Metagenomics unveils the attributes of the alginolytic guilds of sediments from four distant cold coastal environments

Marina N Matos[1], Mariana Lozada[1], Luciano E Anselmino[1], Matías A Musumeci[1], Bernard Henrissat[2,3,4], Janet K Jansson[5], Walter P Mac Cormack[6,7], JoLynn Carroll[8,9], Sara Sjöling[10], Leif Lundgren[11] and Hebe M Dionisi[1]*

[1]Laboratorio de Microbiología Ambiental, Centro para el Estudio de Sistemas Marinos (CESIMAR, CONICET), Puerto Madryn, U9120ACD, Chubut, Argentina

[2]Architecture et Fonction des Macromolécules Biologiques, CNRS, Aix-Marseille Université, 13288 Marseille, France

[3]INRA, USC 1408 AFMB, F-13288 Marseille, France

[4]Department of Biological Sciences, King Abdulaziz University, Jeddah, 21589, Saudi Arabia

[5]Earth and Biological Sciences Directorate, Pacific Northwest National Laboratory, Richland, WA 99352, USA

[6]Instituto Antártico Argentino, Ciudad Autónoma de Buenos Aires, C1064ABR, Argentina

[7]Instituto Nanobiotec, CONICET- Universidad de Buenos Aires, Ciudad Autónoma de Buenos Aires, C1113AAC, Argentina

[8]Akvaplan-niva, Fram – High North Research Centre for Climate and the Environment, NO-9296 Tromsø, Norway

[9]CAGE - Centre for Arctic Gas Hydrate, Environment and Climate, UiT The Arctic University of Norway, N-9037 Tromsø, Norway

[10]School of Natural Sciences and Environmental Studies, Södertörn University, 141 89 Huddinge, Sweden

[11]Stockholm University, SE-106 91 Stockholm, Sweden

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13433

This article is protected by copyright. All rights reserved.

Running title: Alginolytic guilds from cold sediments

***Correspondenc**e: Hebe M. Dionisi, Laboratorio de Microbiología Ambiental, Centro para el Estudio de Sistemas Marinos (CESIMAR, CONICET), Blvd. Brown 2915, U9120ACD, Puerto Madryn, Chubut, Argentina. Tel. +54-280-4883184 extension 1251; Fax +54-280-4883543. e-mail: hdionisi@cenpat-conicet.gob.ar.

Originality-Significance Statement. The alginate degradation potential of high-latitude coastal sediments and the gene pool related to this process, which has a key role in brown algae biomass decomposition, were comprehensively analyzed using a metagenomic approach.

Accepted Article

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13433

This article is protected by copyright. All rights reserved.

Summary

Alginates are abundant polysaccharides in brown algae that constitute an important energy source for marine heterotrophic bacteria. Despite the key role of alginate degradation processes in the marine carbon cycle, little information is available on the bacterial populations involved in these processes. The aim of this work was to gain a better understanding of alginate utilization capabilities in cold coastal environments. Sediment metagenomes from four high-latitude regions of both Hemispheres were interrogated for alginate lyase gene homolog sequences and their genomic context. Sediments contained highly abundant and diverse bacterial assemblages with alginolytic potential, including members of Bacteroidetes and Proteobacteria, as well as several poorly characterized taxa. The microbial communities in Arctic and Antarctic sediments exhibited the most similar alginolytic profiles, whereas brackish sediments showed distinct structures with a higher proportion of novel genes. Examination of the gene neighborhood of the alginate lyase homologs revealed distinct patterns depending on the potential lineage of the scaffolds, with evidence of evolutionary relationships among alginolytic gene clusters from Bacteroidetes and Proteobacteria. This information is relevant for understanding carbon fluxes in cold coastal environments and provides valuable information for the development of biotechnological applications from brown algae biomass.

Accepted Article

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13433

Wiley-Blackwell and Society for Applied Microbiology

This article is protected by copyright. All rights reserved.

Accepted Article

Introduction

Marine macroalgae present high photosynthesis and productivity rates even in highly challenging conditions such as those prevalent in polar environments, and they can contribute to long-term sedimentary carbon sequestration (Chung *et al.*, 2011; Wiencke and Amsler, 2012). Both macroalgae and the microorganisms specialized in the degradation of algal biomass are therefore key components of the carbon cycle in these rapidly changing environments (Arnosti *et al.*, 2013; Quartino *et al.*, 2013; Krause-Jensen and Duarte, 2014; Berlemont and Martiny, 2015). For instance, microbial communities from Arctic sediments are able to depolymerize a broad range of algal polysaccharides at high rates, acting as a final filter before carbon sequestration (Arnosti, 2008; Arnosti, 2014). Despite the key role of polysaccharide utilization for carbon fluxes in coastal ecosystems, the microbial populations participating in the degradation of algal polysaccharides and the mechanisms that they use are still poorly understood (Thomas *et al.*, 2012; Hehemann *et al.*, 2014; Wietz *et al.*, 2015). Metagenomics can provide unique insights into the polysaccharide utilization capabilities of these environments. Furthermore, this approach presents an opportunity for bioprospecting genes associated with polysaccharide utilization processes, which can be used for the development of biotechnological products from macroalgal biomass (Wargacki *et al.*, 2012; Lozada and Dionisi, 2015; Zhu and Yin, 2015).

Alginates are major components of the cell wall of brown algae, constituting up to 40% of their dry weight (Donati and Paoletti, 2009). They are linear polysaccharides composed of 1,4-linked β-D-mannuronate (M) and α-L-guluronate (G), arranged in Poly-M, Poly-G or Poly-MG blocks with 20-30 units each (Donati and Paoletti, 2009). The M/G ratio and the distribution of these blocks depend not only on the algal species, but also on the season, geographic location and type and age of the algal tissues (Kloareg and Quatrano, 1988; Truus *et al.*, 2001). The first step for the utilization of alginate as carbon source by marine bacteria is the depolymerization of these polysaccharides catalyzed by alginate lyases (AL; EC 4.2.2.3 and 4.2.2.11) through β-

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13433

This article is protected by copyright. All rights reserved.

elimination (Zhu and Yin, 2015). In the model mechanism proposed for the marine flavobacterium *Gramella forsetii* KT0803, endolytic AL enzymes attached to the outer membrane catalyze the depolymerization of alginates into oligoalginates (Kabisch *et al.*, 2014). The oligomers are then transferred through the outer membrane with the participation of TonB-dependent receptors (TBDR) and sugar-binding proteins. In the periplasm, the unsaturated oligosaccharides are further degraded to monomers by oligoalginate lyases and transported through the inner membrane, and in the cytoplasm a series of enzymes catalyze the conversion of the transported unsaturated monosaccharides into the metabolic intermediates glyceraldehyde-3-phosphate and pyruvate (Kabisch *et al.*, 2014). A similar process has been recently reported for the gammaproteobacterium *Saccharophagus degradans* 2-40 (Takagi *et al.*, 2016).

AL are remarkably diverse and they are currently classified into seven different polysaccharide lyase (PL) families, according to the classification scheme of the manually curated CAZy database (Carbohydrate-Active enZymes; Lombard *et al.*, 2014). The true diversity of these enzymes is only starting to emerge, as numerous AL sequences still remain unclassified or are yet to be included in the database (Garron and Cygler, 2010). Furthermore, the majority of the AL sequences known to date were identified in cultured bacteria, mostly belonging to the Flavobacteriia and Gammaproteobacteria classes (Zhu and Yin, 2015). Marine strains specialized in the utilization of alginates as substrate for growth contain several AL genes belonging to different PL families within their genomes (Thomas *et al.*, 2012; Mann *et al.*, 2013; Kabisch *et al.*, 2014; Neumann *et al.*, 2015, Takagi *et al.*, 2016), and the number of AL genes per genome has been shown to be related to its alginate degrading capability (Neumann *et al.*, 2015). Recently, studies that used culture-independent approaches started to shed light into the taxa potentially involved in alginate degradation in seawater (Wietz *et al*., 2015) and alginate gel particles (Mitulla *et al*., in press), the biogeographic distribution of AL sequences from *Alteromonas macleodii* ecotypes (Neumann *et al*., 2015), as well as features of alginate-

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13433

This article is protected by copyright. All rights reserved.

degrading consortia growing in anaerobic conditions (Seon *et al.*, 2014; Kita *et al.*, 2016). However, the alginate utilization potential of microbial communities from coastal environments and the genes involved in these processes still remain largely unknown (Neumann *et al.*, 2015).

Brown macroalgae are an important ecosystem component in cold coastal environments (Wiencke and Amsler, 2012). Part of their dead biomass is transferred to surface sediments, where heterotrophic bacteria participate in its decomposition (Hardison *et al.*, 2010). Due to the high alginate content in these macroalgae, we hypothesized that bacterial populations with the ability to use these polysaccharides as carbon source would be abundant in the microbial communities from sediments of these environments. The aim of this work was to increase our understanding of alginate utilization processes in cold coastal environments. Using a multi-level metagenomic approach, we characterized putative AL sequences and their genomic context in sediments from four distant high-latitude environments. The analysis of the gene pool related to alginate utilization suggested that these populations are both remarkably abundant and diverse, and include members of poorly described taxa. In addition, this work sheds light into evolutionary relationships of the alginolytic capabilities across taxa.

Results and Discussion

Study sites

Four high-latitude coastal environments were analyzed in this study (referred as sampling regions, Fig. 1 and Fig. S1). Polar habitats included Advent fjord (Spitsbergen Island, Svalbard Archipelago, Norway), and Potter Cove (25 de Mayo [King George] Island, South Shetland Islands, Antarctica). Subpolar habitats included Ushuaia Bay (Tierra del Fuego Island, Argentina) and Baltic Sea, the only brackish environment analyzed in this study (Table S1). Although many algae species present a bipolar distribution, the macroalgae communities are

This article is protected by copyright. All rights reserved.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13433

Wiley-Blackwell and Society for Applied Microbiology

This article is protected by copyright. All rights reserved.

and Amsler, 2012). The four sampling regions present some level of anthropogenic impact, such as eutrophication and/or chronic contamination with toxic pollutants. More information on these topics can be found in Supporting Information, General Features of the Study Sites. The nested sampling approach included triplicate subtidal sediment samples obtained at two sites in each of the four regions, and 23 of these samples produced the metagenomic dataset.

Abundance of genes encoding putative PL enzymes

Although Pfam domains are not suitable for the classification of CAZymes into specific PL families (Lombard *et al.*, 2014), the abundance of sequences containing domains related to PL enzymes in the gene pool of the metagenomes (relative to the abundance of single-copy genes) can provide an estimation of the prevalence of genes encoding AL and other PL enzymes in these environments. Three of the four Pfam domains found in AL enzymes, PF05426 (Alginate lyase), PF07940 (Heparinase II/III-like protein) and PF08787 (Alginate lyase 2), were abundant in the metagenomes, while PF14592 (Chondroitinase B) was present at lower abundances (Fig. 2). Two of these domains, PF07940 and PF08787, showed significant differences in relative gene abundance among sampling regions (Kruskal-Wallis test, $p = 0.004$ for both domains), which could reflect distinct structures of the alginolytic guilds or differences in the set of AL genes present in members of these guilds. The largest differences were found in the domain PF08787, identified in sequences of the PL7 and PL18 families, both encompassing only AL enzymes (Garron and Cygler, 2010). The relative abundance of this domain varied from 0.12 ± 0.01 copies in Baltic Sea sediments to 0.99 ± 0.28 copies in Antarctic sediments, per single-copy gene (Fig. 2). Pearson correlation analysis revealed a positive correlation between the relative abundance of sequences containing a PF08787 domain and salinity ($r = 0.67$, $p < 0.001$, n = 23). Although less marked, differences in the abundance of this domain were also observed between the two sampling sites from Ushuaia Bay (0.80 ± 0.21 copies in MC [samples ARG01-ARG03] and 0.33 ± 0.09 copies in OR [ARG04-ARG06], per single-copy

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13433

This article is protected by copyright. All rights reserved.

gene). Small differences in salinity were also detected at these sites, probably due to the influence of a freshwater runoff next to OR site (Table S1). A positive correlation was also found between the relative abundance of PF08787 domain and salinity when considering only the samples from Ushuaia Bay (Pearson correlation coefficient $r = 0.85$, $p = 0.032$, n = 6). These results suggest that salinity could be a factor influencing the abundance of AL genes from the PL7 and/or PL18 families in the sediment microbial communities. A previous study also detected a correlation between gene patterns (*alkB*) and salinity in the metagenomes from samples ARG01-06 (Guibert *et al*., 2016). It is important to notice, however, that other environmental factors, such as pollution, could also be influencing the observed gene distribution patters (Supporting Information, General Features of the Study Sites).

The overall abundance of AL genes in the metagenomes was estimated considering the domains PF05426, PF08787 and PF14592. As the latter is also present in chondroitinases, these values can be slightly overestimated. AL gene abundances were significantly higher in the three marine environments than in the Baltic Sea ($p = 0.036$ for all three post-hoc tests, Bonferroni corrected). Remarkably, the overall abundance of AL genes relative to single-copy genes in marine sediments ranged from 1.12 ± 0.18 copies in Svalbard to 1.78 ± 0.46 copies in Antarctica. The high potential for alginate depolymerization in marine sediments could be due to the selection of bacterial populations with this trait as a result of a high brown algae biomass production (Quartino and De Zaixso, 2008; Liuzzi *et al.*, 2011; Wiencke and Amsler, 2012). Sediments from the Arctic archipelago Svalbard were previously shown to depolymerize alginates at high rates, even under anaerobic conditions (Arnosti, 2008, Teske *et al.*, 2011). Although similar extracellular enzymatic activities have not been determined in Antarctic sediments, the high abundance of AL genes in this environment suggests a high potential for alginate depolymerization. Despite this potential, algal biomass decomposition is a slow process at the sampling location in Potter Cove, where insufficient mechanical break-down and low temperatures could be limiting the macroalgae degradation rate (Quartino *et al.*, 2008; Quartino

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13433

This article is protected by copyright. All rights reserved.

et al., 2015). The overall abundance of AL genes in sediments from the Baltic Sea was 0.66 ± 0.04 copies per single-copy gene, suggesting a lesser prevalence of microorganisms with alginolytic potential in sediments of this brackish environment, possibly due to a lower biomass of brown algae (Alexandridis *et al.*, 2012; General Features of the Study Sites in Supporting Information). Homologues of AL genes could remain undetected in this analysis if they are divergent from those used to define the Pfam domains.

A domain found in pectin-degrading enzymes (PF09492) was highly abundant in the sediments and clustered with the four domains from AL enzymes (top cluster, Fig. 2). Reports of pectinolytic marine microorganisms in the literature are scarce (Tuyen *et al.*, 2001), but the presence of organic carbon from terrestrial origin is common in coastal sediments (Wysocki *et al.*, 2008; Sikes *et al.*, 2009) and the ability to utilize plant polysaccharides might represent an advantage for microorganisms inhabiting near-shore environments. Polar and subpolar sediments could therefore represent promising sources for bioprospecting plant biomass-degrading enzymes, as previously suggested for other marine habitats (Vigneron *et al.*, 2014). Other protein families containing pectate lyases, on the other hand, were less abundant and formed a second cluster with domains from other PL enzymes (bottom cluster, Fig. 2).

Ordination and clustering of metagenomes based on shared alginate lyase homolog (ALH) sequences

We identified 2,705 sequences (≥100 amino acids) homologous to AL genes in the assembled metagenomes. In spite of the low level of read assembly in the metagenomes, 13.3% of the dataset were full-length sequences (for a full description of the dataset, see Characteristics of the Alginate Lyase Homolog Sequence Dataset in Supporting Information). As only seven sequences were identified in each of the metagenomes ARG04 and ARG06, these samples were excluded from further analyses. We evaluated the diversity of the 2,961 remaining ALH sequences by grouping them into operational protein units (OPUs) defined at 80% identity at the

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13433

This article is protected by copyright. All rights reserved.

amino acid level. We used a greedy incremental algorithm, which selects the longer sequences as seeds for OPU building (Fu *et al.*, 2012), to reduce the risk of overestimating the number of OPUs in the metagenomic dataset. This analysis identified 690 OPUs, suggesting a high diversity of microbial populations with the potential to depolymerize alginates in these sediments. On ordination plots, the samples retrieved from each region clustered together (Fig. 3), indicating the presence of distinct alginolytic guilds in the different coastal environments. Temperature and salinity correlated with the major ordination axes (Pearson's product-moment correlation coefficients r = 0.77 and 0.93, respectively; p = 0.001 for both correlations), suggesting that these environmental factors could play a role in the structuring of the alginolytic guilds. Similarly, the distribution pattern of sequences encoding chitinases, another enzyme targeting polysaccharides, was found to be strongly influenced by salinity, among other factors (Beier *et al*., 2010). Salinity is a key driver of the structuring of microbial communities, probably through selection processes (Herlemann *et al*., 2011; Thureborn *et al*., 2013). As this study did not include a salinity gradient, however, it is not possible to establish a direct link between the distribution of AL genes and salinity.

Ordination and its superimposed hierarchical clustering analysis showed that the ALH gene pools from both polar environments (Antarctica and Svalbard) were the most similar (Fig. 3). These pools were further related to those from Ushuaia Bay, a subantarctic marine environment. By contrast, the samples from the Baltic Sea, a brackish environment, clustered separately. These datasets shared only 3% and 3.2% of the OPUs with samples from Svalbard and Antarctica, respectively. This result is in agreement with the unique evolutionary lineages found in the Baltic Sea, due to its recent geological history, distinctive environmental conditions and geographic isolation, including both marine and freshwater microorganisms (Johannesson and André, 2006; Herlemann *et al.*, 2011). Despite the differences found in the ALH sequence datasets among sampling regions, four OPUs (representing 7.5% of the dataset) included sequences identified in metagenomes from the four sampling regions. The remarkable

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13433

This article is protected by copyright. All rights reserved.

similarities in this subset of ALH sequences, despite large geographic distances and differences in environmental conditions (Hanson *et al.*, 2012; Zinger *et al.*, 2014), suggests that some alginolytic populations from cold coastal environments could have a widespread geographic distribution. However, the possibility of recent horizontal gene transfer events cannot be excluded, and therefore these closely-related genes could be hosted by unrelated taxa.

Classification of the metagenomic ALH sequences

We assessed the level of relatedness between each metagenomic sequence and its closest CAZyme sequence (Lombard *et al.*, 2014), by comparing the E-values obtained using standalone blast (Altschul *et al.*, 1990) with all PL sequences as a custom database (Fig. 4A). Baltic Sea sediments contained a lower proportion of sequences with highly significant E-values ($< 1e^{-50}$) when compared with those from Antarctica and Svalbard (Kruskal-Wallis test and post-hoc analysis, Bonferroni corrected $p = 0.036$), suggesting the presence of a larger proportion of novel ALH sequences in brackish sediments. Similarly, when the metagenomic sequences were classified using internal CAZy tools (Cantarel *et al.*, 2009), a third of the sequences from the Baltic Sea samples and from ARG05 could not be classified, as they were too distantly related to CAZymes for a reliable assignment (Fig. 4B). Overall, 13% of the ALH sequences could not be classified. These sequences clustered into 184 OPUs (80% identity level), evidencing the high diversity of novel ALH sequences contained in these metagenomes. On the other hand, the majority of the ALH sequences were classified as belonging to the PL17 (30.5%), PL7 (28.2%) and PL6 (22.1%) families (Fig. 4B). AL genes from these families have previously been identified in the genomes of several marine alginolytic bacteria, often in more than one copy (Thomas *et al.*, 2012, Kabisch *et al.*, 2014; Takagi *et al.*, 2016). In the model suggested for the flavobacterium *Gramella forsetii* KT0803, PL7 and PL6 endo-acting AL are located on the outer membrane and PL17 oligoalginate lyases are within the periplasm (Kabisch *et al.*, 2014).

Interestingly, the proportion of sequences from the PL7 family was lower in the metagenomes

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13433

This article is protected by copyright. All rights reserved.

from Baltic Sea sediments and in the sample ARG05. In contrast, samples ARG01-ARG03, retrieved at a 500 m distance from sample ARG05, showed a distribution of PL families more similar to those from Svalbard and Antarctica. These results are in agreement with the lower abundance of its characteristic domain (PF08787) in the total metagenomes from Baltic Sea sediments and ARG04-06 samples, shown in Fig. 2.

Analysis of full-length ALH sequences

We performed a phylogenetic analysis based on the catalytic modules of classified full-length ALH sequences, sequences from the CAZy database and homologous sequences identified in bacterial genomes. Within each PL family, the metagenomic sequences were phylogenetically diverse, as they were widely distributed across different clusters of the trees. Fig. S3 shows the phylogenetic tree of the PL6 family, subfamily 1, and Fig. 5 shows two clusters selected from this tree. Similar clustering patterns were observed in other PL families (data not shown). Several clusters included sequences identified in the metagenomes from both polar environments, sharing high identity values (Fig. S3 and Fig. 5A). Often, these ALH sequences were related to AL genes described in members of the Bacteroidetes (e.g. *Winogradskyella*, *Polaribacter*, *Gramella*) or Proteobacteria (*Alteromonas*, *Pseudoalteromonas*). By contrast, ALH sequences from brackish sediments clustered separately, not only from other metagenomic sequences but also from sequences from bacterial strains (Fig. S3 and Fig. 5B). These results are in accordance with the ordination analysis of the full dataset (Fig. 3).

The gene neighborhoods of the ALH sequences often included additional ALH sequences identified in the dataset (Fig. 5). In addition, they contained other sequences potentially related to alginate utilization, previously described in alginolytic operons from flavobacteria (Thomas *et al.*, 2012; Kabisch *et al.*, 2014). Hypothetical proteins were also often found in these clusters. For instance, two hypothetical proteins located downstream of a putative

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13433

This article is protected by copyright. All rights reserved.

susC gene (which encodes an outer membrane protein associated with polysaccharide utilization, Tang *et al.*, 2012) in scaffolds from Baltic Sea sediments were also found in the genome of the Bacteroidetes *Rhodothermus marinus* DSM 4252 (Fig. 5B) and in the composite genome from an uncultured bacterium belonging to the Ignavibacteriae phylum (NCBI taxon ID 795747, data not shown). The gene order conservation of these three sequences in members of different phyla suggests that these hypothetical proteins could be related to alginate utilization.

A high level of gene order conservation was often found in scaffolds from the same sampling region containing very similar ALH sequences (Fig. 5). Partial shared synteny was also observed among scaffolds containing similar ALH sequences identified in distant sampling regions (Antarctica and Svalbard, Fig. 5A), providing additional evidence that bacterial populations with alginolytic potential could have a broad geographic distribution. Gene order conservation with genomes from bacterial isolates was less frequent, although in some cases the same sequences were found in the gene neighborhood of the ALH sequences, but in a different order (Fig. 5B). An exception was a series of scaffolds from Antarctica, which shared a remarkably conserved synteny with a section of the genome of *Psychromonas arctica* DSM 14288 (Fig. S4). The longest scaffold (~35 Kb) shared 80.8% identity at the nucleotide level with the genome of this gammaproteobacterium, which was isolated from seawater near Svalbard (Groudieva *et al.*, 2003). Although ALH sequences highly similar to those from *P. arctica* DSM 14288 could not be detected in the assembled metagenomes from the other three regions analyzed in this study, sequences sharing 99 - 100% identity with the 16S rRNA genes from this organism were identified in the metagenomes from the four regions (data not shown). Strains belonging to the genus *Psychromonas* have been isolated from several cold marine environments, and it has been proposed that ocean circulation linked to the sinking of cooled seawater in polar regions might facilitate the dispersion of this cold adapted microorganism (Nogi *et al.*, 2002).

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13433

This article is protected by copyright. All rights reserved.

Among the 359 full-length ALH sequences identified in the metagenomes, 50 could not be assigned to a CAZyme class or family. These full-length unclassified sequences formed 24 clusters using an 80% amino acid identity cut-off. In order to assess if these unclassified sequences could code for AL enzymes, we further analyzed a representative sequence from each cluster (Table S3). The majority of these sequences contained conserved Pfam domains characteristic of AL enzymes (mostly PF05426). Furthermore, often their closest matches in blastp searches against the non redundant protein NCBI database were AL sequences from bacterial strains or hypothetical proteins with similar Pfam domains, the majority sharing low to moderate identity values (Table S3). Four unclassified sequences containing a PF05426 domain were selected for *in silico* structure prediction. The three-dimensional modeling of these ALH sequences suggested that they could present a PL5-type fold, which is characteristic of sequences containing this Pfam domain (more information can be found in the Supporting Information, Three-dimensional Models of ALH Sequences; Fig. S5 and S6). Interestingly, in the structural models of the metagenomic sequences NOR15_100118034 and SWE02_100064912, the residues involved in the acid/base mechanism in AL enzymes and some of the amino acids involved in the interaction with the substrate were not identified in the equivalent position (Table S4). In contrast, the predicted structures of two consecutive unclassified ALH sequences from scaffold SWE02_10000683 (Fig. 5B and Fig. S6) showed a higher conservation of key residues (Table S4). These novel metagenomic sequences constitute promising targets for heterologous expression and enzymatic characterization.

Genomic context of ALH sequences in different taxonomic groups

Very limited information is available on the diversity and taxonomic identity of bacterial populations with alginolytic potential from marine environments (Wietz *et al.*, 2015). We analyzed the taxonomic assignment of scaffolds ≥ 4 kb in length that included ALH sequences, as longer sequences are able to provide more reliable taxonomic information (Patil *et al.*, 2012).

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13433

This article is protected by copyright. All rights reserved.

Although only part of the alginolytic guild was represented in this analysis, this set included 206 scaffolds covering a total of 2.2 Mb, and contained 319 ALH sequences (Table S5). We used two methods for the taxonomic binning, a sequence-similarity based method (IMG/M pipeline; Huntemann *et al.*, 2016) and a composition-based method (Patil *et al.*, 2012). The assignments were mostly in agreement at high taxonomic levels (90% at the phylum or superphylum level, Table S5). In fragments from the Baltic Sea, however, almost a third of the scaffolds were classified into different phyla with the two methods and a higher proportion of the scaffolds could not be classified at the phylum level, suggesting that these microorganisms are divergent from those represented in the databases. In the full dataset, twelve percent of the scaffolds could only be affiliated to Bacteria or could not be assigned using the similarity-based method (Table S5). Most of these scaffolds were classified into 10 different phyla using composition-based taxonomic binning, including Bacteroidetes and Proteobacteria, as well as poorly characterized taxa such as Ignavibacteriae, Cloacimonetes, Verrucomicrobia and Planctomycetes (Table S5 and Fig. S7). These phyla were relatively abundant in the sediment microbial communities (Fig. S8A).

Approximately half of the scaffolds binned within the Bacteroidetes phylum using both approaches, the majority of them within the Flavobacteriia class (Table S5 and Fig. S7). Bacteroidetes represented the second most abundant phylum in the sediment microbial communities (Fig. S8A), and the Flavobacteriia class was predominant within this phylum (>90%, Fig. S8B). On the other hand, between a quarter (sequence-similarity based method) and a third (composition-based method) of the scaffolds were assigned to the Proteobacteria phylum (Table S5 and Fig. S7). The majority of these scaffolds were identified in sediments from Antarctica, where two third of these scaffolds binned within the Gammaproteobacteria class and almost half were assigned to the order Alteromonadales. Members of the Gammaproteobacteria class were most abundant in microbial communities from sediments of Antarctica (Fig. S8C), and half of the members of this class belonged to the order Alteromonadales (data not shown).

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13433

This article is protected by copyright. All rights reserved.

These results suggest that members of the Bacteroidetes and Proteobacteria phyla could have an important role in the alginolytic guilds of cold coastal sediments. Gene clusters associated with alginate utilization have been identified in the genomes of several members of these phyla, mostly within the Flavobacteriia and Gammaproteobacteria classes (Thomas *et al.*, 2012). In a recent study that analyzed the structure and metabolic potential of an alginate-degrading anaerobic consortium enriched from coastal sediments, members of the Bacteroidetes and Firmicutes phyla were the dominant members, while populations belonging to the Gammaproteobacteria class were less abundant (Kita *et al.*, 2016). Conversely, only members of the Gammaproteobacteria class (in particular from the order Alteromonadales) were enriched in seawater microcosms after incubation with sodium alginate (Wietz *et al.*, 2015). In a more recent study, the same research group found a higher diversity of taxa colonizing and degrading alginate-gel particles in microcosms with seawater, including members of the Roseobacter clade (Alphaproteobacteria), as well as Cryomorphaceae (Flavobacteriia class) and Saprospiraceae (Sphingobacteriia) families (Mitulla *et al.*, in press). More studies are needed to identify the role of the environmental matrix on the structuring of the alginolytic guild in coastal marine environments. The capability for migration and surface attachment of certain groups could favor their colonization of certain niches such as marine snow or sediments (Cordero and Datta, 2016).

We further analyzed if the genomic context of the ALH sequences from the selected scaffolds varied among the different taxa, based on the potential function of the sequences (significant hits of Pfam domains; Finn *et al.*, 2013) located in codirectional gene clusters. Divergently transcribed genes, even if they were in some cases related to alginate metabolism (for instance *gntR*-like genes found adjacent but in opposite orientation), were not included in this analysis. Clustering analysis based on Bray-Curtis similarity index revealed a high similarity in the genomic context of ALH sequences in scaffolds from Antarctic and Svalbard metagenomes assigned to the Bacteroidetes phylum (Fig. S9). Scaffolds that binned within the

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13433

This article is protected by copyright. All rights reserved.

Bacteroidetes and Proteobacteria phyla from Antarctica, Svalbard and Baltic Sea metagenomes formed a larger cluster. These results support the proposed hypothesis of a common phylogenetic origin for these gene clusters, with Proteobacteria acquiring alginate utilization systems by several independent horizontal gene transfer events from ancestral marine flavobacteria (Thomas *et al.*, 2012). Scaffolds from Svalbard and Baltic Sea metagenomes within the lineage Bacteria also clustered together, despite the large differences in GC content between these two sets of scaffolds (0.56 ± 0.03 and 0.35 ± 0.02, respectively). This clustering was mainly due to a high number of CDS containing the PF00884 domain (Sulfatase) in a third of these scaffolds. Many polysaccharides used by marine bacteria as substrate for growth are sulfated, and removing these groups from the polymer facilitates their degradation (Gerken *et al.*, 2013).

In some of the scaffolds that binned within the Flavobacteriia class, ALH sequences were located within large alginolytic gene clusters, while in other scaffolds the gene neighborhood of ALH was not associated with alginate utilization. The genome of flavobacterial strains isolated from various marine habitats, including cold coastal environments, contain both alginolytic gene clusters and isolated AL genes (Thomas *et al.*, 2012; Mann *et al.*, 2013; Kabisch *et al.*, 2014; Inoue *et al.*, 2014; Xing *et al.*, 2015). The alginolytic gene clusters from the metagenomic scaffolds resembled the highly conserved gene order previously reported in flavobacterial alginolytic clusters (Thomas *et al.*, 2012; Kabisch *et al.*, 2014). However, these clusters contained many deletions, insertions and duplications (Fig. S10), suggesting a higher diversity of gene organizations within this group than currently recognized. Seven Pfam domains related to components of alginolytic clusters were detected in the gene neighborhood of the ALH sequences in the Bacteroidetes scaffolds from Antarctica, Svalbard and Baltic Sea, besides the four domains associated with AL enzymes (PF05426, PF07940, PF08787 and PF14592, Fig. S9): (i) PF00392 (GntR) and PF07729 (FCD), identified in transcriptional regulators of the subfamily FadR of the GntR family (Rigali *et al.*, 2002); (ii) PF07690 (MFS_1), membrane

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13433

This article is protected by copyright. All rights reserved.

transporters belonging to the major facilitator superfamily (Reddy *et al.*, 2012); (iii) PF07883 (Cupin_2), proteins with a cupin-like domain; (iv) PF00106 (adh_short) and PF13561 (adh_short_C2), two of the domains found in members of the large superfamily of short-chain dehydrogenases/reductases (SDR; Persson and Kallberg, 2013); and (v) PF00294 (PfkB), family of carbohydrate kinases that include the enzyme KdgK involved in the cytoplasmic processing of alginate monomers (Cabrera *et al.*, 2010). Another set of Pfam domains were identified in Bacteroidetes scaffolds only from the metagenomes from Antarctica and Svalbard: (i) PF00593 (TonB_dep_Rec) and PF07715 (Plug) from SusC-like proteins (Tang *et al.*, 2012); (ii) PF07980 (SusD) and PF14322 (SusD-like_3) from SusD-like proteins (Tang *et al.*, 2012) and (iii) PF00801 (PKD) in PKD-domain containing proteins, potentially involved in protein–protein interactions in the outer membrane (Bauer *et al.*, 2006). These proteins have been proposed to participate in the transport of oligoalginates through the outer membrane (Kabisch *et al.*, 2014).

The gene organization of the scaffolds from Proteobacteria was highly variable (Table S5), and the ALH sequences were found in short clusters of codirectional genes. These results are in agreement with the limited shared synteny and short length reported in alginolytic gene clusters from strains belonging to the Proteobacteria phylum (Thomas *et al.*, 2012; Neumann *et al.*, 2015). Like in scaffolds assigned to the Bacteroidetes phylum, PF07883 (Cupin_2) was identified in gene clusters from all three regions, suggesting that these proteins could play a key role in alginate degradation in members of both Bacteroidetes and Proteobacteria phyla. In *Z. galactanivorans*, the gene containing this domain (*kdgF*) was found to be differentially expressed in the presence of alginate (Thomas *et al.*, 2012). In addition, the product of this gene has been detected in members of these two phyla growing on alginate: a cupin 2 conserved barrel protein was one of the 23 alginate-specific proteins of the gammaproteobacterium *S. degradans* 2-40 (Takagi *et al.*, 2016), and the KdgF protein was detected in the cytoplasm of the flavobacterium *G. forsetii* KT0803 (Kabisch *et al.*, 2014). Although the function of this protein is not known, it has been recently suggested that it could participate in the first step of the

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13433

This article is protected by copyright. All rights reserved.

cytoplasmic processing (linearization) of the unsaturated monouronic acids (Lee, 2014), which has been proposed to be spontaneous (Thomas *et al.*, 2012). Further studies are needed to determine the role of KdgF in alginate degradation. On the other hand, domains PF00106 (adh_short), PF7715 (Plug) and PF01081 (Aldolase), a family that includes the enzyme phospho-2-dehydro-3-deoxygluconate aldolase (KdgA), were only identified in scaffolds from both polar regions.

Overall, 17% of the sequences from the gene clusters did not have significant Pfam hits. Although some of these sequences could have been incorrectly predicted, other sequences could be related to alginate degradation. For instance, a protein of unknown function encoded in an alginate utilization cluster has been detected in the cytoplasm in *G. forsetii* KT0803 (Kabisch *et al.*, 2014). The expression and characterization of these sequences will be needed to elucidate their role in alginate utilization processes.

Concluding Remarks

The exploitation of metagenomic data to gain ecological and mechanistic insights into the metabolic potential of microbial communities is often hindered due to the highly fragmented nature of this information, and this limitation is even more critical for bioprospecting efforts (Lozada and Dionisi, 2015). In this work, we used different levels of resolution to analyze a highly-complex metagenomic dataset, capitalizing on the available information and accounting for the limitations imposed by the inefficient and uneven read assembly. Evidences of high abundance and diversity of bacteria with the potential to utilize alginates detected in sediments from geographically-distant cold coastal environments suggest that this process is not only highly relevant, but more complex than previously assumed. Although a fraction of the putative alginate lyase gene pool was shared among regions, a highly distinctive guild seems to be

This article has been Accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13433

This article is protected by copyright. All rights reserved.

play an important role in the structuring of these guilds. Besides members of the Bacteroidetes and Proteobacteria phyla, this study also revealed several largely unknown groups of bacteria with the potential to depolymerize alginates in polar and subpolar environments. In particular, sediments of the Baltic Sea represent an attractive target for further exploration, such as the isolation or single-cell genomic analysis of novel polysaccharide-degrading bacteria. The analysis of the genomic context of selected scaffolds confirmed at the community level evolutionary patterns proposed based on comparative genomic analyses (Thomas *et al.*, 2012). The differences in gene organization detected in alginolytic gene clusters suggest recent gene rearrangements in the genome of these microorganisms. These observations raise fundamental questions regarding the dispersal and adaptative capabilities of these microorganisms. Polar marine ecosystems are highly vulnerable to local and global anthropogenic impacts. The effects of climate-induced changes on microbial community structure and function, and their implications in ecosystem function, are still unknown. Identifying the microorganisms and understanding the mechanisms involved, as well as the environmental factors influencing alginate utilization processes, are a prerequisite for modeling the functioning of these rapidly changing ecosystems.

Experimental Procedures

Sample and metadata collection

Four high-latitude coastal environments were analyzed in this study (referred in the text as sampling regions, Fig 1, Table S1 and Fig. S1): (i) Advent Fjord, Spitsbergen, Svalbard Archipelago, Norway (NOR); (ii) Port Värtahamnen, Stockholm, Baltic Sea, Sweden (SWE); (iii) Ushuaia Bay, Tierra del Fuego Island, Argentina (ARG); (iv) Potter Cove, 25 de Mayo (King George) Island, Antarctica (ANT). Triplicate sediment samples (top 5 cm) were collected at each sampling site as described previously (Teeling *et al.*, 2012).

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13433

This article is protected by copyright. All rights reserved.

were collected using cores, and environmental parameters, including temperature, depth and salinity, were determined *in situ* using CTD or multiparameter instruments (Table S1). Sediment samples were stored at -80ºC.

DNA extraction and metagenomic sequencing

Metagenomic DNA was extracted from the sediment samples as previously reported (Mackelprang *et al.*, 2011). Shotgun sequencing of metagenomic DNA was performed using Illumina HiSeq 2000 platform (2×150-bp paired end reads, one lane per sample), at the facilities of the Joint Genome Institute, USA. The dataset includes 23 metagenomes (Table S1), which were annotated using the IMG/M pipeline (Markowitz *et al.*, 2014). The number of reads and the assembly efficiency varied among samples, with <35% of the reads mapping the scaffolds (Table S1). The metagenomes are freely accessible at the IMG server (https://img.jgi.doe.gov/) under accession numbers 3300000118-3300000136, 3300000241-3300000243, and 3300000792.

Relative abundance of Polysaccharide Lyase (PL) sequences

Twelve Pfam domains contained in sequences from different PL families and 12 domains from single-copy genes that code for ribosomal proteins (Kunin *et al.*, 2008) were selected to estimate the relative abundance of PL sequences in the metagenomes (Table S2). The estimated copies of genes containing these domains in the total metagenomes were retrieved from the IMG/M system using the Abundance Profile Overview tool including both unassembled and assembled metagenomes, the latter corrected by read depth. The estimated copies of genes containing Pfam domains from PL were normalized by dividing by the estimated copies of each of the single-copy genes and the values obtained were averaged. Only the domains PF05426, PF08787 and PF14592 were used to estimate the overall abundance of AL genes in the metagenomes, to avoid counting twice members the PL17 family (which contains domains

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13433

This article is protected by copyright. All rights reserved.

PF07940 and PF5426). PF14592 can be detected in both AL and chondroitinases, and therefore the overall abundance could be slightly overestimated. To test for significant differences in gene abundance among sampling regions, Kruskal-Wallis tests were performed using SPSS v. 15 (α = 0.05 after Bonferroni correction for multiple comparisons). Clustering analyses based on Bray Curtis similarity were performed using R package *vegan* (https://cran.r-project.org/web/packages/vegan/index.html).

Identification and analysis of alginate lyase homolog (ALH) sequences

ALH sequences were identified in the assembled metagenomes using blastp (threshold E-value of $1e^{-5}$; Altschul *et al.*, 1990) with sequences representing the PL families containing AL sequences as query: GenBank accession numbers AEW23144, YP_004791784, YP_004791784, AFC88009, WP_007214684, NP_627697, YP_006746939, YP_002800319, YP_001347824, ACN56743, YP_003118157, YP_004738545, BAB03312, BAH79131, NP_624710, BAB19127, AGE48774, BAE81787, ADE10038, NP_357573, YP_528751, ACB87607 and YP_003656270. Protein coding sequences (CDS) with *alginate lyase* as product name were also retrieved. Duplicated sequences identified by both methods or by different query sequences were eliminated. Only sequences $\geq$ 100 amino acids were analyzed. ALH sequences were grouped into operational protein units (OPUs) defined at 80% identity at the amino acid level, using CD-HIT (Fu *et al.*, 2012). Ordination (NMDS) and clustering using Bray-Curtis similarity based on Wisconsin standardized OPU data, and environmental fitting to the metadata matrix were performed in R package *vegan*.

The level of divergence between each metagenomic sequence and the closest PL sequence of the CAZy database (http://www.cazy.org, as of February 2014) was assessed using standalone blast (Altschul *et al.*, 1990), through the E-value of the match. ALH sequences were classified at the family level with the CAZy pipeline, which uses a combination of blast and HMM tools and further subjected to manual curation (Lombard *et al.*, 2014). For the

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13433

This article is protected by copyright. All rights reserved.

phylogenetic analysis, only full-length metagenomic sequences were selected, which represented approximately 13% of the dataset. Phylogenetic trees were built for each PL family, including CAZymes and homologous sequences identified in bacterial genomes deposited at the IMG database (http://img.jgi.doe.gov/) using blastp searches. The PL module of the sequences was detected using dbCAN (Yin *et al.*, 2012), aligned using default parameters in ClustalX 2.1 (Larkin *et al.*, 2007), and manually curated and trimmed. Phylogenetic trees were built using the maximum likelihood algorithm in Mega (Tamura *et al.*, 2011) using WAG (PL6 SF1) or LG (PL17 SF2) substitution models, the most appropriate model in each case. The stability of tree topology was evaluated by bootstrap resampling with 500 replications.

Analysis of scaffolds containing ALH sequences

Phylogenetic lineage, lineage percentage, GC content, gene content and gene organization of the scaffolds were analyzed based on the information obtained from the functional annotation of the metagenomes (Markowitz *et al.*, 2014). Additionally, PhylopythiaS (Patil *et al.*, 2012) was used for the taxonomic binning of the scaffolds. For comparative analysis, gene organization of bacterial strains was retrieved from genomes deposited at the IMG database (Markowitz *et al.*, 2012). For the statistical analysis of gene neighborhood of ALH sequences, a Pfam search (Finn *et al.*, 2013) was performed for all the sequences located codirectional to the identified ALH sequences in scaffolds ≥4 kb. Clustering analysis of the gene content information discriminated by region and by lineage was performed based on Bray-Curtis similarity index calculated for the gene per sample matrix, standardized by Wisconsin double standardization, using R package *vegan*.

Microbial community structure

The analysis of the sediment microbial community structure was based on 16S rRNA gene information obtained from the metagenomes. The relative abundance of blast hits (90% identity) for bacterial sequences in the IMG/M system

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13433

This article is protected by copyright. All rights reserved.

using the tool Phylogenetic Distribution of 16S rRNA Genes for the unassembled and assembled metagenomes.

Acknowledgments

ML, MAM and HML are staff members from The National Research Council of Argentina (CONICET). MNM is a postdoctoral fellow from CONICET. The metagenomic dataset was generated at the Department of Energy-Joint Genome Institute (DOE-JGI) under the Community Sequencing Program (CSP proposal ID 328, project IDs 403959, 404206, 404777-404782, 404786, 404788- 404801). HD and ML were supported by grants from CONICET (Nº 112-200801-01736) and The National Agency for the Promotion of Science and Technology of Argentina (ANPCyT PICT2008 Nº 0468). WMC was supported by grants from the the University of Buenos Aires (UBA 2014-2017 20020130100569BA), the European Commission through the Marie Curie Action IRSES IMCONet (project N° 318718), the Argentinean Antarctic Institute and ANPCyT (PICTO 2010 N° 0124). JKJ was supported by the Pacific Northwest National Laboratory under Contract DE-AC05-76RLO1830. BH was supported by Agence Nationale de la Recherche, grant BIP:BIP (ANR-10-BINF-03-04). JC's research contribution is supported by the Research Council of Norway (grant No. 223259). We thank Krystle Chavarria for her technical support and Ricardo Vera and Horacio Ocariz for their help in sample collection. We would like to dedicate this work to our late colleagues Horacio Ocariz and Leif Lundgren, who left us too early.

The authors declare no conflict of interest.

References

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13433

This article is protected by copyright. All rights reserved.

Alexandridis, N., Oschlies, A., and Wahl, M. (2012). Modeling the effects of abiotic and biotic factors on the depth distribution of *Fucus vesiculosus* in the Baltic Sea. *Mar Ecol-Prog Ser* **463:** 59-72.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. *J Mol Biol* **215**: 403-410.

Arnosti, C. (2008). Functional differences between Arctic seawater and sedimentary microbial communities: contrasts in microbial hydrolysis of complex substrates. *FEMS Microbiol Ecol* **66:** 343-351.

Arnosti, C. (2014). Patterns of microbially driven carbon cycling in the ocean: links between extracellular enzymes and microbial communities. *Adv Oceanog* **2014:** ID 706082.

Arnosti, C., Bell, C., Moorhead, D., Sinsabaugh, R., Steen, A., Stromberger, M., *et al.* (2013). Extracellular enzymes in terrestrial, freshwater, and marine environments: perspectives on system variability and common research needs. *Biogeochemistry* **117:** 5-21.

Bauer, M., Kube, M., Teeling, H., Richter, M., Lombardot, T., Allers, E. *et al.* (2006). Whole genome analysis of the marine Bacteroidetes *'Gramella forsetii'* reveals adaptations to degradation of polymeric organic matter. *Environ Microbiol* **8:** 2201-2213.

Berlemont, R., and Martiny, A.C. (2015). Genomic potential for polysaccharide deconstruction in Bacteria. *Appl Environ Microbiol* **81:** 1513-1519.

Beier S, Jones CM, Mohit V, Hallin S, Bertilsson S (2010). Global phylogeography of chitinase genes in aquatic metagenomes. *Appl Environ Microbiol* **77:** 1101-1106.

Cabrera, R., Babul, J., and Guixé, V. (2010). Ribokinase family evolution and the role of conserved residues at the active site of the PfkB subfamily representative, Pfk-2 from *Escherichia coli*. *Arch Biochem Biophys* **502:** 23-30.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13433

This article is protected by copyright. All rights reserved.

Cantarel, B.L., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009). The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res* **37:** D233-D238.

Chung, I.K., Beardall, J., Mehta, S., Sahoo, D., and Stojkovic, S. (2011). Using marine macroalgae for carbon sequestration: a critical appraisal. *J Appl Phycol* **23:** 877-886.

Cordero, O.X., and Datta, M.S. (2016). Microbial interactions and community assembly at microscales. *Curr Opin Microbiol* **31:** 227-234.

Donati, I., and Paoletti, S. (2009). Material properties of alginates. *Alginates: biology and applications*. Springer. pp 1-53.

Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R. *et al.* (2013). Pfam: the protein families database. *Nucleic Acids Res* **42:** D222-D230.

Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28:** 3150-3152.

Garron, M.-L., and Cygler, M. (2010). Structural and mechanistic classification of uronic acid-containing polysaccharide lyases. *Glycobiology* **20:** 1547-1573.

Gerken, H.G., Donohoe, B., and Knoshaug, E.P. (2013). Enzymatic cell wall degradation of *Chlorella vulgaris* and other microalgae for biofuels production. *Planta* **237:** 239-253.

Guibert, L.M., Loviso, C.L., Borglin, S., Jansson, J., Dionisi, H.M. and Lozada, M. (2016). Diverse bacterial groups contribute to the alkane degradation potential of chronically-polluted Subantarctic coastal sediments. *Microb Ecol* **71**:100-112.

Groudieva, T., Grote, R., and Antranikian, G. (2003). *Psychromonas arctica* sp. nov., a novel psychrotolerant, biofilm-forming bacterium isolated from Spitzbergen. *Int J Syst Evol Microbiol* **53:** 539-545.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13433

This article is protected by copyright. All rights reserved.

Hanson, C.A., Fuhrman, J.A., Horner-Devine, M.C., and Martiny, J.B. (2012). Beyond biogeographic patterns: processes shaping the microbial landscape. *Nat Rev Microbiol* **10:** 497-506.

Hardison, A.K., Canuel, E.A., Anderson, I.C., and Veuger, B. (2010). Fate of macroalgae in benthic systems: carbon and nitrogen cycling within the microbial community. Mar Ecol Prog Ser **414:** 41-55.

Hehemann, J.-H., Boraston, A.B., and Czjzek, M. (2014). A sweet new wave: structures and mechanisms of enzymes that digest polysaccharides from marine algae. *Curr Opin Struct Biol* **28:** 77-86.

Herlemann, D.P.R., Labrenz, M., Jurgens, K., Bertilsson, S., Waniek, J.J., and Andersson, A.F. (2011). Transitions in bacterial communities along the 2000-km salinity gradient of the Baltic Sea. *ISME J* **5:** 1571-1579.

Huntemann, M., Ivanova, N.N., Mavromatis, K., Tripp, H.J., Paez-Espino, D., Tennessen, K. *et al*. (2016). The standard operating procedure of the DOE-JGI Metagenome Annotation Pipeline (MAP v.4). *Stand Genomic Sci* **11:** 17.

Inoue, A., Takadono, K., Nishiyama, R., Tajima, K., Kobayashi, T., and Ojima, T. (2014). Characterization of an alginate lyase, Flalya, from *Flavobacterium* sp. strain UMI-01 and its expression in *Escherichia coli*. *Mar Drugs* **12:** 4693-4712.

Johannesson, K., and André, C. (2006). Life on the margin: genetic isolation and diversity loss in a peripheral marine ecosystem, the Baltic Sea. *Molec Ecol* **15:** 2013-2029.

Kabisch, A., Otto, A., König, S., Becher, D., Albrecht, D., Schüler, M. *et al.* (2014). Functional characterization of polysaccharide utilization loci in the marine Bacteroidetes *Gramella forsetii* KT0803. *ISME J* **8:** 1492-1502.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13433

This article is protected by copyright. All rights reserved.

Kita, A., Miura, T., Kawata, S., Yamaguchi, T., Okamura, Y., Aki, T. *et al.* (2016). Bacterial community structure and predicted alginate metabolic pathway in an alginate-degrading bacterial consortium. *J Biosci Bioeng.* **121:** 286–292.

Kloareg, B., and Quatrano, R. (1988). Structure of the cell walls of marine algae and ecophysiological functions of the matrix polysaccharides. *Oceanography and Marine Biology: an Annual Review.* **26:** 259-315.

Krause-Jensen, D., Duarte, C.M. (2014). Expansion of vegetated coastal ecosystems in the future Arctic. *Front Mar Sci* **1:** 77.

Kunin, V., He, S., Warnecke, F., Peterson, S.B., Martin, H.G., Haynes, M. *et al.* (2008). A bacterial metapopulation adapts locally to phage predation despite global dispersal. *Genome Res* **18:** 293-297.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H. *et al.* (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* **23:** 2947-2948.

Lee, S.H. (2014). Biochemical and structural characterization of a novel enzyme involved in uronic acid metabolism. Department of Biochemistry and Microbiology, Master of Science. University of Victoria: Victoria, Australia.

Liuzzi, M.G, Gappa, J.L., and Piriz, M.L. (2011). Latitudinal gradients in macroalgal biodiversity in the Southwest Atlantic between 36 and 55°S. *Hydrobiologia* **673:** 205-214.

Lombard, V., Ramulu, H.G., Drula, E., Coutinho, P.M., and Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* **42:** D490-D495.

Lozada, M., and Dionisi, H.M. (2015). Microbial Bioprospecting in Marine Environments. In: Kim S-K (ed). *Hb25_Springer Handbook of Marine Biotechnology*. Springer. pp 307-326.

Accepted Article

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13433

This article is protected by copyright. All rights reserved.

Mackelprang, R., Waldrop, M.P., DeAngelis, K.M., David, M.M., Chavarria, K.L., Blazewicz, S.J. *et al.* (2011). Metagenomic analysis of a permafrost microbial community reveals a rapid response to thaw. *Nature* **480:** 368-371.

Mann, A.J., Hahnke, R.L., Huang, S., Werner, J., Xing, P., Barbeyron, T. *et al.* (2013). The genome of the alga-associated marine flavobacterium *Formosa agariphila* KMM 3901T reveals a broad potential for degradation of algal polysaccharides. *Appl Environ Microbiol* **79:** 6813-6822.

Markowitz, V.M., Chen, I.-M.A., Palaniappan, K., Chu, K., Szeto, E., Grechkin, Y. *et al.* (2012). IMG: the integrated microbial genomes database and comparative analysis system. *Nucleic Acids Res* **40:** D115-D122.

Markowitz, V.M., Chen, I.-M.A., Chu, K., Szeto, E., Palaniappan, K., Pillay, M. *et al.* (2014). IMG/M 4 version of the integrated metagenome comparative analysis system. *Nucleic Acids Res* **42**: D568-D573.

Mitulla, M., Dinasquet, J., Guillemette, R., Simon, M., Azam, F., and Wietz, M. (2016). Response of bacterial communities from California coastal waters to alginate particles and an alginolytic Alteromonas macleodii strain. *Environ Microbiol*, in press. DOI: 10.1111/1462-2920.13314.

Neumann, A.M., Balmonte, J.P., Berger, M., Giebel, H.-A., Arnosti, C., Voget, S. *et al.* (2015). Different utilization of alginate and other algal polysaccharides by marine *Alteromonas macleodii* ecotypes. *Environ Microbiol* **17**: 3857-3868.

Nogi, Y., Kato, C., and Horikoshi, K. (2002). *Psychromonas kaikoae* sp. nov., a novel from the deepest piezophilic bacterium cold-seep sediments in the Japan Trench. *Int J Syst Evolut Microbiol* **52:** 1527-1532.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13433

This article is protected by copyright. All rights reserved.

Patil, K.R., Roune, L., and McHardy, A.C. (2012). The PhyloPythiaS web server for taxonomic assignment of metagenome sequences. *PLoS ONE* **7:** e38581.

Persson, B., and Kallberg, Y. (2013). Classification and nomenclature of the superfamily of short-chain dehydrogenases/reductases (SDRs). *Chem-Biol Interact* **202:** 111-115.

Quartino, M., and De Zaixso, A.B. (2008). Summer macroalgal biomass in Potter Cove, South Shetland Islands, Antarctica: its production and flux to the ecosystem. *Polar Biol* **31:** 281-294.

Quartino, M.L., de Zaixso, A.B., and Momo, F.R. (2008). Macroalgal production and the energy cycle of Potter Cove. In: Wiencke, C., Ferreyra, G.A., Abele, D. and Marenssi, S. (eds). The Antarctic ecosystem of Potter Cove, King-George Island (Isla 25 de Mayo). Synopsis of research performed 1999-2006 at the Dallmann Laboratory and Jubany Station. pp 68-74.

Quartino, M.L., Deregibus, D., Campana, G.L., Latorre, G.E.J., and Momo, F.R. (2013). Evidence of macroalgal colonization on newly ice-free areas following glacial retreat in Potter Cove (South Shetland Islands), Antarctica. *PLoS ONE* **8:** e58223.

Quartino, M.L., Vazquez, S.C., Latorre, G.E., and Mac Cormack, W.P. (2015). Possible role of bacteria in the degradation of macro algae *Desmarestia anceps* Montagne (Phaeophyceae) in Antarctic marine waters. *Rev Arg Microbiol* **47**: 274–276.

Reddy, V.S., Shlykov, M.A., Castillo, R., Sun, E.I., and Saier, M.H. (2012). The major facilitator superfamily (MFS) revisited. *FEBS J* **279:** 2022-2035.

Rigali, S.B., Derouaux, A., Giannotta, F., and Dusart, J. (2002). Subdivision of the helix-turn-helix GntR family of bacterial regulators in the FadR, HutC, MocR, and YtrA subfamilies. *J Biol Chem* **277:** 12507-12515.

Accepted Article

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13433

Wiley-Blackwell and Society for Applied Microbiology

This article is protected by copyright. All rights reserved.

Seon, J., Lee, T., Lee, S.C., Pham, H.D., Woo, H.C., and Song, M. (2014). Bacterial community structure in maximum volatile fatty acids production from alginate in acidogenesis. *Biores Technol* **157**: 22-27.

Sikes, E.L., Uhle, M.E., Nodder, S.D., and Howard, M.E. (2009). Sources of organic matter in a coastal marine environment: evidence from n-alkanes and their ^{13}C distributions in the Hauraki Gulf, New Zealand. *Mar Chem* **113:** 149-163.

Takagi, T., Morisaka, H., Aburaya, S., Tatsukami, Y., Kuroda, K., and Ueda, M. (2016). Putative alginate assimilation process of the marine bacterium *Saccharophagus degradans* 2-40 based on quantitative proteomic analysis. *Mar Biotechnol* **18**: 15-23.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molec Biol Evol* **28:** 2731-2739.

Tang, K., Jiao, N., Liu, K., Zhang, Y., and Li, S. (2012). Distribution and functions of TonB-dependent transporters in marine bacteria and environments: implications for dissolved organic matter utilization. *PLoS ONE* **7:** e41204.

Teske, A., Durbin, A., Ziervogel, K., Cox, C., and Arnosti, C. (2011). Microbial community composition and function in permanently cold seawater and sediments from an Arctic fjord of Svalbard. *Appl Environ Microbiol* **77:** 2008-2018.

Thomas, F., Barbeyron, T., Tonon, T., Génicot, S., Czjzek, M., and Michel, G. (2012). Characterization of the first alginolytic operons in a marine bacterium: from their emergence in marine Flavobacteriia to their independent transfers to marine Proteobacteria and human gut Bacteroides. *Environ Microbiol* **14:** 2379-2394.

Accepted Article

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13433

This article is protected by copyright. All rights reserved.

Thureborn P, Lundin D, Plathan J, Poole AM, Sjöberg B-M, Sjöling S (2013). A metagenomics transect into the deepest point of the Baltic Sea reveals clear stratification of microbial functional capacities. *PLoS ONE* **8:** e74983.

Truus, K., Vaher, M., and Taure, I. (2001). Algal biomass from *Fucus vesiculosus* (Phaeophyta): investigation of the mineral and alginate components. *Proc Estonian Acad Sci Chem* **50:** 95-103.

Tuyen, H., Helmke, E., and Schweder, T. (2001). Cloning of two pectate lyase genes from the marine Antarctic bacterium *Pseudoalteromonas haloplanktis* strain ANT/505 and characterization of the enzymes. *Extremophiles* **5:** 35-44.

Vigneron, A., Cruaud, P., Roussel, E.G., Pignet, P., Caprais, J.-C., Callac, N. *et al.* (2014). Phylogenetic and functional diversity of microbial communities associated with subsurface sediments of the Sonora Margin, Guaymas Basin. *PLoS ONE* **9**: e104427.

Wargacki, A.J., Leonard, E., Win, M.N., Regitsky, D.D., Santos, C.N.S., Kim, P.B. *et al.* (2012). An engineered microbial platform for direct biofuel production from brown macroalgae. *Science* **335:** 308-313.

Wiencke, C., and Amsler, C.D. (2012). Seaweeds and their communities in polar regions. *Seaweed Biology*. Springer. pp 265-291.

Wietz, M., Wemheuer, B., Simon, H., Giebel, H.-A., Seibt, M.A., Daniel, R. *et al.* (2015). Bacterial community dynamics during polysaccharide degradation at contrasting sites in the Southern and Atlantic Oceans. *Environ Microbiol* 17: 3822-3831.

Wysocki, L.A., Filley, T.R., and Bianchi, T.S. (2008). Comparison of two methods for the analysis of lignin in marine sediments: CuO oxidation versus tetramethylammonium hydroxide (TMAH) thermochemolysis. *Org Geochem* **39:** 1454-1461.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13433

This article is protected by copyright. All rights reserved.

Xing, P., Hahnke, R.L., Unfried, F., Markert, S., Huang, S., Barbeyron, T. *et al.* (2015). Niches of two polysaccharide-degrading *Polaribacter* isolates from the North Sea during a spring diatom bloom. *ISME J* **9:** 1410-1422.

Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F., and Xu, Y. (2012). dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res* **40:** W445-W451.

Zhu, B., and Yin, H. (2015). Alginate lyase: Review of major sources and classification, properties, structure-function analysis and applications. *Bioengineered* **6**: 125-131.

Zinger, L., Boetius, A., and Ramette, A. (2014). Bacterial taxa-area and distance-decay relationships in marine environments. *Molec Ecol* **23:** 954-964.

Accepted Article

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13433

This article is protected by copyright. All rights reserved.

Figure Legends

Figure 1. Geographic location of the analyzed sampling regions. NOR, Advent Fjord, Spitsbergen, Svalbard Archipelago, Norway; SWE, Port Värtahamnen, Stockholm, Baltic Sea, Sweden; ARG, Ushuaia Bay, Tierra del Fuego Island, Argentina; ANT, Potter Cove, 25 de Mayo (King George) Island, Antarctica. Two sampling sites distanced ~500 m were selected within each coastal environment, and the top 5 cm of subtidal sediments were sampled in triplicate within each site.

Figure 2. Relative abundance of sequences containing Pfam domains from polysaccharide lyases (PL) in the sediment metagenomes. The estimated copies of sequences containing each of the 12 selected Pfam domains in the unassembled and assembled metagenomes (the latter corrected by read depth) was normalized by dividing by the estimated copies of 12 selected single-copy genes (Table S2). Diameters depict the average value of the 12 calculated ratios. The coefficient of variation across all Pfam domains and samples was <18%. For each PL domain, the Pfam accession number followed by the family name abbreviation between parentheses is indicated. NOR, Svalbard Archipelago (green); SWE, Baltic Sea (orange); ARG, Ushuaia Bay (red); ANT, Potter Cove (blue). Cluster analysis was performed using Bray-Curtis index.

Figure 3. Similarity of metagenomes based on shared alginate lyase homolog (ALH) sequences. Results of ordination analysis (non-metric multidimensional scaling, NMDS) and hierarchical clustering are superimposed in the plot. Samples NOR02, NOR05 and NOR08 (sampling site 1), and samples NOR13, NOR15 and NOR18 (sampling site 2) were retrieved from Advent Fjord, Svalbard Archipelago, Norway. Samples SWE02, SWE07 and SWE12 (KBA site), as well as samples SWE21 and SWE26 (KBB site) were retrieved from the Baltic Sea. Samples ARG01-ARG03 (MC site) and ARG05 (OR site) were retrieved from Ushuaia Bay, Argentina. Samples ARG04 and ARG06 were eliminated from the analysis due to the low

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13433

Wiley-Blackwell and Society for Applied Microbiology

This article is protected by copyright. All rights reserved.

number of identified ALH sequences. ANT01-ANT03 (sampling site 1) and ANT04-ANT06 (sampling site 2), were retrieved from Potter cove, 25 de Mayo (King George) Island, Antarctica (Table S1). Sequences were grouped into operational protein units (OPUs) defined at 80% identity at the amino acid level, and the *sample x OPUs* matrix was subjected to Wisconsin double standardization and used for the construction of the distance matrix, using 1-(Bray-Curtis index) as dissimilarity measure. NMDS stress: 0.09. Circles with filled lines, dotted lines and sparse dotted lines show three grouping levels in clustering analysis, corresponding to 0.8, 0.85 and 0.87 dissimilarity, respectively.

Figure 4. Characteristics of the metagenomic ALH sequence dataset. (A) Evaluation of the level of divergence between ALH sequences and CAZymes. Distribution of E-values obtained with their respective best hits in a blastp analysis, grouped in E-value ranges. (B) Classification of the sequences in CAZy families. *Others:* PL12, PL21, and GH. NC: not classified (too distantly related for a reliable assignment). N: number of identified ALH sequences in each assembled metagenome (with at least 100 amino acids).

Figure 5. Phylogenetic analysis of full-length ALH sequences and scaffold gene organization. Two clusters selected from the phylogenetic tree of ALH sequences from the PL6 family are shown in parts (A) and (B) of the figure (the complete tree is available as Fig. S3). For the metagenomic sequences, the name of the scaffolds are shown in distinct colors depending on the sampling region (green, Svalbard Archipelago; orange, Baltic Sea; blue, Antarctica), and the last number/s in gray correspond to the gene number within the scaffold. Isolates are indicated in bold fonts when the AL sequence was included in the CAZy database, and in normal fonts for sequences from genomes that are not included in the database. The ALH sequences used to build the phylogenetic tree are shown in red in the scaffolds on the right of the figure. The gray rectangle indicates the PL6 catalytic module used for constructing the tree. The rest of the colors in the sequences of the scaffolds indicate orthologous groups (top COG hit,

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13433

This article is protected by copyright. All rights reserved.

shown in Fig. S4), except those shown in light yellow that had no COG assignment. Short names of the sequences are indicated in selected scaffolds: *radC*, DNA repair protein RadC; *rraA*, regulator of RNase E activity RraA; *xyll*, xylose isomerase-like TIM barrel; *pl*, PL; *PDG*, 6-phosphogluconate dehydrogenase; *nramp*, Mn^{2+} and Fe^{2+} transporters of the NRAMP family; *kdgF*, cupin domain; *susC*, outer membrane receptor protein; *susD*, SusD family; *pkd*, PKD-domain containing protein; *hyp*, hypothetical protein; *kdgK,* 2-dehydro-3-deoxygluconokinase; *gntR*, GntR family transcriptional regulator; *sdr*, short-chain dehydrogenase/reductase superfamily; *mfs*, major facilitator superfamily transporter; *siae*, sialate O-acetylesterase; *nc*, non-classified ALH; *gh2*, glycoside hydrolase of the GH2 family; tp, long-chain fatty acid transport protein; *G8*, G8-domain containing protein; *kdgA*, 2-keto-3-deoxy-phosphogluconate aldolase; *WD40*, WD40 domain protein beta propeller; *senC*, electron transport protein/SenC; *hk*, histidine kinase; *cgs*, cystathionine gamma-synthase.

Accepted Article

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13433

Wiley-Blackwell and Society for Applied Microbiology

This article is protected by copyright. All rights reserved.

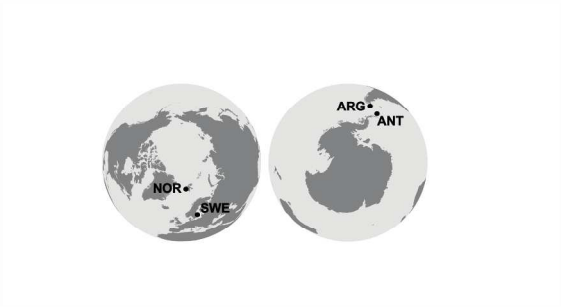

Figure 1. Geographic location of the analyzed sampling regions. NOR, Advent Fjord, Spitsbergen, Svalbard Archipelago, Norway; SWE, Port Värtahamnen, Stockholm, Baltic Sea, Sweden; ARG, Ushuaia Bay, Tierra del Fuego Island, Argentina; ANT, Potter Cove, 25 de Mayo (King George) Island, Antarctica. Two sampling sites distanced ~500 m were selected within each coastal environment, and the top 5 cm of subtidal sediments were sampled in triplicate within each site.
297x420mm (300 x 300 DPI)

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13433

This article is protected by copyright. All rights reserved.

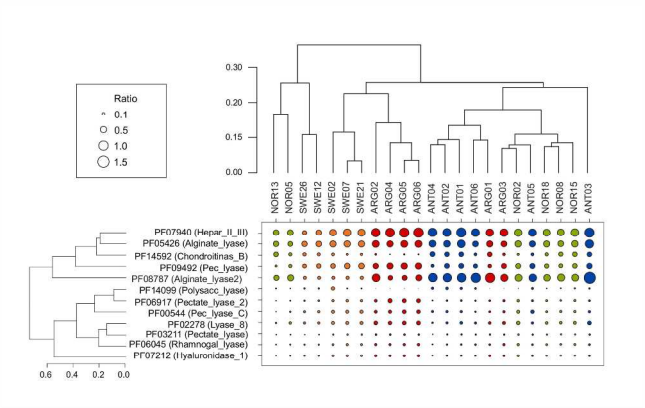

Figure 2. Relative abundance of sequences containing Pfam domains from polysaccharide lyases (PL) in the sediment metagenomes. The estimated copies of sequences containing each of the 12 selected Pfam domains in the unassembled and assembled metagenomes (the latter corrected by read depth) was normalized by dividing by the estimated copies of 12 selected single-copy genes (Table S2). Diameters depict the average value of the 12 calculated ratios. The coefficient of variation across all Pfam domains and samples was <18%. For each PL domain, the Pfam accession number followed by the family name abbreviation between parentheses is indicated. NOR, Svalbard Archipelago (green); SWE, Baltic Sea (orange); ARG, Ushuaia Bay (red); ANT, Potter Cove (blue). Cluster analysis was performed using Bray-Curtis index.
297x420mm (300 x 300 DPI)

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13433

This article is protected by copyright. All rights reserved.

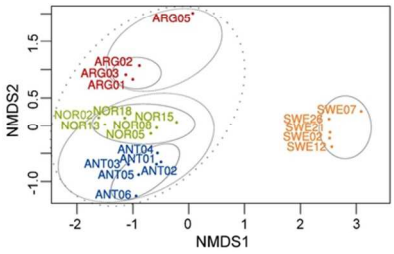

Figure 3. Similarity of metagenomes based on shared alginate lyase homolog (ALH) sequences. Results of ordination analysis (non-metric multidimensional scaling, NMDS) and hierarchical clustering are superimposed in the plot. Samples NOR02, NOR05 and NOR08 (sampling site 1), and samples NOR13, NOR15 and NOR18 (sampling site 2) were retrieved from Advent Fjord, Svalbard Archipelago, Norway. Samples SWE02, SWE07 and SWE12 (KBA site), as well as samples SWE21 and SWE26 (KBB site) were retrieved from the Baltic Sea. Samples ARG01-ARG03 (MC site) and ARG05 (OR site) were retrieved from Ushuaia Bay, Argentina. Samples ARG04 and ARG06 were eliminated from the analysis due to the low number of identified ALH sequences. ANT01-ANT03 (sampling site 1) and ANT04-ANT06 (sampling site 2), were retrieved from Potter cove, 25 de Mayo (King George) Island, Antarctica (Table S1). Sequences were grouped into operational protein units (OPUs) defined at 80% identity at the amino acid level, and the sample x OPUs matrix was subjected to Wisconsin double standardization and used for the construction of the distance matrix, using 1-(Bray-Curtis index) as dissimilarity measure. NMDS stress: 0.09. Circles with filled lines, dotted lines and sparse dotted lines show three grouping levels in clustering analysis, corresponding to 0.8, 0.85 and 0.87 dissimilarity, respectively.
88x37mm (300 x 300 DPI)

Accepted

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13433

This article is protected by copyright. All rights reserved.

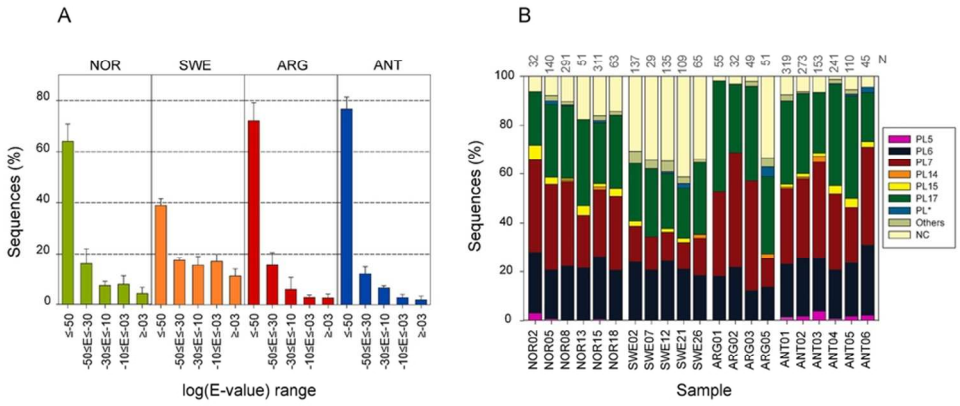

Figure 4. Characteristics of the metagenomic ALH sequence dataset. (A) Evaluation of the level of divergence between ALH sequences and CAZymes. Distribution of E-values obtained with their respective best hits in a blastp analysis, grouped in E-value ranges. (B) Classification of the sequences in CAZy families. Others: PL12, PL21, and GH. NC: not classified (too distantly related for a reliable assignment). N: number of identified ALH sequences in each assembled metagenome (with at least 100 amino acids).
85x41mm (300 x 300 DPI)

Accepted Article

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13433

This article is protected by copyright. All rights reserved.

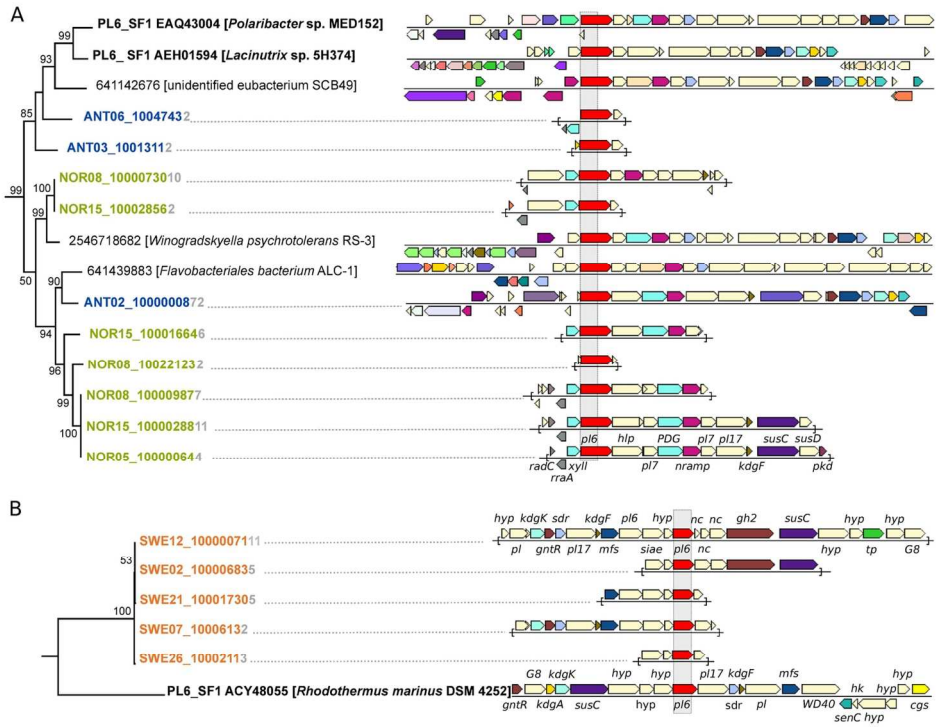

Figure 5. Phylogenetic analysis of full-length ALH sequences and scaffold gene organization. Two clusters selected from the phylogenetic tree of ALH sequences from the PL6 family are shown in parts (A) and (B) of the figure (the complete tree is available as Fig. S3). For the metagenomic sequences, the name of the scaffolds are shown in distinct colors depending on the sampling region (green, Svalbard Archipelago; orange, Baltic Sea; blue, Antarctica), and the last number/s in gray correspond to the gene number within the scaffold. Isolates are indicated in bold fonts when the AL sequence was included in the CAZy database, and in normal fonts for sequences from genomes that are not included in the database. The ALH sequences used to build the phylogenetic tree are shown in red in the scaffolds on the right of the figure. The gray rectangle indicates the PL6 catalytic module used for constructing the tree. The rest of the colors in the sequences of the scaffolds indicate orthologous groups (top COG hit, shown in Fig. S4), except those shown in light yellow that had no COG assignment. Short names of the sequences are indicated in selected scaffolds: radC, DNA repair protein RadC; rraA, regulator of RNase E activity RraA; xyll, xylose isomerase-like TIM barrel; pl, PL; PDG, 6-phosphogluconate dehydrogenase; nramp, Mn2+ and Fe2+ transporters of the NRAMP family; kdgF, cupin domain; susC, outer membrane receptor protein; susD, SusD family; pkd, PKD-domain containing protein; hyp, hypothetical protein; kdgK, 2-dehydro-3-deoxygluconokinase; gntR, GntR family transcriptional regulator; sdr, short-chain dehydrogenase/reductase superfamily; mfs, major facilitator superfamily transporter; siae, sialate O-acetylesterase; nc, non-classified ALH; gh2, glycoside hydrolase of the GH2 family; tp, long-chain fatty acid transport protein; G8, G8-domain containing protein; kdgA, 2-keto-3-deoxy-phosphogluconate aldolase; WD40, WD40 domain protein beta propeller; senC, electron transport protein/SenC; hk, histidine kinase; cgs, cystathionine gamma-synthase.
138x108mm (300 x 300 DPI)

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13433

This article is protected by copyright. All rights reserved.