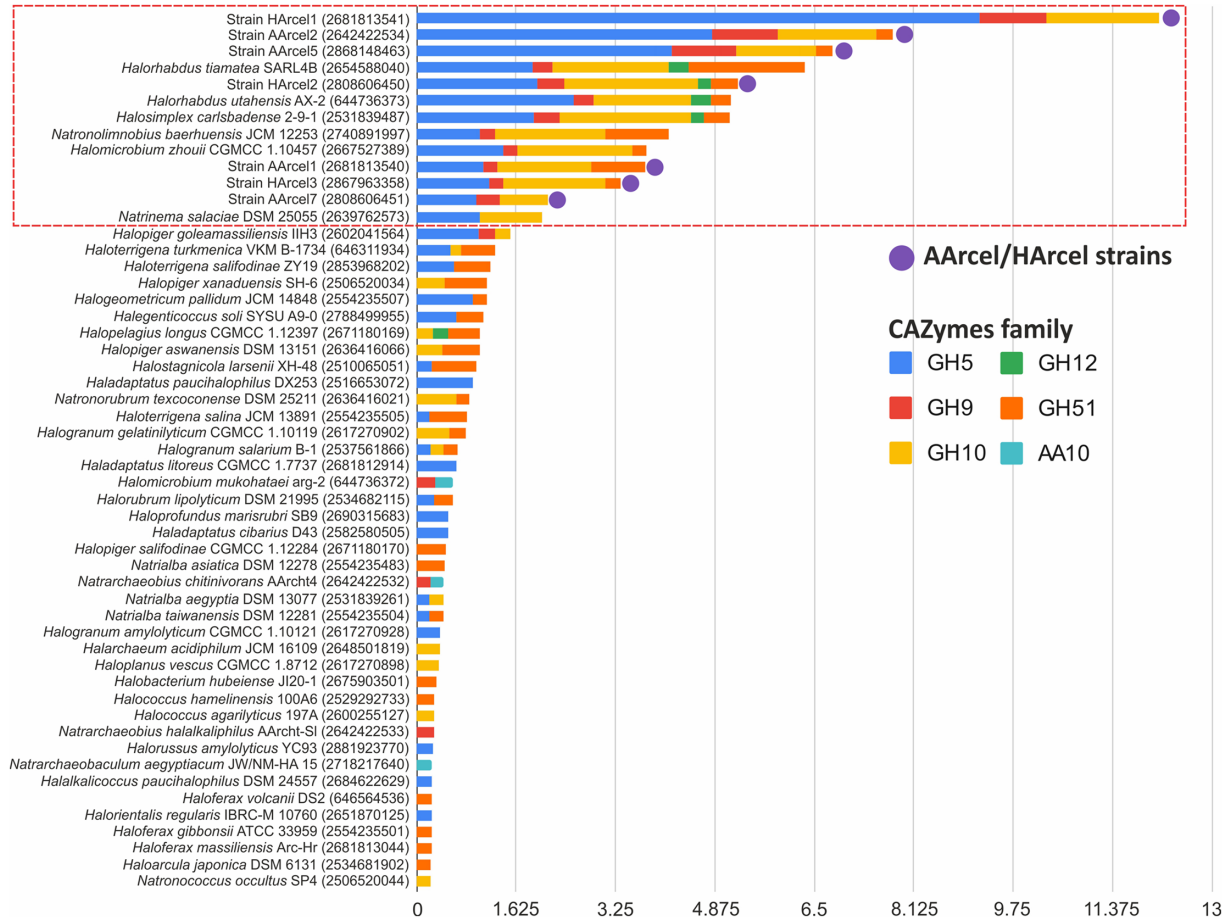

FIGURE 2
Relative abundance (gene number per1 Mbp) of putative endoglucanase genes found in 53 genomes of haloarchaea.

smaller and soluble derivatives as cellobiose, cellooligosacharides or heteropolysaccharides, containing beta-1,4-glucose linkages in their backbone or side chains. In total, 13 strains were found capable of degrading cellulose including all AArcel/HArcel strains, for which an ability to grow on native celluloses was experimentally approved (Sorokin et al., 2015). Besides AArcel and HArcel strains, the following haloarchaea were predicted to be cellulotrophic: *Halosimplex carlsbadense* 2–9-1, *Halorhabdus tiamatea* SARL4B, *Halorhabdus utahensis* AX-2, *Halomicrobium zhouii* CGMCC 1.10457, *Natronolimnobius baerhuensis* JCM 12253 and *Natrinema salaciae* DSM 25055. Three of them, *N. baerhuensis* JCM 12253, *H. carlsbadense* 2–9-1 (JCM 11222) and *H. zhouii* CGMCC 1.10457 (JCM 17095), were acquired from the Japan Collection of Microorganisms (JCM, https://jcm.brc.riken.jp/en/) and their ability to grow on amorphous cellulose was confirmed in our laboratory, while the other three still need to be tested.

While inspecting the reference endoglucanase sets of these cellulotrophic strains, including both AArcel/HArcel with the confirmed growth on cellulose and *de novo* predicted cellulotrophs it became apparent that the true cellulotrophic archaea must possess multiple and variable GH5 and GH10 families glycosidases, as well as at least several representatives from the GH9 family [excluding *Natrinema salaciae* DSM 25055 (2639762573)]. It should be noted that characterized proteins from GH5 and GH9 are mainly endoglucanases, while the majority of GH10 glycosidases are endoxylanases (despite several endoglucanases are also known (Xue et al., 2015; Zhao et al., 2019) being the reason to include this family into the "endoglucanases" set). It is possible that the latter are indeed cellulases in haloarchaea or involved in hemicelluloses decomposition, which might contribute to a better availability of cellulose for cellulases.

All putative endoglucanases encoded in the genomes of cellulotrophic haloarchaea were highly acidic having isoelectric point (pI) values from 3.86 to 4.56 (Figure 3; Supplementary Figure S2) which is linked with high salinity of their environments. Several alkaliphilic strains (AArcel1, AArcel2 and AArcel5) possessed slightly lower median pI values compared with neutrophilic haloarchaea, while AArcel7 and *Natronolimnobius baerhuensis* JCM 12253 had median pI values similar to neutrophiles indicating that environmental pH is not influencing the ratio of charged amino acids in these enzymes.

The genes encoding GH5 family glycosidases were the most numerous GH-encoding genes found in the genomes of 13 proven and predicted cellulotrophic haloarchaea. The number of GH5-encoding genes varied from 5 to 25 per genome. According to PFAM the length of a single GH5 catalytic domain is around 406 amino acids, while the lengths of the GH5-containing proteins in 13 cellulotrophic haloarchaea varied from 326 to 2,101 amino acids (Supplementary Figure S3). The data indicated that many of these proteins contained additional substrate-binding or other yet undetectable domains which can provide novel functionalities (Supplementary Table S2).

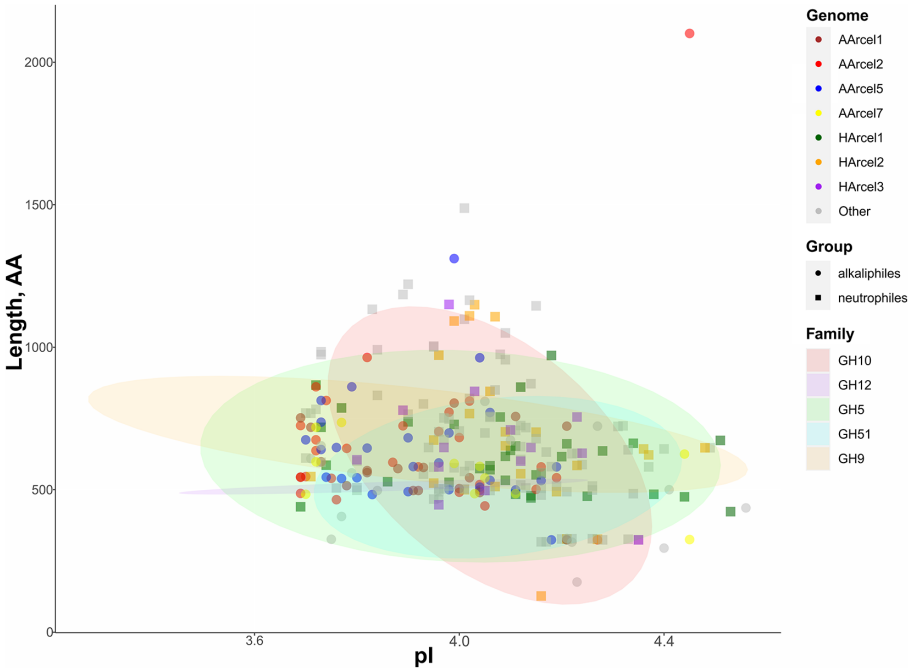

FIGURE 3
Characteristics of putative endoglucanases found in 13 genomes of cellulotrophic haloarchaea. Colors of the dots – genome assignment, shapes of the dots – assignment to neurtrophiles or alkaliphiles (based on literature data), ellipse color – enzyme family.

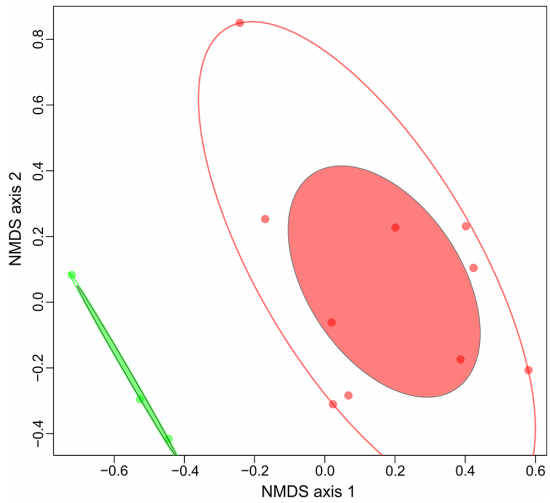

FIGURE 4
NMDS ordination plot (Bray, k=2, stress value=0.1263) of 13 haloarchaeal cellulolytic genomes based on CAZymes sets (with exception of GTs).

Ecological strategies of cellulose-utilizing haloarchaea

In our previous work on polysaccharidolytic haloarchaea (Sorokin et al., 2015) we proposed to divide all strains growing on cellulose into two groups: cellulotrophic and cellulolytic. The first are highly effective cellulose degraders, while the second are opportunists with broader substrate specificities, devouring many different oligo- and polysaccharides including the cellooligosaccharides released due to the action of the first group. In this respect, for 13 haloarchaea which either authoritatively or with high degree of probability being cellulotrophic an attempt has been made to reveal their lifestyle through the comparison of their CAZymes repertoire. Genome clustering of 13 genomes of cellulotrophic haloarchaea with NMDS ordinations was performed based on (i) a complete set of COGs found in the genomes and (ii) a set of CAZymes (excluding glycosyl transferases). Genome clusterization based on COGs gave no results since a relatively similar metabolism in terms of COGs functional categories was observed in all strains. Different results were obtained when CAZymes distribution among the genomes were used for clustering: two clearly separated groups comprised of (i) a compact cluster containing three strains (HArcel1, AArcel2 and AArcel5) and (ii) a larger and more diffused cluster comprising of other ten strains (Figure 4) were observed. We propose that the cellulose-utilizing microorganisms from the first group can be assigned to "specialists" while the second one contained "generalists." Remarkably, the genomes of cellulotrophic specialists are smaller than the generalists: 2.7–3.8 Mb and 4.2–5.1, respectively (Table 1) supporting our assumptions on their behavior.

Moreover, these two groups can be clearly distinguished not only by CAZymes repertoires and genome sizes but also by direct observation of ability to degrade cellulose. When growing on amorphous cellulose specialists form much larger hydrolysis zones in comparison with generalists (Figure 5).

Because of the action of numerous CAZymes cellulose is depolymerized to a single monomer, glucose. The question arose whether the central carbohydrate metabolism of cellulotrophic strains is uniform or the opposite is true. *In silico* reconstruction of the glucose/cellobiose/cellooligosaccharides import and glucose oxidation pathways in AArcel/HArcel strains with confirmed capability to grow on cellulose showed that glucose was transported into the cells by two

10.3389/fmicb.2023.1112247

different transport systems: (a) porters (superfamily 2.A according to TCDB) and ATP-binding cassette (ABC) transporters (superfamily 3.A.1 according to TCDB). Genes of phosphotransferase transport system (PTS) were absent in all genomes. In turn, cellooligosaccharides could be transported into the cells *via* ABC transporters as it was described for hyperthermophilic archaea (Koning et al., 2001). The number of genes encoding presumable carbohydrate transport systems components varied greatly between the genomes (Figure 6; Supplementary Table S3): HArcel1 possessed only 4 transporters (2 porters and 2 ABC transporters), while in the genome of AArcel7 35 transporters-encoding genes (9 porters and 26 ABC transporters) were found. A general observation is that the strains affiliated to specialists have less number of transporters than the generalists.

Genome analysis (Figure 7) revealed that glucose is metabolized *via* canonical-like glycolysis with ADP-phosphofructokinase (strains AArcel1 and AArcel7), haloarchaeal type of glycolysis (strains HArcel1 and HArcel3) or semi-phosphorylative Entner-Doudoroff pathway (all strains with exception of strain HArcel1).

Strain AArcel1 oxidizes glucose *via* glycolysis with ADP-phosphofructokinase as well as by complete semi-phosphorylative Entner-Doudoroff (KDPG) pathway. In the genomes of two closely related strains, AArcel5 and AArcel2, the genes encoding ADP-phosphofructokinase or 1-phosphofructokinase were absent indicating both glycolysis variants cannot be functional in this microorganism. Still, the genes of all KDPG pathway enzymes were found in the genomes of these haloarchaea. The glycolysis with ADP-phosphofructokinase as well as semi-phosphorylative KDPG pathway were predicted for AArcel7. Strain HArcel1 probably catabolized glucose only *via* glycolysis with phosphoglucomutase (performed the conversion of fructose-6-phosphate to fructose-1-phosphate, Lowry and Passonneau, 1969) and 1-phosphofructokinase and lacked KDPG pathway because KDPG aldolase gene was not found in the genome. Strain HArcel2 did not possess any variant of glycolysis due to the absence of the ADP-phosphofructokinase and 1-phosphofructokinase genes. Glucose was oxidized *via* KDPG-pathway in this microorganism. The genes encoding 1-phosphofructokinase and phosphoglucomutase were found in the genome of strain HArcel3 and thus it can utilize glucose *via* glycolysis like strain HArcel1. Complete semi-phosphorylative KDPG pathway was also predicted for this strain. Enzymes catalyzed common reactions for both the glycolysis and the KDPG pathway were present while glyceraldehyde-3-phosphate ferredoxin oxidoreductase (GAPOR), which often found in hyperthermophilic archaea, was absent in all studied strains. A gene of nonphosphorylating glyceraldehyde-3-phosphate dehydrogenase (GAPN) was found only in the AArcel7.

Summarizing the distribution of glucose oxidation pathways among the studied cellulotrophic haloarchaea it appears that specialists possessed only one glucose oxidation pathway, either

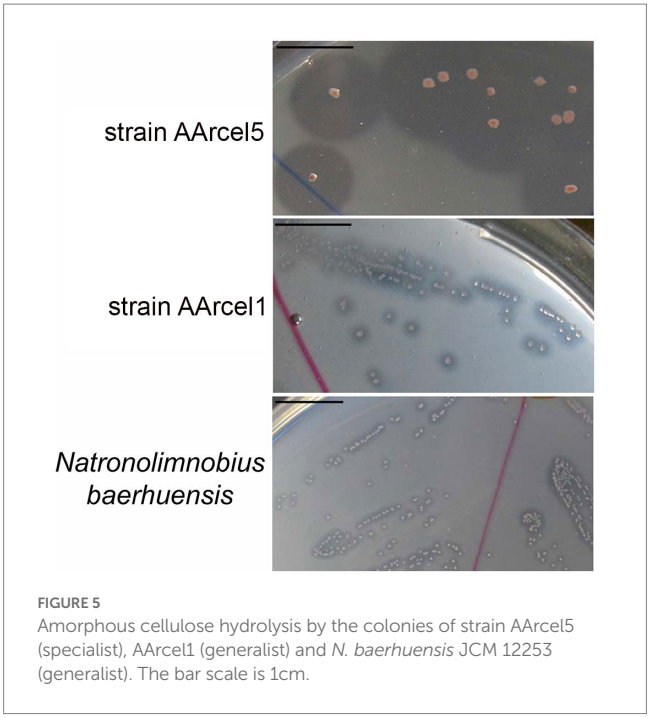

FIGURE 5
Amorphous cellulose hydrolysis by the colonies of strain AArcel5 (specialist), AArcel1 (generalist) and *N. baerhuensis* JCM 12253 (generalist). The bar scale is 1cm.

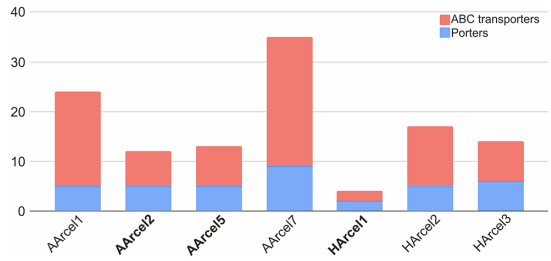

FIGURE 6
Number of putative transport system involved in carbohydrate transport AArcel/HArcel strains. Specialists are in bold.

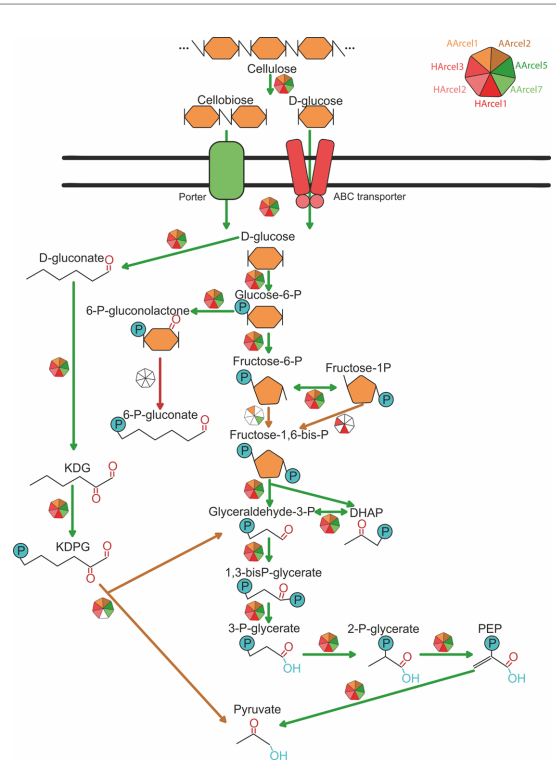

FIGURE 7
Glucose import and catabolism pathways found in AArcel/HArcel strains.

glycolysis (HArcel1) or KDPG (AArcel2 and AArcel5). Generalists, in turn, possess two pathways (AArcel1, AArcel7 and HArcel3) with the only exception – HArcel2, oxidizing glucose *via* KDPG pathway. This seems to be associated with narrower metabolism of specialists. These results are in accordance with other findings, distinguished these two groups: specialists characterized by a narrow specialization on cellulose degradation, smaller genomes, larger repertoire of genes encoding putative endoglucanases and lower number and variety in sugar transporters. Generalists include less specialized strains with a much broader substrate spectrum, larger genomes encoding lower number of cellulases but higher number and variability of sugar transporters.

Conclusion

The capacity of halophilic archaea to degrade various recalcitrant polysaccharides is of considerable interest for the understanding of their role in the mineralization of organic compounds in hypersaline environments and for search of extremely halo(alkali)stable extracellular CAZymes, attractive for the production of biofuel from lignocellulosic wastes since the pre-treatment step of this process is accomplished either with alkali or ionic liquids (Zavrel et al., 2009).

Large-scale analysis of all known CAZymes families containing cellulases encoded in the high-quality genomes of cultivated haloarchaea allowed to predict putative cellulotrophic strains. Since the dataset included the genomes of haloarchaea for which growth on and degradation of cellulose were experimentally confirmed and which therefore can be used as positive markers, these predictions allowed to propose a set of CAZymes-encoding genes indicative of the potential cellulotrophic lifestyle with a high degree of probability. Experimental validation of three out of seven cellulotrophic strains for which this property was not shown before confirmed their ability to grow on cellulose. The CAZymes patterns characteristic to cellulotrophic haloarchaea can serve as a tool for the comparative genomics-based identifying other haloarchaea carrying this trait.

Finally, genomic analysis followed by experimental verification of cellulase activity allowed dividing the cellulotrophic haloarchaea into two groups differed in strategies of cellulose utilization - specialists and generalists. The groups differed in efficiency of cellulose hydrolysis, CAZyme profiles, genome sizes, as well as in variability of mechanisms of import and central metabolism of sugars. Both groups are capable of growth on cellulose but specialists are more effective in cellulose degradation while generalists are more flexible to environmental changes, particularly to the changes in nutrient sources.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary material.

Author contributions

AE, YU, and IK analyzed the genomes and run phylogenetic analysis. IE and AK were responsible for DNA isolation and genome sequencing libraries preparation. DS performed microbiological experiments. IK supervised the study. AE, IK, and DS wrote the manuscript. All authors contributed to the article and approved the submitted version.

Funding

The work was supported by the Ministry of Science and Higher Education of the Russian Federation.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2023.1112247/full#supplementary-material

References

Altekar, W., and Rangaswamy, V. (1991). Ketohexokinase (ATP: D-fructose 1-phosphotransferase) initiates fructose breakdown via the modified EMP pathway in halophilic archaebacteria. *FEMS Microbiol. Lett.* 83, 241–246. doi: 10.1111/j.1574-6968.1991.tb04471.x

Artzi, L., Bayer, E. A., and Moraïs, S. (2017). Cellulosomes: bacterial nanomachines for dismantling plant polysaccharides. *Nat. Rev. Microbiol.* 15, 83–95. doi: 10.1038/nrmicro.2016.164

Boutet, E., Lieberherr, D., Tognolli, M., Schneider, M., Bansal, P., Bridge, A. J., et al. (2016). UniProtKB/Swiss-Prot, the manually annotated section of the UniProt KnowledgeBase: how to use the entry view. *Methods Mol. Biol.* 1374, 23–54. doi: 10.1007/978-1-4939-3167-5_2

Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P., and Parks, D. H. (2019). GTDB-Tk: a toolkit to classify genomes with the genome taxonomy database. *Bioinformatics* 36, 1925–1927. doi: 10.1093/bioinformatics/btz848

Chen, I. A., Chu, K., Palaniappan, K., Pillay, M., Ratner, A., Huang, J., et al. (2019). IMG/M v. 5.0: an integrated data management and comparative analysis system for microbial genomes and microbiomes. *Nucleic Acids Res.* 47, D666–D677. doi: 10.1093/nar/gky901

Conway, J. M., Pierce, W. S., Le, J. H., Harper, G. W., Wright, J. H., Tucker, A. L., et al. (2016). Multidomain, surface layer-associated glycoside hydrolases contribute to plant polysaccharide degradation by caldicellulosiruptor species. *J. Biol. Chem.* 291, 6732–6747. doi: 10.1074/jbc.M115.707810

Drula, E., Garron, M.-L., Dogan, S., Lombard, V., Henrissat, B., and Terrapon, N. (2022). The carbohydrate-active enzyme database: functions and literature. *Nucleic Acids Res.* 50, D571–D577. doi: 10.1093/nar/gkab1045

Elferink, M. G., Albers, S. V., Konings, W. N., and Driessen, A. J. (2001). Sugar transport in *Sulfolobus solfataricus* is mediated by two families of binding protein-dependent ABC transporters. *Mol. Microbiol.* 39, 1494–1503. doi: 10.1046/j.1365-2958.2001.02336.x

Elhenawy, W., Debelyy, M. O., and Feldman, M. F. (2014). Preferential packing of acidic glycosidases and proteases into Bacteroides outer membrane vesicles. *MBio* 5, e00909–e00914. doi: 10.1128/mBio.00909-14

Gardner, J. G., and Keating, D. H. (2010). Requirement of the type II secretion system for utilization of cellulosic substrates by *Cellvibrio japonicus*. *Appl. Environ. Microbiol.* 76, 5079–5087. doi: 10.1128/AEM.00454-10

Gardner, J. G., and Schreier, H. J. (2021). Unifying themes and distinct features of carbon and nitrogen assimilation by polysaccharide-degrading bacteria: a summary of four model systems. *Appl. Microbiol. Biotechnol.* 105, 8109–8127. doi: 10.1007/s00253-021-11614-2

Gavrilov, S. N., Stracke, C., Jensen, K., Menzel, P., Kallnik, V., Slesarev, A., et al. (2016). Isolation and characterization of the first xylanolytic hyperthermophilic euryarchaeon Thermococcus sp. strain 2319x1 and its unusual multidomain glycosidase. *Front. Microbiol.* 7:552. doi: 10.3389/fmicb.2016.00552

Johnsen, U., Selig, M., Xavier, K. B., Santos, H., and Schonheit, P. (2001). Different glycolytic pathways for glucose and fructose in the halophilic archaeon *Halococcus saccharolyticus*. *Arch. Microbiol.* 175, 52–61. doi: 10.1007/s002030000237

Koning, S. M., Elferink, M. G., Konings, W. N., and Driessen, A. J. (2001). Cellobiose uptake in the hyperthermophilic archaeon *Pyrococcus furiosus* is mediated by an inducible, high-affinity ABC transporter. *J. Bacteriol.* 183, 4979–4984. doi: 10.1128/JB.183.17.4979-4984.2001

Koren, S., Walenz, B. P., Berlin, K., Miller, J. P., Bergman, N. H., and Philippy, A. (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* 27, 722–736. doi: 10.1101/gr.215087.116

Kozlowski, L. P. (2021). IPC 2.0 : prediction of isoelectric point and pKa dissociation constants. *Nucleic Acids Res.* 49, W285–W292. doi: 10.1093/nar/gkab295

Lee, L. L., Hart, W. S., Lunin, V. V., Alahuhta, M., Bomble, Y. J., Himmel, M. E., et al. (2019). Comparative biochemical and structural analysis of novel cellulose binding proteins (tapirins) from xxtremely thermophilic *Caldicellulosiruptor* species. *Appl. Environ. Microbiol.* 85, e01983–e01918. doi: 10.1128/AEM.01983-18

Letunic, I., and Bork, P. (2019). Interactive tree of life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* 47, W256–W259. doi: 10.1093/nar/gkz239

Loman, N. J., Quick, J., and Simpson, J. T. (2015). A complete bacterial genome assembled de novo using only nanopore sequencing data. *Nat. Methods* 12, 733–735. doi: 10.1038/nmeth.3444

Lowry, O. H., and Passonneau, J. V. (1969). Phosphoglucomutase kinetics with the phosphates of fructose, glucose, mannose, ribose, and galactose. *J. Biol. Chem.* 244, 910–916. doi: 10.1016/s0021-9258(18)91872-7

Mandeep, , Liu, H., and Shukla, P. (2021). Synthetic biology and biocomputational approaches for improving microbial endoglucanases toward their innovative applications. *ACS Omega* 6, 6055–6063. doi: 10.1021/acsomega.0c05744

Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., and Tyson, G. W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 25, 1043–1055. doi: 10.1101/gr.186072.114

Pfennig, N., and Lippert, K. D. (1966). Über das vitamin B12-bedürfnis phototropher schwefelbakterien. *Arch. Microbiol.* 55, 245–256.

Pickl, A., Johnsen, U., Archer, R. M., and Schönheit, P. (2014). Identification and characterization of 2-keto-3-deoxygluconate kinase and 2-keto-3-deoxygalactonate kinase in the haloarchaeon *Haloferax volcanii*. *FEMS Microbiol. Lett.* 361, 76–83. doi: 10.1111/1574-6968.12617

Pickl, A., Johnsen, U., and Schönheit, P. (2012). Fructose degradation in the haloarchaeon *Haloferax volcanii* involves a bacterial type phosphoenolpyruvate-dependent phosphotransferase system, fructose-1-phosphate kinase, and class II fructose-1,6-bisphosphate aldolase. *J. Bacteriol.* 194, 3088–3097. doi: 10.1128/JB.00200-12

Rodriguez-R, L., and Konstantinidis, K. (2016). The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. *PeerJ Prepr.* 4:e1900v1. doi: 10.7287/peerj.preprints.1900

Saier, M. H., Reddy, V. S., Tamang, D. G., and Västermark, A. (2014). The transporter classification database. *Nucleic Acids Res.* 42, D251–D258. doi: 10.1093/nar/gkt1097

Sorokin, D. Y., Elcheninov, A. G., Toshchakov, S. V., Bale, N. J., Sinninghe Damsté, J. S., Khijniak, T. V., et al. (2019b). *Natrarchaeobius chitinivorans* gen. Nov., sp. nov., and *Natrarchaeobius halalkaliphilus* sp. nov., alkaliphilic, chitin-utilizing haloarchaea from hypersaline alkaline lakes. *Syst. Appl. Microbiol.* 42, 309–318. doi: 10.1016/j.syapm.2019.01.001

Sorokin, D. Y., Khijniak, T. V., Elcheninov, A. G., Toshchakov, S. V., Kostrikina, N. A., Bale, N. J., et al. (2019a). *Halococcoides cellulosivorans* gen. Nov., sp. nov., an extremely halophilic cellulose-utilizing haloarchaeon from hypersaline lakes. *Int. J. Syst. Evol. Microbiol.* 69, 1327–1335. doi: 10.1099/ijsem.0.003312

Sorokin, D. Y., Khijniak, T. V., Kostrikina, N. A., Elcheninov, A. G., Toshchakov, S. V., Bale, N. J., et al. (2018). *Natronobiforma cellulositropha* gen. Nov., sp. nov., a novel haloalkaliphilic member of the family *Natrialbaceae* (class *Halobacteria*) from hypersaline alkaline lakes. *Syst. Appl. Microbiol.* 41, 355–362. doi: 10.1016/j.syapm.2018.04.002

Sorokin, D. Y., Kublanov, I. V., Gavrilov, S. N., Rojo, D., Roman, P., Golyshin, P. N., et al. (2016). Elemental sulfur and acetate can support life of a novel strictly anaerobic haloarchaeon. *ISME J.* 10, 240–252. doi: 10.1038/ismej.2015.79

Sorokin, D. Y., Messina, E., Smedile, F., La Cono, V., Hallsworth, J. E., and Yakimov, M. M. (2021). Carbohydrate-dependent sulfur respiration in halo(alkali)philic archaea. *Environ. Microbiol.* 23, 3789–3808. doi: 10.1111/1462-2920.15421

Sorokin, D. Y., Messina, E., Smedile, F., Roman, P., Damsté, J. S. S., Ciordia, S., et al. (2017). Discovery of anaerobic lithoheterotrophic haloarchaea, ubiquitous in hypersaline habitats. *ISME J.* 11, 1245–1260. doi: 10.1038/ismej.2016.203

Sorokin, D. Y., Toshchakov, S. V., Kolganova, T. V., and Kublanov, I. V. (2015). Halo(natrono)archaea isolated from hypersaline lakes utilize cellulose and chitin as growth substrates. *Front. Microbiol.* 6:942. doi: 10.3389/fmicb.2015.00942

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033

Teufel, F., Almagro Armenteros, J. J., Johansen, A. R., Gíslason, M. H., Pihl, S. I., Tsirigos, K. D., et al. (2022). SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nat. Biotechnol.* 40, 1023–1025. doi: 10.1038/s41587-021-01156-3

Uusi-Tarkka, E.-K., Skrifvars, M., and Haapala, A. (2021). Fabricating sustainable all-cellulose composites. *Appl. Sci.* 11:10069. doi: 10.3390/app112110069

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., et al. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963. doi: 10.1371/journal.pone.0112963

Williams, T. J., Allen, M. A., Liao, Y., Raftery, M. J., and Cavicchioli, R. (2019). Sucrose metabolism in haloarchaea: reassessment using genomics, proteomics, and metagenomics. *Appl. Environ. Microbiol.* 85, e02935–e02918. doi: 10.1128/AEM.02935-18

Xue, X., Wang, R., Tu, T., Shi, P., Ma, R., Luo, H., et al. (2015). The N-terminal GH10 domain of a multimodular protein from *Caldicellulosiruptor bescii* is a versatile xylanase/β-glucanase that can degrade crystalline cellulose. *Appl. Environ. Microbiol.* 81, 3823–3833. doi: 10.1128/AEM.00432-15

Zavrel, M., Bross, D., Funke, M., Büchs, J., and Spiess, A. C. (2009). High-throughput screening for ionic liquids dissolving (ligno-)cellulose. *Bioresour. Technol.* 100, 2580–2587. doi: 10.1016/j.biortech.2008.11.052

Zhang, H., Yohe, T., Huang, L., Entwistle, S., Wu, P., Yang, Z., et al. (2018). dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* 46, W95–W101. doi: 10.1093/nar/gky418

Zhao, F., Cao, H.-Y., Zhao, L.-S., Zhang, Y., Li, C.-Y., Zhang, Y.-Z., et al. (2019). A novel subfamily of endo-β-1,4-glucanases in glycoside hydrolase family 10. *Appl. Environ. Microbiol.* 85, e01029–e01019. doi: 10.1128/AEM.01029-19

Zhong, C., Ukowitz, C., Domig, K. J., and Nidetzky, B. (2020). Short-chain cello-oligosaccharides: intensification and scale-up of their enzymatic production and selective growth promotion among probiotic bacteria. *J. Agric. Food Chem.* 68, 8557–8567. doi: 10.1021/acs.jafc.0c02660