scientific reports

Check for updates

OPEN

Insights into putative alginate lyases from epipelagic and mesopelagic communities of the global ocean

Mariana Lozada[1] & Hebe M. Dionisi[2]✉

Alginate lyases and oligoalginate lyases catalyze the cleavage of the glycosidic bonds of alginate, an acidic polysaccharide synthesized by brown algae and other organisms. These enzymes are highly diverse, currently classified into 15 families of the Carbohydrate-Active Enzyme (CAZy) database. We explored the structural and taxonomic diversity, the biogeographic distribution of the genes and transcripts, and the potential environmental drivers of putative alginate-degrading enzymes from picoplanktonic communities of the upper layers of the global ocean. The identified sequences were first analyzed using sequence similarity networks to assess their relationship with CAZy members. Sequences related to the PL5, PL6, PL7, PL17, and PL38 families had higher gene and transcript abundances, with temperature being a key driver of the structuring of the community members carrying putative alginate lyase genes. PL5 homologs included variants in a key residue of the active site, and sequences assigned to '*Candidatus* Pelagibacter' showed high gene and transcript abundances that negatively correlated with inorganic phosphorus concentrations. Sequences assigned to Flavobacteriia and/or Gammaproteobacteria classes dominated the PL6, PL7, and PL17 families, in particular those closely related to sequences from uncultured Polaribacter and *Alteromonas australica*. In the PL38 family, while sequences assigned to taxa from the Planctomycetota, Verrucomicrobiota, and Bacteroidota phyla showed the highest relative gene abundance at most regions and depths, high expression levels were observed at high latitudes in sequences assigned to Eukaryota (e.g., *Phaeocystis antarctica*). Overall, the putative enzymes uncovered in this study could be involved in various physiological processes, including alginate assimilation and biosynthesis.

Keywords Alginate lyases, Epipelagic layer, Gene abundance, Mesopelagic zone, *Tara* Ocean expeditions, Transcript abundance

Brown algae (Phaeophyceae) constitute a diverse class of photoautotrophic organisms that provide key ecosystem functions in coastal marine ecosystems. Approximately a quarter of the oceans' coastal areas are covered by large forests of brown algae, creating unique habitats that are essential for sustaining communities with a high biodiversity[1]. Brown algae dominate carbon dioxide fixation in coastal ecosystems and positively affect secondary productivity in coastal areas, providing biomass for direct consumption by herbivores and through the release of algal detritus[2]. The influence of brown algae is not restricted to coastal areas, as pelagic species of the genus *Sargassum* are distributed in the open ocean[3], and dislodged fragments from coastal species can be transported long distances[4]. A cell wall rich in polysaccharides surrounds their cells, playing key structural and functional roles[5]. The polysaccharide alginate is the main component of the cell wall, accounting for up to 40% of the total dry weight[6]. As alginate constitutes an important fraction of the carbon fixed by brown algae, its breakdown and ultimate assimilation by marine microorganisms is a key step of the carbon cycle in the global ocean[7].

Alginate is a linear polysaccharide containing (1→4)-linked β-(1,4)-D-manuronate (M) and its epimer α-(1,4)-L-guluronate (G) arranged as MM, GG, or MG blocks. Both endolytic and exolytic enzymes with different specificities (EC 4.2.2.3, EC 4.2.2.11, EC 4.2.2.26, EC 4.2.2.-) catalyze the cleavage of the glycosidic bonds of alginate through a β-elimination mechanism. These enzymes have been identified in bacteria, fungi, mollusks, and viruses[8]. They are required for the complete depolymerization of alginate before its assimilation,

[1]Instituto de Biología de Organismos Marinos (IBIOMAR-CONICET), Boulevard Brown 2915 (U9120ACD), Puerto Madryn, Chubut, Argentina. [2]Centro para el Estudio de Sistemas Marinos (CESIMAR-CONICET), Boulevard Brown 2915 (U9120ACD), Puerto Madryn, Chubut, Argentina. ✉email: hdionisi@cenpat-conicet.gob.ar

and they also participate in the alginate biosynthesis pathway from both brown algae[9] and bacteria[10]. Alginate-depolymerizing enzymes are highly diverse, and as of November 2024, they are included in the PL5, PL6, PL7, PL8, PL14, PL15, PL17, PL18, PL31, PL34, PL36, PL38, PL39, PL41, and PL44 families of polysaccharide lyases (PLs) of the Carbohydrate-Active Enzymes (CAZy) database[11,12]. The three-dimensional structures or models of members of the 15 families indicated that they can be included in one of four fold groups: $(\alpha/\alpha)_n$ barrel, parallel β-helix, β-jelly roll, and $(\alpha/\alpha)_6$ toroid + anti-parallel β-sheet, suggesting evolutionary relationships among members of these families[13,14]. Despite the structural differences, remarkable similarities can be observed in their active sites, suggesting an evolutionary convergence in both catalytic and functional features[15].

As new families of alginate lyases continue to be created[16] and more sequences and activities are added to the CAZy database[12,17], it is clear that the diversity and catalytic mechanisms of alginate-depolymerizing enzymes are yet to be fully explored. This information is relevant for revealing the key players and the main environmental drivers of alginate assimilation in the marine environment, for uncovering structure–function relationships, and for facilitating their biotechnological applications[8]. Although recent studies analyzed the CAZyome of the microbial communities from marine habitats[18–20], few studies focused on the less prevalent PL enzymes and, in particular, alginate lyases[21]. Consequently, our knowledge of the diversity of these enzymes in microbial communities from different marine habitats is still limited. In this study, we used a gene-centric approach based on the analysis of the PL families that include alginate-degrading enzymes in epipelagic and mesopelagic microbial communities, using the multi-omics information generated in the *Tara* Oceans expeditions (2009–2013)[22,23]. We hypothesized that at least part of the putative alginate lyases would be divergent from those deposited in the CAZy database, mostly containing sequences from cultured microorganisms. Therefore, we chose a permissive threshold for sequence identification, followed by the assessment of the level of relatedness with members of these families. Using this approach, we were able to uncover a high taxonomic and structural diversity of actively transcribed enzymes potentially involved in alginate degradation or other physiological processes, and the main environmental factors that correlated with their gene and transcript abundances.

Material and methods

Sequence identification

Sequences homologous to members of CAZy families that include alginate-depolymerizing enzymes (Table 1) were searched in the Ocean Gene Atlas 2.0 web server (OGA)[24]. The Ocean Microbial Reference Gene Catalogue with data from the Arctic Ocean (OM-RGC.v2) was searched in OGA through the OM-RGCv2 + metaG dataset[23]. The search was performed using Hidden Markov Models (HMMs) of the targeted families downloaded from the V.12 dbCAN database[25]. For the recently created family PL44, not available in the dbCAN database at the moment of the search, an HMM was created from the alignment of the sequences deposited in CAZy (www.cazy.org), using HMMER[26]. For all families, the E-value cutoff was set at 10^{-10}, and the relative gene abundance was set as a percent of mapped reads. The alignment table, sequences, taxonomy, environmental parameters, and gene abundance matrix were retrieved from the web server for each search. The same procedure was used to search the metatranscriptomes (OM-RGCv2 + metaT dataset) to obtain transcript abundance information (also set as percent of mapped reads).

PL5 homolog sequences from Pelagibacteraceae genomes were identified in the IMG/M system[27] using the PF05426 (Alginate Lyase domain). PL38 homolog sequences were searched in Eukaryotic Genomes (SMAGs)[28] and the MetaGenomic Transcriptomes (MGT)[29] datasets by blastp[30] at the OGA web server (threshold 10^{-100}, query amino acid sequence ID OM-RGC.v2.001480294).

Fold	HMM	Number of identified sequences[a]	Sequences clustering with family members (%)[b]	Figures
$(\alpha/\alpha)_n$ Barrel	PL5	373	89.8	1
	PL38	241	98.3	
Parallel β-helix	PL6	478	90.4	2
	PL31	124	4	
	PL41	135	0	
	PL44	146	2.1	
β-Jelly roll	PL7	400	98	3
	PL14	137	83.2	
	PL18	15	100	
	PL36	84	1.2	
$(\alpha/\alpha)_6$ Toroid + anti-parallel β-sheet	PL8	118	97.5	4
	PL15	101	92.1	
	PL17	220	96.8	
	PL34	19	73.7	
	PL39	143	40.6	

Table 1. Number of sequences identified in the Ocean Microbial Reference Gene Catalogue (OM-RGC.v2). [a]Sequence identification was performed using the HMMs of the dbCAN3 database in the OGA web server[24], or with a HMM designed with members of the PL44 family (E-value cutoff 10^{-10}). [b]Percentage of the identified sequences present in clusters of the SSN that also include CAZy members of the same family.

Sequence similarity networks

Sequence Similarity Networks (SSNs) were built from the catalytic modules of the targeted families using the Enzyme Similarity Tool (EFI-EST)[31]. Each SSN assessed the relationships between sequences of CAZy members and the homologs identified in the OM-RGC.v2 reference dataset. First, the accession numbers of the members of each CAZy family were obtained from the CAZy database[12]. The GenBank sequences were retrieved using the Batch Entrez tool of the National Center for Biotechnology Information (NCBI) website (www.ncbi.nlm.nih.gov/sites/batchentrez). Sequences deposited in the Mycocosm (https://mycocosm.jgi.doe.gov/) or the Phycocosm (https://phycocosm.jgi.doe.gov) databases of the Joint Genome Institute (JGI) were retrieved using their search tools or from the Conserved Unique Peptide Patterns (CUPP) online platform[32]. Sequences were discriminated by subfamily (SF), if defined in the CAZy database. Unique sequences from the members of each PL family and the homolog sequences retrieved using the respective HMM were first aligned using ClustalX 2.1 (> 1000 sequences)[33] or ClustalWS in Jalview 2.10.5.0 (< 1000 sequences)[34]. The sequences from characterized enzymes containing only a single domain were used to select a region in the alignment corresponding to the catalytic module. The amino acid sequences from the catalytic module of all sequences were then retrieved from the selected section of the alignment and used to construct the SSN.

Each SSN included a group of PL families with a similar fold (Table 1). The SSN was finalized using a series of cutoffs of their alignment scores (10^{-1} increments, starting with 10^{-15}), and the connection of the nodes containing members of different families was assessed in Cytoscape v. 3.10.1[35]. The chosen cutoff for the finalization was the E-value at which the members from different families were first disconnected in the SSN. The information of the sequences included in each cluster or subcluster was obtained from Cytoscape, and the taxonomy as reported in the OGA web server[24] was analyzed separately at the class level for each cluster or subcluster.

Three-dimensional modeling and structural analyses

Sequences from different clusters and subclusters of the SSNs were selected for modeling. Although only the putative catalytic module was included in the SSN, the full-length sequence was used for modeling in AlphaFold 3[36]. Model 0 was visualized in ChimeraX 1.8[37] and superimposed to the structures from the analyzed CAZy families using the pairwise structure comparison function. The predicted aligned error (PAE) graph and the TM score (pTM) of the models are shown next to each model. The models were further analyzed using the Protein Data Bank (PDB) search function in the DALI server[38], using the region of the catalytic module as query. The results were reported as matches against PDB25, which is a representative subset of the PDB[38]. The signal peptide was predicted using DeepTMHMM[39].

Relative abundance of genes and transcripts

The abundance of genes and transcripts (OM-RGCv2 + metaG and OM-RGCv2 + metaT datasets, respectively) was normalized as a percent of mapped reads[24]. The abundance matrices (0.22–1.6 or 0.22–3 µm size range[22]) of each selected CAZy family were first curated by eliminating the sequences not included in the cluster/s of the SSN with members of the same family, considered as potential unspecific identifications. The abundance matrices from each family were analyzed in the nine geographic regions: Arctic Ocean (AO), North Atlantic Ocean (NAO), North Pacific Ocean (NPO), Indian Ocean (IO), Mediterranean Sea (MS), Red Sea (RS), South Atlantic Ocean (SAO), South Pacific Ocean (SPO), and Southern Ocean (SO). In addition, sample depth was considered in the analyses: upper layer zone (SRF), deep chlorophyll maximum layer (DCM), and mesopelagic zone (MES) (median of 5, 50 and 550 m, respectively[23,24]). Relative abundance of genes and transcripts was reported as average ± standard deviation at each depth layer and the geographic region. Significant differences in gene and transcript abundances across depths were analyzed by the Kruskal–Wallis rank sum test followed by contrasts (Dunn's test). The statistical analyses were carried out in R packages phyloseq v. 1.46.0[40] and dplyr v. 1.1.4[41]. Taxonomy-based analyses were based on the classification of the OGA web server[24], except when indicated. In these cases, the assigned taxonomy was compared to the results of blastp analyses against the NCBI non-redundant protein sequence database[30]. If the closest matches differed from the assigned taxonomy due to the recent addition of genomes from uncultured microorganisms (sharing protein identity values > 60%), the assignment was changed.

Ordination and statistical analyses

Non-metric multidimensional scaling (NMDS) analysis was performed for both gene and transcript abundances in SRF, DCM, MES, marine epipelagic mixed layer (MIX), and marine water layer (ZZZ) samples. As in previous analyses, only sequences that clustered with members of the same family were included. The analysis was based on a Bray–Curtis dissimilarity matrix after Wisconsin normalization and square root transformation of the abundance matrix. Environmental variables were fit onto ordinations using the data obtained in situ during the *Tara* Oceans expedition, in order to evaluate which variables significantly correlated with community patterns. Correlations between individual environmental variables and abundances of specific sequence groups (clusters or subclusters of an SSN) were also performed. The analyses were carried out using stats v. 4.5.0, dplyr v. 1.1.4[41], and vegan v. 2.6.8[42], in R environment version 4.3.3[43].

Results

Sequence identification

We searched for putative alginate lyase sequences in the prokaryote-enriched reference catalog, including 47 million non-redundant protein-coding sequences created from the dataset of the *Tara* Oceans expeditions (2009–2013)[23]. This subset includes the 0.22–3 µm size fraction (mostly free-living microorganisms) from epipelagic and mesopelagic communities. The HMMs built from sequences of the 15 PL families that include alginate-

depolymerizing enzymes identified 2734 sequences with a cutoff E-value of 10^{-10} (Table 1). The relatively high cutoff value was chosen due to the high proportion of sequences identified at E-values between 10^{-10} and 10^{-15} in some of the families (Supplementary Figs. S1, S2, S3, S4), suggesting the identification of highly divergent sequences.

Sequence Similarity Networks (SSNs) were used to assess the specificity of the identifications by analyzing their relationship with members of the respective families from the CAZy database. The 15 targeted PL families can be divided into four fold groups: $(\alpha/\alpha)_n$ barrel, parallel β-helix, β-jelly roll, and $(\alpha/\alpha)_6$ toroid + anti-parallel β-sheet (Table 1). The families from each fold group were analyzed in the same SSN, in order to assess potential cross-identifications. After an initial all-by-all blastp analysis of the sequences, the results were visualized as a network where a node (circle) represents each sequence and the edges (lines) connect the nodes when sequence similarity surpasses a user-defined threshold[31]. As expected, high SSN threshold values resulted in clusters of the network that included CAZy members of more than one family, which separated when increasing the stringency by lowering the threshold value. The selected cutoff was the highest threshold value that resulted in the separation of members of the CAZy database from different families into separate clusters. In the four SSNs, we evaluated the fraction of sequences identified with each HMM that clustered with CAZymes from the same family (Table 1).

Sequence and structural diversity

PL5 and PL38 families

The distantly related PL5 and PL38 families present an $(\alpha/\alpha)_n$ barrel fold, with n having values of 6 and 7, respectively[44,45]. The PL5 family mostly includes M-specific alginate lyases, and the more recently created PL38 family includes enzymes with alginate and polyglucuronic acid as substrates[13]. The E-value distribution of the identification was very different in the two families, as PL5 homolog sequences had predominantly high e-values (10^{-15}–10^{-10}), while the alignments of PL38 homolog sequences included both high and low E-values (Supplementary Fig. S1). These results suggest a closer evolutionary distance between family members and their homologs in the PL38 family with respect to the PL5 family.

An SSN was built including the catalytic modules of members from both families and their homologs. CAZy members of the PL5 and PL38 families formed distinct clusters with thresholds ≤ 10^{-21}. At a 10^{-21} threshold value, the SSN contained three clusters, one including PL38 and PL38 homolog sequences, one including PL5 and PL5 homologs, and one containing only PL5 homolog sequences (Fig. 1a). PL38 homologs presented a high taxonomic diversity, including sequences assigned to Bacteria and Eukaryota. Bacteria mostly included classes of the Bacteroidota, Pseudomonadota, and Verrucomicrobiota phyla, although more than 40% of the sequences were not assigned at the phylum level (Fig. 1b). Sequences assigned to Eukaryota included the Phaeocystales (*Phaeocystis antarctica*), the Cyathomonadacea (*Goniomonas pacifica*) and the Acanthoecida (*Acanthoeca*) orders, as well as the Chlorophyta phylum.

Most PL5 homologs clustered with members of the PL5 family, despite the relatively high E-values obtained in their identification (Fig. 1a, Table 1). In cluster 2, the larger subcluster (named 2a) contained most PL5 members, including the PL5 subfamily 1 (PL5_SF1) sequences, but only 15% of the identified PL5 homologs. The metagenomic sequences in subcluster 2a were mostly assigned to various orders of Gammaproteobacteria (e.g., Oceanospirillales and Pseudomonadales) and Alphaproteobacteria (e.g., Caulobacterales, Sphingomonadales, and Rhodobacterales) (Fig. 1b). Three sequences from this subcluster were also identified with the PL38 HMM, although with higher E-values (PL5 HMM 10^{-64}–10^{-77}; PL38 HMM 10^{-11}–10^{-12}). Subcluster 2b, on the other hand, included members of the PL5 family without an assigned subfamily (PL5_NSF) from strains of the *Agrobacterium*, *Rhizobium*, *Jeongeupella*, and *Microvirga* genera, as well as 75% of the metagenomic sequences (Fig. 1a). The majority of the metagenomic sequences of subcluster 2b (97.1%) were assigned to '*Candidatus* (*Ca.*) Pelagibacter' of the Alphaproteobacteria class (Fig. 1b). Cluster 3 of the SSN only contained PL5 homolog sequences, which were also assigned either to *'Ca.* Pelagibacter' (82.4%) or only to the class level to Alphaproteobacteria (17.6%).

Metagenomic sequences from the different clusters and subclusters were modeled and compared with PDB structures (Supplementary Fig. S1). Z-score values showed that the models of the metagenomic sequences were more similar to structures from the same PL family in cluster 1 and subcluster 2a (47.3 and 46, respectively) than in subcluster 2b and cluster 3 (32.6 and 24.2, respectively). In a pairwise structure comparison analysis, a Z-score value above 20 indicates a high probability of a common ancestor[46]. The amino acid in the position of the highly conserved active site residue His192 (A1-III from *Sphingomonas* sp. A1 numbering[44]) varied in some of the PL5 homolog sequences (encircled in Fig. 1a). All sequences from cluster 3 presented a Lys residue, and 10 metagenomic sequences from subcluster 2a contained a Gln residue (Supplementary Fig. S1).

PL6, PL31, PL41, and PL44 families

The parallel β-helix fold has been observed in several PL families[47], four of them including alginate-degrading enzymes: the well-characterized PL6 family[48] and the more recently created PL31[49], PL41[9], and PL44[16] families. In the case of the families without known structures, the characterized enzymes were modeled. The models of the alginate lyase SjAly from *Saccharina japonica* (PL41, NCBI Acc. Numb. BBH10631) and Aly44A from *Bacillus* sp. FJAT-50079 (PL44, NCBI Acc. Numb. WP_213137828) showed similarities with the structures from members of the PL6 and PL31 families, with Z-score values ≥ 24.7 (Supplementary Fig. S5). These results support including these families in the same SSN as PL6 and PL31 members.

The identification of PL6 homologs showed a wide E-value distribution, while the PL31, PL41, and PL44 families mostly presented relatively high E-values indicating a higher divergence with sequences from these families from the CAZy database (Supplementary Fig. S2). When analyzed in an SSN, the catalytic modules of members of the four families separated into distinct clusters with threshold values ≤ 10^{-26}, with PL31 and

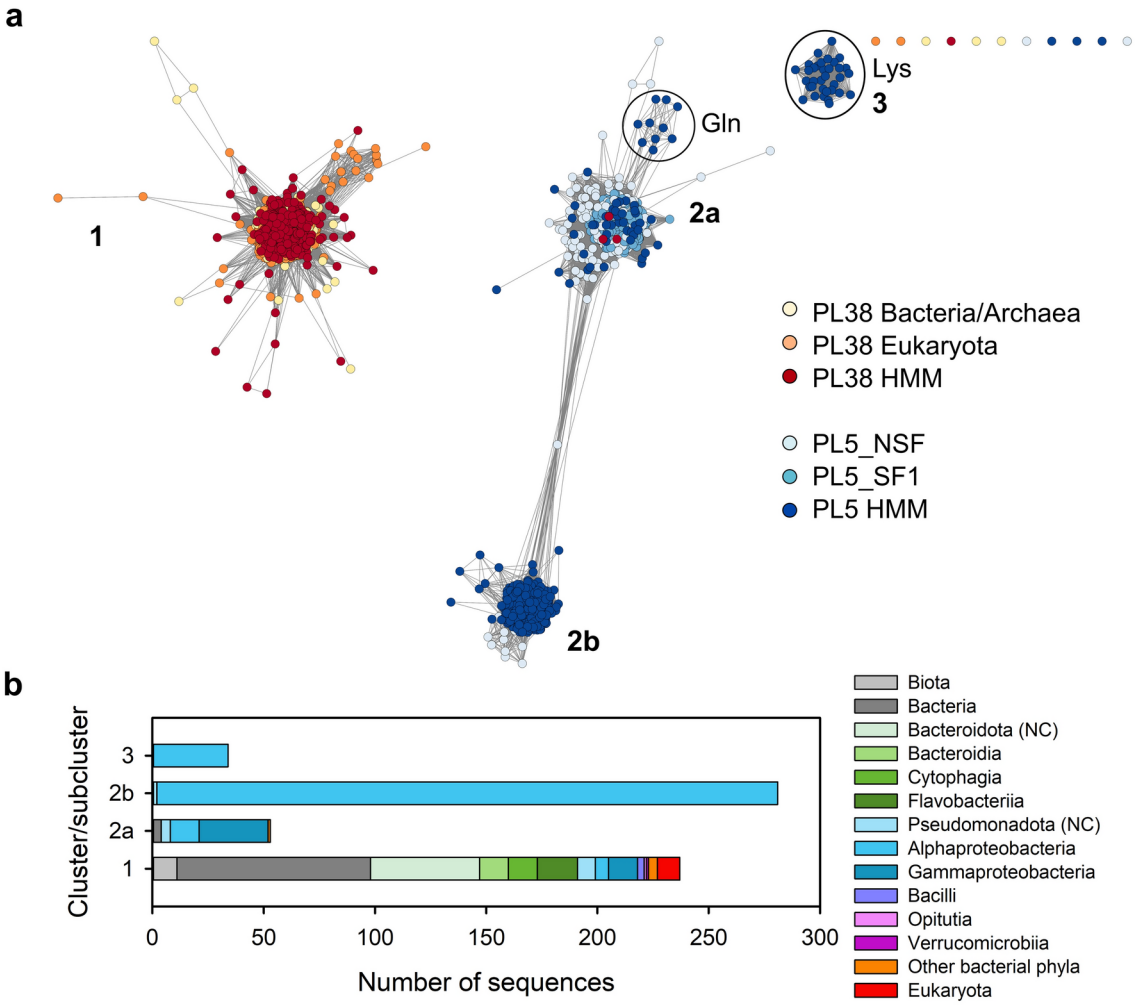

Fig. 1. Diversity of PL5 and PL38 homologs. (**a**) Sequence Similarity Network (SSN) constructed from the catalytic module of PL5 and PL38 members of the CAZy database and homolog sequences identified in the OM-RGC_v2_metaG dataset in the Ocean Gene Atlas 2.0 web server (OGA)[24] using HMM of these families from the V.12 dbCAN database[25]. The SSN was constructed using a threshold value of 10^{-21} in the Enzyme Similarity Tool (EFI-EST)[31], and visualized in Cytoscape[34]. The circles group sequences with a variation in the residue located in the His192 active site position (A1-III from *Sphingomonas* sp. A1 numbering[44]). More information can be found in Supplementary Fig. S1. (**b**) Taxonomic classification of the metagenomic sequences included in each cluster or subcluster, as assigned in OGA[24]. NC, not assigned at the class level; SF, subfamily.

PL41 members showing the closest relationship. In an SSN with a threshold value of 10^{-26}, only PL6 homolog sequences clustered in high proportion with members of the same family (Table 1; Fig. 2a). As similar results were obtained after relaxing the threshold value to 10^{-15}, the majority of the metagenomic sequences identified with the PL31, PL41, and PL44 HMMs were highly divergent.

Most PL6 homologs clustered with members of the PL6 family without an assigned subfamily (PL6_NSF) and from subfamilies 1 and 2 (PL6_SF1, PL6_SF2) (Fig. 2a, subcluster 1a). The metagenomic sequences from subcluster 2a were assigned to various taxa, mostly from the Flavobacteriia and Gammaproteobacteria classes (66% from the Alteromonadales order) (Fig. 2b). The model of a metagenomic sequence from subcluster 1a showed a high Z-score when compared to the closest structure, the alginate lyase BcAlyPL6 from *Bacteroides clarus* (PL6_SF1; PDB 7dmk, Z-score 47.4, Supplementary Fig. S2). The modeled sequence, and most metagenomic sequences in the subcluster, presented the highly conserved Lys and Arg residues in the positions of the proposed catalytic residues of this enzyme[50]. A smaller group of PL6 homologs, all assigned to the Alteromonadales order, clustered with PL6_SF3 members (Fig. 2a,b, subcluster 1b). As some members of this subfamily[51], a model of a metagenomic sequence from subcluster 1b presented Gln and Glu residues in the positions of the catalytic Lys and Arg residues of the structure, respectively (Supplementary Fig. S2). The Gln residue was fully conserved in the alignment of the sequences of subcluster 1b, while the Glu position was mostly conserved (with some sequences containing an Arg residue).

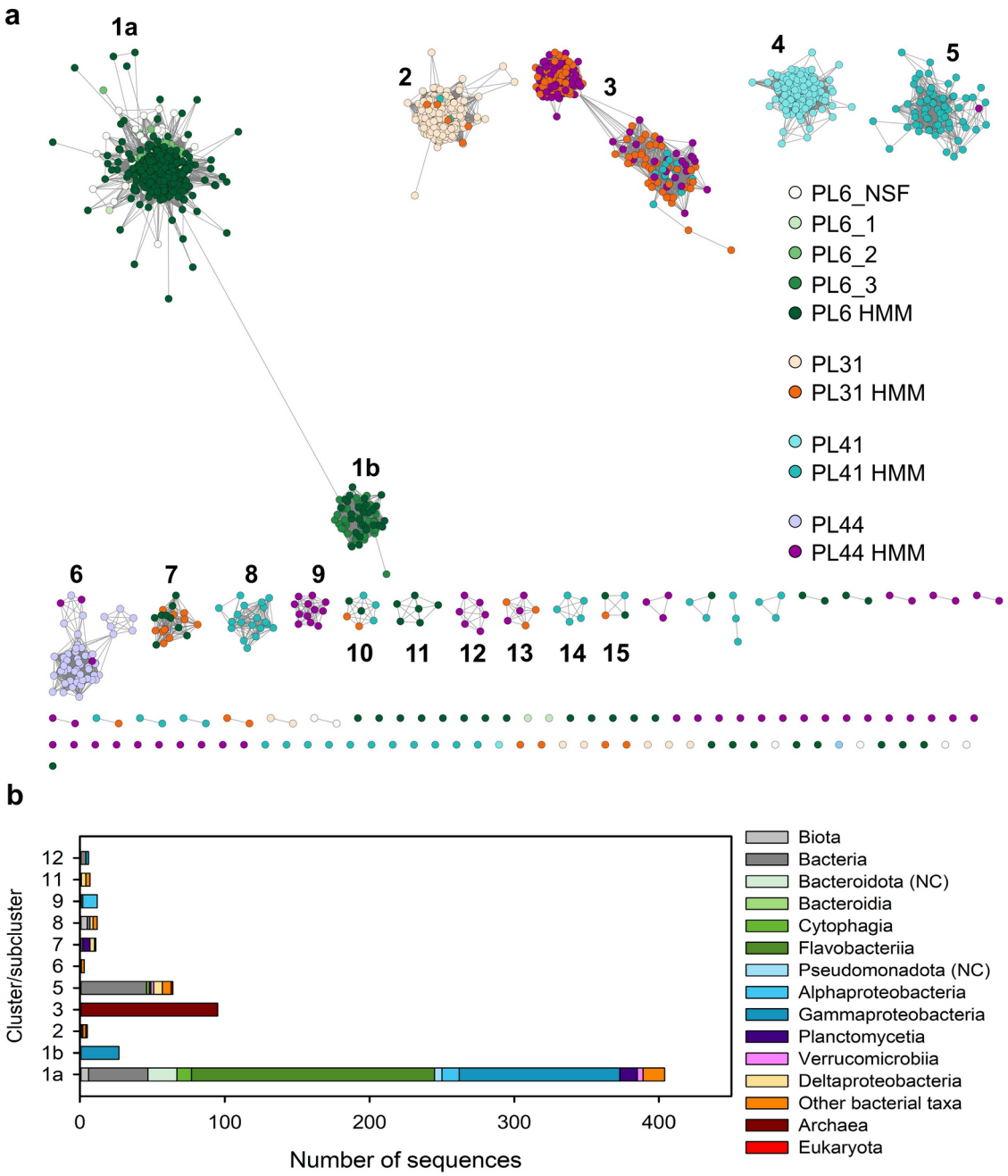

Fig. 2. Diversity of PL6, PL31, PL41, and PL44 homologs. (**a**) SSN of the catalytic modules of members of PL6, PL31, PL41, and PL44 families of the CAZy database, and homologs identified in the OM-RGC_v2_metaG dataset in OGA[24] using HMM of these families from the V.12 dbCAN database[25], with the exception of the HMM of the PL44 family that was created from sequences from the CAZy database. The SSN was constructed using a threshold value of 10^{-26}. PL6_1, PL6_2, PL6_3, and PL6_NSF correspond to the subfamilies 1, 2, and 3 of the PL6 family and sequences not assigned to a subfamily, respectively. More information can be found in Supplementary Fig. S2. (**b**) Taxonomic assignment of the metagenomic sequences included in each cluster. NC, not assigned at the class level.

Only 4% of the PL31 homologs identified in the metagenomes clustered with members of the PL31 family, located in cluster 2 of the SSN (Table 1, Fig. 2a). These sequences were assigned to various taxa, including Flavobacteriia and Gammaproteobacteria (*Aquimarina* and *Cellvibrio*, respectively). Two of the PL31 homolog sequences were also identified with the HMM of the PL41 family, although with higher E-values (PL31 HMM: 10^{-71} and 10^{-54}; PL41 HMM: 10^{-12} and 10^{-23}). The overall fold of a PL31 homolog from cluster 2 of the SSN was similar to the only determined structure of the family, the alginate lyase PsAly from *Paenibacillus* sp. FPU-7 (PDB 6kfn[52], Z-score 45.1). The model showed conserved residues in the positions of the proposed catalytic

residues (Tyr184 and Lys221, PsAly numbering, Supplementary Fig. S2). Members of the PL44 family formed cluster 6 of the SSN (Fig. 2a), which included only three metagenomic sequences assigned to *Cellvibrio* and *Mycobacterium* genera. A model of a sequence from this cluster showed close structural similarities with the model of Aly44A from *Bacillus* sp. FJAT-50079 (Z-score 54.7, Supplementary Fig. S2).

The rest of the clusters only included metagenomic sequences, often identified with more than one HMM. The majority of the sequences identified with the PL31 HMM formed various clusters (e.g., clusters 3, 7, 10, and 13, Fig. 2a). Cluster 3 included sequences from Euryarchaeota, which presented a complex multimodular architecture with various parallel β-helix domains. Sequences from cluster 7 (most assigned to the planctomycete *Rhodopirellula*) were also identified with the HMM of the PL6 family. A model of a sequence from this cluster showed similar Z-scores when compared with the structures from members of the PL31 (PDB 6kfn) and PL6 (PDB 7dmk) families (Supplementary Fig. S2). PL41 homolog sequences were mostly included in clusters 5 and 8, and PL44 homolog sequences were included in clusters 9 and 12 (Fig. 2a). Models of sequences from clusters 9 and 12 showed distant relationships with Aly44A (Supplementary Fig. S2). Cluster 11 only contained PL6 homolog sequences, although a model from this cluster was closer to the structure from the PL31 family (Supplementary Fig. S2). Overall, the HMMs of the most recently created families identified a low number of sequences that clustered with members of these families.

PL7, PL14, PL18, and PL36 families

These families include enzymes with a catalytic module with a β-jelly roll fold, the majority of them with alginate lyase (EC 4.2.2.3, EC 4.2.2.11) or oligoalginate lyase (EC 4.2.2.26) activities[53–55]. The other activity identified in members of these families is endo-β-1,4-glucuronan lyase (EC 4.2.2.14), so far detected in enzymes of PL7_SF4, PL14_SF1, and PL14_SF5. In an SSN built with the catalytic modules of the members of these families and their homologs, PL14 members separated from PL36 members with an E-value cutoff of 10^{-25}, while PL18 members separated from a cluster containing most of the PL7 members at an E-value cutoff of 10^{-31}. At this cutoff value, however, the cluster with PL18 members also included a small group of sequences of the PL7 family without an assigned subfamily, forming a distinct subcluster (NCBI Acc. Numb. WPO37025, QWG10357, QWG02653, QWG02648, AZQ63727, ANQ49913, ANQ49908, BDD00988, BDD00985, BDD00931, BCD95952, and WNJ19702; Fig. 3a, cluster 8). PL7 and PL18 sequences from cluster 8 only split at an e-value cutoff of 10^{-72}, suggesting that these PL7 members are closer to PL18 members than to the rest of the PL7 family.

Figure 3a shows the SSN constructed using a cutoff of 10^{-31}. Most of the sequences from PL7 family members were located in cluster 2, which presented a large subcluster with more than half of the identified PL7 homologs and members of the multispecific PL7_SF5 subfamily (subcluster 2a). These metagenomic sequences were mostly assigned to the classes Flavobacteriia (74% Flavobacteriales) and Gammaproteobacteria (69% Alteromonadales) (Fig. 3b). Subcluster 2b included sequences from PL7_SF1 and PL7_SF4, which separated from PL7_NSF members (Fig. 3a). Only nine sequences clustered with PL7_SF1 members, mostly assigned to the *Pseudomonas* genus, and only one sequence (assigned to Ascomycota) clustered with members of the PL7_SF4 subfamily, which thus far only includes glucuronan lyases. Cluster 4 contained most of the remaining PL7 homologs (mostly assigned to the Flavobacteriaceae family) and PL7_NSF members. None of the metagenomic sequences clustered with PL7_SF2 sequences (cluster 3), and only 12 metagenomic sequences clustered with PL7_SF3 members (cluster 5). Clusters 4, 13, and 14 included both PL7_NSF members and PL7 homologs, and clusters 11 and 12 included PL7_NSF sequences and some PL7_SF5 and PL7_SF3 members, respectively. These metagenomic sequences were mostly assigned to the Flavobacteriia class (Fig. 3b).

Three-dimensional models of the PL7 homolog sequences from the different clusters showed a β-jelly roll fold similar to structures from members of the PL7 family (Supplementary Fig. S3). More than 90% of PL7 homolog sequences had the highly conserved QIH residues previously reported to be related to a preference for polyG substrates, while less than 3% presented the QVH motif observed in enzymes with a preference for polyM[53,56]. The PL7 homolog sequence assigned to Ascomycota (subcluster 2b) presented a Lys residue in the position of the conserved His (Supplementary Fig. S3). The position of the Tyr residue acting as a general acid (Y541 in AlyQ from *Persicobacter* sp. CCB-QB2, PDB 5xnr) was conserved in 99.4% of the metagenomic sequences where this region of the protein was available in the alignment.

Cluster 1 of the SSN included sequences from PL14 members from SF1, SF3, SF4, SF5, and NSF, as well as the majority of the sequences identified with the HMM of the PL14 family (Fig. 3a). The taxonomic assignment of these sequences included both Eukaryota and Bacteria, with 10 bacterial classes (Fig. 3b). Seventy-two percent of the PL14 homolog sequences from cluster 1 were also identified with the HMM of the PL36 family, although with higher E-values ($\geq 3\times10^{-24}$ for PL36, $\leq 6\times10^{-25}$ for PL14). In fact, the models of sequences located in different subclusters of cluster 1 showed similarities with structures from members of both families (Supplementary Fig. S3). Similarly, the only PL36 homolog that clustered with members of the PL36 family was identified with both PL36 (7×10^{-43}) and PL14 (5×10^{-38}) HMMs with significant E-values (Fig. 3a, cluster 7), although the model showed a higher Z-score with the structure of the PL36 family (Supplementary Fig. S3). Discriminating between PL14 and PL36 homolog sequences was therefore complex in this dataset. Two additional clusters contained only PL14 homolog sequences not identified with the PL36 HMM, assigned to both Bacteria and Eukaryota (clusters 18 and 21; Fig. 3).

Members of the PL18 family and all the identified PL18 homolog sequences were located in cluster 8 of the SSN (Fig. 3a). The majority of the metagenomic sequences were assigned to the Gammaproteobacteria class (Fig. 3b), with 73% of the sequences assigned to the Alteromonadales order. The model of a PL18 homolog showed high Z-score values with the two structures from members of this family (PDB 1J1T, 36.9; PDB 4Q8K, 36.8; Supplementary Fig. S3). Furthermore, the Arg219, Gln257, His259, Tyr347, Tyr353, and Lys349 residues proposed as key for the catalysis (Aly-SJ02 numbering[57]) were conserved in all the metagenomic sequences when the information was available.

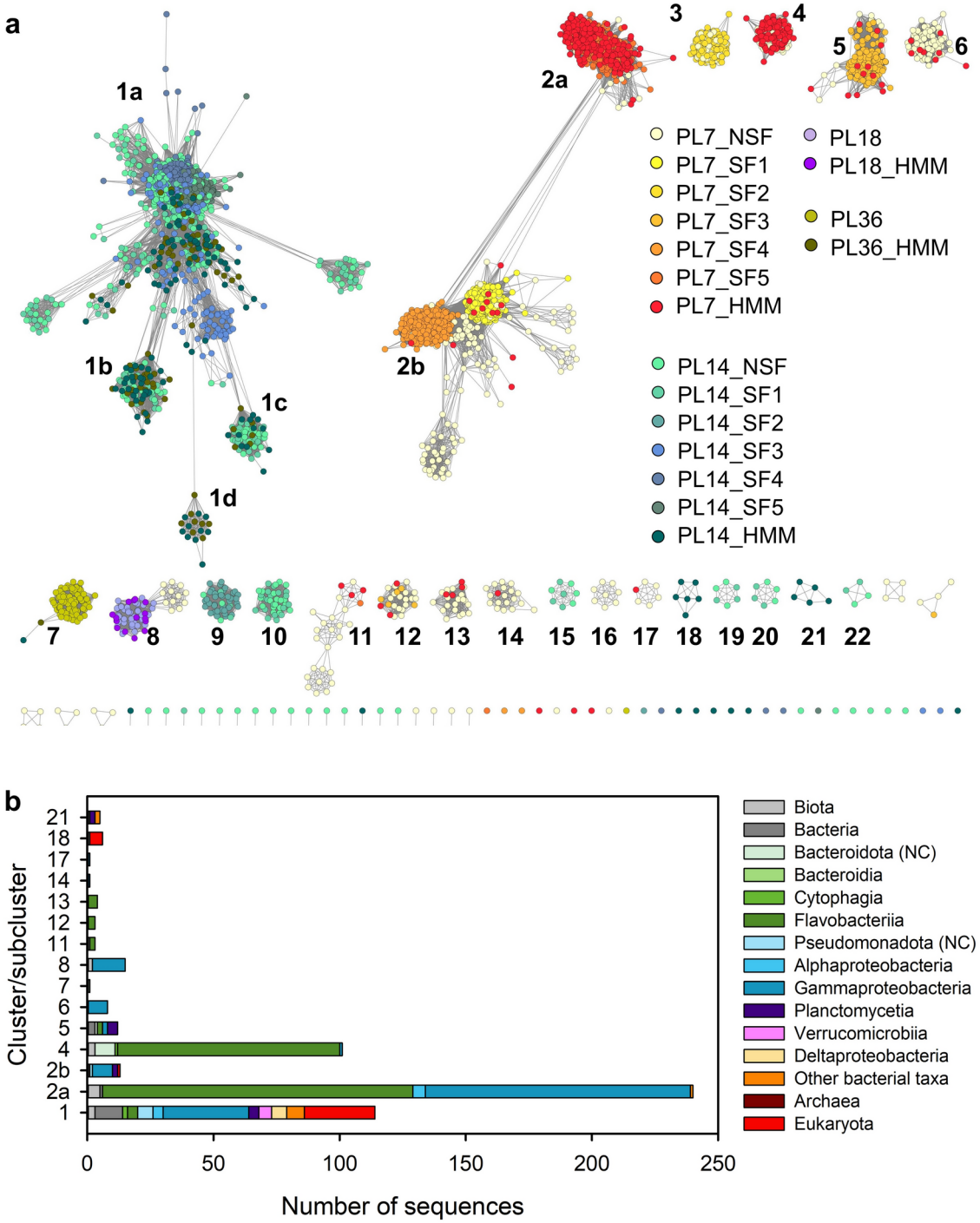

Fig. 3. Diversity of PL7, PL14, PL18, and PL36 homologs. (**a**) SSN of the catalytic modules of members of the PL7, PL14, PL18, and PL36 families of the CAZy database, and homologs identified in the OM-RGC_v2_metaG dataset in OGA[24] using HMM of these families from the V.12 dbCAN database[25]. The SSN was constructed using a threshold value of 10^{-31}. SF, subfamily; NSF, sequences not assigned to a subfamily. More information can be found in Supplementary Fig. S3. (**b**) Taxonomic assignment of the metagenomic sequences included in each cluster. NC, not assigned at the class level.

PL8, PL15, PL17, PL34, and PL39 families

The last group of PL families that include alginate depolymerizing enzymes presents a $(\alpha/\alpha)_6$ toroid + anti-parallel β-sheet fold. In SSNs, an E-value cutoff ≤ 10^{-24} separated the members of the PL39 family from a cluster containing both PL15 and PL34 family members, while a cutoff ≤ 10^{-30} was needed in order to separate the last two families. The SSN shown in Fig. 4a was constructed using a 10^{-30} threshold value. Sequences from PL8 family members were located in clusters 1 (PL8_NSF, SF1, SF3, SF4) and 3 (PL8_NSF and SF2) of the SSN,

and both clusters included metagenomic sequences. PL8 is a multi-specific family, and only one out of the 34 characterized enzymes of this family presented alginate-depolymerizing activity (PL8_SF4[58]). Sequences from cluster 1 were mostly assigned to the Bacteroidia and Cytophagia classes (Fig. 4b), and were almost exclusively connected to PL8_NSF, PL8_SF1, or PL8_SF3 members (Supplementary Fig. S4). One model, however, was closest to the structure from the M-specific alginate lyase from *Paradendryphiella salina* (PDB 7r2x, Z-score 40.5; Supplementary Fig. S4). Cluster 3 contained PL8_SF2 members and metagenomic sequences mostly assigned to the Flavobacteriia class (Fig. 4b, Supplementary Fig. S4). A model from a sequence from cluster 3 was most closely related to the structure of the chondroitin ABC lyase I from *Proteus vulgaris* (PL8_SF2; PDB 1hn0; Supplementary Fig. S4). These results suggest that the majority of the PL8 homologs identified in the dataset were not related to alginate degradation processes.

Most PL15 homolog sequences were located in cluster 4 of the SSN, which also included sequences from PL15_NSF, PL15_SF1, and PL15_SF2 members. Half of the metagenomic sequences were connected to PL15_SF2 members (exo-heparin lyases, EC 4.2.2.-) in an SSN with a 10^{-50} threshold value (Supplementary Fig. S4). Most models of sequences from this cluster showed higher Z-score values when compared to a structure of a PL15_SF2 enzyme (PDB 6lja; Z-score 40.1–47.1), and only one to the structure of an oligoalginate lyase from the PL15_SF1 (PDB 3afl; Z-score 56.1; Supplementary Fig. S4). These results suggest that the majority of the

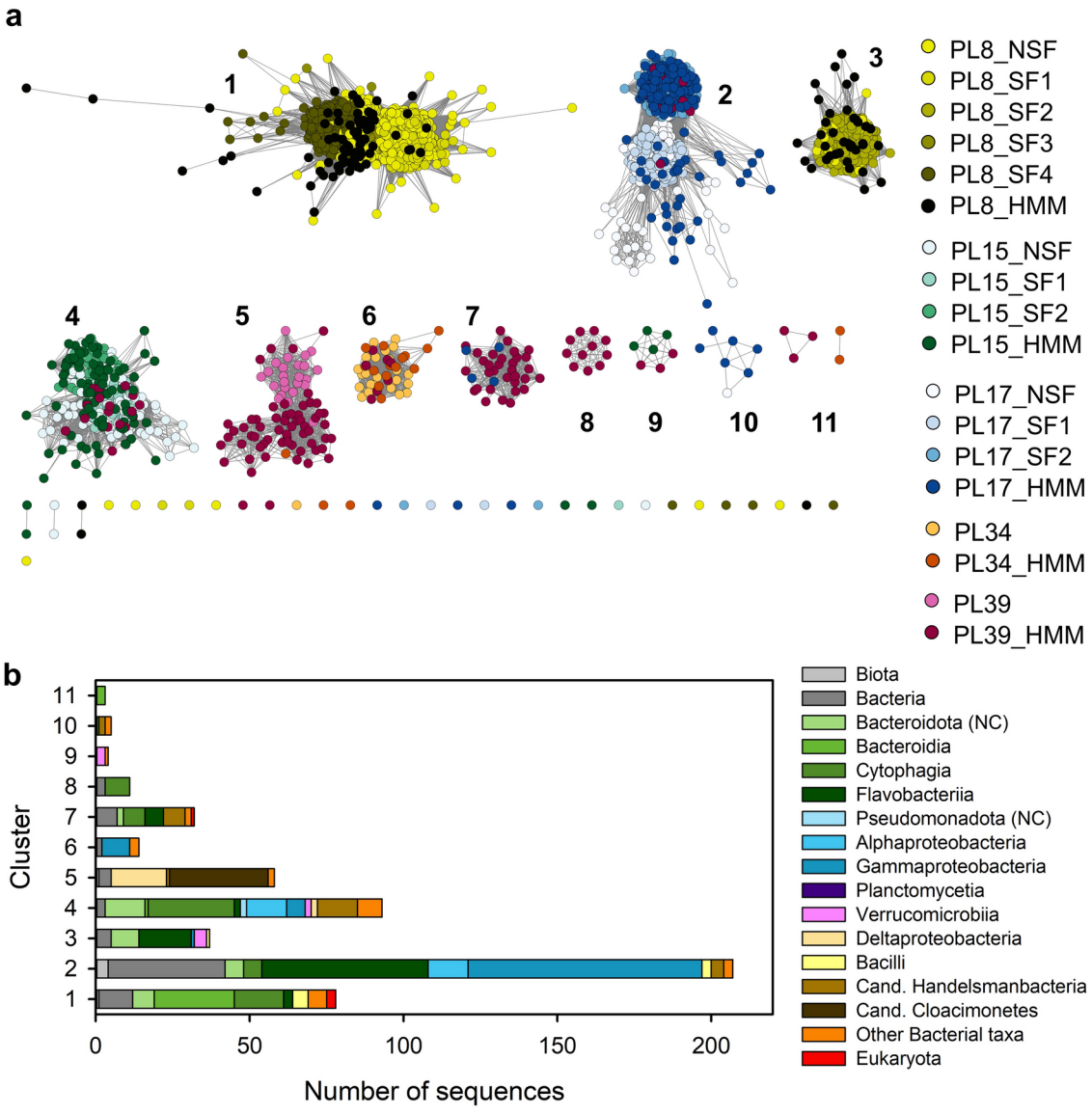

Fig. 4. Diversity of PL8, PL15, PL17, PL34, and PL39 homologs. (**a**) SSN of the catalytic modules of members of PL8, PL15, PL17, PL34, and PL39 families of the CAZy database and homologs identified in the OM-RGC_v2_metaG dataset in OGA[24] using HMM of these families from the V.12 dbCAN database[25]. The SSN was constructed using a threshold value of 10^{-30}. SF, subfamily; NSF, sequences not assigned to a subfamily. More information can be found in Supplementary Fig. S4. (**b**) Taxonomic assignment of the metagenomic sequences included in each cluster. NC, not assigned at the class level.

PL15 homolog sequences do not encode alginate depolymerizing enzymes. Almost 20% of the PL15 homolog sequences were also identified with the PL39 HMM, although almost exclusively with higher E-values (10^{-47}–10^{-12} for PL15, 10^{-28}–10^{-11} for PL39; Fig. 4a and Supplementary Fig. S4). Furthermore, Z-score values of the four analyzed models were lower when compared with the structure of a PL39 member than a PL15 member (Supplementary Fig. S4).

Cluster 2 of the SSN included most members of the PL17 family and PL17 homolog sequences (Fig. 4a). The model of a metagenomic sequence from cluster 2 had a Z-score of 36.1 in a pairwise comparison with the structure from Alya3 of *Zobellia galactanivorans* DsijT (PDB 7bjt; Supplementary Fig. S4), with a preference for cleaving oligomannuronate[59]. Most characterized members of the PL17 family have alginate lyase or oligoalginate lyase activities, and only two PL17_NSF members have substrates other than alginate, although to our knowledge no member of the SF1 has been characterized to date[60]. When the cluster was analyzed in an SSN built with a more restrictive E-value threshold (10^{-50}), almost 80% of the metagenomic sequences (mostly assigned to Gammaproteobacteria or Flavobacteriia) were located in a subcluster containing PL17_SF2 and PL17_NSF members (Supplementary Fig. S4). The rest of the metagenomic sequences clustered with PL17_SF1 and PL17_NSF members and were mostly assigned to the domain level or as *Ca.* phyla.

The PL34 family, with only 28 members, includes one characterized enzyme, the endo-alginate lyase OPIT5_18480 from Opitutaceae bacterium TAV5[49]. PL34 members and PL34 homologs were included in cluster 6 of the SSN (most sequences assigned to Gammaproteobacteria, mostly Alteromonadales, Fig. 4). Half of the PL34 homolog sequences were also identified with the PL39 HMM, although with higher E-values ($> 10^{-18}$ for PL39, $< 10^{-23}$ for PL34). A comparison of the model of one of the most divergent PL34 homolog sequences and the model of OPIT5_18480 (AlphaFold identifier AF-W0J4H8-F1) had a Z-score value of 32 (Supplementary Fig. S4).

Sequences identified with the PL39 HMM were included in eight clusters of the SSN, often with sequences identified with other HMMs (Fig. 4a). This could be due to the relationship of the PL39 family to several other PL families[61]. Only cluster 5 included PL39 members, and 40% of the PL39 homolog sequences (HMM alignment E-values 10^{-171}–10^{-11}; Fig. 4a). These sequences were mostly assigned to Deltaproteobacteria and to *'Ca.* Cloacimonetes' (Fig. 4b). Models of sequences from this cluster had the highest Z-score with the only known structure of the PL39 family (Supplementary Fig. S4). Cluster 7, on the other hand, included PL39 homolog sequences (HMM alignment E-values 10^{-20}–10^{-11}) and sequences identified with the PL17 HMM (10^{-12}–10^{-11}). Models from cluster 7 were more closely related to a structure from PL12_SF2 (4fnv[62]; Supplementary Fig. S4), suggesting unspecific identifications. PL39 homolog sequences from cluster 8 (HMM alignment E-values 10^{-25}–10^{-12}) were assigned to the Cytophagia class of Bacteroidota or to Bacteria (Fig. 4b). Models of sequences from this cluster presented Z-score values between 25.9 and 28.2 in pairwise alignments with the structure of the alginate lyase from the PL39 family (PDB 6jp4), but similar values with structures from members of other families (Supplementary Fig. S4). Therefore, the affiliation of these sequences is not clear.

Relative abundance of genes and transcripts

We selected the homologs that clustered with members of the same family in the SSNs (Figs. 1, 2, 3 and 4) in order to assess the abundance and biogeographic distribution of the genes and transcripts in the epipelagic and mesopelagic picoplanktonic communities. The PL31, P36, PL41, and PL44 families were excluded, since very few homolog sequences clustered with members of these families (Table 1). The relative gene and transcript abundances were analyzed in samples of the SRF, DCM, and MES depth layers, where most data points were available. Homolog genes from all the analyzed families were detected at almost all geographic regions and depths, indicating a rather widespread distribution of organisms carrying these genes (Supplementary Table S1). Homologs of the PL5, PL6, PL7, PL17, and PL38 families presented the highest relative gene abundances, and transcript levels were also generally higher in these families. However, transcript rarely exceeded gene abundances, suggesting low gene expression levels (Supplementary Tables S1 and S2). In addition, transcript levels showed a higher variability with respect to relative gene abundances.

The relative gene abundance of PL5 homologs included in both subclusters 2a and 2b (Fig. 1a and Supplementary Table S1) was significantly higher in metagenomes from the epipelagic layer (SRF and DCM) when compared to those from the MES zone (Kruskal–Wallis rank sum test followed by contrasts, p-value SRF-MES $= 3 \times 10^{-14}$, p-value DCM-MES $= 7 \times 10^{-8}$, p-value DCM-SRF $= 0.76$). Transcript abundances were also higher at SRF and DCM depths, but the difference was less prominent (Supplementary Table S2). In the SSN, PL5 homologs showed a clustering pattern that separated sequences assigned to different taxa, as well as variants in the highly conserved His position of the active site (Fig. 1 and Supplementary Fig. S1). We analyzed the relative gene and transcript abundances of sequences included in subcluster 2a (containing Gln or His residues), subcluster 2b, and cluster 3 (Lys variant) separately (Supplementary Tables S3 and S4). Although PL5 homologs from subcluster 2a containing the archetypical His residue presented low gene abundances, transcript levels were relatively high, in particular at higher depths, indicating high expression levels. The sequences with the highest transcript abundances were assigned to the *Halomonas* (ID 003006528, 022558976) and *Pseudomonas* (ID 005486491, 005133091, 005908986) genera. Sequence 003006528, which showed the highest transcript levels, shared 100% protein identity with a putative alginate lyase from *Halomonas* sp. S2151 (IMG/M gene ID 2712496321). As both alg8 and alg44 genes are present in the genome of this organism (IMG/M gene IDs 2712496323 and 2712496324; mannuronan synthases, EC:2.4.1.33), these putative alginate lyases are probably playing a role in exopolysaccharide biosynthesis. Similarly, sequence 005486491 shared 100% identity with an alginate lyase family protein identified in the genome of *Pseudomonas* sp. 5Ae-yellow (NCBI Acc. Numb. WP_182248321), which also carries genes for alginate biosynthesis (e.g., NCBI Acc. Numb. WP_310734963, WP_259364809, and WP_259364810). Sequences from subcluster 2a with a Gln residue in the conserved His position of the active site showed low gene and transcript abundances (Supplementary Tables S3 and S4).

Subcluster 2b sequences, with an archetypical His residue in the active site, were the major contributors to the high relative gene and transcript abundances of PL5 homolog sequences, although cluster 3 sequences (containing a Lys residue) were also detected in metagenomes and metatranscriptomes (Supplementary Tables S3 and S4). These sequences were mostly assigned to *'Ca.* Pelagibacter', and blastp searches showed that the closest matches in GenBank were putative alginate lyases from genomes of these organisms in 89.3% of the sequences from subcluster 2b and 59% of the sequences from cluster 3. To gain insight into the prevalence of PL5 homolog genes in these organisms, we searched 2656 genomes from *Pelagibacter* and Pelagibacteraceae bacterium deposited at the IMG/M system for sequences with a PF05426 (Alginate Lyase) domain. Only 266 genomes contained sequences with this domain, indicating that they are not common in Pelagibacterales. Most sequences were identified in genomes of single-cell analyses and of *'Ca.* Pelagibacter sp.' strains RS39 and IMCC9063, and *Pelagibacter ubique* strains HTCC7211 and HTCC7214. An SSN of the PL5 family including the newly identified genomic sequences showed that 83% of the genomic sequences grouped with those from subcluster 2b, while 8% grouped with cluster 3 sequences (threshold 10^{-21}, Supplementary Fig. S6). Most genomes contained either subcluster 2b or cluster 3 sequences, with the exception of three genomes containing both genes located in close proximity (Supplementary Fig. S6). No other genes related to either alginate degradation or biosynthesis could be identified in the genomes of these organisms, which precluded inferring a potential physiological role of these enzymes. Overall, these results suggest that the PL5 putative enzymes could be involved in different physiological processes, including alginate biosynthesis.

Besides the PL5 family, the highest gene and transcript abundances were observed in PL6, PL7, and PL17 homolog sequences, showing a similar pattern across depths and regions (Supplementary Tables S1 and S2). The relative gene abundances of PL6, PL7, and PL17 families were analyzed in the context of their taxonomic composition. Flavobacteriia and/or Gammaproteobacteria dominated at most depths and geographic regions with few exceptions (Fig. 5). Gene abundances of PL6 homologs were relatively high in the epipelagic layer of the Arctic Ocean (Supplementary Table S1), in particular of sequences very similar (>97% identity) to sequences from genomes of uncultured Flavobacteriia and Gammaproteobacteria, such as uncultured Polaribacter (e.g., ID 001348749, 000429872, 000554781) and Porticoccaceae bacterium (ID 000606555, 000736871, 000588545). On the other hand, the most abundant genes in temperate geographic regions corresponded to homologs sharing >99% identity with sequences from *Alteromonas australica* (e.g., ID 000660703, 000322103, and 000628968), or 85% identity with a sequence from a Metagenome Assembled Genome (MAG) from a Flavobacteriaceae bacterium (ID 000513847). PL6 homolog sequences assigned to *A. australica*, which shared <80% identity with putative alginate lyases from *Alteromonas macleodii*, also showed the highest transcript levels, with 37% of the total relative transcript abundance of the PL6 family. PL7 homolog sequences assigned to *Alteromonas australica* (ID 001123987, 001894804, 005116300) were also very abundant genes and transcripts, representing 27% of the total relative abundance for the family in the metatranscriptomes at the three considered depths. PL17 homolog sequences also included homologs closely related (>99% protein identity) to sequences from the same taxa, including *A. australica* (ID 000599442) and *Polaribacter* (ID 001740408). A similar pattern of relative gene and transcript abundances was observed at the sequence level for sequences assigned to *Polaribacter* and *A. australica* from the PL6, PL7, and PL17 families (Fig. 6).

Genes coding for PL6 homolog sequences closely related to hypothetical proteins from Verrucomicrobia (e.g., from genomes from uncultured Opitutaceae and Puniceicoccaceae) were also found in most regions and depths (Fig. 5). However, a similar pattern was not observed in the PL7 and PL17 families. A high proportion of genes coding for PL6 and PL17 homologs at the DCM layer of the Red Sea were closely related to hypothetical proteins from MAGs of Gemmatimonadota bacterium (e.g., ID 001869608, 003908987), although these sequences also shared high identity with hypothetical proteins from '*Ca.* Latescibacterota' bacterium. Genes coding for PL6 and PL17 homologs closely related to sequences from '*Ca.* Neomarinimicrobiota' (e.g., ID 000657829, 013800311, 016837349) were also detected in relatively high proportion in some layers and regions (Fig. 5).

Homolog sequences of the PL38 family showed low relative gene abundances but a high taxonomic diversity (Fig. 5 and Supplementary Table S1). The most abundant PL38 homologs assigned to bacterial taxa included sequences that shared >80% amino acid sequence identity with putative alginate lyases from Flavobacteriaceae bacterium (ID 004117946, 003699212, 004181783) or *Polaribacter filamentus* (ID 003748930). Genes coding for sequences with >96% identity with sequences from Planctomycetaceae bacterium (ID 011577659), as well as Limisphaerales bacterium (ID 004264011) and Puniceicoccaceae bacterium (ID 000426398) from the Verrucomicrobiota phylum also presented high abundances. These genes showed high relative transcript abundances in samples from the Arctic Ocean (but not from the Southern Ocean). Sequence with ID 004281798 (assigned to Bacteria), which shared 98.2% identity with a putative alginate lyase from an Opitutales bacterium (NCBI Acc. Numb. MBO25657) had the highest transcript abundances in samples from temperate regions. These results suggest the relevance of putative PL38 enzymes from members of the Planctomycetota and Verrucomicrobiota phyla in polysaccharide biodegradation processes (alginate, polyglucuronic acid, or other yet-to-be identified substrates in this multispecific family).

PL38 homolog sequences assigned to Eukaryota showed high gene abundances in samples from the epipelagic layer of the Arctic and the Southern Oceans. The Arctic Ocean samples were enriched in genes coding for PL38 homologs assigned to Cryptophyceae and Choanoflagellata, while the Southern Ocean presented higher abundance of genes assigned to Phaeocystales, in particular to *P. antarctica*. PL38 homolog sequences assigned to Eukaryota shared only 12–26% identity at the amino acid level with the only characterized PL38 enzyme from an Eukaryota, the exo-β-1,4-glucuronan lyase from *Trichoderma parareesei* (NCBI Acc. Numb. OTA06285). Remarkably, the five PL38 homolog sequences assigned to *P. antarctica* accounted for 60% of the overall transcript abundance for the family. In particular, a partial sequence (ID 008024536) showed the highest gene expression levels, with 55 ± 17 times higher relative transcript abundance than gene abundance in TARA_082, TARA_084, and TARA_085 sampling sites (South Atlantic and Southern Oceans). This sequence did not have

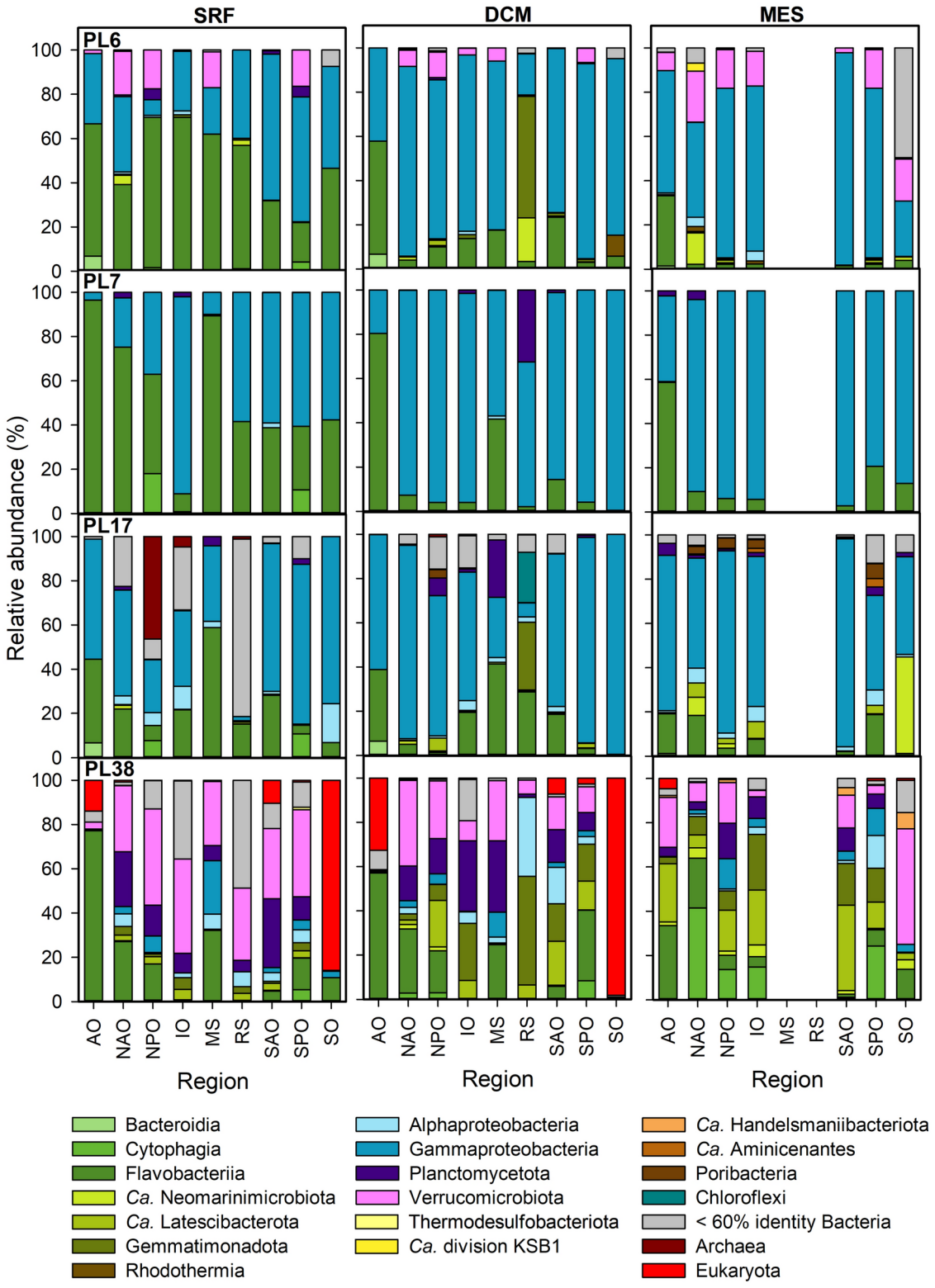

Fig. 5. Relative gene abundances of PL6, PL7, PL17, and PL38 homologs. Relative gene abundances are shown discriminated by depth layer (SRF, surface; DCM, deep chlorophyll maximum layer; MES, mesopelagic zone), geographic region (AO, Arctic Ocean; NAO, North Atlantic Ocean; NPO, North Pacific Ocean; IO, Indian Ocean; MS, Mediterranean Sea; RS, Red Sea; SAO, South Atlantic Ocean; SPO, South Pacific Ocean; SO, Southern Ocean), and taxonomic affiliation (as indicated below the graph). Sequences not classified at the class level in the OGA web server were analyzed by blastp and the taxon of the first match was considered, if identity was > 60%. When protein identity was < 60%, gene abundances were indicated as gray bars.

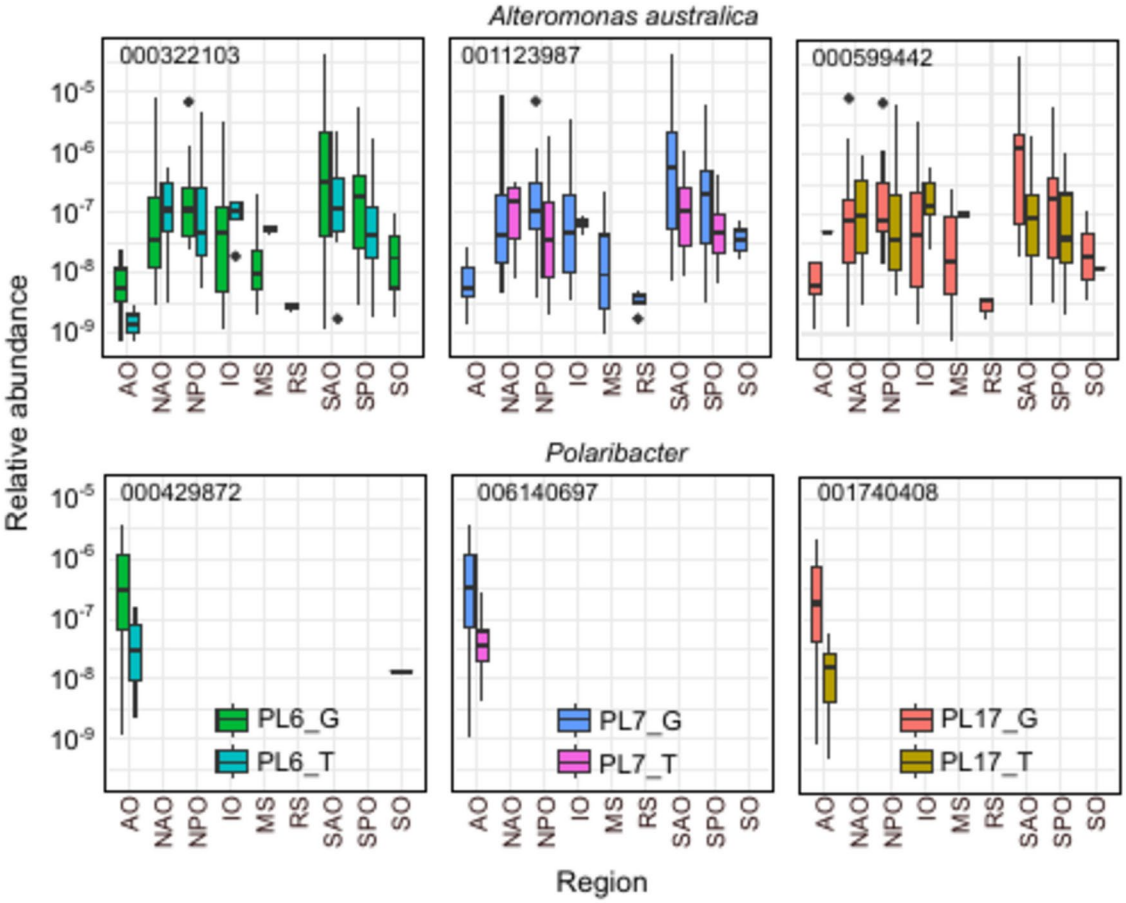

Fig. 6. Relative gene and transcript abundances of selected PL6, PL7, and PL17 homologs. Top, sequences assigned to *A. australica*; bottom, sequences assigned to *Polaribacter*. Left, PL6 homologs; center, PL7 homologs; right, PL17 homologs. Numbers indicate the sequence IDs from the OM-RGC.v2 dataset. The relative gene (G) and transcript (T) abundances were normalized as percent of mapped reads, and discriminated by the geographic region: AO, Arctic Ocean; NAO, North Atlantic Ocean; NPO, North Pacific Ocean; IO, Indian Ocean; MS, Mediterranean Sea; RS, Red Sea; SAO, South Atlantic Ocean; SPO; South Pacific Ocean; SO, Southern Ocean. The relative abundances are shown as boxplots, where horizontal lines correspond to median values, boxes to interquartile ranges and whiskers to 1.5 times the standard deviation.

close relatives in GenBank (blastp, 33% identity, 87% coverage with NCBI Acc. Numb. KAJ8613241). We searched the datasets of eukaryotic genomes and transcriptomes (SMAG and MGT) for sequences sharing high identity values with those assigned to *P. antarctica* using blastp (threshold 10^{-100}, query amino acid sequence ID 001480294). The sequences identified in these eukaryotic genomes also showed higher gene and transcript abundances in the Southern Ocean, a pattern that was observed at all size fractions (Supplementary Fig. S7). The sequences identified in the SMAG dataset, assigned to Phaeocystaceae, were closely related to the sequences assigned to *P. antarctica* and identified with the PL38 HMM (Supplementary Fig. S7). In particular, a sequence from the assembled genome TARA_AOS_82_MAG_00183 (*Phaeocystis*) showed 100% amino acid identity with sequence 008024536.

Environmental drivers of gene and transcript abundances

The environmental factors potentially driving the observed patterns in gene and transcript abundances were investigated for sequences clustering with members of the PL5, PL6, PL7, PL17, and PL38 families, which showed higher abundance levels (Supplementary Tables S1 and S2). Ordination analyses (NMDS) included samples from all depth layers: SRF, DCM, MES, MIX, and ZZZ. A clear partition was observed in the ordinations of genes of PL5 homologs, separating samples from different regions and depths (Supplementary Fig. S8). In particular, the samples from the Arctic and Southern Oceans (as well as the southernmost samples of the South Atlantic Ocean) separated from those from the warmer Mediterranean Sea and Indian Ocean over the NMDS1 axis. A strong correlation was found between the ordination pattern and temperature ($r^2 = 0.8656$, $p = 0.001$, Supplementary Fig. S8), suggesting that the distribution of the microorganisms carrying these genes is influenced by temperature. In PL5 homolog sequences, the NMDS analysis showed a partitioning of the samples from the deeper layer (MES) from those of epipelagic environments (e.g., SRF and DCM) over the NMDS2 axis, and the correlation with depth categories was significant ($r^2 = 0.2241$, $p = 0.001$). Other environmental variables showing

significant correlations include O_2, inorganic phosphate (Pi), latitude, and total C ($r^2 = 0.6714$, 0.4890, 0.4803, and 0.2685, respectively, $p = 0.001$), as well as the categories Biome, Region, and Province ($r^2 = 0.5086$, 0.5247, 0.6432, respectively), indicating a strong biogeographic signal in the observed gene abundance patterns. The transcript abundance-based ordination, in contrast, did not show a clear partition pattern. In this case, a close clustering of the samples was observed (Supplementary Fig. S8), suggesting a more homogeneous pattern of transcript abundances with less influence of environmental factors (Supplementary Table S2).

When only considering PL5 homolog sequences from subcluster 2b, a significant correlation was also found between gene abundance distributions and temperature (Spearman's $\rho = 0.34$, p-value $< 2.2 \times 10^{-16}$). Transcript abundances and gene expression levels (defined as abundance of transcripts/genes) showed significant (although weaker) correlation patterns with temperature ($\rho = 0.08$, p-value $< 2.2 \times 10^{-16}$, and $\rho = -0.17$, p-value $= 2.2 \times 10^{-16}$, respectively). Therefore, sequences assigned to '*Ca.* Pelagibacter' seem to be at least partially responsible for the general patterns displayed by PL5 homolog sequences as a function of temperature (Supplementary Fig. S8). Homologs from this subcluster closely related to sequences from '*Ca.* Pelagibacter' or Pelagibacterales bacterium genomes (87.43 ± 8.41% amino acid sequence identity, 99.51 ± 1.16% coverage) showed distinct distribution patterns of relative gene abundance as a function of temperature (Supplementary Fig. S9). In most sequences, gene abundances increased with temperature. However, in some cases, the optimum temperature range was surprisingly low.

On the other hand, significant negative correlations were found between seawater Pi concentrations and gene and transcript abundances of subcluster 2b sequences (Fig. 7a,b). In both cases, low relative abundances were found at concentrations > 2 μm l^{-1}, while high abundances and high variability were observed at lower Pi concentrations (< 0.25 μm l^{-1}). A similar result was found when the correlation was performed with the 20 sequences from cluster 3 closely related to putative alginate lyase sequences from *'Ca.* Pelagibacter' (Fig. 7c,d), but not when all the sequences in the cluster were included (p-value = 0.2932). Pi concentrations were significantly higher in MES samples than in the other layers (Kruskal–Wallis rank sum test, p-value < 0.009, Fig. 7e), where relative gene and transcript abundances of subcluster 2b sequences were lower (Supplementary Tables S3 and S4). These results suggest that the bacterial populations carrying these genes (including the novel Lys variant closely related to sequences from '*Ca.* Pelagibacter') are thriving at limiting levels of Pi.

Relative gene abundances of PL6, PL7, and PL17 homolog sequences showed a partitioning of Arctic Ocean samples (from all depth layers) from the rest of the regions on the NMDS1 axis. Similar to the PL5 family, the strongest correlation in the three families was found with temperature ($r^2 = 0.8651$, 0.6715, and 0.7072 for PL6, PL7, and PL17, respectively, Supplementary Fig. S8). The ordination of PL6 homolog gene abundances showed significant correlations with depth ($r^2 = 0.2247$, $p = 0.001$), Pi concentrations ($r^2 = 0.4861$, $p = 0.001$), and total C ($r^2 = 0.2684$, $p = 0.001$). However, while correlations with Pi and total C were observed for PL7 and PL17 gene abundances, correlations with depth were not significant (Supplementary Fig. S8). In contrast to the PL5 family, the ordination patterns of the PL6, PL7, and PL17 families were similar for gene and transcript abundances, with samples from all layers of the Arctic Ocean partitioning from the rest of the samples. The differences observed between the Arctic and the Southern Ocean suggest the presence of other environmental influences besides temperature, although a low number of samples were analyzed from the Southern Ocean.

The NMDS analysis of gene abundances of PL38 homolog sequences showed a partitioning by region, with the NMDS1 axis separating samples from all layers of the Arctic and Southern Oceans from the rest of the samples (Supplementary Fig. S8). In the case of samples from temperate regions, some level of separation by depth was observed along the NMDS2 axis, partitioning SRF samples on top, DCM samples in the center, and MES samples in the bottom of the graph. Several environmental variables showed significant correlations with the ordination based on gene abundances, including latitude, temperature, and depth (Supplementary Fig. S8). As in most other families, a strong biogeographic pattern was evident, with significant correlations by biome, region, and province. No clear pattern was observed, however, for the ordination of the samples based on transcript abundances, which was mostly influenced by the expression of genes assigned to Eukaryota at high latitudes.

Discussion

Accuracy of the sequence identification varied among families

In this study, we used HMMs for the identification of putative alginate lyase sequences in a prokaryote-enriched reference sequence catalog of the upper layers of the global ocean[23], as they represent a powerful approach for the detection of remote homologs[63]. Reaching high detection accuracy is challenging, in particular with CAZymes due to the high sequence diversity encompassed within each family[12]. For instance, annotation errors were recurrent in certain families of glycoside hydrolases using HMMs[64]. The HMMs targeting CAZy families are based on sequences mostly identified in genomes from cultured microorganisms[25], which can cause an inherent bias towards taxa with easy-to-culture members. In fact, a recent analysis of MAGs from different environments revealed an important taxonomic bias in the CAZy database, with 49% of the phyla found to contain CAZymes yet to be included in CAZy[65]. To increase the sensitivity, we used a permissive cutoff value for the sequence identification (10^{-10}), followed by an analysis of the identified sequences to assess their relationship with CAZy members. Commonly used E-value thresholds during the identification of CAZyme sequences are 10^{-15}[19,25] to 10^{-20}[65]. However, increasing the stringency in the identification step resulted in an important reduction in the number of identified sequences in several of the analyzed families, and no clear cutoff value could be defined that allowed the separation of potentially unspecific identifications from sequences clustering with members of the targeted families.

SSNs and comparative structural analyses were used to assess the relationships between the identified sequences and members of targeted families[36,38]. SSNs are useful tools to visualize sequence similarities for a large number of sequences and to assess relationships considering other lines of evidence, such as taxonomic,

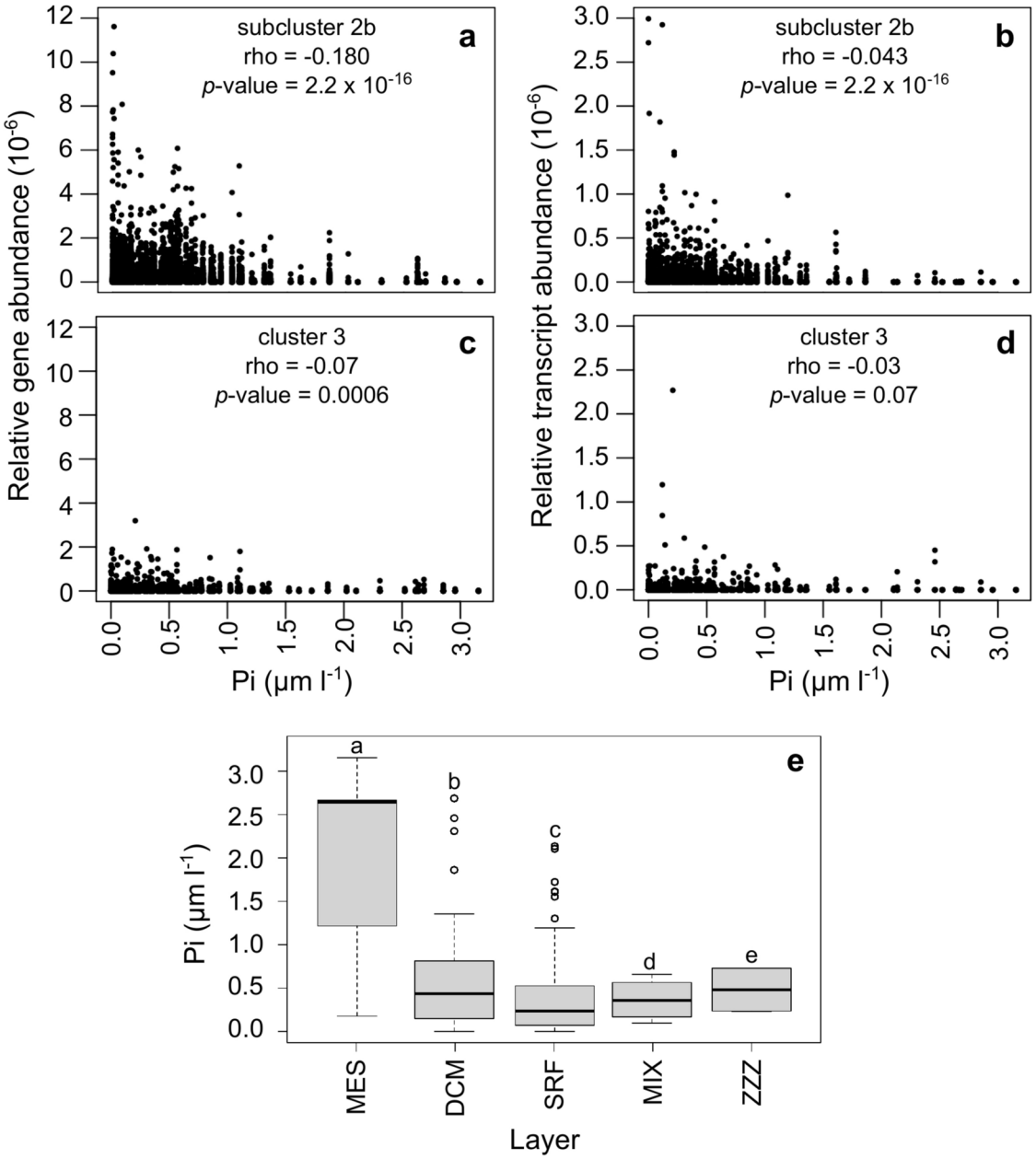

Fig. 7. Relative abundances of PL5 homologs from subcluster 2b and cluster 3 as a function of inorganic phosphate (Pi) concentrations. (**a**) Relative gene abundances of subcluster 2b sequences. (**b**) Relative transcript abundances of subcluster 2b sequences. (**c**) Relative gene abundances of cluster 3 sequences. (**d**) Relative transcript abundances of cluster 3 sequences. In the case of cluster 3 analyses, the information of those closely related to sequences from Pelagibacterales were included. Relative gene and transcript abundances were normalized as percent of mapped reads. Inside each plot, the analyzed dataset and the results of the correlation analyses between relative abundances and Pi concentrations (Spearman's rank correlation) are shown. (**e**) Boxplot of seawater Pi levels as a function of the depth layer: MES, mesopelagic zone; DCM, deep chlorophyll maximum layer; SRF, upper layer zone; MIX, marine epipelagic mixed layer; ZZZ, marine water layer. Median values are indicated as horizontal lines, interquartile ranges as boxes with whiskers extending up to 1.5 times the standard deviation. Different letters on top of the boxplot indicate significant differences ($P < 0.05$).

structural, and functional information[66]. We included PL families with a similar fold in the same SSN as they could descend from a common ancestor, although it can also be the result of convergent evolution[47,67]. As expected, permissive threshold values resulted in clusters including CAZy members from more than one family, which separated when increasing the stringency. The E-value cutoff needed to separate members from the different PL families ranged from 10^{-21} to 10^{-31}, probably reflecting different evolutionary distances between distantly related families[68]. Similar threshold values have been suggested for obtaining informative connections

within members of a superfamily[66], and to separate subfamilies within a CAZy family[69]. The approach chosen in this study allowed the identification of sequences with novel structural characteristics and taxonomic affiliations, while still clearly related to CAZy members.

While the catalytic mechanism is considered a common feature for all members of a CAZy family, substrates and products can vary[12,17]. Out of the 15 PL families that include alginate-degrading enzymes (as of November 2024), eight are multi-specific, and four (i.e., PL8, PL15, PL31, and PL38; http://www.cazy.org/) include half or less of their characterized members that act on alginate. Furthermore, our knowledge of the catalytic potential of some of these families is still limited, due to the low number of characterized enzymes. After the comparative analyses, 2019 sequences (74% of the identified sequences) from 11 PL families were considered for further analyses. The PL31, PL41, PL36, and PL44 were excluded as very few sequences clustered with sequences from the same family in the SSN. It is possible that enzymes from these families are not common in the picoplanktonic community, at least in the free-living fraction targeted in this study. The most abundant genes and transcripts were related to the most extensively studied families (PL5, PL6, PL7, and PL17). In these families, >90% of the characterized members showed alginate-depolymerizing activities, supporting their potential alginate lyase activity.

Multiple PL6, PL7, and PL17 homologs were assigned to *A. australica* and '*Ca.* Polaribacter'

Genes from the PL6, PL7, and PL17 families are often located in polysaccharide utilization loci in alginate-assimilating bacteria, mainly in members of the Bacteroidota and Pseudomonadota phyla[70]. When analyzing the genes and transcripts from the PL6, PL7, and PL17 families, it was evident that those with higher abundances presented similar taxonomic assignments and biogeographic patterns. This was the case of sequences closely related to putative alginate lyases from *A. australica*. The three PL6 homolog sequences assigned to this organism shared <43% identity among them, but each had >90% identity with one of three sequences from *A. australica* H 17 (two as CAZy members and NCBI Acc. Numb. WP_044059062). Similarly, the three PL7 (<51% identity) and the single PL17 homolog sequences assigned to this species shared ≥99% identity with the sequences from *A. australica* H 17 (https://www.cazy.org/b3621.html), with alginate-degrading activity[71]. The high identity values of seven sequences from three PL families to sequences from the same species suggest that they belong to the same organism and show the high coverage of the analyzed dataset[72]. The metagenomic sequences assigned to *A. australica* showed a broad geographic distribution. *Alteromonas* are ubiquitous marine Gammaproteobacteria that play a key role in the aerobic degradation of complex carbohydrates, such as alginate[73]. This genus was found to be predominant in the CAZyome of samples from an estuarine coastal environment[18]. In contrast to our results, *A. macleodii* dominated the microbial community in microcosm experiments amended with alginate[74]. *A. macleodii* and *A. australica* were confirmed as different species by genomic analyses[75].

In the Arctic Ocean, the most abundant PL6, PL7, and PL17 homolog sequences were assigned to both Flavobacteriia and Gammaproteobacteria classes. These sequences were most closely related to putative alginate lyases from single-cell genomes and MAGs. For instance, the three PL6 homolog sequences closely related to putative alginate lyases from *Polaribacter* (sharing only <25% identity among them) had >98% identity with sequences from yet uncultured *Polaribacter* derived from single-cell analyses, obtained from seawater at <6 °C (East Boothbay, ME, USA[76]). Interestingly, these sequences shared <70% identity with PL6 family members from isolates belonging to the *Polaribacter* genus (https://www.cazy.org/PL6.html). Previous studies showed that these organisms increased its abundance in bacterial communities from the Arctic Ocean amended with alginate[77], and high levels of transcription of CAZymes targeting algal polysaccharides from Polaribacer were detected in high-latitude environments of the Atlantic Ocean[19]. In spite of the potential relevance of these organisms in alginate biodegradation processes in high latitude environments, to our knowledge no alginate lyase enzyme from members of this genus has been characterized.

PL5 homologs presented high gene and transcript abundances

The high abundance of genes and transcripts of PL5 homologs in the picoplanktonic community was unexpected, as most characterized members of this family have been identified in pathogenic or soil bacteria, where they are part of the alginate biosynthesis pathway[78,79]. The small group of PL5 homolog sequences that clustered with the majority of the PL5 members in subcluster 2a of the SSN showed low gene abundance but relatively high expression levels, and could be involved in alginate biosynthesis. The production of an alginate exopolysaccharide could be important for the survival of marine bacteria. For instance, *Pseudomonas aeruginosa* SW1, isolated from degraded seaweed, produced alginate and contained alginate biosynthesis pathway genes encoded in its genome[80], and alginate production increased at low temperature in a strain of *Pseudomonas mandelii* isolated from marine sediments of Antarctica[81]. An example of a PL5 alginate lyase from a marine bacterium is AlyA from *Cobetia marina* N-1 (NCBI Acc. Numb. BAA3396, subcluster 2a), isolated from decayed seaweed[82]. This enzyme shares high identity with various putative alginate lyases from *Pseudomonas* (e.g., NCBI Acc. Numb. WP_024011675, 99.73% identity, 100% coverage).

The more divergent PL5 homolog sequences, included in subcluster 2b of the SSN, were mostly assigned to *'Ca.* Pelagibacter' and were closely related to sequences from genomes of these organisms. These sequences showed high gene and transcript levels, in particular in the epipelagic layer. A second group of sequences included in cluster 3 had lower but widespread gene and transcript abundances. In a study of the CAZyome of microbial communities from an estuary, *'Ca.* Pelagibacter' was one of the genera that most contributed to CAZyme gene abundances[18]. This study also showed a high abundance of genes coding for ABC transporters in these organisms, supporting a potential for complex carbohydrate degradation[83]. Genes related to carbohydrate metabolism, however, are not consistently found in these organisms and could define niche partitioning[84]. This was the case for PL5 homolog sequences in Pelagibacter genomes mostly obtained from single-cell analyses[84,85],

which were only present in 10% of the analyzed genomes. The putative PL5 genes were also identified in four isolates, including *'Ca.* Pelagibacter' sp. HTCC7211[86–88].

The metagenomic sequences related to the PL5 family included two variants in one of the key residues of the active site, His192 (A1-III from *Sphingomonas* sp. A1 numbering[44]). It has been proposed that the highly conserved His residue acts stabilizing the carboxyl group of the substrate in M-specific alginate lyases of the PL5 family, and its mutation for Ala greatly reduces the enzymatic activity[89,90]. A conserved Tyr residue (Tyr246 in A1-III), on the other hand, acts as both donor and acceptor in β-elimination reactions with a *syn* configuration[91]. All sequences from cluster 3 of the SSN presented a Lys residue in this position. Lys, like His, is titratable with a positively charged side chain and has been found to perform the highest number of functions in lyases (EC 2) both in reactant and spectator catalytic roles[92,93]. Furthermore, an association has been observed between the electrostatic stabilizer role and Lys in a global analysis of residue conservation in enzymes[93]. Depending on the geometry of the substrate-active site interaction, Lys could function as a base in an *anti*-configuration, as it is the case for the catalytic His residue of Smlt1473 from *Stenotrophomonas maltophilia* with hyaluronic acid[90,94]. The second variant included only ten metagenomic sequences from subcluster 2a with a Gln residue in the His192 position, but these sequences showed very low gene and transcript levels.

PL38 homolog sequences from Eukaryota showed high expression levels in high-latitude regions

The identification of sequences assigned to Eukaryota with the HMM of the PL38 family was not unexpected, as sequences from eukaryotic organisms represent approximately 12% of the prokaryote-enriched dataset[95]. The activities detected to date in the four characterized enzymes of the PL38 family include endo MM-, GG-, MG-, and GM-specific alginate lyase (EC 4.2.2.3; 4.2.2.11; 4.2.2.-), endo-β-1,4-glucuronan lyase (EC 4.2.2.14), and exo-β-1,4-glucuronan lyase (EC 4.2.2.-)[45,96–98]. Alginate and polyglucuronic acid are components of the cell wall of brown and green algae species, respectively[4,99], and they can also be synthesized by other organisms, such as bacteria[80,99]. Although PL38 homolog sequences showed relatively low gene abundances, some presented remarkably high transcript levels, in particular at high latitudes. This is the case of sequences assigned to *P. antarctica* (e.g., ID 008024536), which showed the highest gene expression levels in the South Atlantic Ocean and in the Southern Ocean, with up to 78 times higher transcript than gene relative abundances. The life cycle of some of these organisms includes multiple morphotypes, both as solitary cells and as colonies confined in a mucus matrix composed of heteropolysaccharides that can contain low levels of glucuronic acids[100–102]. It is possible that these enzymes could be used for heteropolysaccharide remodeling, or alternatively for alginate or polyglucuronate degradation. As other phytoplanktonic protists, *Phaeocystis* spp. display metabolic flexibility, with the potential for mixotrophy as a strategy for supplementing autotrophic growth under unfavorable conditions[103]. A recently published heterotrophy index suggests that *Phaeocystis* spp. mixotrophy is more common in polar areas[104], although the physiological role of these enzymes in *P. antarctica* could not be established in this study.

Multiple environmental drivers are affecting the structure of the population assemblages

We analyzed the environmental factors that could be influencing the distribution of organisms containing the identified putative alginate lyase genes and their expression. Temperature was a key driver in structuring the microbial populations containing these genes, as has been observed at the microbial community level[72,95,105]. Phosphate and ammonium concentrations were significantly correlated with sample partitioning in ordination analyses in all the analyzed PL families, although the influence of nitrite and nitrate was not clear. Nutrient concentrations have also been found to be important factors determining biogeographic patterns, in particular in small-sized organisms[95]. Depth-related patterns, in particular between epipelagic and mesopelagic zones, were evident in the gene relative abundances, although this could be related to phosphate levels increasing with depth[106,107]. The abundance of genes and transcripts of PL homolog sequences was generally higher in the epipelagic layer than in the mesopelagic layer. In a study including various high molecular weight biopolymers, the rate and/or range of enzymatic activities was higher in the epipelagic layer and decreased with depth[108]. The structural complexity of the polysaccharides also varied with depth, constituting an additional factor controlling the permanence and fate of these compounds in the different layers of the ocean[109].

In the case of PL5 homolog sequences, we further explored the relationship between relative gene abundances of individual sequences from subcluster 2b closely related to sequences from '*Ca.* Pelagibacter' and temperature. Remarkably, the optimum temperature ranges varied greatly among sequences, the most frequent being 20–30 °C, but were also < 15 °C in some sequences. The greatest relative contribution of '*Ca.* Pelagibacter' to the bacterial community in the open ocean has been observed at high temperatures and low chlorophyll concentration[110,111]. Niche differentiation between major SAR11 subgroups has been observed, in which phylotypes display transitions in abundance that are strongly correlated with temperature and latitude[112,113]. The results of this work showed various distinct distribution patterns along the temperature gradient, reinforcing the ecotype hypothesis.

Phosphorus is a key environmental variable that strongly influences ocean productivity and microbial community composition, with Pi being its most bioavailable form[114]. Although Pi concentration varies in time and space, it is often found at extremely low concentrations in the surface of the ocean in mid-latitude regions, while higher concentrations are found in high-latitude environments and at higher depths[106]. The ecological relevance of Pi and the previously reported three-fold induction of a putative alginate lyase gene in the strain HTCC7211 under phosphorus-limiting conditions[115] prompted us to assess a potential correlation between Pi concentrations and both gene and transcript abundances of sequences assigned to this genus. A negative correlation was found in both cases for PL5 homologs from the His (subcluster 2b) and Lys (cluster 3) variants. As transcript abundances were proportional to gene abundances, gene expression levels (transcript levels relative

to gene abundances) were not higher at low Pi concentrations, indicating population-level selection rather than an effect on expression levels. Phosphate limitations were found to select for certain bacterial populations with the capability for dissolved organic phosphorus utilization, as is the case in various Alphaproteobacteria, including *Pelagibacter* sp. HTCC7211[116]. Another strategy for reducing the phosphorus requirement is the synthesis of glycoglycerolipids (including the acidic monoglucuronic acid diacylglycerol) as a replacement of cellular membrane phospholipids[88]. Further work is needed to determine the substrates of the identified putative enzymes and to elucidate if these polysaccharide lyases are involved in the response of these organisms to phosphate limitation.

Polysaccharides represent an important fraction of the organic matter in the ocean, constituting a complex mixture with different degradability and uneven distribution[7]. Brown algae species are an important source of alginate in both coastal and open ocean environments[3,117]. Alginate concentrations are higher in the upper layers of the ocean, decreasing with depth[109]. They are mostly found as part of the particulate organic matter due to the low solubility of alginate in seawater[118]. However, they are also expected to be present as part of the dissolved organic matter pool, albeit at lower concentrations[19]. In this study, we analyzed the putative alginate lyase sequences and the microbial populations containing these genes in the free-living fraction of the picoplanktonic communities. We found significant correlations between community structure and total C, suggesting some level of influence of organic matter on some of the organisms carrying these genes. These organisms could be benefitting from occasional pulses of soluble alginate present in the seawater, which could be aided by the partial degradation of these polysaccharides by extracellular alginate lyase enzymes produced in the particles[119–121]. A similar approach targeting the free-living community was used to study the carbohydrate degradation potential and polysaccharide content in an Arctic marine environment[19]. Other studies have focused on the response of both free-living and particle-attached communities to alginate addition in experimental systems[121,122]. Although important differences were found in the response to sodium alginate addition in mesocosms constructed from microbial communities from distant Pacific Ocean sites, a larger diversity of organisms was enriched in the particle-attached fraction than in the free-living communities as a response to this amendment[121]. Other populations presenting a similar response in both fractions seem to have both lifestyles[121]. Therefore, the diverse microbial populations with alginate degradation potential uncovered in this study through a multi-omics approach are one component of the alginate degradation cascade involving multiple fractions of the ocean microbiome.

Conclusions

In this study, we describe the sequence space of homologs of alginate-degrading enzymes from the upper layers of the global ocean. Structural comparative analyses were used to select for sequences with characteristics compatible with members of the targeted families, after relaxed identification conditions. This curated dataset revealed the diverse taxonomic groups containing these enzymes, as well as specific depth and biogeographic patterns in gene and transcript abundances. Significant correlations were identified among the biological and environmental information associated with the dataset, evidencing the main drivers of the distribution of organisms carrying and expressing these genes. Future research needs include the study of substrate preferences of more members of these families, in particular those that have been created more recently. This information would allow defining substrate-specific subfamilies and better assess the potential physiological role of these enzymes.

Data availability

The datasets used in this study are part of the OGA web server, and are publicly available at https://tara-oceans.mio.osupytheas.fr/ocean-gene-atlas. The SSN files for Cytoscape corresponding to Figs. 1 to 4 and Supplementary Fig. S6 are publicly available under DOI https://doi.org/10.6084/m9.figshare.c.7583477.

Received: 18 December 2024; Accepted: 4 March 2025
Published online: 08 March 2025

References

1. Bringloe, T. T. et al. Phylogeny and evolution of the brown algae. *Crit. Rev. Plant Sci.* **39**, 281–321. https://doi.org/10.1080/07352689.2020.1787679 (2020).
2. Krumhansl, K. A. & Scheibling, R. E. Production and fate of kelp detritus. *Mar. Ecol. Prog. Ser.* **467**, 281–302. https://doi.org/10.3354/meps09940 (2012).
3. Huffard, C., Von Thun, S., Sherman, A., Sealey, K. & Smith, K. Pelagic *Sargassum* community change over a 40-year period: Temporal and spatial variability. *Mar. Biol.* **161**, 2735–2751. https://doi.org/10.1007/s00227-014-2539-y (2014).
4. Krause-Jensen, D. & Duarte, C. M. Substantial role of macroalgae in marine carbon sequestration. *Nat. Geosci.* **9**, 737–742. https://doi.org/10.1038/NGEO2790 (2016).
5. Deniaud-Bouët, E., Hardouin, K., Potin, P., Kloareg, B. & Hervé, C. A review about brown algal cell walls and fucose-containing sulfated polysaccharides: Cell wall context, biomedical properties and key research challenges. *Carbohydr. Polym.* **175**, 395–408. https://doi.org/10.1016/j.carbpol.2017.07.082 (2017).
6. Schiener, P., Black, K. D., Stanley, M. S. & Green, D. H. The seasonal variation in the chemical composition of the kelp species *Laminaria digitata*, *Laminaria hyperborea*, *Saccharina latissima* and *Alaria esculenta*. *J. Appl. Phycol.* **27**, 363–373. https://doi.org/10.1007/s10811-014-0327-1 (2015).
7. Arnosti, C. et al. The biogeochemistry of marine polysaccharides: sources, inventories, and bacterial drivers of the carbohydrate cycle. *Ann. Rev. Mar. Sci.* **13**, 81–108. https://doi.org/10.1146/annurev-marine-032020-012810 (2021).
8. Cheng, D., Jiang, C., Xu, J., Liu, Z. & Mao, X. Characteristics and applications of alginate lyases: A review. *Int. J. Biol. Macromol.* **164**, 1304–1320. https://doi.org/10.1016/j.ijbiomac.2020.07.199 (2020).
9. Inoue, A. & Ojima, T. Functional identification of alginate lyase from the brown alga *Saccharina japonica*. *Sci. Rep.* **9**, 4937. https://doi.org/10.1038/s41598-019-41351-6 (2019).

10. Das, S. Genetic regulation, biosynthesis and applications of extracellular polysaccharides of the biofilm matrix of bacteria. *Carbohydr. Polym.* **291**, 119536. https://doi.org/10.1016/j.carbpol.2022.119536 (2022).
11. Garron, M.-L. & Cygler, M. Uronic polysaccharide degrading enzymes. *Curr. Opin. Struct. Biol.* **28**, 87–95. https://doi.org/10.1016/j.sbi.2014.07.012 (2014).
12. Garron, M.-L. & Henrissat, B. The continuing expansion of CAZymes and their families. *Curr. Opin. Chem. Biol.* **53**, 82–87. https://doi.org/10.1016/j.cbpa.2019.08.004 (2019).
13. Li, Q., Zheng, L., Guo, Z., Tang, T. & Zhu, B. Alginate degrading enzymes: An updated comprehensive review of the structure, catalytic mechanism, modification method and applications of alginate lyases. *Crit. Rev. Biotechnol.* **41**, 953–968. https://doi.org/10.1080/07388551.2021.1898330 (2021).
14. Garron, M.-L. & Cygler, M. Structural and mechanistic classification of uronic acid-containing polysaccharide lyases. *Glycobiology* **20**, 1547–1573. https://doi.org/10.1093/glycob/cwq122 (2010).
15. Pandey, S., Berger, B. W. & Acharya, R. Structural analyses of substrate–pH activity pairing observed across diverse polysaccharide lyases. *Biochemistry* **62**, 2775–2790. https://doi.org/10.1021/acs.biochem.3c00321 (2023).
16. Zhou, J. et al. Discovery and characterization of a novel poly-mannuronate preferred alginate lyase: The first member of a new polysaccharide lyase family. *Carbohydr. Polym.* **343**, 122474. https://doi.org/10.1016/j.carbpol.2024.122474 (2024).
17. Lombard, V., Henrissat, B. & Garron, M.-L. CAZac: An activity descriptor for carbohydrate-active enzymes. *Nucleic Acids Res.* https://doi.org/10.1093/nar/gkae1045 (2024).
18. Sun, C.-C. et al. Polymeric carbohydrates utilization separates microbiomes into niches: Insights into the diversity of microbial carbohydrate-active enzymes in the inner shelf of the Pearl River Estuary, China. *Front. Microbiol.* **14**, 1180321. https://doi.org/10.3389/fmicb.2023.1180321 (2023).
19. Priest, T., Vidal-Melgosa, S., Hehemann, J.-H., Amann, R. & Fuchs, B. M. Carbohydrates and carbohydrate degradation gene abundance and transcription in Atlantic waters of the Arctic. *ISME Commun.* **3**, 130. https://doi.org/10.1038/s43705-023-00324-7 (2023).
20. López-Sánchez, R. et al. Metagenomic analysis of carbohydrate-active enzymes and their contribution to marine sediment biodiversity. *World J. Microbiol. Biotechnol.* **40**, 95. https://doi.org/10.1007/s11274-024-03884-5 (2024).
21. Matos, M. N. et al. Metagenomics unveils the attributes of the alginolytic guilds of sediments from four distant cold coastal environments. *Environ. Microbiol.* **18**, 4471–4484. https://doi.org/10.1111/1462-2920.13433 (2016).
22. Pesant, S. et al. Open science resources for the discovery and analysis of *Tara* Oceans data. *Sci. Data* **2**, 1–16. https://doi.org/10.1038/sdata.2015.23 (2015).
23. Salazar, G. et al. Gene expression changes and community turnover differentially shape the global ocean metatranscriptome. *Cell* **179**, 1068–1083. https://doi.org/10.1016/j.cell.2019.10.014 (2019).
24. Vernette, C. et al. The Ocean gene Atlas v2.0: Online exploration of the biogeography and phylogeny of plankton genes. *Nucleic Acids Res.* **50**, W516–W526. https://doi.org/10.1093/nar/gkac420 (2022).
25. Zheng, J. et al. dbCAN3: Automated carbohydrate-active enzyme and substrate annotation. *Nucleic Acids Res.* **51**, W115–W121. https://doi.org/10.1093/nar/gkad328 (2023).
26. Eddy, S. R. A new generation of homology search tools based on probabilistic inference. In *Genome Informatics 2009: Genome Informatics Series* Vol. 23, 205–211 (World Scientific, 2009).
27. Chen, I.-M.A. et al. The IMG/M data management and analysis system v.7: Content updates and new features. *Nucleic Acids Res.* **51**, D723–D732. https://doi.org/10.1093/nar/gkac976 (2023).
28. Delmont, T. et al. Functional repertoire convergence of distantly related eukaryotic plankton lineages revealed by genome-resolved metagenomics. *Cell Genom.* **28**, 100123. https://doi.org/10.1016/j.xgen.2022.100123 (2021).
29. Vorobev, A. et al. Transcriptome reconstruction and functional analysis of eukaryotic marine plankton communities via high-throughput metagenomics and metatranscriptomics. *Genome Res.* **30**, 647–659. https://doi.org/10.1101/gr.253070.119 (2020).
30. Johnson, M. et al. NCBI BLAST: A better web interface. *Nucleic Acids Res.* **36**, W5–W9. https://doi.org/10.1093/nar/gkn201 (2008).
31. Zallot, R., Oberg, N. & Gerlt, J. A. The EFI web resource for genomic enzymology tools: Leveraging protein, genome, and metagenome databases to discover novel enzymes and metabolic pathways. *Biochemistry* **58**, 4169–4182. https://doi.org/10.1021/acs.biochem.9b00735 (2019).
32. Barrett, K., Hunt, C. J., Lange, L., Grigoriev, I. V. & Meyer, A. S. Conserved unique peptide patterns (CUPP) online platform 2.0: Implementation of+ 1000 JGI fungal genomes. *Nucleic Acids Res.* **51**, W108–W114. https://doi.org/10.1093/nar/gkad385 (2023).
33. Larkin, M. A. et al. Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948. https://doi.org/10.1093/bioinformatics/btm404 (2007).
34. Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M. & Barton, G. J. Jalview version 2—A multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191. https://doi.org/10.1093/bioinformatics/btp033 (2009).
35. Su, G., Morris, J. H., Demchak, B. & Bader, G. D. Biological network exploration with Cytoscape 3. *Curr. Prot. Bioinf.* **47**, 8–13. https://doi.org/10.1002/0471250953.bi0813s47 (2014).
36. Abramson, J. et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500. https://doi.org/10.1038/s41586-024-07487-w (2024).
37. Goddard, T. D. et al. UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25. https://doi.org/10.1002/pro.3235 (2018).
38. Holm, L., Laiho, A., Törönen, P. & Salgado, M. DALI shines a light on remote homologs: One hundred discoveries. *Protein Sci.* **32**, e4519. https://doi.org/10.1002/pro.4519 (2023).
39. Hallgren, J. et al. DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks. *BioRxiv* https://doi.org/10.1101/2022.04.08.487609 (2022).
40. McMurdie, P. J. & Holmes, S. phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* **8**, e61217. https://doi.org/10.1371/journal.pone.0061217 (2013).
41. Broatch, J. E., Dietrich, S. & Goelman, D. Introducing data science techniques by connecting database concepts and dplyr. *J. Stat. Educ.* **27**, 147–153. https://doi.org/10.1080/10691898.2019.1647768 (2019).
42. Vegan: Community ecology package. *R package version 2.6-8* (2024). https://CRAN.R-project.org/package=vegan
43. R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing, Vienna, Austria* (2024). https://www.R-project.org/
44. Yoon, H.-J., Hashimoto, W., Miyake, O., Murata, K. & Mikami, B. Crystal structure of alginate lyase A1-III complexed with trisaccharide product at 2.0 Å resolution. *J. Mol. Biol.* **307**, 9–16. https://doi.org/10.1006/jmbi.2000.4509 (2001).
45. Rønne, M. E. et al. Three alginate lyases provide a new gut *Bacteroides ovatus* isolate with the ability to grow on alginate. *Appl. Environ. Microbiol.* **89**, e01185-23. https://doi.org/10.1128/aem.01185-23 (2023).
46. Holm, L. Using Dali for protein structure comparison. In *Structural Bioinformatics: Methods and Protocols* Vol. 2112 (ed. Gáspári, Z.) 29–42 (Springer Nature, 2020).
47. Burnim, A. A., Dufault-Thompson, K. & Jiang, X. The three-sided right-handed β-helix is a versatile fold for glycan interactions. *Glycobiology* **34**, cwae037. https://doi.org/10.1093/glycob/cwae037 (2024).
48. Mathieu, S., Henrissat, B., Labre, F., Skjåk-Bræk, G. & Helbert, W. Functional exploration of the polysaccharide lyase family PL6. *PLoS One* **11**, e0159415. https://doi.org/10.1371/journal.pone.0159415 (2016).

49. Helbert, W. et al. Discovery of novel carbohydrate-active enzymes through the rational exploration of the protein sequences space. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 6063–6068. https://doi.org/10.1073/pnas.1815791116 (2019).
50. Wang, B., Dong, S., Li, F.-L. & Ma, X.-Q. Structural basis for the exolytic activity of polysaccharide lyase family 6 alginate lyase BcAlyPL6 from human gut microbe *Bacteroides clarus*. *Biochem. Biophys. Res. Commun.* **547**, 111–117. https://doi.org/10.1016/j.bbrc.2021.02.040 (2021).
51. Xue, Z. et al. A novel alginate lyase: Identification, characterization, and potential application in alginate trisaccharide preparation. *Mar. Drugs* **20**, 159. https://doi.org/10.3390/md20030159) (2022).
52. Itoh, T. et al. Structural and biochemical characterisation of a novel alginate lyase from *Paenibacillus* sp. Str. FPU-7. *Sci. Rep.* **9**, 14870. https://doi.org/10.1038/s41598-019-51006-1 (2019).
53. Yamasaki, M., Ogura, K., Hashimoto, W., Mikami, B. & Murata, K. A structural basis for depolymerization of alginate by polysaccharide lyase family-7. *J. Mol. Biol.* **352**, 11–21. https://doi.org/10.1016/j.jmb.2005.06.075 (2005).
54. Ogura, K. et al. Crystal structure of family 14 polysaccharide lyase with pH-dependent modes of action. *J. Biol. Chem.* **284**, 35572–35579. https://doi.org/10.1074/jbc.M109.068056 (2009).
55. Dong, F. et al. Alginate lyase Aly36B is a new bacterial member of the polysaccharide lyase family 36 and catalyzes by a novel mechanism with lysine as both the catalytic base and catalytic acid. *J. Mol. Biol.* **431**, 4897–4909. https://doi.org/10.1016/j.jmb.2019.10.023 (2019).
56. Uchimura, K., Miyazaki, M., Nogi, Y., Kobayashi, T. & Horikoshi, K. Cloning and sequencing of alginate lyase genes from deep-sea strains of *Vibrio* and *Agarivorans* and characterization of a new *Vibrio* enzyme. *Mar. Biotechnol.* **12**, 526–533. https://doi.org/10.1007/s10126-009-9237-7 (2010).
57. Dong, S. et al. Molecular insight into the role of the N-terminal extension in the maturation, substrate recognition, and catalysis of a bacterial alginate lyase from polysaccharide lyase family 18. *J. Biol. Chem.* **289**, 29558–29569. https://doi.org/10.1074/jbc.M114.584573 (2014).
58. Pilgaard, B., Vuillemin, M., Holck, J., Wilkens, C. & Meyer, A. S. Specificities and synergistic actions of novel PL8 and PL7 alginate lyases from the marine fungus *Paradendryphiella salina*. *J. Fungi* **7**, 80 (2021).
59. Jouanneau, D. et al. Structure–function analysis of a new PL17 oligoalginate lyase from the marine bacterium *Zobellia galactanivorans* DsijT. *Glycobiology* **31**, 1364–1377 (2021).
60. Mathieu, S. et al. Ancient acquisition of "alginate utilization loci" by human gut microbiota. *Sci. Rep.* **8**, 8075. https://doi.org/10.1038/s41598-018-26104-1 (2018).
61. Ji, S. et al. The molecular basis of endolytic activity of a multidomain alginate lyase from *Defluviitalea phaphyphila*, a representative of a new lyase family, PL39. *J. Biol. Chem.* **294**, 18077–18091. https://doi.org/10.1074/jbc.RA119.010716 (2019).
62. Dong, W., Lu, W., McKeehan, W. L., Luo, Y. & Ye, S. Structural basis of heparan sulfate-specific degradation by heparinase III. *Protein Cell* **3**, 950–961. https://doi.org/10.1007/s13238-012-2056-z (2012).
63. Eddy, S. R. Profile hidden Markov models. *Bioinformatics (Oxford, England)* **14**, 755–763. https://doi.org/10.1093/bioinformatics/14.9.755 (1998).
64. Rossi, M. F., Mello, B. & Schrago, C. G. Performance of hidden markov models in recovering the standard classification of glycoside hydrolases. *Evol. Bioinform.* **13**, 1176934317703401. https://doi.org/10.1177/1176934317703401 (2017).
65. López-Mondéjar, R., Tláskal, V., Da Rocha, U. & Baldrian, P. Global distribution of carbohydrate utilization potential in the prokaryotic tree of life. *Msystems* **7**, e0082922. https://doi.org/10.1128/msystems.00829-22 (2022).
66. Copp, J. N., Anderson, D. W., Akiva, E., Babbitt, P. C. & Tokuriki, N. Exploring the sequence, function, and evolutionary space of protein superfamilies using sequence similarity networks and phylogenetic reconstructions. In *Methods in Enzymology* Vol. 620 315–347 (Elsevier, 2019). https://doi.org/10.1016/bs.mie.2019.03.015
67. Jenkins, J., Mayans, O. & Pickersgill, R. Structure and evolution of parallel β-helix proteins. *J. Struct. Biol.* **122**, 236–246. https://doi.org/10.1006/jsbi.1998.3985 (1998).
68. Atkinson, H. J., Morris, J. H., Ferrin, T. E. & Babbitt, P. C. Using sequence similarity networks for visualization of relationships across diverse protein superfamilies. *PLoS One* **4**, e4345. https://doi.org/10.1371/journal.pone.0004345 (2009).
69. Hornung, B. V. H. & Terrapon, N. An objective criterion to evaluate sequence-similarity networks helps in dividing the protein family sequence space. *PLoS Comp. Biol.* **19**, e1010881. https://doi.org/10.1371/journal.pcbi.1010881 (2023).
70. Zhang, L., Li, X., Zhang, X., Li, Y. & Wang, L. Bacterial alginate metabolism: An important pathway for bioconversion of brown algae. *Biotechnol. Biofuels* **14**, 1–18. https://doi.org/10.1186/s13068-021-02007-8 (2021).
71. Barbeyron, T., Zonta, E., Le Panse, S., Duchaud, E. & Michel, G. *Alteromonas fortis* sp. nov., a non-flagellated bacterium specialized in the degradation of iota-carrageenan, and emended description of the genus *Alteromonas*. *Int. J. Syst. Evol. Microbiol.* **69**, 2514–2521. https://doi.org/10.1099/ijsem.0.003533 (2019).
72. Sunagawa, S. et al. *Tara* Oceans: Towards global ocean ecosystems biology. *Nat. Rev. Microbiol.* **18**, 428–445. https://doi.org/10.1038/s41579-020-0364-5 (2020).
73. Henríquez-Castillo, C. et al. Metaomics unveils the contribution of *Alteromonas* bacteria to carbon cycling in marine oxygen minimum zones. *Front. Mar. Sci.* **9**, 993667. https://doi.org/10.3389/fmars.2022.993667 (2022).
74. Mitulla, M. et al. Response of bacterial communities from California coastal waters to alginate particles and an alginolytic *Alteromonas macleodii* strain. *Environ. Microbiol.* **18**, 4369–4377. https://doi.org/10.1111/1462-2920.13314 (2016).
75. López-Pérez, M., Gonzaga, A., Ivanova, E. P. & Rodriguez-Valera, F. Genomes of *Alteromonas australica*, a world apart. *BMC Genomics* **15**, 1–13. https://doi.org/10.1186/1471-2164-15-483 (2014).
76. Munson-McGee, J. H. et al. Decoupling of respiration rates and abundance in marine prokaryoplankton. *Nature* **612**, 764–770. https://doi.org/10.1038/s41586-022-05505-3 (2022).
77. Jain, A. et al. Response of bacterial communities from Kongsfjorden (Svalbard, Arctic Ocean) to macroalgal polysaccharide amendments. *Mar. Environ. Res.* **155**, 104874. https://doi.org/10.1016/j.marenvres.2020.104874 (2020).
78. Greninger, A. L., Streithorst, J., Golden, J. A., Chiu, C. Y. & Miller, S. Complete genome sequence of sequential *Pandoraea apista* isolates from the same cystic fibrosis patient supports a model of chronic colonization with *in vivo* strain evolution over time. *Diagn. Microbiol. Infect. Dis.* **87**, 1–6. https://doi.org/10.1016/j.diagmicrobio.2016.10.013 (2017).
79. Vazquez, A., Moreno, S., Guzmán, J., Alvarado, A. & Espín, G. Transcriptional organization of the *Azotobacter vinelandii algGXLVIFA* genes: Characterization of *algF* mutants. *Gene* **232**, 217–222. https://doi.org/10.1016/S0378-1119(99)00119-5 (1999).
80. Dharshini, R. S., Manickam, R., Curtis, W. R., Rathinasabapathi, P. & Ramya, M. Genome analysis of alginate synthesizing *Pseudomonas aeruginosa* strain SW1 isolated from degraded seaweeds. *Antonie Van Leeuwenhoek* **114**, 2205–2217. https://doi.org/10.1007/s10482-021-01673-w (2021).
81. Vásquez-Ponce, F. et al. Alginate overproduction and biofilm formation by psychrotolerant *Pseudomonas mandelii* depend on temperature in Antarctic marine sediments. *Electron. J. Biotechnol.* **28**, 27–34. https://doi.org/10.1016/j.ejbt.2017.05.001 (2017).
82. Kraiwattanapong, J., Tsuruga, H., Ooi, T. & Kinoshita, S. Cloning and sequencing of a *Deleya marina* gene encoding for alginate lyase. *Biotechnol. Lett.* **21**, 169–174. https://doi.org/10.1023/A:1005435725903 (1999).
83. Carini, P., Steindler, L., Beszteri, S. & Giovannoni, S. J. Nutrient requirements for growth of the extreme oligotroph '*Candidatus* Pelagibacter ubique'HTCC1062 on a defined medium. *ISME J.* **7**, 592–602. https://doi.org/10.1038/ismej.2012.122 (2013).
84. Grote, J. et al. Streamlining and core genome conservation among highly divergent members of the SAR11 clade. *MBio* **3**, 10–1128. https://doi.org/10.1128/mbio.00252-12 (2012).

85. Chang, T., Gavelis, G. S., Brown, J. M. & Stepanauskas, R. Genomic representativeness and chimerism in large collections of SAGs and MAGs of marine prokaryoplankton. *Microbiome* **12**, 126. https://doi.org/10.1186/s40168-024-01848-3 (2024).
86. Stingl, U., Tripp, H. J. & Giovannoni, S. J. Improvements of high-throughput culturing yielded novel SAR11 strains and other abundant marine bacteria from the Oregon coast and the Bermuda Atlantic Time Series study site. *ISME J* **1**, 361–371. https://doi.org/10.1038/ismej.2007.49 (2007).
87. Wei, T., Quareshy, M., Zhang, Y.-Z., Scanlan, D. J. & Chen, Y. Manganese is essential for PlcP metallophosphoesterase activity involved in lipid remodeling in abundant marine heterotrophic bacteria. *Appl. Environ. Microbiol.* **84**, e01109-01118. https://doi.org/10.1128/AEM.01109-18 (2018).
88. Wei, T. et al. A glycolipid glycosyltransferase with broad substrate specificity from the marine bacterium "*Candidatus* Pelagibacter sp." strain HTCC7211. *Appl. Environ. Microbiol.* **87**, e00326-21. https://doi.org/10.1128/AEM.00326-21 (2021).
89. Pandey, S., Mahanta, P., Berger, B. W. & Acharya, R. Structural insights into the mechanism of pH-selective substrate specificity of the polysaccharide lyase Smlt1473. *J. Biol. Chem.* **297**, 101014. https://doi.org/10.1016/j.jbc.2021.101014 (2021).
90. Dash, P. & Acharya, R. Distinct modes of hidden structural dynamics in the functioning of an allosteric polysaccharide lyase. *ACS Cent. Sci.* **8**, 933–947. https://doi.org/10.1021/acscentsci.2c00277 (2022).
91. Zhu, B. & Yin, H. Alginate lyase: Review of major sources and classification, properties, structure-function analysis and applications. *Bioengineered* **6**, 125–131. https://doi.org/10.1080/21655979.2015.1030543 (2015).
92. Holliday, G. L., Mitchell, J. B. & Thornton, J. M. Understanding the functional roles of amino acid residues in enzyme catalysis. *J. Mol. Biol.* **390**, 560–577. https://doi.org/10.1016/j.jmb.2009.05.015 (2009).
93. Ribeiro, A. J., Tyzack, J. D., Borkakoti, N., Holliday, G. L. & Thornton, J. M. A global analysis of function and conservation of catalytic residues in enzymes. *J. Biol. Chem.* **295**, 314–324. https://doi.org/10.1074/jbc.REV119.006289 (2020).
94. MacDonald, L. C. & Berger, B. W. A polysaccharide lyase from *Stenotrophomonas maltophilia* with a unique, pH-regulated substrate specificity. *J. Biol. Chem.* **289**, 312–325. https://doi.org/10.1074/jbc.M113.489195 (2014).
95. Richter, D. J. et al. Genomic evidence for global ocean plankton biogeography shaped by large-scale current systems. *Elife* **11**, e78129. https://doi.org/10.7554/eLife.78129 (2022).
96. Kikuchi, M. et al. A bacterial endo-β-1, 4-glucuronan lyase, CUL-I from *Brevundimonas* sp. SH203, belonging to a novel polysaccharide lyase family. *Protein Expr. Purif.* **166**, 105502. https://doi.org/10.1016/j.pep.2019.105502 (2020).
97. Sun, X.-K. et al. Degradation of alginate by a newly isolated marine bacterium *Agarivorans* sp. B2Z047. *Mar. Drugs* **20**, 254. https://doi.org/10.3390/md20040254 (2022).
98. Pilgaard, B. et al. Discovery of a novel glucuronan lyase system in *Trichoderma parareesei*. *Appl. Environ. Microbiol.* **88**, e01819-01821. https://doi.org/10.1128/AEM.01819-21 (2022).
99. Elboutachfaiti, R., Delattre, C., Petit, E. & Michaud, P. Polyglucuronic acids: Structures, functions and degrading enzymes. *Carbohydr. Polym.* **84**, 1–13. https://doi.org/10.1016/j.carbpol.2010.10.063 (2011).
100. Smith, W. O. Jr. & Trimborn, S. *Phaeocystis*: A global enigma. *Annu. Rev. Mar. Sci.* **16**, 417–441. https://doi.org/10.1146/annurev-marine-022223-025031 (2024).
101. Alderkamp, A.-C., Buma, A. G. & van Rijssel, M. The carbohydrates of *Phaeocystis* and their degradation in the microbial food web. *Biogeochemistry* **83**, 99–118. https://doi.org/10.1007/s10533-007-9078-2 (2007).
102. Van Rijssel, M., Janse, I., Noordkamp, D. & Gieskes, W. An inventory of factors that affect polysaccharide production by *Phaeocystis globosa*. *J. Sea Res.* **43**, 297–306. https://doi.org/10.1016/S1385-1101(00)00013-7 (2000).
103. Stoecker, D. K. & Lavrentyev, P. J. Mixotrophic plankton in the polar seas: A pan-arctic review. *Front. Mar. Sci.* **5**, 292. https://doi.org/10.3389/fmars.2018.00292 (2018).
104. Allen, A. et al. Genome-resolved biogeography of Phaeocystales, cosmopolitan bloom-forming algae. *ResearchSquare*. https://doi.org/10.21203/rs.3.rs-4339559/v1 (2024).
105. Yu, Z., Yang, J., Liu, L., Zhang, W. & Amalfitano, S. Bacterioplankton community shifts associated with epipelagic and mesopelagic waters in the Southern Ocean. *Sci. Rep.* **5**, 12897. https://doi.org/10.1038/srep12897 (2015).
106. Levitus, S., Conkright, M. E., Reid, J. L., Najjar, R. G. & Mantyla, A. Distribution of nitrate, phosphate and silicate in the world oceans. *Prog. Oceanogr.* **31**, 245–273. https://doi.org/10.1016/0079-6611(93)90003-V (1993).
107. Junger, P. C. et al. Global biogeography of the smallest plankton across ocean depths. *Sci. Adv.* **9**, eadg9763. https://doi.org/10.1126/sciadv.adg9763 (2023).
108. Brown, S. A., Balmonte, J. P., Hoarfrost, A., Ghobrial, S. & Arnosti, C. Depth-related patterns in microbial community responses to complex organic matter in the western North Atlantic Ocean. *Biogeosciences* **19**, 5617–5631. https://doi.org/10.5194/bg-19-5617-2022 (2022).
109. Lloyd, C. C. et al. Correlations among carbohydrate inventories, enzyme activities, and microbial communities in the western North Atlantic Ocean. *EGUsphere* **2024**, 1–33. https://doi.org/10.5194/egusphere-2024-615 (2024).
110. Lefort, T. & Gasol, J. M. Global-scale distributions of marine surface bacterioplankton groups along gradients of salinity, temperature, and chlorophyll: a meta-analysis of fluorescence *in situ* hybridization studies. *Aquat. Microb. Ecol.* **70**, 111–130 (2013).
111. Eggleston, E. M. & Hewson, I. Abundance of two *Pelagibacter ubique* bacteriophage genotypes along a latitudinal transect in the North and South Atlantic Oceans. *Front. Microbiol.* **7**, 1534. https://doi.org/10.3389/fmicb.2016.01534 (2016).
112. Brown, M. V. et al. Global biogeography of SAR11 marine bacteria. *Mol. Syst. Biol.* **8**, 595. https://doi.org/10.1038/msb.2012.28 (2012).
113. Kraemer, S., Ramachandran, A., Colatriano, D., Lovejoy, C. & Walsh, D. A. Diversity and biogeography of SAR11 bacteria from the Arctic Ocean. *ISME J.* **14**, 79–90. https://doi.org/10.1038/s41396-019-0499-4 (2020).
114. Villarreal-Chiu, J. F., Quinn, J. P. & McGrath, J. W. The genes and enzymes of phosphonate metabolism by bacteria, and their distribution in the marine environment. *Front. Microbiol.* **3**, 19. https://doi.org/10.3389/fmicb.2012.00019 (2012).
115. Carini, P., White, A. E., Campbell, E. O. & Giovannoni, S. J. Methane production by phosphate-starved SAR11 chemoheterotrophic marine bacteria. *Nat. Commun.* **5**, 4346. https://doi.org/10.1038/ncomms5346 (2014).
116. Sosa, O. A., Repeta, D. J., DeLong, E. F., Ashkezari, M. D. & Karl, D. M. Phosphate-limited ocean regions select for bacterial populations enriched in the carbon–phosphorus lyase pathway for phosphonate degradation. *Environ. Microbiol.* **21**, 2402–2414. https://doi.org/10.1111/1462-2920.14628 (2019).
117. Baker, P. et al. Potential contribution of surface-dwelling *Sargassum* algae to deep-sea ecosystems in the southern North Atlantic. *Deep Sea Res. Pt. II: Top. Stud. Oceanogr.* **148**, 21–34. https://doi.org/10.1016/j.dsr2.2017.10.002 (2018).
118. Abka-Khajouei, R. et al. Structures, properties and applications of alginates. *Mar. Drugs* **20**, 364. https://doi.org/10.3390/md20060364 (2022).
119. D'Souza, G. et al. Cell aggregation is associated with enzyme secretion strategies in marine polysaccharide-degrading bacteria. *ISME J.* **17**, 703–711. https://doi.org/10.1038/s41396-023-01385-1 (2023).
120. Hehemann, J.-H. et al. Adaptive radiation by waves of gene transfer leads to fine-scale resource partitioning in marine microbes. *Nat. Commun.* **7**, 1–10. https://doi.org/10.1038/ncomms12860 (2016).
121. Balmonte, J. P., Giebel, H. A., Arnosti, C., Simon, M. & Wietz, M. Distinct bacterial succession and functional response to alginate in the South, Equatorial, and North Pacific Ocean. *Environ. Microbiol.* **26**, e16594. https://doi.org/10.1111/1462-2920.16594 (2024).

122. Bunse, C., Koch, H., Breider, S., Simon, M. & Wietz, M. Sweet spheres: Succession and CAZyme expression of marine bacterial communities colonizing a mix of alginate and pectin particles. *Environ. Microbiol.* **23**, 3130–3148. https://doi.org/10.1111/1462-2920.15536 (2021).

Acknowledgements
ML and HMD are staff members of the National Research Council of Argentina (CONICET).

Author contributions
H.M.D. designed the study; M.L. and H.M.D. performed the bioinformatic analyses and prepared the figures and tables; H.M.D. wrote the original draft of the manuscript; M.L. and H.M.D reviewed and edited the manuscript. Both authors have read and agreed to the published version of the manuscript.

Funding
This work was supported by a grant from the National Research Council of Argentina-CONICET (PIP 2023-11220220100242CO).

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-92960-3.

Correspondence and requests for materials should be addressed to H.M.D.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.