Computational and Structural Biotechnology Journal 23 (2024) 406–416

Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

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

Discovery of a novel marine Bacteroidetes with a rich repertoire of carbohydrate-active enzymes

Beihan Chen [a,b], Guohua Liu [a], Quanrui Chen [a], Huanyu Wang [a], Le Liu [a], Kai Tang [a,*]

[a] State Key Laboratory of Marine Environmental Science, College of Ocean and Earth Science, Fujian Key Laboratory of Marine Carbon Sequestration, Xiamen University, Xiamen, China
[b] School of Oceanography, Shanghai Jiao Tong University, Shanghai, China

ARTICLE INFO

Keywords:
Bacteroidetes
Polysaccharide metabolism
Carbohydrate-active enzymes
Type IX secretion system
Comparative genomic analysis

ABSTRACT

Members of the phylum Bacteroidetes play a key role in the marine carbon cycle through their degradation of polysaccharides via carbohydrate-active enzymes (CAZymes) and polysaccharide utilization loci (PULs). The discovery of novel CAZymes and PULs is important for our understanding of the marine carbon cycle. In this study, we isolated and identified a potential new genus of the family Catalimonadaceae, in the phylum Bacteroidetes, from the southwest Indian Ocean. Strain TK19036, the type strain of the new genus, is predicted to encode CAZymes that are relatively abundant in marine Bacteroidetes genomes. *Tunicatimonas pelagia* NBRC 107804^T, *Porifericola rhodea* NBRC 107748^T and *Catalinimonas niigatensis* NBRC 109829^T, which exhibit 16 S rRNA similarities exceeding 90% with strain TK19036, and belong to the same family, were selected as reference strains. These organisms possess a highly diverse repertoire of CAZymes and PULs, which may enable them to degrade a wide range of polysaccharides, especially pectin and alginate. In addition, some secretory CAZymes in strain TK19036 and its relatives were predicted to be transported by type IX secretion system (T9SS). Further, to the best of our knowledge, we propose the first reported "hybrid" PUL targeting alginates in *T. pelagia* NBRC 107804^T. Our findings provide new insights into the polysaccharide degradation capacity of marine Bacteroidetes, and suggest that T9SS may play a more important role in this process than previously believed.

1. Introduction

The Bacteroidetes phylum is a diverse group of bacteria that are found in a wide range of environments, including marine systems [1]. Bacteroidetes are known for their ability to degrade complex polysaccharides, making them important players in the marine carbon cycle [2]. The ability of Bacteroidetes to degrade polysaccharides is due to their possession of carbohydrate-active enzymes (CAZymes) [2]. CAZymes consist of six major classes involved in carbohydrate degradation, modification and biosynthesis: Glycoside Hydrolase (GH), Glycosyltransferase (GT), Polysaccharide Lyase (PL) Carbohydrate Esterase (CE), Auxiliary Activities (AA) and Carbohydrate Binding Module (CBM) [3,4]. Bacteroidetes also possess polysaccharide utilization loci (PULs). PULs are genetic regions that encode the genes for CAZymes, as well as other proteins that are involved in the transport of carbohydrates, such as SusC- and SusD-like transporter proteins [5].

Recently, it has been shown that some CAZymes in Bacteroidetes are secreted by a protein transport system called the type IX secretion system (T9SS) [6]. T9SS is a secretion system that is exclusive to the Bacteroidetes [6–8]. It is involved in a variety of functions, including gliding motility, the degradation of biopolymers, and the secretion of CAZymes [9,10]. The secretion of CAZymes by T9SS has led to the discovery of two novel polysaccharide degradation mechanisms in Bacteroidetes [6]. The first mechanism is a non-PUL mechanism, which relies solely on T9SS to transport large multi-domain CAZymes. The second mechanism is a hybrid PUL system, which uses both T9SS and PULs to degrade complex carbohydrates [6]. PULs that harbor type A CTD-containing CAZymes constitute "hybrid" PULs [6]. Other type A CTD-containing CAZymes, that are not located at predicted PULs, may function as a "PUL-free" mechanism, that could enhance the polysaccharide degrading capabilities of marine Bacteroidetes [5,6]. However, the participation and contribution of T9SS in marine-derived polysaccharide degradation is still poorly understood.

In this study, we isolated and identified a novel strain of Bacteroidetes from the Southwest Indian Ocean. We then analyzed the composition of CAZymes, PULs, and T9SS components in this novel

* Corresponding author.
E-mail address: tangkai@xmu.edu.cn (K. Tang).

https://doi.org/10.1016/j.csbj.2023.12.025
Received 15 June 2023; Received in revised form 20 December 2023; Accepted 23 December 2023
Available online 26 December 2023
2001-0370/© 2023 The Authors. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

strain and three related Bacteroidetes. Our findings provide new insights into the mechanisms by which Bacteroidetes degrade complex polysaccharides, and has implications for our understanding of the role of Bacteroidetes in the marine environment.

2. Materials and methods

2.1. Isolation and identification

Seawater samples were collected from the Longqi hydrothermal zone of the Southwest Indian Ocean (49.64796°E, 37.78083°S), at a depth of about 500 m. Before laboratory use, samples were mixed with sterilized glycerol to a final concentration of 15% (v/v) and preserved at 4 °C. One hundred μL of the sample was then spread on marine agar 2216E, and cultivated at 28 °C for a week. A single pink colony was picked and purified by three rounds of streak plating. The purified isolate was assigned as TK19036 and studied further. Gram stain test was performed using a Gram stain kit (Hopebio, China; HB8278). Cell morphology was observed using transmission electron microscopy (TEM) (Tecnai G2 Spirit BioTwin) after 5 days of growth. Gliding motility was tested using 1% (w/v) agar plates.

A full-length 16S rRNA gene sequence was extracted using ContEst16S (https://www.ezbiocloud.net/tools/contest16s) [11] from the whole genome shotgun sequence, and aligned using NCBI online server Blast (in 16S/ITS database). Digital DNA–DNA hybridization (dDDH), average nucleotide identity (ANI) and average amino acid identity (AAI) values were calculated using the online toolkits GGDC 3.0 (https://ggdc.dsmz.de/ggdc.php) [12], ANI Calculator (https://www.ezbiocloud.net/tools/ani) [13] and AAI Calculator (http://enve-omics.ce.gatech.edu/aai/) [14], respectively. The latest Genome Taxonomy Database associated taxonomic classification toolkit (GTDB-Tk v2.2.6) [15] was used to construct a multigene-based phylogenomic tree on a Linux system, with all parameters set to default values.

2.2. Genome sequencing and assembly

Reference strains, *Tunicatimonas pelagia* NBRC 107804^T [16], *Porifericola rhodea* NBRC 107748^T [17] and *Catalinimonas niigatensis* NBRC 109829^T [18], were purchased from the Biological Resource Centner, NITE, Japan. For genome sequencing and assembly of *Tunicatimonas* sp. TK19036, *P. rhodea* NBRC 107748^T and *C. niigatensis* NBRC 109829^T, the library construction was carried out in accordance with the manufacturers' protocols for the Magen Hipure Soil DNA Kit and the SMRTbell® Express Template Preparation Kit 2.0. Sequencing was performed using the Illumina Hiseq X Ten (GENEWIZ, China) and Oxford Pacbio Sequel II platforms (GENEWIZ, China), following the manufacturer's protocols for the NovaSeq 6000 S4 Reagent Kit and NovaSeq Xp 4-Lane Kit. Pacbio reads were then de novo assembled using HGAP4 [19], and the genomes were recorrected by previous Illumina data using Pilon [20]. Genome sequencing and assembly of *T. pelagia* NBRC 107804^T were similar with those described above, except that the third generation sequencing technology of Nanopore (Personalbio, China) and the de novo assembly software of Unicycler [21] were utilized. The whole genome shotgun data for all isolates mentioned above were deposited in the NCBI genome database.

2.3. Genome annotation

Genomic features of strain TK19036, *T. pelagia* NBRC 107804^T, *P. rhodea* NBRC 107748^T and chromosome of *C. niigatensis* NBRC 109829^T were annotated by Prokka v1.14.6, a rapid prokaryotic genome annotation software [22]. Prokka invokes Prodigal [23], RNAmmer [24], Aragorn [25], SignalP [26] and Infernal [27] to predict coding sequences (CDS), ribosomal RNA genes (rRNA), transfer RNA genes, signal leader peptides and non-coding RNA, respectively. Then Prokka produces a set of standard-compliant output files. The output faa, gff and gbk files were used as input files for the following analyses, unless otherwise specified.

2.4. Predictions of CAZymes and CGCs

Carbohydrate-active enzymes and CAZyme-containing gene clusters were automatically annotated by an online server, dbCAN3 (https://bcb.unl.edu/dbCAN2/blast.php) [28,29]. To maintain data consistency, the protein sequences predicted by Prokka were also submitted to the dbCAN3 server. All submissions used HMMER (E-Value < 1e-15, coverage > 0.35) to search for CAZyme family annotation under the dbCAN CAZyme domain HMM database and CAZyme subfamily annotation under the dbCAN-sub HMM database, and DIAMOND (E-Value < 1e-102) to search for BLAST hits in the CAZy database [4]. As dbCAN HMM only detects homology in the N-terminal part ($NADH^+$ domain) without homology to the catalytic GH109 domain, to eliminate false positives, we conducted a search for the catalytic GH109 domain using the InterPro web service (https://www.ebi.ac.uk/interpro/) [30]. Subsequent analyses exclusively utilized the filtered GH109 results. The CGC-Finder server (Distance <= 2, signature genes = CAZyme + TC), within dbCAN3, was used to identify potential CAZyme-containing gene clusters (CGCs) [31], with transporters (TC) searched in the Transporter Classification Database (TCDB) [32]. In addition, we utilized the dbCAN-seq (https://bcb.unl.edu/dbCAN_seq/) [31,33] and dbCAN-PUL (https://bcb.unl.edu/dbCAN_PUL/) [34] databases to conduct a comprehensive investigation of carbohydrate substrates, bacterial taxonomy and CGCs. The dbCAN-seq database was updated in 2022, during which 9421 MAGs from four ecological environments (human gut, human oral, cow rumen and marine) were added. Considering the strong dependence on sulfatases in the degradation of polysaccharides derived from marine phytoplankton, we also annotated genes encoding sulfatase families and subfamilies in the genomes using the online server Sulfatlas HMM (https://sulfatlas.sb-roscoff.fr/sulfatlashmm/) [35,36].

2.5. Identification of T9SS

The components of T9SS are listed in Table S1, and the corresponding protein sequences were downloaded. The genome of *Cytophaga hutchinsonii* ATCC 33406^T was used as reference [37]. Proteins predicted in the genomes of strain TK19036 and related strains were set as the local database by Blast Zone function in TBtools [38], then searched for T9SS components using the blastp method, with an E-value of 1e-10. Amino-terminal secretory signal peptides were predicted by the updated SignalP 6.0 (https://services.healthtech.dtu.dk/services/SignalP-6.0/) [39]. To identify proteins potentially secreted through T9SS, we first downloaded the hmm files of the domain families TIGR04183 and TIGR04131, and then searched for the two types of carboxy-terminal domains (CTDs) in the proteins predicted in the genomes of strain TK19036 and related strains using HMMER 3.3.2 (http://hmmer.org/), with the default E-value of 1e-2.

2.6. Sole-carbon-source cultivation experiment

Pectin (Aladdin, 9000–69-5) and alginate (Aladdin, 9005–38-3) were chosen as test polysaccharide substrates. Cells were harvested by centrifugation (4000 rpm, 10 min) at the exponential stage, and washed three times with sterile artificial seawater (KCl 0.3 g, $MgSO_4{\cdot}H_2O$ 0.5 g, $CaCl_2{\cdot}H_4O_2$ 0.038 g, NH_4Cl 0.3 g, K_2HPO_4 0.3 g, NaCl 35 g, Milli-Q water 1 L, pH 7.8–8.0) before inoculation. Cultivation experiments in a sole-carbon-source medium (artificial seawater 1 L, carbon source 30 mM, vitamin 1 mL, trace metal 1 mL) were conducted in triplicate at 26 °C for 12 days, with D-glucose as the sole carbon source in the control group. Daily samples (200 μL) were measured for optical density at 600 nm (OD600) to monitor cell growth.

Table 1
Genome and identification information of the four species.

	TK19036	[a]*T. pelagia* NBRC 107804[T]	*C. niigatensis* NBRC 109829[T]	*P. rhodea* NBRC 107748[T]
Isolate source	seawater	sea anemone	blackish lake sediment	marine sponge
Genome size (bp)	8204,189	7171,967	7134,482	5487,627
G + C content (%)	45.20	44.92	42.04	41.20
Sequencing depth(x)	197	134	304	298
N50 (bp)	7777	12204	6901	7865
Total genes	6683	6145	5969	4547
Protein coding genes	6631	6097	5923	4501
tRNA genes	45	41	39	39
rRNA genes	6	6	6	6
16S rRNA similarities (%)	(100.00)	94.32	90.74	90.56
[b]ANI values (%)	(100.00)	71.82	69.44	68.51
[b]AAI values (%)	(100.00)	69.78	62.23	61.09
[b]dDDH values (%)	(100.00)	18.00	18.00	17.80

[a] Only chromosome sequence compared here.
[b] Genome of strain TK19036 used as query sequence.

2.7. Protein structure prediction and search

Protein structures were predicted using AlphaFold2 [40,41]. Then, the targeted protein structures were searched through Foldseek (https://search.foldseek.com/) [42]. Furthermore, the structures of the query and the targeted proteins were aligned using US-align (https://zhanggroup.org/US-align/) [43]. Protein structures were visualized by Pymol v2.4 [44].

2.8. Heterologous expression and enzymatic assays

The pectate lyase gene (FBLHDAGO_00238) and alginate lyase gene (FBLHDAGO_04016) (Table S2) were synthesized by Sangon Biotech (Sangon, China). The synthesized genes were inserted into pET28a expression vectors followed by transformation into *Escherichia coli* BL21 (DE3). The protein expression and purification followed previously reported protocols [45]. To determine the enzymatic activity of pectate lyase, 25 μL of purified enzyme was added to 75 μL of 0.2% pectin solution (0.2 M Glycine-NaOH buffer, 0.06 M CaCl2, pH 9.4), and the volume of the reaction was adjusted to 200 μL by adding the same buffer. Subsequently, the change in absorbance at 235 nm was measured over 15 min at 45 °C [46,47]. The enzymatic activity unit (U/mL) was defined as the amount of enzyme solution required to cleave pectin to produce 1 μmoL of unsaturated polygalacturonic acid in 1 min. The molar extinction coefficient for calculation was 4600 mL/(mmol cm) [47]. Alginate lyase cleaves alginate to produce reducing substances that discolor ferricyanide solution. Its activity was determined by incubating 1.2 mL of 0.2% sodium alginate (Tris-HCl 50 mM, $CaCl_2$ 1 mM, pH 8.0) with 120 μL of the purified enzyme at 25 °C for 10 min, and then transferring 40 μL to 200 μL of ferricyanide solution ($K_3[Fe(CN)_6]$ 1.5 g/L, Na_2CO_3 24 g/L, NaOH 5 mM) which stopped the enzymatic reaction. The solution was boiled for 10 min and cooled to room temperature. The absorbance of 200 μL reaction mixture was measured at 415 nm [48]. Definition of enzyme activity unit (U/mL): the amount of enzyme solution required to cleave sodium alginate to form 1 μmoL of reduced carbon-carbon double bonds per minute.

3. Results and discussion

3.1. Description of Catalimonadaceae gen. nov., sp. nov. TK19036

Colonies of strain TK19036 were dark pink after 4–5 days of growth on marine agar 2216E plates at 28 °C. Cells were long rods, approximately 0.7–0.8 μm wide and 3.3–6.0 μm long (Fig. S1). Gram staining was negative. The top three 16 S rRNA gene similarities with strain

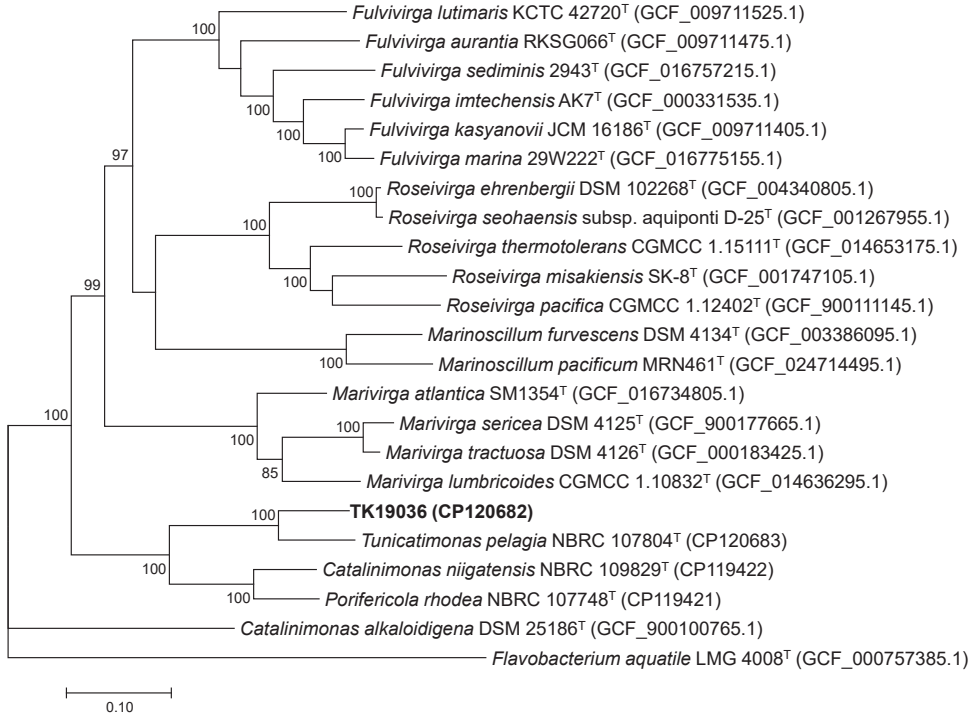


Fig. 1. GTDB-Tk phylogenetic tree of strain TK19036 and related species based on 120 bacterial marker genes. Values under 70% are hidden. Bar, 0.10 substitutions per nucleotide base. NBCI accession numbers of sequences used in tree-construction are labeled in parentheses.

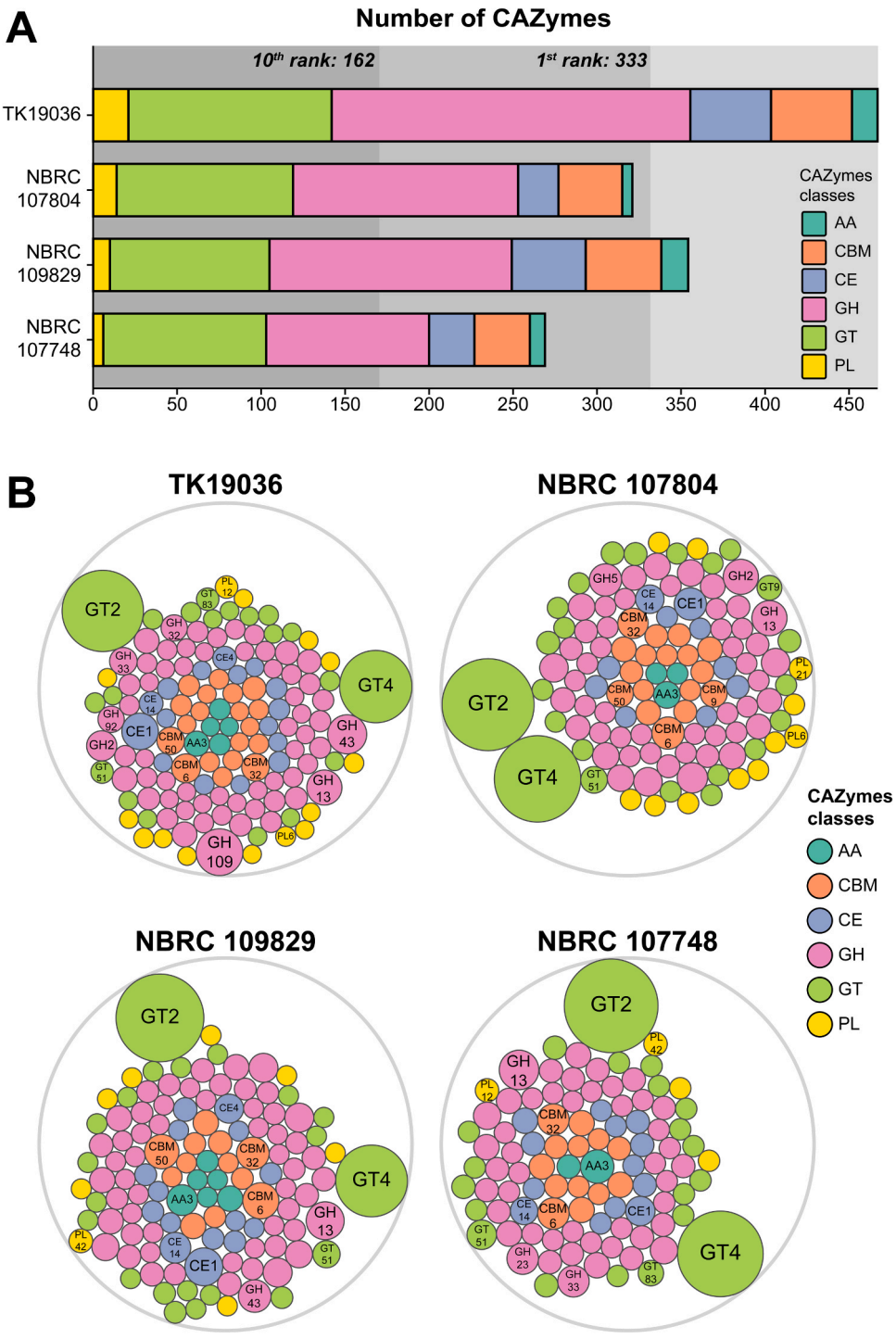


Fig. 2. Annotated CAZymes. A. Number of CAZymes annotated in the four marine Bacteroidetes strains, compared with records in the dbCAN-seq database. B. CAZymes family distribution of the four marine Bacteroidetes strains. Detailed information can be found in Table S5. AA, Auxiliary Activities; CBM, Carbohydrate Binding Module; CE, Carbohydrate Esterase; GH, Glycoside Hydrolase; GT, Glycosyltransferase; PL, Polysaccharide Lyase.

TK19036 were *T. pelagia* NBRC 107804^T, *P. rhodea* NBRC 107748^T and *C. niigatensis* NBRC 109829^T of 94.32%, 90.74% and 90.56%, respectively. Other hits had similarities below 90.00%. Base on the 16 S rRNA gene similarities, *T. pelagia* NBRC 107804^T, *P. rhodea* NBRC 107748^T and *C. niigatensis* NBRC 109829^T were chosen as references, to be compared with the strain TK19036 as follows. ANI, AAI and dDDH values are listed in Table 1. 16 S rRNA gene similarities, ANI and AAI values between strain TK19036 and the closest relative, *T. pelagia* NBRC 107804^T, were significantly lower than the classical thresholds for genus delineation at 94.5% [49], 73.98% (95% CI, 73.34–74.62%) [50] and 65–72% [51], respectively. Digital DDH values were all $\leq$ 70%, supporting a clear species boundary [52]. As shown in Fig. 1, strain TK19036 forms a deep branch within the family Catalimonadaceae and is closely related to *T. pelagia* NBRC 107804^T, surrounded by the other genera. Based on the available information, strain TK19036 is proposed to represent a novel genus within the family Catalimonadaceae.

3.2. Genome features of isolates from the family Catalimonadaceae

The four marine Bacteroidetes strains used in this study had genome sizes ranging from 5.49 to 8.2 Mb, with GC contents ranging from 41.20 to 45.20% (Table 1). Strain TK19036 had the largest genome size, the highest GC content, and the largest number of predicted protein-coding genes. Only *T. pelagia* NBRC 107804^T was found to have three plasmids. However, none of the genes on these plasmids were annotated as CAZymes. Therefore, they were not included in subsequent analyses.

3.3. Large CAZymes reservoirs in the four marine Bacteroidetes strains

There were 465, 319, 353 and 267 CAZymes in total predicted by dbCAN3 (Fig. 2A) in the genomes of strain TK19036, *T. pelagia* NBRC 107804^T, *P. rhodea* NBRC 107748^T and *C. niigatensis* NBRC 109829^T, respectively. Strain TK19036 harbored relatively higher number of CAZymes compared with other marine Bacteroidetes genomes. The most abundant CAZymes in TK19036 were GHs (215), followed by GTs (121), CBMs [49], CEs [47], PLs [21] and AAs [14], respectively.

GH enzymes catalyze hydrolysis of glycosidic bonds to generate hemiacetals, and are classified by catalytic mechanism. GH13 enzymes accounted for the high proportion of GH families, at between 6.6–12.6%, in the genomes of strain TK19036 and the reference strains (Fig. 2B). GH13 has been divided into 47 subfamilies. GH subfamilies GH13_3, GH13_9, GH13_10, GH13_11, GH13_16, GH13_26, GH13_30, CBM48 + GH13_31 and CBM48 + GH13_38 were shared among the four marine Bacteroidetes strains (Table S3). Some GH13 subfamilies contained CBM48 as starch-binding domains [53]. Only CBM48 was annotated in the GH13 subfamilies of the four marine Bacteroidetes strains (Table S3). CBM48 can bind to various linear or cyclic α-glucans derived from starch and glycogen containing both α-1,4- and α-1,6-linkages [54]. CBM48s from the pullulanase subfamily (GH13_8–14) are widespread in prokaryotes [55]. GH13_9, GH13_10 and GH13_11 were annotated beside CBM48, possibly because of the requirement for the hydrolysis of amylopectin, which is common in marine environments [56].

GT enzymes catalyze glycosidic bond formation during glycoside synthesis. GT4 and GT2 accounted for the highest proportion in the GT families of strain T19036 and the reference strains, followed by GT51 (Fig. 2B). GT4 and GT2 are thought to be the ancestral origins of several other GH families [57]. They constitute a basic, but significant, portion of marine clade genomes, accounting for over 50% of GTs (GT2/GTs, 30%; GT4/GTs, 28%) in the marine environment in the dbCAN-seq database. GT51 enzymes, also known as murein polymerase (EC 2.4.1.129), are involved in the synthesis of murein in both Gram-positive and Gram-negative bacteria [58]. Therefore, GT51 plays an important role in the GT families of Gram-negative bacteria, such as strain TK19036 and the reference strains.

The ratio of GH + GT to total protein-coding genes in strain *T. pelagia* NBRC 107804^T, *C. niigatensis* NBRC 109829^T and *P. rhodea* NBRC 107748^T were 5.0%, 3.9%, 4.0% and 4.3%, respectively. Strain TK19036 had the highest ratio of GH + GT to total protein-coding genes. For other bacterial taxa, the GH+GT to total protein coding genes ratio is often between 1% to 3% [59]. The higher GH + GT to total protein-coding genes ratio in strain TK19036 suggests that the family Catalimonadaceae of marine Bacteroidetes has a greater potential for polysaccharide metabolism.

CE enzymes release acyl or alkyl groups attached by ester linkage to carbohydrates and facilitate the degradation of complex polysaccharides [60], such as pectin and alginate [61]. CE1 and CE14 were the most abundant CE families in the genomes of the four marine Bacteroidetes strains (Fig. 2B). CE1 and CE14 are major families in CE, and are common among bacteria [62]. They are recognized for facilitating the degradation of xylan, a component of plant cell walls, and chitin, a constituent of crustacean shells, respectively [62,63]. It is presumed that strain TK19036 and the reference strains could utilize CE1 enzymes to facilitate the hydrolysis of recalcitrant polysaccharides.

CBMs are typically found in association with other catalytic

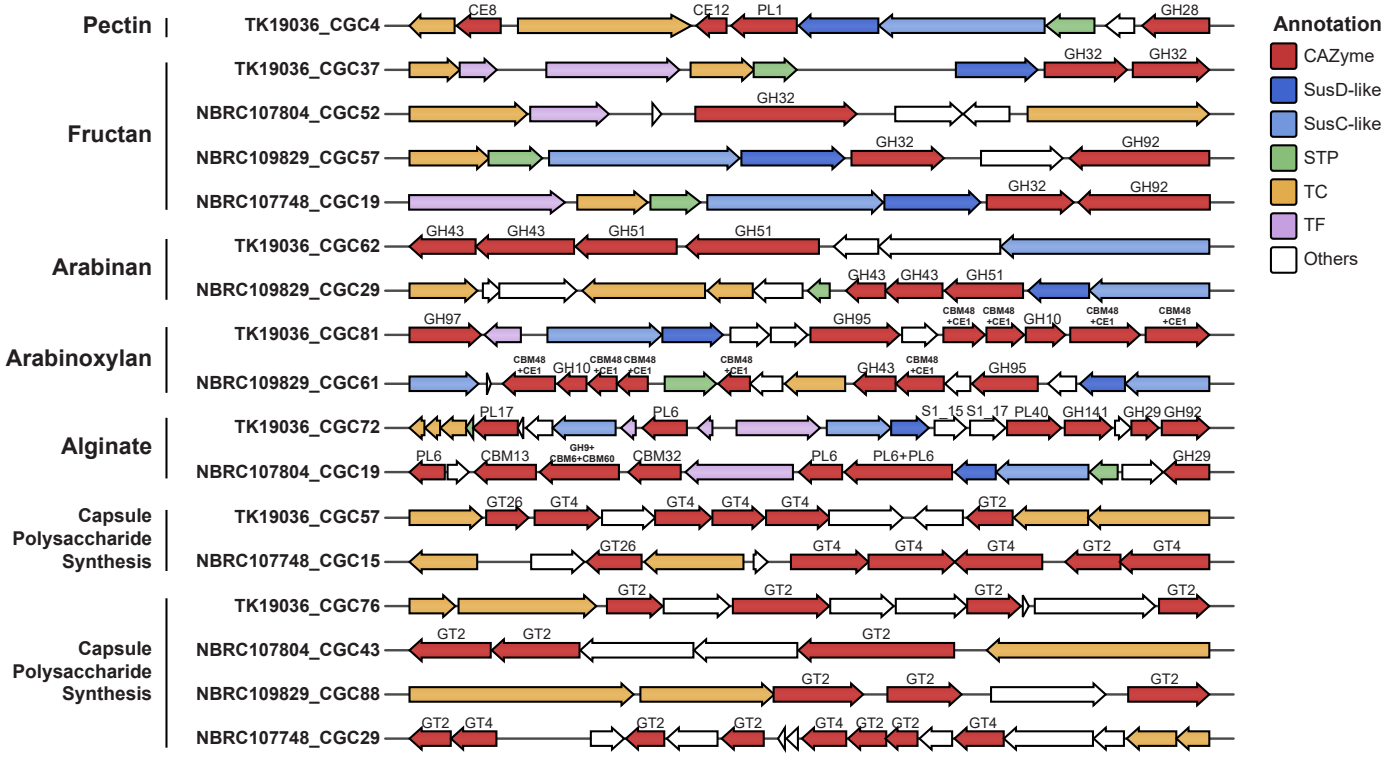


Fig. 3. Part of the predicted CGCs and PULs in the genomes of the strain TK19036 and reference strains. Annotations indicate predicted coding DNA sequence (CDS): CAZymes, carbohydrate-active enzymes; SusC-like, SusC-like transporter protein; SusD-like, SusD-like transporter protein; STP, signal transduction protein; TC, transporter; TF, transcription factor; other. Scales for each CDS are not consistent.

CAZymes, and can bind a variety of carbohydrates [64]. CBM50, CBM6, and CBM32 were the most abundant CBM families in the genomes of the four marine Bacteroidetes strains (Fig. 2B). These CBMs are involved in the degradation of universal polysaccharides in the marine environment, such as peptidoglycan, xylan, and mannan [65–67].

PL enzymes cleave uronic acid-containing polysaccharides by a β-elimination mechanism, instead of a hydrolysis mechanism [68]. This is a different mechanism than the one used by most other CAZymes, and it allows PL enzymes to degrade a wider range of polysaccharides. The families PL8, PL12, PL35 and PL42 were shared among the four marine Bacteroidetes strains (Table S4). These PLs are involved in the degradation of alginate, glycosaminoglycan, pectin, chondroitin sulfate, and gum arabic [69–72]. Strain TK19036 possessed the largest number and variety of PL families among the four Bacteroidetes strains. Additionally, PL11 + CBM13, PL17 + PL17, PL22, PL29, PL33, and PL40 were predicted to be present only in the genome of strain TK19036 (Table S4). These PLs are involved in the degradation of phytoplankton-derived polysaccharides, such as rhamnogalacturonan, alginate, oligogalacturonide, and ulvan [73–75]. This suggests that strain TK19036 may be more specialized in the degradation of algal polysaccharides. Notably, CBM13 has been reported to enhance the alginate catalytic efficiency of PL11 [76].

AA are redox-active enzymes that potentially assist GH, PL, and CE enzymes [77,78]. Two free-living marine Bacteroidetes strains, TK19036 and *C. niigatensis* NBRC 109829^T, contain more varied AA families and larger CAZymes counts than the two parasitic strains, *T. pelagia* NBRC 107804^T and *P. rhodea* NBRC 107748^T (Table 1 and Fig. 2). This implies that free-living Bacteroidetes must possess a wider variety of AAs to deal with the more complex polysaccharide components in their environment. AA3, which consist of four subfamilies (AA3_1, AA3_2, AA3_3, AA3_4), assists in lignocellulose degradation with its reaction products [78,79]. It has been relatively little studied in marine bacteria [78–81]. AA3 accounted for the largest proportion of AA enzymes among the four Bacteroidetes (Table S5), indicating their potential activities in lignocellulose metabolism. Additionally, the prediction of AA3_4 being shared among the four Bacteroidetes provides initial evidence for the presence of AA3_4 coding genes in marine Bacteroidetes genomes, warranting further investigation [82]. However, caution needs to be exercised in the analysis of the AA families as the CAZy database did not include bacterial sequences for these families [4]. Our study indicates that AAs in bacteria need to be explored further and updated within the CAZy database.

In summary, strain TK19036 and the reference strains had numerous CAZymes in their genomes, which suggests that they have a great potential for utilizing marine-derived polysaccharides. Strain TK19036 is especially notable for its high CAZymes content, which is the highest of any known marine Bacteroidetes strain.

3.4. Diverse substrates of PULs in the four marine Bacteroidetes strains

Statistical results of PULs in the genomes of strain TK19036 and the reference strains predicted by CGC-Finder are listed in Table S6 [31]. Among them, CGCs containing at least one complete SusC-like and SusD-like pair are grouped as PULs, except for capsule polysaccharide synthesis. Consistent with CAZymes analysis, strain TK19036 harbored the greatest number and highest density of PULs (Table S6).

In terms of pectin degradation, only TK19036_CGC4 was annotated by both dbCAN3 and protein structure search to have a full set of pectin degrading enzymes, including pectate lyase (PL1), pectin methylesterase (CE8), rhamnogalacturonan acetylesterase (CE12) and exopolygalacturonase (GH28) (Fig. 3, Table S7). The potential pectin degrading pathway, involving the above CAZymes expressed by the PUL TK19036_CGC4, is shown in Fig. S4. To verify the bioinformatic predictions, a growth experiment was conducted using pectin as the sole carbon source, during which growth was observed (Fig. S2). The recombinant pectate lyase coded by FBLHDAGO_00238, located at

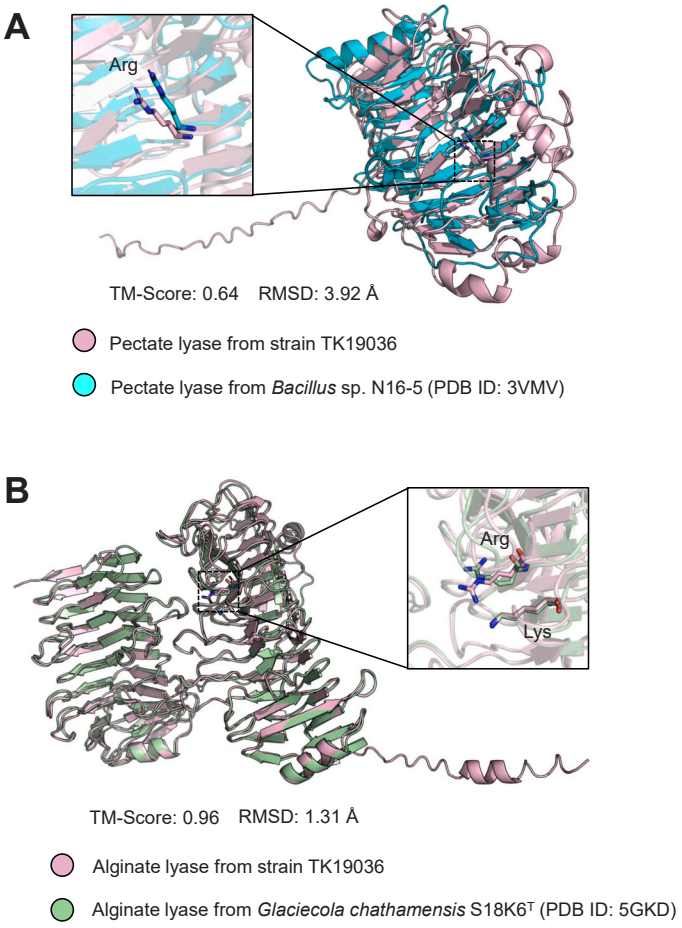


Fig. 4. Protein structural analysis. A. Structure alignment of the pectate lyase from the strain TK19036 (FBLHDAGO_00238) with that from the strain *Bacillus* sp. N16–5. B. Structure alignment of the alginate lyase from the strain TK19036 (FBLHDAGO_04016) with that from the *G. chathamensis* S18K6^T. Structures of the pectate lyase and the alginate lyase from the strain TK19036 were predicted by Alphafold2. Conserved catalytic residues are shown as sticks. RMSD and TM-Score values are presented.

TK19036_CGC4, shares a moderate RMSD value of 3.92 Å with a verified pectate lyase (PDB ID: 3VMV) (Fig. 4). Also, in the pectate lyase from the strain TK19036, Arg223 is the conserved catalytic residue, as compared to that in the similar enzyme (Fig. 4) [83]. To verify the ability to cleave pectin, it was assayed and found to have an enzymatic activity of 0.33 U/mL (Fig. S3 A, Table S5).

In terms of fructan degradation, GH32 was found to be a shared family among the fructan-degrading PULs/CGCs of all four marine Bacteroidetes strains (Fig. 3). Levan-type fructans are a type of marine bacterial exopolysaccharide [84]. Their universal distribution and the ease of their utilization in diverse marine environments is indicated by this finding.

In terms of arabinan and arabinoxylan degradation, TK19036_CGC62 and NBRC109829_CGC29 were predicted to degrade arabinan, and TK19036_CGC81 and NBRC109829_CGC61 were predicted to degrade arabinoxylan (Fig. 3). These predictions were based on the presence of GH43, GH51, GH10, GH95, and CBM48 + CE1 in these PULs (Fig. 3). As mentioned above, CE1 enzymes can enhance the hydrolysis of arabinoxylan by GHs. In the PULDB database, the CE1 + CBM48 combination appeared in the form of CE6 + CBM48 + CE1. Our study provides a possible new combination that does not contain CE6.

In terms of marine-specific alginate degradation, TK19036_CGC72

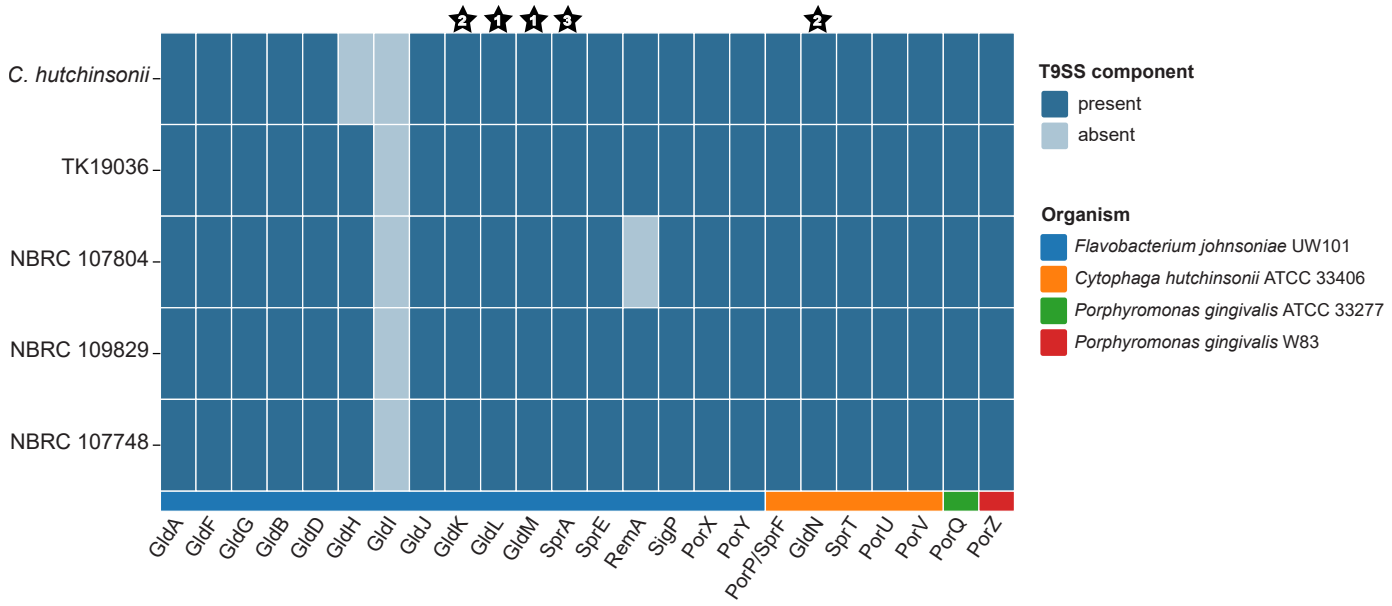


Fig. 5. T9SS components in the genomes of the strain TK19036 and related strains. Black stars indicate the three conserved subsystems of T9SS: [1] An inner membrane molecular motor composed of the PorL (GldL) and PorM (GldM) subunits, involved in energy sensing. [2] A periplasmic translocation motor complex composed the protein PorN (GldN) and the outer-membrane-associated lipoprotein PorK (GldK), can be activated in response to energy sensing. [3] An outer-membrane-spanning translocon termed SprA, that transports substrate proteins from this large channel [93].

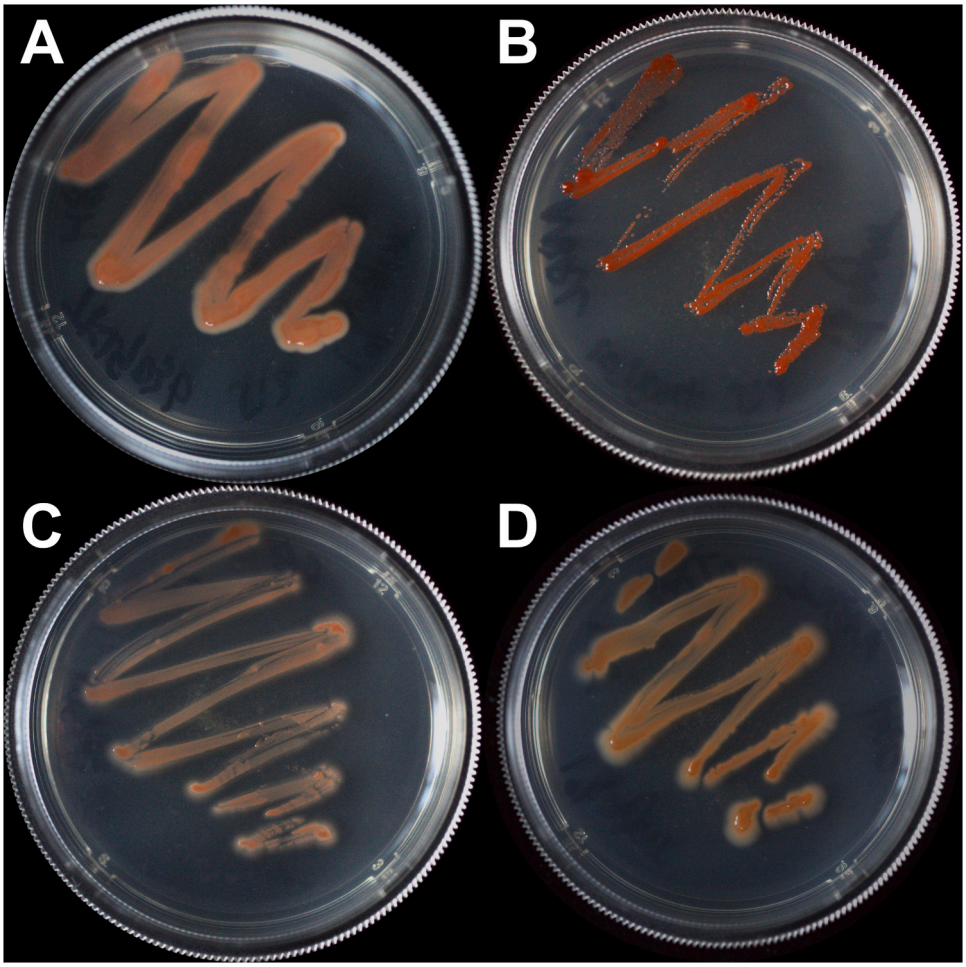


Fig. 6. Growth of the four marine Bacteroides strains on the 1% (w/v) agar plates. A, TK19036; B, *T. pelagia* NBRC 107804^{T}; C, *C. niigatensis* NBRC 109829^{T}; D, *P. rhodea* NBRC 107748^{T}.

Table 2
CAZymes containing type-A CTD in the genomes of strain TK19036 and the reference strains. -, not detected; Y, yes.

Strain	Locus_tag	CAZymes	Whether located at PULs	Whether have Sec signal peptide
TK19036	FBLHDAGO_00421	GH5_39	-	Y
	FBLHDAGO_02085	CE1	-	Y
	FBLHDAGO_02444	CE1	-	Y
	FBLHDAGO_02842	GH55 +CBM13	-	Y
	FBLHDAGO_03478	CBM4 +CBM4 +GH10	-	Y
	FBLHDAGO_04685	PL11 +CBM13	Y	-
	FBLHDAGO_05028	CE1	-	Y
	FBLHDAGO_05037	CE1	-	-
	FBLHDAGO_05216	CBM57	Y	Y
	FBLHDAGO_06596	CBM3	-	Y
NBRC 107804	NKNBAIDM_00253	CBM57	-	Y
	NKNBAIDM_01590	CBM6	-	Y
	NKNBAIDM_01591	GH74 +GH74 +CBM6	-	Y
	NKNBAIDM_01656	CBM13	Y	Y
	NKNBAIDM_01657	GH9 +CBM6 +CBM60	Y	Y
	NKNBAIDM_01658	CBM32	Y	Y
	NKNBAIDM_01660	PL6	Y	Y
	NKNBAIDM_01661	PL6 +PL6	Y	Y
	NKNBAIDM_01714	GH28	Y	Y
	NKNBAIDM_02197	PL31	-	Y
	NKNBAIDM_02203	GH51	Y	Y
	NKNBAIDM_02204	GH74 +GH74	-	Y
	NKNBAIDM_02899	GH13_10	Y	Y
	NKNBAIDM_02982	CE1	-	Y
	NKNBAIDM_05068	CE1 +CBM32	Y	Y
	NKNBAIDM_05069	CE1	-	-
	NKNBAIDM_05730	GH5_39	-	Y
NBRC 109829	BOEALPJH_00632	CBM60 +CBM60 +CBM60 +CBM60	-	-
	BOEALPJH_02253	CBM57	-	Y
	BOEALPJH_02685	PL1_2	-	Y
NBRC 107748	KHLMAHLP_00538	CBM32	-	Y
	KHLMAHLP_01532	CBM57	-	Y
	KHLMAHLP_00833	CBM60	-	Y
	KHLMAHLP_03241	GH136	-	-
	KHLMAHLP_01818	CBM32	-	Y
	KHLMAHLP_00835	CBM57 +CBM57	Y	Y

and NBRC107804_CGC19 were predicted to degrade alginate by both dbCAN3 and protein structure search (Fig. 3, Table S7). These predictions were based on the presence of PL6 and GH29 in these CGCs (Fig. 3). PL6, known as alginate lyase, were only annotated in strain TK19036 and *T. pelagia* NBRC 107804^T, indicating a possible lateral gene transfer event. However, the protein identities calculated by blastp between the PL6s in TK19036_CGC72 and in NBRC107804_CGC19 were 25.4–26.4% with p-values less than 1e-10, which indicated a distant relationship. To test the ability of strain TK19036 to utilize alginate, a growth experiment was conducted using this substrate as the sole carbon source, during which growth was observed (Fig. S2). The recombinant alginate lyase encoded by FBLHDAGO_04016, located at TK19036_CGC72, shares a low RMSD value of 1.31 Å with a verified PL6 family alginate lyase (PDB ID: 5GKD) (Fig. 4). Also, in the alginate lyase from the strain TK19036, Arg265 and Lys244 are conserved catalytic residues, as compared to those in the similar enzyme (Fig. 4) [85]. To verify the ability to cleave alginate, it was tested and found to have an enzymatic activity of 1.83 U/mL (Fig. S3 B, Table S5). Additionally, sulfatase subfamilies S1_15 and S1_17 were annotated within the alginate-degrading PUL TK19036_CGC72. These sulfatases may target chondroitin sulfate [86] and alpha-carrageenan [87].

In terms of capsule polysaccharides synthesis, two types of gene clusters were identified in our study. The first type of capsule polysaccharide synthesis gene cluster only contained GT4 and GT2 of CAZymes (Fig. 3). This type has been widely reported in genomes of Proteobacteria and Bacteroidetes [88,89]. The second type of capsule polysaccharide synthesis gene cluster contained not only GT4 and GT2, but also GT26 (Fig. 3). However, in the dbCAN-PUL database this type was only reported in the genome of a symbiotic Bacteroidetes from the human intestine [88]. Similarly, in the dbCAN-seq database, there were only 6 records of this type, out of 580 records of marine capsule polysaccharides synthesis gene clusters. These 6 records were all derived from assemblies of Proteobacteria. Our study reported the second type of capsule polysaccharide synthesis gene cluster in two marine Bacteroidetes strains, TK19036 and *T. pelagia* NBRC 107804^T (Fig. 3). This suggests that these strains may be involved in the synthesis of specific capsule polysaccharides.

In summary, PULs that target divergent polysaccharide substrates were identified in strain TK19036 and the reference strains. These PULs may be involved in the degradation of biopolymers in diverse marine environments.

3.5. *New evidence for T9SS-associated polysaccharide metabolism in marine Bacteroidetes*

The three conserved subunits in T9SS were identified in the genomes of TK19036 and the references strains (Fig. 5). Consistent with these findings, strain TK19036, *P. rhodea* NBRC 107748^T and *C. niigatensis* NBRC 109829^T were observed to have gliding motility (Fig. 6). Whereas *T. pelagia* NBRC 107804^T did not display motility (Fig. 6), possibly due to the lack of motility adhesin RemA in this strain (Fig. 5) [90].

Most large multi-domain proteins secreted by T9SS are characterized by amino-terminal signal peptides that tag them for transport across the cytoplasmic membrane, these include type A carboxy-terminal domains (CTDs) or type B CTDs, which belong to the families TIGR04183 and TIGR04131, respectively [9]. In our study, Type A CTD in CAZymes were annotated in all four marine Bacteroidetes (Table 2). In the four strains, Type A CTD-containing CAZymes accounted for 0.8–5.3% of CAZymes, and 9.4–22.4% of the predicted Type A CTD-containing proteins. However, type B CTD in CAZymes were only annotated in *C. niigatensis* NBRC 109829^T and *P. rhodea* NBRC 107748^T (e.g., BOEALPJH_04580 and KHLMAHLP_00538). This rarity of type B CTD

was consistent with previous studies [91,92].

The genome of *T. pelagia* NBRC 107804^T was predicted to have the most type A CTD-containing CAZymes (Table 2). This suggests that the lack of RemA motility adhesins may not influence the function of T9SS to secrete CAZymes for polysaccharide degradation. In other words, the polysaccharide metabolic capabilities of non-motile Bacteroidetes may have been underestimated. Moreover, NKNBAIDM_01656 (CBM13), NKNBAIDM_01657 (GH9 +CBM6 +CBM60), NKNBAIDM_01658 (CBM32), NKNBAIDM_01660 (PL6) and NKNBAIDM_01661 (PL6 +PL6) were located at the PUL NBRC107804_CGC19 (Fig. 3). The predicted structure of their Sec peptide and type A CTD, as shown in Fig. S5, also suggests their capability to be secreted across both the inner and outer membranes [9]. This suggests that NBRC107804_CGC19 is a "hybrid" PUL. A prediction of the polysaccharide degradation process is as follows: CBM13, GH9 +CBM6 +CBM60, CBM32 and PL6 are first secreted to the outer membrane. Then they cleave or hydrolyze complex alginates to oligosaccharides. Finally, the oligosaccharides are transported to the periplasm by SusC/D-like pairs to be further hydrolyzed by other periplasm-located CAZymes. Notably, to the best of our knowledge, NBRC107804_CGC19 is the first reported "hybrid" PUL targeting alginates.

In summary, strain TK19036 and the reference strains were identified to have three conserved components from T9SS. Domain analysis indicated that T9SS could facilitate the degradation of marine-derived polysaccharides via both "PUL-free" and "hybrid" PUL mechanisms. Especially, to the best of our knowledge, we proposed the first reported "hybrid" PUL targeting alginates. However, all of the genomic findings require further confirmation by transcriptomic, proteomic and other multi-omic methods.

4. Conclusions

In conclusion, we isolated the novel marine Bacteroidetes strain TK19036, which was proposed to represent a novel genus within the family Catalimonadaceae. We found that strain TK19036 had the largest number of CAZymes among currently known marine Bacteroidetes genomes, and was capable of degrading pectin and alginate. Comparative genome analysis revealed the presence of diverse PULs and core T9SS components in strain TK19036 and its relatives in the family Catalimonadaceae, suggesting that they are capable of sensing, binding, and utilizing complex carbohydrates.

CRediT authorship contribution statement

Beihan Chen: Conceptualization, Methodology, Investigation, Formal analysis, Writing – original draft. **Guohua Liu:** Investigation, Software, Writing – review & editing. **Quanrui Chen:** Investigation, Writing – review & editing. **Huanyu Wang:** Investigation, Writing – review & editing. **Le Liu:** Investigation, Writing – review & editing. **Kai Tang:** Supervision, Funding acquisition, Resources, Writing – review & editing.

Declaration of Competing Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

This study was supported by the National Natural Science Foundation of China project (92251306, 42188102, 42076160, 42276120) and International Science Partnership Program of the Chinese Academy of Sciences (121311KYSB20190029). We appreciate Shicong Xiao for obtaining the seawater sample, Jianing Ye and Huaying Lin for the isolation of strain TK19036, Dan Lin for bacterial culture purchase, and Shujing Liu for setting up the operating environment of GTDB-Tk. We are also grateful to the reviewers for their constructive suggestions and insightful comments on the manuscript.

Data availability statement

The GenBank accession numbers for the full-length 16S rRNA gene sequence and the whole genome shotgun sequence of strain TK19036 are OL468765 and CP120682 respectively. The GenBank accession numbers for the whole genome shotgun sequences of *Tunicatimonas pelagia* NBRC 107804^T, *Porifericola rhodea* NBRC 107748^T and *Catalinimonas niigatensis* NBRC 109829^T are CP120683-CP120686, CP119421 and CP119422, respectively.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.csbj.2023.12.025.

References

[1] Alonso C, Warnecke F, Amann R, Pernthaler J. High local and global diversity of Flavobacteria in marine plankton. Environ Microbiol 2007;9:1253–66.
[2] Lapébie P, Lombard V, Drula E, Terrapon N, Henrissat B. Bacteroidetes use thousands of enzyme combinations to break down glycans. Nat Commun 2019;10 (1):7.
[3] Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res 2014;42: D490–5.
[4] Drula E, Garron M-L, Dogan S, Lombard V, Henrissat B, et al. The carbohydrate-active enzyme database: functions and literature. Nucleic Acids Res 2021;50: D571–7.
[5] Grondin JM, Tamura K, Déjean G, Abbott DW, Brumer H. Polysaccharide Utilization Loci: fueling microbial communities. J Bacteriol 2017:199. https://doi.org/10.1128/JB.00860-00816.
[6] McKee LS, La Rosa SL, Westereng B, Eijsink VG, Pope PB, et al. Polysaccharide degradation by the Bacteroidetes: mechanisms and nomenclature. Environ Microbiol Rep 2021;13:559–81.
[7] Lasica AM, Ksiazek M, Madej M, Potempa J. The Type IX Secretion System (T9SS): highlights and recent insights into its structure and function. Front Cell Infect Microbiol 2017;7. https://doi.org/10.3389/fcimb.2017.00215.
[8] McBride MJ, Zhu Y. Gliding motility and por secretion system genes are widespread among members of the phylum bacteroidetes. J Bacteriol 2013;195: 270–8.
[9] Veith PD, Glew MD, Gorasia DG, Reynolds EC. Type IX secretion: the generation of bacterial cell surface coatings involved in virulence, gliding motility and the degradation of complex biopolymers. Mol Microbiol 2017;106:35–53.
[10] Lauber F, Deme JC, Lea SM, Berks BC. Type 9 secretion system structures reveal a new protein transport mechanism. Nature 2018;564:77–82.
[11] Lee I, Chalita M, Ha SM, Na SI, Yoon SH, et al. ContEst16S: an algorithm that identifies contaminated prokaryotic genomes using 16S RNA gene sequences. Int J Syst Evol Microbiol 2017;67:2053–7.
[12] Meier-Kolthoff JP, Carbasse JS, Peinado-Olarte RL, Göker M. TYGS and LPSN: a database tandem for fast and reliable genome-based classification and nomenclature of prokaryotes. Nucleic Acids Res 2021;50:D801–7.
[13] Yoon SH, Ha SM, Lim J, Kwon S, Chun J. A large-scale evaluation of algorithms to calculate average nucleotide identity. Antonie Van Leeuwenhoek 2017;110: 1281–6.
[14] R-R LM, K KT. The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. PeerJ Prepr 2016:4. https://doi.org/10.7287/peerj.preprints.1900v7281.
[15] Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk v2: memory friendly classification with the genome taxonomy database. Bioinformatics 2022;38: 5315–6.
[16] Yoon J, Oku N, Park S, Katsuta A, Kasai H. Tunicatimonas pelagia gen. nov., sp. nov., a novel representative of the family Flammeovirgaceae isolated from a sea anemone by the differential growth screening method. Antonie Van Leeuwenhoek 2012;101:133–40.
[17] Yoon J, Oku N, Park S, Kasai H, Yokota A. *Porifericola rhodea* gen. nov., sp. nov., a new member of the phylum Bacteroidetes isolated by the bait-streaked agar technique. Antonie Van Leeuwenhoek 2011;100:145–53.
[18] Yoon J, Adachi K, Kasai H. *Catalinimonas niigatensis* sp. nov., a novel member of the family Catalimonadaceae within the phylum Bacteroidetes isolated from lake sediment. J Gen Appl Microbiol 2014;60:33–7.
[19] Chin C-S, Alexander DH, Marks P, Klammer AA, Drake J, et al. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat Methods 2013;10:563–9.
[20] Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 2014;9. https://doi.org/10.1371/journal.pone.0112963.

[21] Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol 2017;13. https://doi.org/10.1371/journal.pcbi.1005595.

[22] Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics 2014; 30:2068–9.

[23] Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinf 2010:11. https://doi.org/10.1186/1471-2105-1111-1119.

[24] Lagesen K, Hallin P, Rødland EA, Staerfeldt HH, Rognes T, et al. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res 2007; 35:3100–8.

[25] Laslett D, Canback B. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res 2004;32:11–6.

[26] Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods 2011;8:785–6.

[27] Kolbe DL, Eddy SR. Fast filtering for RNA homology search. Bioinformatics 2011; 27:3102–9.

[28] Zheng J, Ge Q, Yan Y, Zhang X, Huang L, et al. dbCAN3: automated carbohydrate-active enzyme and substrate annotation. Nucleic Acids Res 2023;51:W115–21.

[29] Zhang H, Yohe T, Huang L, Entwistle S, Wu P, et al. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. Nucleic Acids Res 2018;46: W95–101.

[30] Paysan-Lafosse T, Blum M, Chuguransky S, Grego T, Pinto BL, et al. InterPro in 2022. Nucleic Acids Res 2023;51. D418-d427.

[31] Huang L, Zhang H, Wu P, Entwistle S, Li X, et al. dbCAN-seq: a database of carbohydrate-active enzyme (CAZyme) sequence and annotation. Nucleic Acids Res 2017;46:D516–21.

[32] Saier MH, Reddy VS, Moreno-Hagelsieb G, Hendargo KJ, Zhang Y, et al. The transporter classification database (TCDB): 2021 update. Nucleic Acids Res 2021; 49. D461-d467.

[33] Zheng J, Hu B, Zhang X, Ge Q, Yan Y, et al. dbCAN-seq update: CAZyme gene clusters and substrates in microbiomes. Nucleic Acids Res 2022;51:D557–63.

[34] Ausland C, Zheng J, Yi H, Yang B, Li T, et al. dbCAN-PUL: a database of experimentally characterized CAZyme gene clusters and their substrates. Nucleic Acids Res 2020;49. D523-D528.

[35] Barbeyron T, Brillet-Guéguen L, Carré W, Carrière C, Caron C, et al. Matching the diversity of sulfated biomolecules: creation of a classification database for sulfatases reflecting their substrate specificity. PLoS One 2016;11. https://doi.org/10.1371/journal.pone.0164846.

[36] Stam M, Lelièvre P, Hoebeke M, Corre E, Barbeyron T, et al. SulfAtlas, the sulfatase database: state of the art and new developments. Nucleic Acids Res 2022;51: D647–53.

[37] Gao L, Su Y, Song W, Zhang W, Qi Q, et al. A type IX secretion system substrate involved in crystalline cellulose degradation by affecting crucial cellulose binding proteins in *Cytophaga hutchinsonii*. Appl Environ Microbiol 2022;88:e0183721.

[38] Chen C, Wu Y, Li J, Wang X, Zeng Z, et al. TBtools-II: A" One for All, All for One" bioinformatics platform for biological big-data mining. Mol Plant 2023;16: 1733–42.

[39] Teufel F, Almagro Armenteros JJ, Johansen AR, Gíslason MH, Pihl SI, et al. SignalP 6.0 predicts all five types of signal peptides using protein language models. Nat Biotechnol 2022;40:1023–5.

[40] Jumper J, Evans R, Pritzel A, Green T, Figurnov M, et al. Highly accurate protein structure prediction with AlphaFold. Nature 2021;596:583–9.

[41] Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, et al. ColabFold: making protein folding accessible to all. Nat Methods 2022;19:679–82.

[42] van Kempen M, Kim SS, Tumescheit C, Mirdita M, Lee J, et al. Fast and accurate protein structure search with Foldseek. Nat Biotechnol 2023. https://doi.org/10.1038/s41587-023-01773-0.

[43] Zhang C, Shine M, Pyle AM, Zhang Y. US-align: universal structure alignments of proteins, nucleic acids, and macromolecular complexes. Nat Methods 2022;19: 1109–15.

[44] Delano WL. Pymol: an open-source molecular graphics tool. CCP4 Newsl. Protein Crystallogr 2002;40:82–92.

[45] Liu L, Chen X, Ye J, Ma X, Han Y, et al. Sulfoquinovose is a widespread organosulfur substrate for Roseobacter clade bacteria in the ocean. Isme J 2023;17: 393–405.

[46] Albersheim P, Killias U. Studies relating to the purification and properties of pectin transeliminase. Arch Biochem Biophys 1962;97:107–15.

[47] Wang H, Li J, Liu L, Li X, Jia D, et al. Increased production of alkaline polygalacturonate lyase in the recombinant Pichia pastoris by controlling cell concentration during continuous culture. Bioresour Technol 2012;124:338–46.

[48] Mathieu S, Henrissat B, Labre F, Skjåk-Bræk G, Helbert W. Functional exploration of the polysaccharide lyase family PL6. PLoS One 2016;11. https://doi.org/10.1371/journal.pone.0159415.

[49] Yarza P, Yilmaz P, Pruesse E, Glöckner FO, Ludwig W, et al. Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. Nat Rev Microbiol 2014;12:635–45.

[50] Barco RA, Garrity GM, Scott JJ, Amend JP, Nealson KH, et al. A genus definition for bacteria and archaea based on a standard genome relatedness index. mBio 2020:11. https://doi.org/10.1128/mBio.02475-02419.

[51] Konstantinidis KT, Tiedje JM. Prokaryotic taxonomy and phylogeny in the genomic era: advancements and challenges ahead. Curr Opin Microbiol 2007;10:504–9.

[52] Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, et al. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. Int J Syst Evol Microbiol 2007;57:81–91.

[53] Christiansen C, Abou Hachem M, Janecek S, Viksø-Nielsen A, Blennow A, et al. The carbohydrate-binding module family 20-diversity, structure, and function. Febs J 2009;276:5006–29.

[54] Janecek S., Svensson B. (2021) "Carbohydrate Binding Module Family 48" in CAZypedia, available at URL http://www.cazypedia.org/. accessed 21 November 2023.

[55] Janeček Š, Svensson B, MacGregor EA. Structural and evolutionary aspects of two families of non-catalytic domains present in starch and glycogen binding proteins from microbes, plants and animals. Enzym Micro Technol 2011;49:429–40.

[56] Wolf MJ, Coe A, Dove LA, Zawadowicz MA, Dooley K, et al. Investigating the heterogeneous ice nucleation of sea spray aerosols using prochlorococcus as a model source of marine organic matter. Environ Sci Technol 2019;53:1139–49.

[57] Martinez-Fleites C, Proctor M, Roberts S, Bolam DN, Gilbert HJ, et al. Insights into the synthesis of lipopolysaccharide and antibiotics through the structures of two retaining glycosyltransferases from family GT4. Chem Biol 2006;13:1143–52.

[58] van Heijenoort J. Formation of the glycan chains in the synthesis of bacterial peptidoglycan. Glycobiology 2001;11. 25r-36r.

[59] Davies GJ, Gloster TM, Henrissat B. Recent structural insights into the expanding world of carbohydrate-active enzymes. Curr Opin Struct Biol 2005;15:637–45.

[60] Armendáriz-Ruiz M, Rodríguez-González JA, Camacho-Ruíz RM, Mateos-Díaz JC. Carbohydrate Esterases: an overview. Methods Mol Biol 2018;1835:39–68.

[61] Mohammed ASA, Naveed M, Jost N. Polysaccharides; classification, chemical properties, and future perspective applications in fields of pharmacology and biological medicine (A Review of Current Applications and Upcoming Potentialities). J Polym Environ 2021;29:2359–71.

[62] Nakamura AM, Nascimento AS, Polikarpov I. Structural diversity of carbohydrate esterases. Biotechnol Res Innov 2017;1:35–51.

[63] Nakamura T, Yonezawa Y, Tsuchiya Y, Niiyama M, Ida K, et al. Substrate recognition of N,N′-diacetylchitobiose deacetylase from Pyrococcus horikoshii. J Struct Biol 2016;195:286–93.

[64] Boraston AB, Bolam DN, Gilbert HJ, Davies GJ. Carbohydrate-binding modules: fine-tuning polysaccharide recognition. Biochem J 2004;382:769–81.

[65] Steen A, Buist G, Leenhouts KJ, El Khattabi M, Grijpstra F, et al. Cell wall attachment of a widely distributed peptidoglycan binding domain is hindered by cell wall constituents. J Biol Chem 2003;278:23874–81.

[66] Fernandes AC, Fontes CM, Gilbert HJ, Hazlewood GP, Fernandes TH, et al. Homologous xylanases from *Clostridium thermocellum*: evidence for bi-functional activity, synergism between xylanase catalytic modules and the presence of xylan-binding domains in enzyme complexes. Biochem J 1999;342(Pt 1):105–10.

[67] Mizutani K, Fernandes VO, Karita S, Luís AS, Sakka M, et al. Influence of a mannan binding family 32 carbohydrate binding module on the activity of the appended mannanase. Appl Environ Microbiol 2012;78:4781–7.

[68] Lombard V, Bernard T, Rancurel C, Brumer H, Coutinho PM, et al. A hierarchical classification of polysaccharide lyases for glycogenomics. Biochem J 2010;432: 437–44.

[69] Pilgaard B, Vuillemin M, Holck J, Wilkens C, Meyer AS. Specificities and synergistic actions of novel PL8 and PL7 alginate lyases from the marine fungus *Paradendryphiella salina*. J Fungi (Basel) 2021;7. https://doi.org/10.3390/jof7020080.

[70] Garron M-L, Cygler M. Uronic polysaccharide degrading enzymes. Curr Opin Struct Biol 2014;28:87–95.

[71] Helbert W, Poulet L, Drouillard S, Mathieu S, Loiodice M, et al. Discovery of novel carbohydrate-active enzymes through the rational exploration of the protein sequences space. Proc Natl Acad Sci 2019;116:6063–8.

[72] Kondo T, Kichijo M, Maruta A, Nakaya M, Takenaka S, et al. Structural and functional analysis of gum arabic l-rhamnose-α-1,4-d-glucuronate lyase establishes a novel polysaccharide lyase family. J Biol Chem 2021;297:101001.

[73] Jensen MH, Otten H, Christensen U, Borchert TV, Christensen LL, et al. Structural and biochemical studies elucidate the mechanism of rhamnogalacturonan lyase from *Aspergillus aculeatus*. J Mol Biol 2010;404:100–11.

[74] Park D, Jagtap S, Nair SK. Structure of a PL17 family alginate lyase demonstrates functional similarities among exotype depolymerases. J Biol Chem 2014;289: 8645–55.

[75] Reisky L, Préchoux A, Zühlke MK, Bäumgen M, Robb CS, et al. A marine bacterial enzymatic cascade degrades the algal polysaccharide ulvan. Nat Chem Biol 2019; 15:803–12.

[76] Li S, Yang X, Bao M, Wu Y, Yu W, et al. Family 13 carbohydrate-binding module of alginate lyase from *Agarivorans* sp. L11 enhances its catalytic efficiency and thermostability, and alters its substrate preference and product distribution. FEMS Microbiol Lett 2015:362. https://doi.org/10.1093/femsle/fnv1054.

[77] Levasseur A, Drula E, Lombard V, Coutinho PM, Henrissat B. Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. Biotechnol Biofuels 2013:6. https://doi.org/10.1186/1754-6834-1186-1141.

[78] Sützl L, Laurent CVFP, Abrera AT, Schütz G, Ludwig R, et al. Multiplicity of enzymatic functions in the CAZy AA3 family. Appl Microbiol Biotechnol 2018;102: 2477–92.

[79] Yu X, Zhao Y, Yu J, Wang L. Recent advances in the efficient degradation of lignocellulosic metabolic networks by lytic polysaccharide monooxygenase. Acta Biochim Biophys Sin 2023;55:529–39.

[80] Brott S, Nam KH, Thomas F, Dutschei T, Reisky L, et al. Unique alcohol dehydrogenases involved in algal sugar utilization by marine bacteria. Appl Microbiol Biotechnol 2023;107:2363–84.

[81] Chernysheva N, Bystritskaya E, Stenkova A, Golovkin I, Nedashkovskaya O, et al. Comparative genomics and CAZyme genome repertoires of marine *Zobellia amurskyensis* KMM 3526^T and *Zobellia laminariae* KMM 3676^T. Mar Drugs 2019;17. https://doi.org/10.3390/md17120661.

[82] Kostelac A, Sützl L, Puc J, Furlanetto V, Divne C, et al. Biochemical characterization of pyranose oxidase from streptomyces canus—towards a better understanding of pyranose oxidase homologues in bacteria. Int J Mol Sci 2022;23. https://doi.org/10.3390/ijms232113595.
[83] Zheng Y, Huang C-H, Liu W, Ko T-P, Xue Y, et al. Crystal structure and substrate-binding mode of a novel pectate lyase from alkaliphilic *Bacillus* sp. N16-5. Biochem Biophys Res Commun 2012;420:269–74.
[84] El Halmouch Y, Ibrahim HAH, Dofdaa NM, Mabrouk MEM, El-Metwally MM, et al. Complementary spectroscopy studies and potential activities of levan-type fructan produced by *Bacillus paralicheniformis* ND2. Carbohydr Polym 2023;311. https://doi.org/10.1016/j.carbpol.2023.120743.
[85] Xu F, Dong F, Wang P, Cao H-Y, Li C-Y, et al. Novel molecular insights into the catalytic mechanism of marine bacterial alginate lyase AlyGC from polysaccharide lyase family 6. J Biol Chem 2017;292:4457–68.
[86] Ndeh D, Baslé A, Strahl H, Yates EA, McClurgg UL, et al. Metabolism of multiple glycosaminoglycans by *Bacteroides thetaiotaomicron* is orchestrated by a versatile core genetic locus. Nat Commun 2020:11. https://doi.org/10.1038/s41467-41017-01832-41466.
[87] Ficko-Blean E, Préchoux A, Thomas F, Rochat T, Larocque R, et al. Carrageenan catabolism is encoded by a complex regulon in marine heterotrophic bacteria. Nat Commun 2017:8. https://doi.org/10.1038/s41467-41017-01832-41466.
[88] Patrick S, Blakely GW, Houston S, Moore J, Abratt VR, et al. Twenty-eight divergent polysaccharide loci specifying within- and amongst-strain capsule diversity in three strains of *Bacteroides fragilis*. Microbiol (Read) 2010;156:3255–69.
[89] Shashkov AS, Cahill SM, Arbatsky NP, Westacott AC, Kasimova AA, et al. *Acinetobacter baumannii* K116 capsular polysaccharide structure is a hybrid of the K14 and revised K37 structures. Carbohydr Res 2019;484:107774.
[90] McBride MJ, Nakane D. *Flavobacterium* gliding motility and the type IX secretion system. Curr Opin Microbiol 2015;28:72–7.
[91] Kulkarni SS, Zhu Y, Brendel CJ, McBride MJ. Diverse C-Terminal sequences involved in *Flavobacterium johnsoniae* protein secretion. J Bacteriol 2017:199. https://doi.org/10.1128/JB.00884-00816.
[92] Gorasia DG, Seers CA, Heath JE, Glew MD, Soleimaninejad H, et al. Type B CTD proteins secreted by the type IX secretion system associate with PorP-like proteins for cell surface anchorage. Int J Mol Sci 2022;23. https://doi.org/10.3390/ijms23105681.
[93] Song L, Perpich JD, Wu C, Doan T, Nowakowska Z, et al. A unique bacterial secretion machinery with multiple secretion centers. Proc Natl Acad Sci 2022;119. https://doi.org/10.1073/pnas.2119907119.