Microb Ecol
DOI 10.1007/s00248-017-1028-5

ENVIRONMENTAL MICROBIOLOGY

Metagenomic Analysis of Subtidal Sediments from Polar and Subpolar Coastal Environments Highlights the Relevance of Anaerobic Hydrocarbon Degradation Processes

Fernando Espínola[1] · Hebe M. Dionisi[1] · Sharon Borglin[2] · Colin J. Brislawn[3] · Janet K. Jansson[3] · Walter P. Mac Cormack[4,5] · JoLynn Carroll[6] · Sara Sjöling[7] · Mariana Lozada[1]

Received: 6 May 2017 / Accepted: 27 June 2017

Abstract In this work, we analyzed the community structure and metabolic potential of sediment microbial communities in high-latitude coastal environments subjected to low to moderate levels of chronic pollution. Subtidal sediments from four low-energy inlets located in polar and subpolar regions from both Hemispheres were analyzed using large-scale 16S rRNA gene and metagenomic sequencing. Communities showed high diversity (Shannon's index 6.8 to 10.2), with distinct phylogenetic structures (<40% shared taxa at the Phylum level among regions) but similar metabolic potential in terms of sequences assigned to KOs. Environmental factors (mainly salinity, temperature, and in less extent organic pollution) were drivers of both phylogenetic and functional traits. Bacterial taxa correlating with hydrocarbon pollution included families of anaerobic or facultative anaerobic lifestyle, such as Desulfuromonadaceae, Geobacteraceae, and Rhodocyclaceae. In accordance, biomarker genes for anaerobic hydrocarbon degradation (*bamA*, *ebdA*, *bcrA*, and *bssA*) were prevalent, only outnumbered by *alkB*, and their sequences were taxonomically binned to the same bacterial groups. BssA-assigned metagenomic sequences showed an extremely wide diversity distributed all along the phylogeny known for this gene, including *bssA* sensu stricto, *nmsA*, *assA*, and other clusters from poorly or not yet described variants. This work increases our understanding of microbial community patterns in cold coastal sediments, and highlights the relevance of anaerobic hydrocarbon degradation processes in subtidal environments.

Keywords Cold environments · Subtidal sediments · Hydrocarbons · Anaerobic biodegradation · Community structure · Metagenomics · Biomarker genes

Electronic supplementary material The online version of this article (doi:10.1007/s00248-017-1028-5) contains supplementary material, which is available to authorized users.

✉ Mariana Lozada
lozada@cenpat-conicet.gob.ar

[1] Laboratorio de Microbiología Ambiental, Centro para el Estudio de Sistemas Marinos (CESIMAR, CONICET), Centro Nacional Patagónico, Blvd. Brown 2915, U9120ACD, Puerto Madryn, Chubut, Argentina

[2] Energy Geosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

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

[4] Instituto Nanobiotec, Facultad de Farmacia y Bioquímica, Universidad de Buenos Aires, CONICET, Buenos Aires, Argentina

[5] Instituto Antártico Argentino, Buenos Aires, Argentina

[6] Akvaplan-niva, Fram—High North Research Centre for Climate and the Environment, and ARCEx—Research Centre for Arctic Petroleum Exploration, Department of Geosciences, UiT The Arctic University of Norway, N-9037 Tromsø, Norway

[7] School of Natural Sciences and Environmental Studies, Södertörn University, Huddinge, Sweden

Introduction

High-latitude regions are increasingly suffering the effects of various environmental stressors derived directly or indirectly from human activities. Climate change, as well as local human activities, are rapidly affecting these environments, with unknown effects at a global scale [1]. Of particular concern is the exploration and exploitation of oil reserves in Arctic and Subantarctic regions, as well as the growth in scientific and tourism activities in Antarctica, which leads to higher risks of oil pollution in these sensitive environments [2]. Despite

Published online: 12 July 2017

similar present environmental conditions, high-latitude regions from both Hemispheres have different geological, oceanographic, and climatic histories, which have influenced their biological history [3]. This is also true for microbial life: besides being tightly coupled to their environment, microorganisms are known to display biogeographic patterns resulting in unique assemblages at each site [3–5]. Therefore, an in-depth study of the microbial communities inhabiting comparable polar or subpolar regions in the Northern and Southern Hemispheres can provide insights into the ways in which microorganisms adapt to these environments and how they cope with anthropogenic impacts.

Coastal sediments provide important ecosystem services, playing significant roles in nutrient cycling, as well as carbon turnover and sequestration [6]. These habitats receive hydrocarbon inputs from various sources (biogenic, diagenetic, petrogenic, pyrogenic) through multiple mechanisms such as oil release, land runoff, urban discharges, and atmospheric deposition. Hydrocarbons are highly hydrophobic molecules with tendency to adsorb to the sediment matrix, reducing their bioavailability and persisting for several years, in particular in fine-grained sediments [7, 8]. In sediments, hydrocarbon biodegradation is a complex process due to the high diversity of microbial populations involved, interacting among themselves and with the environment in a complex fashion [9]. Subtidal habitats are of special concern, as they become anoxic at low depths due to respiration of organic matter [10]. Anaerobic hydrocarbon biodegradation, less understood at the molecular level than its aerobic counterpart, is most likely to occur under these conditions [11–13]. Over the last few years, a series of studies analyzed the response of sediment microbial communities to acute pollution events such as the Deepwater Horizon spill [14–19] and the Prestige oil spill [12, 20]. The integration of these results produced a wealth of knowledge that can be applied to marine environments in general and to coastal sediments in particular [21–23]. Hydrocarbon-degrading populations have been less well studied in environments receiving low or moderate levels of chronic pollution, and existing studies have been mostly limited to the analysis of phylogenetic and functional marker genes [24–29]. Recently, the metagenomic analysis of harbor coastal sediments from the Mediterranean basin revealed that temperature is a major environmental factor controlling hydrocarbon-degrading potential [30].

The aim of this work was to uncover the phylogenetic structure and functional adaptations to both natural environmental factors and anthropogenic stressors of sediment microbial communities from cold coastal environments. To achieve this goal, we carried out deep metagenomic and amplicon sequencing analyses in four distant polar or subpolar regions from both Hemispheres chronically exposed to different levels of anthropogenic activity. The results of this work provide insights into the diversity and composition of the microbial communities inhabiting sediments at remote, high latitude locations, and uncover relationships between community structure, metabolic potential, and environmental variables. In addition, this work indicates specific taxa potentially associated with hydrocarbon biodegradation in these environments, and reveals the abundance and diversity of biomarker genes for these processes.

Methods

Sampling and Metadata Collection

Four high-latitude coastal environments were analyzed in this study: Adventfjord, Svalbard Archipelago, Norway (NOR); Port Värtahamnen, Stockholm, Sweden (SWE); Ushuaia Bay, Tierra del Fuego Island, Argentina (ARG); and Potter Cove, 25 de Mayo (King George) Island, Antarctic Peninsula (ANT) (Fig. 1a). Two stations per region distanced approximately 500 m were sampled. At each station, three core samples were obtained, separated by 3–5 m (Fig. S0, Table S1). Surficial (top 5 cm) subtidal (10 to 50 m bathymetry) sediments were sampled with cores. Metadata obtained in situ included temperature, depth, conductivity, and salinity using conductivity-temperature-depth (CTD) or multiparameter instruments. Sediment samples were stored at −80 °C until use in chemical or molecular analyses. Hydrocarbons and other organic compounds were determined by gas chromatography–mass spectrometry (GC/MS), as previously described [16]. When more than one fraction of a core sample was used for hydrocarbon analysis, concentration values were averaged (Table S1). Measured compounds included aliphatic hydrocarbons, polycyclic aromatic hydrocarbons (PAHs), and other organic compounds (Table S2).

Illumina 16S rRNA Gene Amplicon Sequencing (I-Tags)

Metagenomic DNA was purified using a previously reported methodology [31], performing one to five extractions per core (Table S3). Large-scale sequencing of 16S rRNA gene amplicons was performed as part of the international microbial diversity initiative Earth Microbiome Project (www.earthmicrobiome.org/) [32]. Gene fragments were amplified using primers 515F (5′GTGCCAGCMGCCGCGGTA3′) and 806R (5′GGACTACHVGGGTWTCTAAT3′), designed to amplify prokaryotes (bacteria and archaea), according to standard EMP Protocols (http://www.earthmicrobiome.org/emp-standard-protocols/16s/). Sequence datasets were obtained from all samples, with the exception of two samples from Ushuaia Bay (ARG01 and ARG02) and one from Potter Cove (ANT01) which failed in the amplification/sequencing process (Table S3). Preprocessing was performed according to Caporaso et al. [33]. Briefly, reads were truncated immediately after the first Q3 quality score, and truncated reads which resulted in 75% or less of the original sequence length were

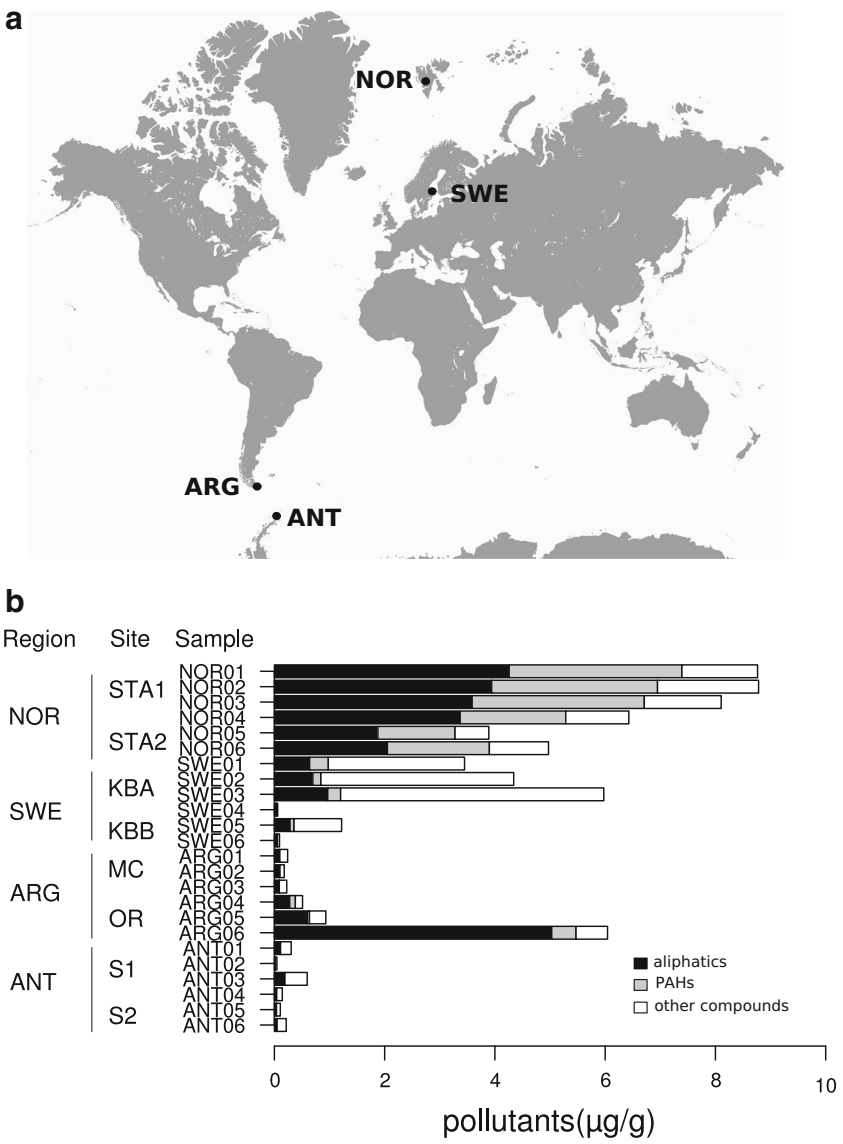


Fig. 1 **a** Geographic location of the four analyzed sampling regions. *NOR* Adventfjord, Spitsbergen, Svalbard Archipelago, Norway; *SWE* Port Värtahamnen, Stockholm, Baltic Sea, Sweden; *ARG* Ushuaia Bay, Tierra del Fuego Island, Argentina; *ANT* Potter Cove, 25 de Mayo (King George) Island, Antarctic Peninsula. Two sites distanced ~500 m were selected within each coastal environment and sampled in triplicate. For details, see Fig. S0 and Table S1. **b** Organic compounds content of the sediment samples. Values are in expressed in micrograms per gram of wet sediment

eliminated. Sequences with ambiguous bases were also eliminated. Closed-reference operational taxonomic unit (OTU) picking was performed at 97% identity threshold in QIIME 1.9.1 against the Greengenes reference database pre-clustered at 97% identity [34]. In addition, an open reference OTU picking strategy was carried out, in which reads that did not have hits with reference sequences using the closed-reference strategy were clustered de novo. Sequences assigned to chloroplasts and mitochondria were eliminated. The resulting OTU table was rarefied to more than 130,000 sequences per sample, and processed in QIIME to build OTU taxonomic assignment tables.

Bioinformatic and Statistical Analyses of I-Tags

Diversity (Shannon) and richness (Chao1) estimators were calculated from the resulting OTU tables in QIIME. Ordinations based on Bray-Curtis similarity index and weighted UniFrac were performed for OTU tables in R package *vegan* (www.r-project.org/vegan). When more than one community structure dataset was available for a core sediment sample (Table S3), corresponding values were averaged to be included in ordinations and other analyses. Environmental fitting against ordinations was performed with the *envfit* function in *vegan*. For heatmaps construction from the OTU table, data were transformed by square root and standardized using Wisconsin method in R. Canonical correspondence analysis for OTUs assigned at the family level was also performed using *vegan*. Individual correlations of bacterial families were calculated using Pearson's product-moment correlation in R environment. Analyses of the 97% OTUs were performed on the totality of the data, while analyses at the Class and Family level were carried out considering taxa with relative abundances ≥1%.

Shotgun Sequencing and Functional Annotation of Metagenomes

Shotgun sequencing of sediment 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 [35]. A total of 23 metagenomes were obtained from the 24 core samples (one of the Baltic Sea samples failed, Table S3). Both the assembled and unassembled fractions of the metagenomes were annotated through the Integrated Microbial Genomes with Microbiome Samples pipeline (IMG/M, https://img.jgi.doe.gov/m/) [36].

Bioinformatic and Statistical Analyses of Metagenomes

KEGG Orthology (KO) Terms and Pathways system [37] was used to compare the metagenomes in terms of their potential functions. The number of coding sequences (CDS) assigned to KOs were downloaded from IMG/M for all metagenomes, including assembled and unassembled fractions, with the option "estimated gene copies," which corrects for read depth in assembled metagenomes. Next, KO abundances were normalized by dividing the total count for each KO term by the total number of reads assigned to KOs in each metagenome dataset, in order to account for differences in sequencing depth. For module- and pathway-level analyses, abundances of KOs were summed for each pathway level (level 2, general categories; level 3, pathways; level 4, modules) to calculate the relative abundances per sample. Abundances were further normalized by dividing by the total sum of abundances for each sample, in order to account for KOs that map to more than one pathway. For comparative analyses with metagenomes from different environments, KO tables were downloaded from IMG/M for a number of publicly available metagenomes (for accession numbers, see Fig. S6). KO abundances were assigned to pathways as described above, analyzed by constructing distance matrices based on these functional categories, and further performing non-metric multidimensional scaling (NMDS) ordination of metagenomes in *vegan*.

Statistical analyses for testing significant differences among pathways from the metagenomes of this work (Kruskal-Wallis test) were performed with STAMP software, using eta-squared as effect size estimator [38]. Multiple test correction was performed by Benjamini-Hochberg false-discovery rate test, and post hoc comparisons were performed by Games-Howell method [38]. Regularized canonical correlation analysis (RCCA) based on KEGG modules or pathways considering environmental and chemical metadata was performed using R package *CCA* (www.r-project.org/CCA). A version of the partial Mantel's test developed in *vegan*, which finds the correlation between two matrices conditioned on a third matrix and tests its significance by permutation [39], was used to test the hypothesis that the biological (metagenomic) data were significantly correlated with the measured environmental variables, controlling for the variation attributed to geographic distance, and vice versa.

For biomarker gene analysis, the estimated gene copies of KOs from key enzymes related to aerobic and anaerobic hydrocarbon degradation were obtained from the KO table. These were K00496 (AlkB, alkane-1-monooxygenase, [EC:1.14.15.3]), K14579 (NdoB, alpha subunit of naphthalene 1,2-dioxygenase, [EC:1.14.12.12]), K11943 (NidA, large subunit of PAH dioxygenase, [EC:1.13.11.-]), K14599 (DbfA, alpha subunit of dibenzofuran dioxygenase, first step in aerobic degradation of fluorene, [EC:1.14.12.-]), K00446 (XylE, catechol 2,3-dioxygenase, [EC:1.13.11.2], catechol meta-cleavage), K00450 (GtdA, gentisate 1,2-dioxygenase, [EC:1.13.11.4]), K15760 (TmoA, protein A of toluene monooxygenase system, [EC:1.14.13.-]), K03268 (TodC1, alpha subunit of benzene/toluene/chlorobenzene dioxygenase, [EC:1.14.12.3 1.14.12.11 1.14.12.-]), K15757 (XylM, hydroxylase subunit of xylene monooxygenase, [EC:1.14.13.-]), K14748 (EtbAa; alpha subunit of ethylbenzene dioxygenase, [EC:1.14.12.-]), K07540 (BssA, alpha subunit of benzylsuccinate synthase, [EC:4.1.99.11], participating in anaerobic toluene degradation), K10700 (EbdA, alpha subunit of ethylbenzene dehydrogenase, [EC:1.17.99.2]), K04114 (BcrA, subunit A of benzoyl-CoA reductase, [EC:1.3.7.8], catalyzing the central benzoyl-CoA degradation step shared by toluene, ethylbenzene, and metylnaphthalene degradation pathways), and K07539 (BamA or Oah, 6-oxo-cyclohex-1-ene-carbonyl-CoA hydrolase, [EC:3.7.1.-], catalyzing the ring cleavage step, also central in anaerobic degradation of toluene, ethylbenzene, and metylnaphthalene) [40]. To calculate relative abundances, estimated gene copies (assembled and unassembled fractions of the metagenomes as calculated by IMG) were normalized dividing by the total sequences assigned to KOs in each sample. Individual correlations with chemical compounds and other statistical analyses were performed in R environment. Taxonomic assignments of sequences corresponding to biomarker genes were performed by BLAST search against in-house built reference databases, followed by Least Common Ancestor algorithm (MEGAN, v. 5.11.3) with default parameters, except for min_score 35 (recommended for short sequences) and top_percent 10 [41]. In addition, the metagenomic sequences assigned to the KOs K00496 (AlkB) and K07540 (BssA) were further phylogenetically placed onto a reference tree containing representative sequences. In the case of AlkB, reference sequences were obtained from a previous analysis performed by Nie et al. [42], while for BssA, deduced amino acid sequences, including homologous genes *bssA* (sensu stricto), *assA*, and *nmsA* from pure cultures and environmental surveys, were selected [43]. The analysis was performed using Evolutionary Placement Algorithm, with RAxML-EPA, which assigns short non-overlapping sequences independently on to a fixed phylogeny by maximum likelihood [44].

Data Availability

All amplicon sequencing data have been made public through the QIITA portal under Study ID 1198 (www.microbio.me/emp, https://qiita.ucsd.edu/study/description/1198) as well as through ENA (Study ERP016557, samples ERS1262253-ERS1262313). The metagenomes are available at IMG (https://img.jgi.doe.gov/) under accession numbers 3300000118–3300000136, 3300000241–3300000243, and 3300000792.

Results and Discussion

Characteristics of the Study Sites and Sediment Samples

The four regions analyzed in this work are polar or subpolar coastal environments located in the Northern or Southern Hemispheres (Fig. 1a): (1) Adventfjord, Svalbard Archipelago, Norway (NOR) (polar); (2) Port Värtahamnen, Baltic Sea, Sweden (SWE) (subpolar); (3) Ushuaia Bay, Tierra del Fuego Island, Argentina (ARG) (subpolar); and (4) Potter Cove, 25 de Mayo (King George) Island, Antarctica (ANT) (polar). Svalbard, Ushuaia Bay, and Potter Cove are marine environments, while the Baltic Sea is a brackish environment (Table 1, Table S1). Although most of these regions are located in remote locations with low population densities, they are exposed to variable levels of chronic pollution due to anthropogenic activities. The sediment samples, obtained in triplicate at two sites per region, showed differences in organic pollutant levels (Table 1, Table S1). Up to two orders of magnitude differences were found in hydrocarbon pollution levels among samples, although they were only moderately polluted (<10 μg/g of hydrocarbons, Fig. 1b).

Adventfjord, Svalbard Archipelago, is located near Longyearbyen, a small but active town whose main economic activities are divided between tourism, coal mining, and research infrastructures. High PAH, chlorinated hydrocarbons, and BTEX concentrations have been observed at this site [45, 46]. In accordance, samples obtained in this region were the most polluted (Fig. 1b). In contrast, samples retrieved from Potter Cove, near an Antarctic scientific base, were the least polluted. This site is only slightly impacted by anthropogenic activities, yet it is exposed to occasional accidental diesel spills [47].

The Baltic Sea is a heavily impacted area as a result of multiple anthropogenic activities. Samples were obtained in Port Värtahamnen, in Stockholm, Sweden, which is among the most polluted areas within this enclosed water body [48]. A distinct pattern in organic compound distribution was observed in these sediment samples in which a variety of organic compounds apart from hydrocarbons were present at higher levels (Fig. 1b, for details of the measured compounds see Table S2). The samples from this region also presented low oxygen levels at the sediment–water interface (Table S1). Samples from the two stations in Ushuaia Bay, Argentina, were very different in their pollution pattern: near the Orion Plant jetty used for loading and offloading refined petroleum products (OR site), the sediments showed higher and more variable hydrocarbon levels than those next to the city commercial pier (MC site). Chronic hydrocarbon pollution has previously been detected in Ushuaia Bay sediments, in particular near the Orion Plant jetty [7, 26]. The sheltered nature of Ushuaia Bay shores results in low-energy conditions that might prevent pollutant dispersion [49]. A more detailed description of the sampling sites is available elsewhere [35].

Diversity

A total of 61 million 16S rRNA gene sequences were obtained from all of the analyzed samples, accounting for more than 130,000 sequences per dataset (Table S3) that were used to calculate diversity estimators and to analyze community composition. At a cutoff of 10^5 sequences, coverage ranged from 98.9 ± 0.0% (SWE) to 99.6 ± 0.1% (ANT). Shannon diversity index ranged from 6 to 8.7 using a conservative approach (closed reference OTU picking). When an open reference OTU picking strategy was applied [50], a higher diversity was estimated in all cases (Shannon index from 6.8 to 10.2, with coverage values ranging from 96.9 ± 0.1% to 99.2 ± 0.2%) (Fig. 2). The estimated diversity is among the highest observed for microbial communities, and comparable

Table 1 General characteristics of the sediments by region

Region	Abbreviated name	Temperature (°C)	Salinity (g/L)	Total organic pollutants (μg/g ws)
Adventfjord, Svalbard, Norway	NOR	6.4 ± 0	34.4 ± 0.05	6.8 ± 2.1
Port Värtahamnen, Baltic Sea, Sweden	SWE	11.3 ± 0.1	$4.74 \pm 2 \times 10^{-5}$	2.5 ± 2.4
Ushuaia Bay, Tierra del Fuego Island, Argentina	ARG	8.6 ± 0.1	29.3 ± 0.1	1.4 ± 2.3
Potter Cove, 25 de Mayo (King George) Island, Antarctica	ANT	0.30 ± 0.2	34.10 ± 0.05	0.24 ± 0.2

ws wet sediment

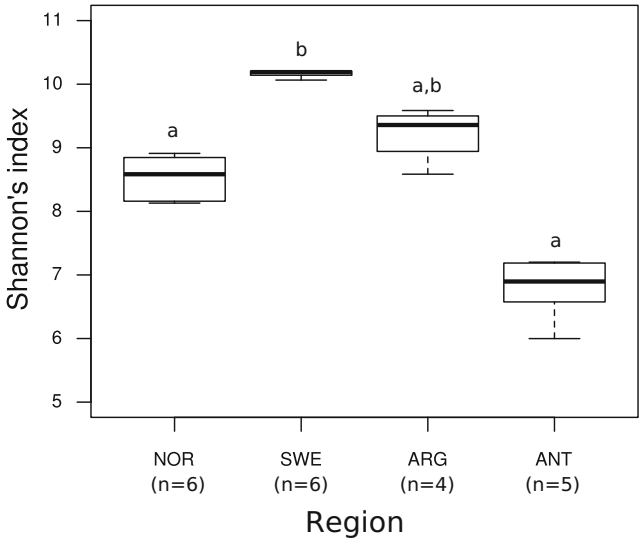


Fig. 2 Alpha diversity estimation of sediment microbial communities per region. *NOR* Adventfjord, Spitsbergen, Svalbard Archipelago, Norway; *SWE* Port Värtahamnen, Stockholm, Baltic Sea, Sweden; *ARG* Ushuaia Bay, Tierra del Fuego Island, Argentina; *ANT* Potter Cove, 25 de Mayo (King George) Island, Antarctic Peninsula. Shannon diversity index was calculated based on OTUs defined at 97% sequence identity (open reference OTU picking method) obtained from the I-tag dataset. In the boxplots, *boxes* represent the standard deviation and *whiskers* 95% confidence intervals, for values obtained from six core samples (three per site), with the exception of ARG (two amplicon sequencing runs failed for station MC) and ANT (one failure for station S1). For further details, see Table S3

to other marine and freshwater sediments [51]. A high richness of Bacteria was estimated for all samples, in particular in sediments of the Baltic Sea brackish environment (Chao1 > 10,000 OTUs, Fig. S1A). For Archaea, however, estimated richness was significantly lower (less than 200 OTUs, Fig. S1B). These results are in agreement with other studies, which found that in coastal and estuarine sediments distributed across salinity and sulfate gradients, only a few groups of Archaea were dominant [52]. In coastal sediments of Antarctica, Archaea were found to contribute only a small proportion of the total carbon flow [53]. However, it must be taken into account that the primer pair used in this study (515F-806R [54]) holds biases against Crenarachaeota/Thaumarchaeota, which could be affecting richness estimates for these groups. In 2015, an updated version of the primers was released, which overcomes this and other biases [55].

Although only eight sites from four regions were sampled, limiting the power of the analysis, there was a positive correlation between diversity (Shannon's index) and temperature measured in the sediment–water interface (Pearson's product-moment correlation $r = 0.96$, p value ≤ 0.0001). In addition, a weaker but significant negative correlation was found between diversity and salinity ($r = -0.74$, p value ≤ 0.0001). Recently, a negative correlation between richness and temperature was reported in chronically polluted temperate coastal sediments from the Mediterranean basin [30]. In contrast, various studies including a global meta-analysis of marine samples suggested the existence of a diversity latitudinal gradient, with higher richness at lower latitudes and decreasing toward the poles [5]. The sites analyzed by Bargiela and collaborators were placed along a latitudinal gradient ranging from 44° to 30°, which could be affected by local factors, whereas the sampling regions analyzed in this work comprised a latitudinal range of 78° to 54°, and in both Hemispheres. This difference in scale could explain the different results obtained in this study, which are more consistent with global-scale trends.

No correlation was detected between diversity and pollutant levels or other measured environmental variables (data not shown). In coastal sediments affected by the Deepwater Horizon spill, diversity and evenness were negatively correlated with total petroleum hydrocarbons, which is in accordance with the strong selection of certain populations after an acute pollution event [16]. The moderate levels of chronic pollution detected in this work did not seem to affect the diversity of the microbial communities. The presence of microbial communities with extremely high diversity in chronically polluted environments has important implications regarding their role in ecosystem services, as communities with higher diversity are expected to have increased stability, being more capable of responding adequately to disturbances [56, 57].

Community Structure

Distinct patterns of community structure were observed, with the communities generally grouping according to sampling site (Fig. 3, Fig. S2). The most abundant bacterial phylum was Proteobacteria, although different proteobacterial classes were dominant in sediments from different regions. In Svalbard sediments for instance, the Deltaproteobacteria were the most abundant, while in Potter Cove, Gammaproteobacteria were dominant (Fig. S2). Interestingly, the Deltaproteobacteria followed the same trend as the pollution gradient (higher in Svalbard samples, medium in Baltic Sea and Ushuaia Bay samples, and lower in Potter Cove samples; Fig. 1b), while the Gammaproteobacteria followed an opposite trend. We found that members of the Flavobacteria and Alphaproteobacteria were more abundant in sediments from polar regions than those from subpolar regions. In addition, Betaproteobacteria, commonly found at low salinities [58], were abundant in the brackish Baltic Sea samples that have a lower salinity compared to the samples from the marine coasts. We compared these results, which are based on large-scale 16S rRNA gene amplicon sequencing, with those previously reported, based on assignment of metagenomic shotgun sequences to 16S rRNA genes in the same sediment samples [35]. Although similar results were observed by the two methods for dominant bacterial groups (e.g., same trends per region in Deltaproteobacteria and Gammaproteobacteria), a large fraction of the 16S rRNA

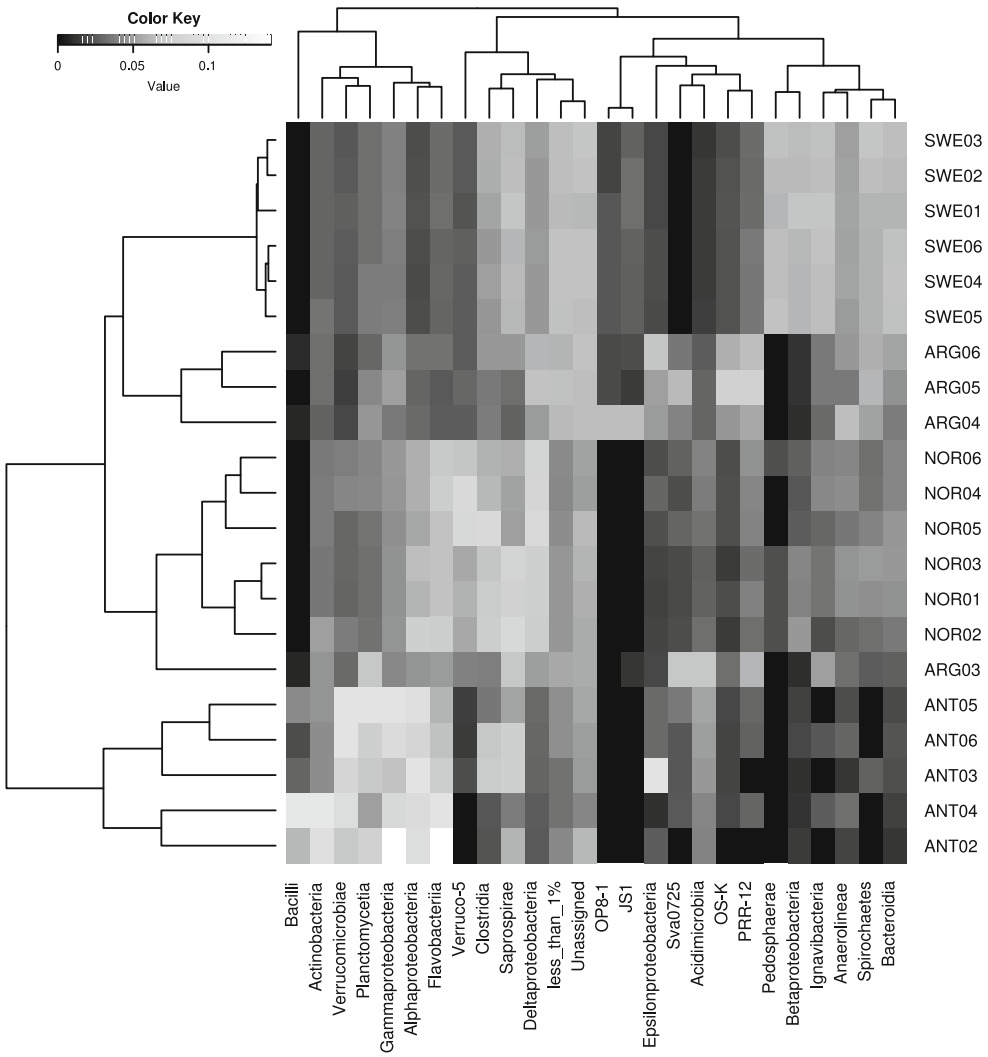


Fig. 3 Bacterial community structure (Class level). The heatmap was constructed based on I-tags. Only taxa with relative abundance ≥1% are indicated; the others were grouped as "≤1%". Samples ARG01 and ARG02 from Ushuaia Bay station MC and ANT01 from Potter Cove station S1 were not available due to failure during amplicon sequencing. For further details, see Table S3

sequences obtained from the metagenomes could not be assigned even at the Phylum level [35]. This is probably due to the fact that, unlike amplicon sequencing datasets, which are constrained within hypervariable regions due to primer usage, metagenomic sequences are more variable in length and cover different sections of the 16S rRNA gene, limiting the power of the phylogenetic inference. Furthermore, most of them are short reads, as they were obtained from the unassembled fraction of metagenomes. In addition, the number of sequences obtained per sample was one order of magnitude larger by amplicon sequencing (more than 10^5 sequences, Table S3 vs. 10^4 [35]). The higher coverage allowed the detection of low-abundance groups from candidate bacterial phyla, such as JS1 or PRR-12 (Fig. S2), which are poorly represented in databases and for which limited information is yet available with the exception of metagenomic or single-cell genomics studies [59, 60].

At the OTU level, significant correlations were found between community structure and environmental factors (non-metric multidimensional scaling, NMDS, Fig. 4). Salinity (highly correlated with axis 1), temperature (dividing polar and subpolar samples along both axes), and organic compounds (associated with samples suffering organic pollution) were the main factors found to correlate with community structure (Fig. 4). Similar results were obtained for ordinations based on weighted UniFrac (Fig. S3). Despite the large geographic distance between the two polar sampling regions, these communities were positioned closely in the ordination space, indicating similar structure (Fig. 4, Fig. S3). Although Arctic and Antarctic regions have been exposed to different geological and evolutionary events, the dynamic of climate drivers can be comparable, and therefore microorganisms subjected to similar environmental stressors could be selected in sites that are geographically distant, resulting in a bipolar distribution [3, 5, 61, 62].

The contribution of different environmental factors toward explaining bacterial community structure was statistically analyzed by canonical correspondence analysis (CCA) after merging OTUs at the Family level (Fig. 5). Among the tested

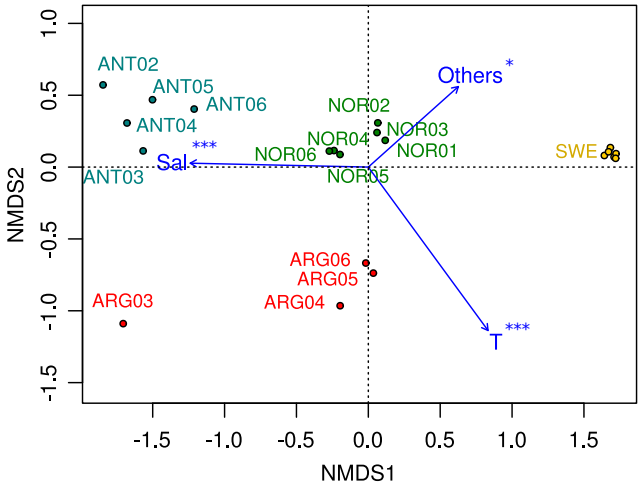


Fig. 4 Ordination of microbial communities based on OTUs defined at 97% sequence identity. Non-metric multidimensional scaling based on I-tag data, standardized by Wisconsin method and transformed by square root before analysis. Stress: 0.02. Environmental variables found to be significantly correlated with ordination patterns are represented as vectors (***$p < 0.001$, *$p < 0.05$). Two samples from Ushuaia Bay station MC and one sample from Potter Cove station S1 were not available due to failure during amplicon sequencing. For further details, see Table S3

environmental variables, the best model using temperature, salinity, depth, and pollution explained more than 75% of the variance ($p = 0.001$). The first canonical axis (CCA 1, associated with temperature and salinity) explained 48.6% of the constrained variation while the second axis (CCA 2, mostly related to depth and hydrocarbon pollution) explained 27.2% (corresponding to 36.1 and 20.2% of the total variation, respectively) (Fig. 5a). According to conventional CCA interpretation, a right-angled projection of a point representing a response variable (species) onto an arrow representing an explanatory variable (environmental variable) indicates the species optimum along that environmental gradient [63]. For instance, various Gammaproteobacterial families are interpreted as likely to reach their maximum at low temperature values (points on negative values of the temperature vector, such as 33, 34, 38, 40, 43, and 45; Fig. 5a). These families (Pyschromonadaceae, Colwelliacae, Alteromondaceae, and Moraxellaceae, also called *Alteromonas*-like bacteria [64]) are known to have several members with psychrophilic or psychrotolerant characteristics [65, 66]. Accordingly, these were higher in Antarctic sites (Fig. S4). On the other hand, families of the Betaproteobacteria Class, such as the Rhodocyclaceae and Comamonadaceae (14 and 15, respectively; Fig. 5a), were predicted to be more abundant at low salinity values (mainly explained by Baltic Sea samples, SWE; Fig. S4). Members of these families have been previously identified throughout aquatic environments, including the Baltic Sea [51, 67, 68]. Low oxygen concentrations found at the water–sediment interface in the Baltic Sea samples (Table S1), consistent with the high levels of hypoxia and eutrophication encountered in the entire area [48, 68], could also explain the high abundance of members of these families, which hold many nitrate reducers and denitrifiers. Moreover, from these two families, the most abundant genus identified in the Baltic Sea samples and almost absent in all other sites was *Dechloromonas* (Rhodocyclaceae) (data not shown). These are chemoorganotrophs that can grow on a number of substrates and use multiple electron acceptors for respiration including oxygen, chlorate, perchlorate, and nitrate [69].

Various families of the Deltaproteobacteria class covaried with vectors for anthropogenic pollutants (Fig. 5a, b), and three taxa belonging to the Desulfuromonadales (Deltaproteobacteria) were linearly correlated with the PAHs vector: these included the Desulfuromonadaceae ($r = 0.96$, Bonferroni-corrected p value $= 8.5 \times 10^{-10}$), Pelobacteraceae ($r = 0.97$, corrected $p = 2.41 \times 10^{-11}$), "unclassified Desulfuromonadales" ($r = 0.96$, corrected $p = 6.11 \times 10^{-10}$), and "unclassified Deltaproteobacteria" ($r = 0.97$, corrected $p = 4.04 \times 10^{-11}$) (Fig. S5). The correlations were slightly lower, but still highly significant, when the total sum (aliphatics, PAHs, and other compounds) was taken into consideration (data not shown). This result was expected due to the presence of various hydrocarbon structures within each pollution source resulting in highly correlated organic compounds among themselves. Together, these taxa accounted for more than 15% of the communities in the most polluted samples, mainly contributed by the "unclassified Desulfuromonadales," with more than 10% of the I-tag reads in Svalbard (NOR) samples (Fig. S5). Members of the Deltaproteobacteria have been reported to increase in subtidal sediments as a result of acute events of oil inputs [21], and they could be participating in anaerobic hydrocarbon biodegradation in more polluted locations. Still, variables that have not been measured but are related to hydrocarbon content could be responsible for these differences, even in a multicausal way. For example, higher levels of organic matter respiration (including the compounds measured in this work and/or others) could steeply decrease oxygen concentrations along the first centimeters of the sediment at the most polluted sites, contributing to the selection of bacterial groups with anoxic lifestyle. Similarly, a finer sediment granulometry could impair oxygen penetration and contribute in the accumulation of organic compounds through adsorption to the sediment matrix [7].

In samples from marine environments (Svalbard, Ushuaia Bay, and Potter Cove), in addition to the groups previously mentioned, the following taxa were positively correlated with hydrocarbon content: Geobacteraceae (Deltaproteobacteria, Desulfuromonadales); Hydrogenophilaceae (Betaproteobacteria, Hydrogenophilales), Comamonadaceae (Betaproteobacteria, Burkholderiales), and Rhodocyclaceae (Betaproteobacteria, Rhodocyclales); and unclassified Betaproteobacteria (Table S5). Interestingly, all these families were also found by CCA to have their maximum relative abundances near high pollution values but with higher

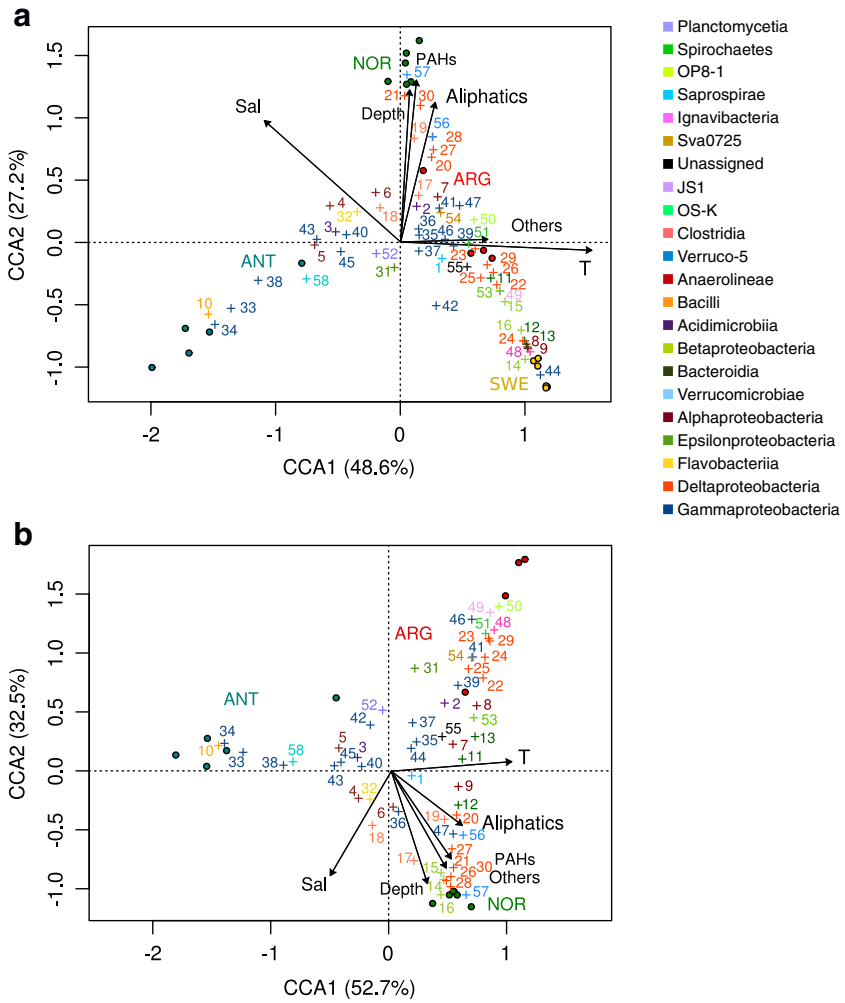


Fig. 5 Canonical correspondence analysis of bacterial families based on taxonomic assignment of I-tags. **a** Analysis based on the complete dataset. **b** Analysis only including samples from marine environments (NOR-ARG-ANT). Only families with abundances ≥1% were used for the analysis. Bacterial families are colored based on their assignment at the Class level. *1* Saprospiraceae, *2* koll13, *3* unclassified Acidimicrobiales, *4* Rhodobacteraceae, *5* Phyllobacteriaceae, *6* Sphingomonadaceae, *7* unclassified GCA004, *8* Anaerolinaceae, *9* unclassified_Anaerolineae, *10* unclassified Bacillales, *11* unclassified Bacteroidales, *12* SB-1, *13* VC21_Bac22, *14* Rhodocyclaceae, *15* Comamonadaceae, *16* Hydrogenophilaceae, *17* Lachnospiraceae, *18* Acidaminobacteraceae, *19* Christensenellaceae, *20* Desulfobulbaceae, *21* Desulfuromonadaceae, *22* Desulfobacteraceae, *23* unclassified Myxococcales, *24* Desulfarculaceae, *25* unclassified deltaproteobacteria_1, *26* Geobacteraceae, *27* unclassified deltaproteobacteria_2, *28* unclassified deltaproteobacteria_3, *29* Syntrophobacteriaceae, *30* Pelobacteraceae, *31* Helicobacteriaceae, *32* Flavobacteriaceae, *33* Psychromonadaceae, *34* Moraxellaceae, *35* Piscirickettsiaceae, *36* Marinicellaceae, *37* OM60, *38* Colwelliaceae, *39* unclassified Chromatiales, *40* Thiotrichaceae, *41* unclassified Thiohalorhabdales, *42* unclassified gammaproteobacteria_1, *43* Alteromonadaceae, *44* Sinobacteraceae, *45* unclassified gammaproteobacteria_2, *46* Chromatiaceae, *47* unclassified Oceanospirillales, *48* Ignavibacteriaceae, *49* unclassified SB-45, *50* unclassified HMMVPog-54, *51* unclassified OS-K, *52* Pirellulaceae, *53* Spirochaetaceae, *54* unclassified_Sva0725, *55* unassigned, *56* unclassified WCHB1–41, *57* WCHB1–25, *58* Verrucomicrobiaceae. Complete names of the families and corresponding class affiliations are detailed in Table S4. Two samples from Ushuaia Bay station MC and one sample from Potter Cove station S1 were not available due to failure during amplicon sequencing. For further details, see Table S3

preference for low salinity, suggesting a synergistic effect (Fig. 5a, b).

Metabolic Potential of Cold Sediment Metagenomes

We further aimed to infer how the metabolic potential of the sediment microbial communities varied across the analyzed locations, as well as to gain insight into the relevance of different environmental factors on these features. We analyzed metagenomic datasets generated from the same core samples described above (Table S3), based on the functional annotation of the protein CDS into KO (KEGG Orthology) terms and pathways. The size of the total metagenomes (assembled and unassembled) ranged from 1.4 to 16 gigabases (Table S3). The estimated gene copies of sequences assigned to the 17,948 KOs available at IMG were mapped into KEGG pathways

and modules, and normalized to avoid biases due to different metagenome sizes (for details see "Methods" section). In contrast to the phylogenetic structure, which exhibited clear distribution patterns among sites and regions even at the Phylum level (Figs. 3, 4, and 5), the metabolic potential of these communities was remarkably similar (Fig. 6). Interestingly, when the metabolic traits of the sediment metagenomes were compared with datasets from other studies annotated in the same pipeline, they were also found to be more similar to the ones from other high-complexity environments (marine sediments, freshwater sediments, and soils) than to those from environments with a lower diversity, such as pelagic or human-associated microbiome (Fig. S6). In this analysis, however, we did not find any evidence of a higher similarity with other communities from cold regions. For instance, cold sediment samples from this work were as similar to permafrost and peat soils from Alaska as to grassland soils from Northern California (data not shown).

The most abundant functional categories in the sediment metagenomes were related to carbohydrate, amino acid, and energy metabolism (Table S6). Of the 148 analyzed pathways (level 3), only 15 showed significant differences among sampling regions (corrected $p < 0.05$ and effect size higher than 0.8, Kruskal-Wallis test, Table S6). Similarly, out of 225 modules (level 4), only 29 showed significant differences (Table S7). Moreover, among samples, there was a higher proportion of shared KOs than shared microbial phyla (data not shown). Overall, these results are in accordance with the hypothesis of environmentally driven adaptations of phylogenetically distinct microbial communities into core metabolic functions, which has been proposed for other habitats [11, 70, 71]. However, it must be noted that the same results could be obtained if the functions in KEGG are not discriminating enough for these environments.

No significant correlation was observed between functional traits (modules) and the geographic distance matrix (Mantel's test $R = -0.03$, $p = 0.67$), in accordance with the previous results that showed that these metagenomes were all very similar in functional potential despite being far apart. In contrast, a significant correlation between community structure and geographic distance was observed (based on the 16S rRNA gene sequence data described above, Mantel's $R = 0.38$, $p = 4 \times 10^{-4}$), indicating that there are limits to the dispersion of microorganisms as a consequence of geographic barriers.

The previous analyses indicated that, although similar in their core functions, these metagenomes still presented some differences. For instance, according to the results of the non-parametric analysis of variance and post hoc comparisons, there were significant differences in some KEGG pathways among regions (Table S7), although no relation with environmental factors can be inferred from this test. To better visualize the environmental variables shaping the potential functional attributes of these communities, we conducted a regularized canonical correlation analysis (RCCA, [71, 72]) using the estimated gene copies of KOs mapped into KEGG modules. According to this analysis, salinity and temperature were the main functional drivers of the metabolic potential of these communities, with the majority of modules distributed along this axis (Fig. S7). These variables were themselves inversely correlated, and correlating with canonical variate (CV) 1 with $r = -0.89$ and 0.95 for salinity and temperature, respectively. It has been previously proposed that salinity is a major

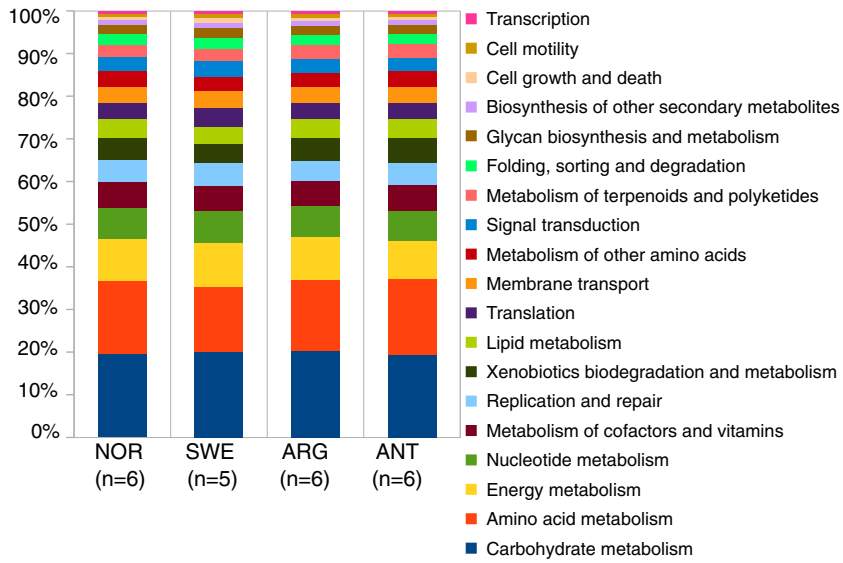


Fig. 6 Distribution of potential functions in cold sediment metagenomes. The relative abundance for each sample was defined as the sum of the total number of CDS assigned to KEGG pathways (level 2), normalizing over total number of assignments to KEGG pathways for that sample, to compensate for coverage differences among samples. Only pathways with values ≥0.1% were considered for the analysis. *n* = number of metagenomes analyzed per sampling region. In the case of the Baltic Sea region, SWE04 sample from station KBB was not available due to metagenome sequencing failure, and therefore *n* = 5. For further details, see Table S3

determinant of microbial community structure and function [67, 73, 74], and that temperature affects both phylogenetic and functional patterns in marine sediments [30]. In Potter Cove, Hernández et al. [75] also reported that salinity and temperature were the main environmental factors affecting dominance and richness of marine archaeal communities [75]. Interestingly, modules belonging to energy metabolism showed a pronounced variation along CV 1. For instance, nitrate reduction and denitrification (Nit_529, 530, and 531) as well as sulfate reduction (Sul_176 and 595) correlated negatively with CV 1 (Fig. S7 i). This can be explained by a more prevalent nutrient metabolism in colder regions [71]. In contrast, a positive correlation with CV1 was observed for nitrification (Nit_528) and various modules of methanogenesis (Meth_356, 357, 563, 567), suggesting a temperature-dependent enhancement of these functions (Fig. S7 i). Most cofactor and vitamin biosynthesis-related modules were also positively correlated to CV1 (Fig. S7 ii), in accordance with previous observations in the marine environment [72].

Canonical Variate 2 correlated positively with pollutants and other organic compounds vectors (r = 0.89, r = 0.82, and r = 0.63 for aliphatics, PAHs, and "other compounds," respectively). However, in this case, fewer metabolic modules were positioned along this axis with correlation values higher than 0.5, such as aromatic compound degradation modules AroD_534, 539, and 623, corresponding to naphthalene, *p*-cumate, and phthalate degradation, respectively (Fig. S7 iv). This indicates a more restricted and specific influence of these factors on the general metabolic potential of these communities. The less evident effect could be due to the relatively low levels of pollutants present in the analyzed samples (Fig. 1).

Biomarker Genes for Pollutant Degradation

Due to our previous findings, we hypothesized that there could be only subtle signals regarding microbial adaptation to pollution in these metagenomes. Therefore, we further focused the analyses on specific genes that are considered biomarkers for hydrocarbon degradation, as they are known to participate in key steps of these processes. Fourteen hydrocarbon-degradation biomarker genes were selected based on KO terms, including those that code for enzymes involved in degradation of PAHs, other aromatic compounds including BTEX, and alkanes (for details see the "Methods" section). Sequences coding for alkane-1-monooxygenase (AlkB) were in general the most abundant (Fig. 7). AlkB was previously found to be the most abundant biomarker in the metagenomic analysis of polluted Mediterranean sediments, followed by sequences assigned to catechol- and gentisate-degrading enzymes [30]. This result was attributed to the fact that alkanes are the most readily degradable hydrocarbon compounds. In the Mediterranean samples, however, relative abundances of AlkB-coding sequences were two orders of magnitude higher than those found in this work (0.68

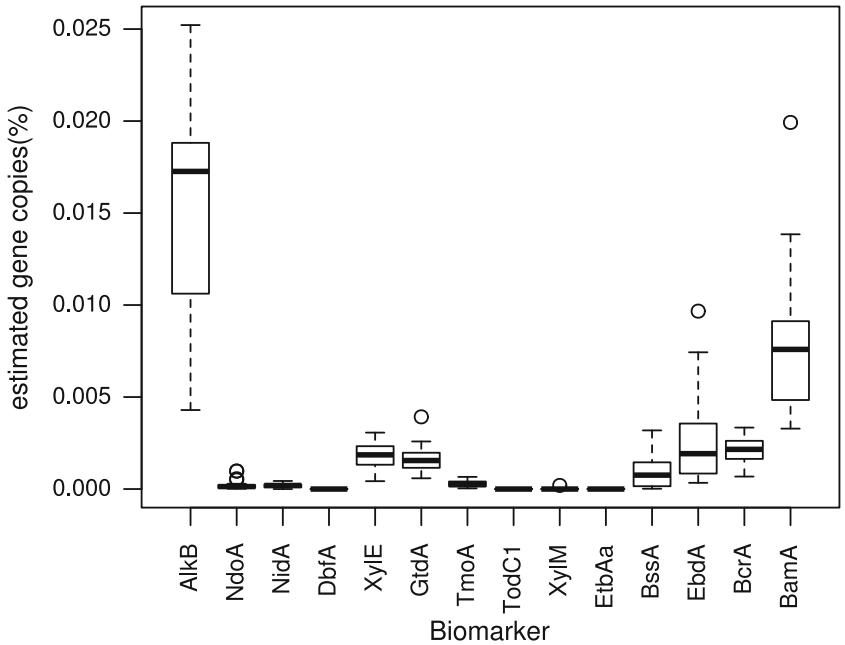
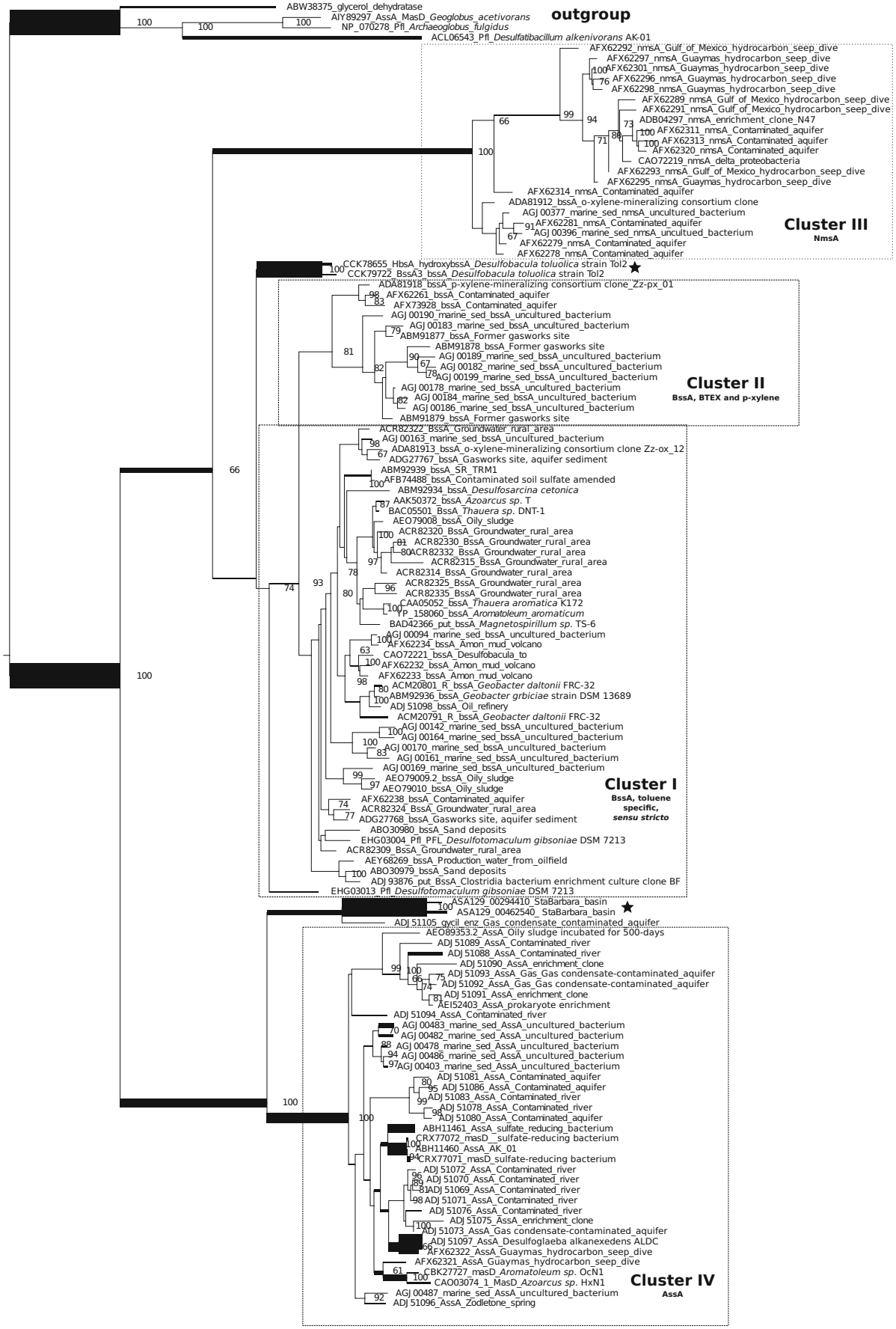


Fig. 7 Relative abundance of biomarker genes for hydrocarbon degradation. Estimated gene copies of sequences assigned to KOs corresponding to enzymes potentially involved in hydrocarbon degradation were divided by the total read count assigned to KOs. *AlkB* alkane-1-monooxygenase, *NdoB* naphthalene 1,2-dioxygenase subunit alpha, *NidA* PAH dioxygenase large subunit, *DbfA* dibenzofuran dioxygenase alpha subunit, *XylE* catechol 2,3-dioxygenase, *GtdA* gentisate dioxygenase, *TmoA* toluene monooxygenase system protein A, *TodC1* benzene/toluene dioxygenase alpha subunit, *XylM* xylene monooxygenase, *EtbAa* ethylbenzene dioxygenase alpha subunit, *NmsA* naphthyl-2-methylsuccinate synthase alpha subunit, *BssA* benzylsuccinate synthase, *EbdA* ethylbenzene hydroxylase subunit alpha, *BcrA* benzoyl-CoA reductase subunit A, *BamA* 6-oxo-cyclohex-1-ene-carbonyl-CoA hydrolase


outgroup
ABW38375_glycerol_dehydratase
AIY89297_AssA_MasD_Geoglobus_acetivorans
NP_070278_Pfl_Archaeoglobus_fulgidus
ACL06543_Pfl_Desulfatibacillum alkenivorans AK-01
Cluster III
NmsA
AFX62292_nmsA_Gulf_of_Mexico_hydrocarbon_seep_dive
AFX62297_nmsA_Guaymas_hydrocarbon_seep_dive
AFX62301_nmsA_Guaymas_hydrocarbon_seep_dive
AFX62296_nmsA_Guaymas_hydrocarbon_seep_dive
AFX62298_nmsA_Guaymas_hydrocarbon_seep_dive
AFX62289_nmsA_Gulf_of_Mexico_hydrocarbon_seep_dive
AFX62291_nmsA_Gulf_of_Mexico_hydrocarbon_seep_dive
ADB04297_nmsA_enrichment_clone_N47
AFX62311_nmsA_Contaminated_aquifer
AFX62313_nmsA_Contaminated_aquifer
AFX62320_nmsA_Contaminated_aquifer
CAO72219_nmsA_delta_proteobacteria
AFX62293_nmsA_Gulf_of_Mexico_hydrocarbon_seep_dive
AFX62295_nmsA_Guaymas_hydrocarbon_seep_dive
AFX62314_nmsA_Contaminated_aquifer
ADA81912_bssA_o-xylene-mineralizing consortium clone
AGJ 00377_marine_sed_nmsA_uncultured_bacterium
AFX62281_nmsA_Contaminated_aquifer
AGJ 00396_marine_sed_nmsA_uncultued_bacterium
AFX62279_nmsA_Contaminated_aquifer
AFX62278_nmsA_Contaminated_aquifer
Cluster II
BssA, BTEX and p-xylene
CCK78655_HbsA_hydroxybssA_Desulfobacula toluolica strain Tol2
CCK79722_BssA3_bssA_Desulfobacula toluolica strain Tol2
ADA81918_bssA_p-xylene-mineralizing consortium clone_Zz-px_01
AFX62261_bssA_Contaminated_aquifer
AFX73928_bssA_Contaminated_aquifer
AGJ 00190_marine_sed_bssA_uncultured_bacterium
AGJ 00183_marine_sed_bssA_uncultured_bacterium
ABM91877_bssA_Former gasworks site
ABM91878_bssA_Former gasworks site
AGJ 00189_marine_sed_bssA_uncultured_bacterium
AGJ 00182_marine_sed_bssA_uncultured_bacterium
AGJ 00199_marine_sed_bssA_uncultured_bacterium
AGJ 00178_marine_sed_bssA_uncultured_bacterium
AGJ 00184_marine_sed_bssA_uncultured_bacterium
AGJ 00186_marine_sed_bssA_uncultured_bacterium
ABM91879_bssA_Former gasworks site
Cluster I
BssA, toluene specific, sensu stricto
ACR82322_BssA_Groundwater_rural_area
AGJ 00163_marine_sed_bssA_uncultured_bacterium
ADA81913_bssA_o-xylene-mineralizing consortium clone Zz-ox_12
ADG27767_bssA_Gasworks site, aquifer sediment
ABM92939_bssA_SR_TRM1
AFB74488_bssA_Contaminated soil sulfate amended
ABM92934_bssA_Desulfosarcina cetonica
AAK50372_bssA_Azoarcus sp. T
BAC05501_BssA_Thauera sp. DNT-1
AEO79008_bssA_Oily_sludge
ACR82320_BssA_Groundwater_rural_area
ACR82330_BssA_Groundwater_rural_area
ACR82332_BssA_Groundwater_rural_area
ACR82315_BssA_Groundwater_rural_area
ACR82314_BssA_Groundwater_rural_area
ACR82325_BssA_Groundwater_rural_area
ACR82335_BssA_Groundwater_rural_area
CAA05052_bssA_Thauera aromatica K172
YP_158060_bssA_Aromatoleum_aromaticum
BAD42366_put_bssA_Magnetospirillum sp. TS-6
AGJ 00094_marine_sed_bssA_uncultured_bacterium
AFX62234_bssA_Amon_mud_volcano
CAO72221_bssA_Desulfobacula_to
AFX62232_bssA_Amon_mud_volcano
AFX62233_bssA_Amon_mud_volcano
ACM20801_R_bssA_Geobacter daltonii FRC-32
ABM92936_bssA_Geobacter grbiciae strain DSM 13689
ADJ 51098_bssA_Oil_refinery
ACM20791_R_bssA_Geobacter daltonii FRC-32
AGJ 00142_marine_sed_bssA_uncultured_bacterium
AGJ 00164_marine_sed_bssA_uncultured_bacterium
AGJ 00170_marine_sed_bssA_uncultured_bacterium
AGJ 00161_marine_sed_bssA_uncultured_bacterium
AGJ 00169_marine_sed_bssA_uncultured_bacterium
AEO79009.2_bssA_Oily_sludge
AEO79010_bssA_Oily_sludge
AFX62238_bssA_Contaminated_aquifer
ACR82324_BssA_Groundwater_rural_area
ADG27768_bssA_Gasworks site, aquifer sediment
ABO30980_bssA_Sand deposits
EHG03004_Pfl_PFL_Desulfotomaculum gibsoniae DSM 7213
ACR82309_BssA_Groundwater_rural_area
AEY68269_bssA_Production_water_from_oilfield
ABO30979_bssA_Sand deposits
ADJ 93876_put_BssA_Clostridia bacterium enrichment culture clone BF
EHG03013_Pfl_Desulfotomaculum gibsoniae DSM 7213
Cluster IV
AssA
ASA129_00294410_StaBarbara_basin
ASA129_00462540_ StaBarbara_basin
ADJ 51105_gycil_enz_Gas_condensate_contaminated_aquifer
AEO89353.2_AssA_Oily sludge incubated for 500-days
ADJ 51089_AssA_Contaminated_river
ADJ 51088_AssA_Contaminated_river
ADJ 51090_AssA_enrichment_clone
ADJ 51093_AssA_Gas_Gas condensate-contaminated_aquifer
ADJ 51092_AssA_Gas_Gas condensate-contaminated_aquifer
ADJ 51091_AssA_enrichment_clone
AEI52403_AssA_prokaryote enrichment
ADJ 51094_AssA_Contaminated_river
AGJ 00483_marine_sed_AssA_uncultured_bacterium
AGJ 00482_marine_sed_AssA_uncultured_bacterium
AGJ 00478_marine_sed_AssA_uncultured_bacterium
AGJ 00486_marine_sed_AssA_uncultured_bacterium
AGJ 00403_marine_sed_AssA_uncultured_bacterium
ADJ 51081_AssA_Contaminated_aquifer
ADJ 51086_AssA_Contaminated_aquifer
ADJ 51083_AssA_Contaminated_river
ADJ 51078_AssA_Contaminated_river
ADJ 51080_AssA_Contaminated_aquifer
ABH11461_AssA_sulfate_reducing_bacterium
CRX77072_masD_sulfate-reducing bacterium
ABH11460_AssA_AK_01
CRX77071_masD_sulfate-reducing bacterium
ADJ 51072_AssA_Contaminated_river
ADJ 51070_AssA_Contaminated_river
ADJ 51069_AssA_Contaminated_river
ADJ 51071_AssA_Contaminated_river
ADJ 51076_AssA_Contaminated_river
ADJ 51075_AssA_enrichment_clone
ADJ 51073_AssA_Gas condensate-contaminated_aquifer
ADJ 51097_AssA_Desulfoglaeba alkanexedens ALDC
AFX62322_AssA_Guaymas_hydrocarbon_seep_dive
AFX62321_AssA_Guaymas_hydrocarbon_seep_dive
CBK27727_masD_Aromatoleum sp. OcN1
CAO03074_1_MasD_Azoarcus sp. HxN1
AGJ 00487_marine_sed_AssA_uncultured_bacterium
ADJ 51096_AssA_Zodletone_spring

◀ **Fig. 8** Evolutionary placement analysis of BssA sequences. Metagenomic sequences were independently assigned on to a fixed phylogenetic tree by maximum likelihood using RaxML [44]. Clusters correspond to the ones previously defined by Acosta-González et al. [43]. Divergent clusters containing metagenomic sequences from this work are indicated with *stars*

to 1.36% for the Mediterranean vs. 0.004 to 0.025% in this work), in accordance with the much higher level of anthropogenic impact of the Mediterranean environment [116–260,000 ppm (μg/g) of total petroleum hydrocarbons, one to four orders of magnitude higher than the hydrocarbon concentrations in samples from this work].

In order to gain some insight into the structure of the fraction of the community with alkane-degrading potential, a total of 6908 metagenomic AlkB sequences longer than 70 amino acids from all regions were analyzed by phylogenetic placing onto a reference tree (Fig. S8). Of these, 96% could be unequivocally assigned to one of the clusters defined in a previous study by Nie and collaborators [42]. For all regions, the major fraction of the sequences (47% on average) belonged to Cluster V (Fig. S8). This is a diverse cluster typically composed of putative *alkB* genes identified in genomes from members of the Bacteroidetes, which had not been previously detected in PCR-based studies [42]. Our results are in accordance with metagenomic analyses carried out in Ushuaia Bay [25], and in other metagenomes [42], and suggest that these previously underestimated populations could be important members of the hydrocarbon-degrading community in cold coastal environments. Sequences assigned to Cluster IV (similar to archetypical AlkB1 from *Alcanivorax borkumensis*) were also abundant in all regions (28% on average). Interestingly, sequences assigned to Cluster VII, identified in genomes from members of the Rhodobacterales (Alphaproteobacteria), were more abundant in polar regions (Potter Cove and Svalbard, approximately 20%), while in the Baltic Sea metagenomes, sequences assigned to Cluster I (typically described in Actinobacteria) were instead more abundant (17%).

The second most abundant biomarkers in the sediment samples were those encoding enzymes putatively involved in anaerobic hydrocarbon biodegradation (BssA, BamA, EbdA, BcrA), as well as XylE and GtdA (catechol 2,3-dioxygenase and gentisate 1,2-dioxygenase) (Fig. 7). Notably, in sediments from the Baltic Sea (both sites) and Ushuaia Bay (site OR), sequences assigned to BamA (performing the ring cleavage step funneling the anaerobic degradation of toluene, ethylbenzene, and methylnaphthalene [40]), were similar in abundance or even higher than AlkB (Fig. S9). Despite the important differences in natural environmental factors such as salinity, these two environments have high levels of nutrient load and suffer eutrophication [76, 77], suggesting a role of organic matter and nutrient availability for microorganisms in the consumption of available oxygen.

The abundances of CDS encoding BssA and NdoB enzymes were found to significantly correlate with pollutant levels (Table S8). Other markers were not correlated or were inversely correlated with pollution, such as the ones targeting the more volatile compounds (TodC1, benzene/toluene dioxygenase alpha subunit, EbdA, ethylbenzene dehydrogenase) (Table S8). On the other hand, 2-naphthoyl-CoA reductase and dihydronaphthoyl-CoA reductase, two enzymes belonging to the Old Yellow Enzyme family that have recently been reported to be involved in the anaerobic degradation of naphthalene [78], have not yet been assigned to KOs and therefore they could not be comparatively analyzed with the other biomarkers. Sequences assigned to the corresponding COG (COG1902, 2,4-dienoyl-CoA reductase or related NADH-dependent reductases) were very abundant in the metagenomes (data not shown), but this COG is not specific as it includes other enzymes as well.

The taxonomic assignment of biomarker gene sequences suggested the dominance of distinct microbial groups according to the potential metabolic condition of the enzymatic reaction (aerobic or anaerobic). In this way, sequences related to aerobic degradation (AlkB, NdoB and NidA, XylE, GtdA) were mainly assigned to Alphaproteobacteria and Gammaproteobacteria, and to a lesser extent to Betaproteobacteria and Actinobacteria (Table S9). These results are in accordance with our previous studies (PCR-based) of functional and phylogenetic biomarker gene diversity in the subantarctic environment of Ushuaia Bay, where these groups were also found to be prevalent [24, 26, 79]. In contrast, biomarkers for anaerobic degradation (BssA, EbdA, BcrA, BamA) showed a different taxonomic distribution. BssA sequences were assigned mainly to members of the Deltaproteobacteria (Desulfuromonadales and Geobacteraceae), and in some cases to Clostridia (Svalbard and Ushuaia Bay, Table S9). Notably, the families that were identified in functional-biomarker gene analyses were the same ones that were detected to be associated with pollution based on a phylogenetic-based approach (e.g., CCA in Fig. 5, individual correlations with pollutants, Fig. S5, Table S5). These results support the hypothesis that these families could have a key role in the degradation processes of both aliphatic and aromatic compounds. Nevertheless, it should be noted that these results will be largely influenced by biases in the databases, leading to the underestimation of the diversity at these sites. Hydrocarbon-degradation genes are extremely diverse and have been poorly characterized so far in these environments, and therefore it is possible that highly divergent sequences (such as many of those found in marine sediments) were missed not only during the taxonomic assignment of biomarker genes but also in the previous step, the functional assignment of metagenomic sequences, especially those with short length.

Metagenomic sequences assigned by KO to the BssA bin (3326 coding sequences for all samples) were further analyzed

by phylogenetic placement against a reference tree that included sequences from all *bssA*-like homologs (*assA*, *bssA* sensu stricto, *nmsA*). These homologous genes code enzymes that catalyze the anaerobic degradation of various hydrocarbon compounds by a general activation mechanism involving fumarate addition [43]. Among the sequences that could be effectively placed on to specific clusters (Fig. 8), the majority (1580) belonged to the AssA cluster (alkylsuccinate synthase, catalyzing alkane degradation). Sequences assigned to NmsA (naphthyl-2-methylsuccinate synthase, participating in methylnapthalene degradation) were also found (138 sequences). A total of 414 sequences clustered with archetypical BssA sequences (Fig. 8), which were evenly distributed along the two predefined clusters for this gene (cluster I, toluene-specific, *bssA* sensu stricto, and cluster II, with substrates including BTEX including *p*-xylene) [43]. In addition, 245 sequences were placed within a divergent cluster (Fig. 8, star) formed by hydroxybenzylsuccinate synthase genes (*hbsA*, participating in *p*-cresol degradation) and an additional *bssA* homolog (of unknown function) from *Desulfobacula toluolica* (Desulfobacteraceae). This organism is a sulfate-reducing bacterium isolated from marine sediments, capable of a wide range of biological transformations of aromatic compounds [43, 80].

Interestingly, 337 sequences (mainly from Svalbard samples, data not shown) were placed in a cluster containing hitherto functionally unknown metagenomic sequences from methane-oxidizing archaeal communities in coastal marine sediments of Santa Barbara Basin (California) (Fig. 8, star). Many of these protein sequences were similar to a hypothetical protein from *Desulfosarcina* sp. Bu5 (Deltaproteobacteria, Desulfobacterales), the only pure culture known so far capable of gaseous alkane oxidation under sulfate-reducing conditions [81]. Although culture-based evidence is limited, there is in situ evidence that indicates this process is widespread in the marine environment [82–84]. It has been proposed that not only oxidation of butane and propane but also of methane could be theoretically possible via fumarate addition by glycil-radical enzymes under sulfate-reducing conditions. Chemical evidence of this process has been observed in hydrocarbon-contaminated environments, although it has not fully been proved to date [85, 86]. The presence of non-archetypical gaseous alkane-oxidizing enzymes in polar sediments could have important environmental consequences [87]. More studies are needed in this direction in order to increase our understanding of carbon and sulfur cycles in these environments.

While *bssA* is an already established biomarker and has been widely analyzed in community studies [43, 88], for other genes (*bamA*, e*bdA*) no environmental information was previously available in natural microbial communities [89], and this work is the first to analyze the prevalence of these biomarker genes in the coastal environments. The results from this work highlight the role of microorganisms with an anaerobic lifestyle in pollutant biodegradation processes and raises questions of whether appropriate depuration rates could be reached, as anaerobic are in general less efficient than aerobic degradation processes [11].

Conclusion

The combination of 16S rRNA gene sequencing and metagenomics, and the broad geographic scale used in this study allowed to increase our understanding of the community structure and metabolic potential of coastal sediments from high-latitude regions. In addition to well-known natural environmental constraints such as salinity and temperature, anthropogenic stressors (namely pollution) were found to be factors shaping the phylogenetic structure of these communities, even at low pollution levels. Our analyses showed that anaerobic biodegradation, a process that has yet to be fully elucidated at the molecular and ecological level, could be driving natural attenuation of hydrocarbon pollution in these environments. The results from this work provide a starting point for integrating ecological insights involving phylogenetic, functional, and environmental data in these vulnerable environments. Further studies will be needed to model the capabilities of key yet-uncultured microbial populations, and to monitor the expression of relevant genes in controlled conditions, in order to integrate this information at different levels, building a more robust and direct link between structure and function.

Acknowledgments M.L. and H.M.D. are staff members from The Argentinean National Research Council (CONICET), and F.E. is a doctoral fellow from CONICET. The amplicon dataset was generated by the Earth Microbiome Project (www.earthmicrobiome.org/). 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). Sampling was funded 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), the Argentinean Antarctic Institute, the Universidad de Buenos Aires (UBA 20020100100378), ANPCyT (PICT-O 0124), and the Research Council of Norway (grant no. 228107). This research was also partly funded under the Laboratory Directed Research and Development Program at PNNL, a multi-program national laboratory operated by Battelle for the U.S. Department of Energy under Contract DE-AC05-76RL01830. We would like to thank Ricardo Vera, Horacio Ocariz, and Alejandro Ulrich for their help in sample collection.

Compliance with Ethical Standards

Conflict of Interest The authors declare no conflict of interest.

References

1. Doney SC, Ruckelshaus M, Emmett Duffy J, Barry JP, Chan F, English CA, et al. (2012) Climate change impacts on marine ecosystems Annu. Rev. Mar. Sci. 4:11–37
2. Atlas RM (2010) Microbial bioremediation in polar environments: current status and future directions. In: Bej AK, Aislabie J, Atlas RM (eds) Polar microbiology: the ecology, biodiversity and bioremediation potential of microorganisms in extremely cold environments. CRC Press, Boca Raton, pp. 373–391
3. Ghiglione J-F, Galand PE, Pommier T, Pedrós-Alió C, Maas EW, Bakker K, et al. (2012) Pole-to-pole biogeography of surface and deep marine bacterial communities Proc. Natl. Acad. Sci. 109: 17633–17638
4. Hanson CA, Fuhrman JA, Horner-Devine MC, Martiny JBH (2012) Beyond biogeographic patterns: processes shaping the microbial landscape Nat. Rev. Microbiol. 10:497–506
5. Sul WJ, Oliver TA, Ducklow HW, Amaral-Zettler LA, Sogin ML (2013) Marine bacteria exhibit a bipolar distribution Proc. Natl. Acad. Sci. 110:2342–2347
6. Huettel M, Berg P, Kostka JE (2014) Benthic exchange and biogeochemical cycling in permeable sediments Annu. Rev. Mar. Sci. 6:23–51
7. Commendatore MG, Nievas ML, Amin O, Esteves JL (2012) Sources and distribution of aliphatic and polyaromatic hydrocarbons in coastal sediments from the Ushuaia Bay (Tierra del Fuego, Patagonia, Argentina) Mar. Environ. Res. 74:20–31
8. León VM, García I, Martínez-Gómez C, Campillo JA, Benedicto J (2014) Heterogeneous distribution of polycyclic aromatic hydrocarbons in surface sediments and red mullet along the Spanish Mediterranean coast Mar. Pollut. Bull. 87:352–363
9. Cravo-Laureau C, Duran R (2014) Marine coastal sediments microbial hydrocarbon degradation processes: contribution of experimental ecology in the omics'era Front. Microbiol. 5:39
10. Reese BK, Witmer AD, Moller S, Morse JW, Mills HJ (2013) Molecular assays advance understanding of sulfate reduction despite cryptic cycles Biogeochemistry 118:307–319
11. Acosta-González A, Marqués S (2016) Bacterial diversity in oil-polluted marine coastal sediments Energy Biotechnol. Environ. Biotechnol 38:24–32
12. Acosta-González A, Rosselló-Móra R, Marqués S (2013a) Characterization of the anaerobic microbial community in oil-polluted subtidal sediments: aromatic biodegradation potential after the prestige oil spill Environ. Microbiol. 15:77–92
13. Meckenstock RU, Mouttaki H (2011) Anaerobic degradation of non-substituted aromatic hydrocarbons Curr. Opin. Biotechnol. 22:406–414
14. Kappell AD, Wei Y, Newton RJ, Van Nostrand JD, Zhou J, McLellan SL, Hristova KR (2014) The polycyclic aromatic hydrocarbon degradation potential of Gulf of Mexico native coastal microbial communities after the Deepwater Horizon oil spill Front. Microbiol. 5:205
15. Kostka JE, Prakash O, Overholt WA, Green SJ, Freyer G, Canion A, et al. (2011) Hydrocarbon-degrading bacteria and the bacterial community response in Gulf of Mexico beach sands impacted by the Deepwater Horizon oil spill Appl. Environ. Microbiol. 77: 7962–7974
16. Lamendella R, Strutt S, Borglin S, Chakraborty R, Tas N, Mason OU, et al. (2014) Assessment of the deepwater horizon oil spill impact on Gulf Coast microbial communities Front. Microbiol. 5: 130
17. Mason OU, Scott NM, Gonzalez A, Robbins-Pianka A, Bælum J, Kimbrel J, et al. (2014) Metagenomics reveals sediment microbial community response to deepwater horizon oil spill ISME J. 8: 1464–1475
18. Rodriguez-R LM, Overholt WA, Hagan C, Huettel M, Kostka JE, Konstantinidis KT (2015) Microbial community successional patterns in beach sands impacted by the deepwater horizon oil spill ISME J. 9:1928–1940
19. Scott NM, Hess M, Bouskill NJ, Mason OU, Jansson JK, Gilbert JA (2014) The microbial nitrogen cycling potential is impacted by polyaromatic hydrocarbon pollution of marine sediments Front. Microbiol. 5:108
20. Alonso-Gutiérrez J, Figueras A, Albaigés J, Jiménez N, Viñas M, Solanas AM, Novoa B (2009) Bacterial communities from shoreline environments (Costa da Morte, northwestern Spain) affected by the Prestige oil spill Appl. Environ. Microbiol. 75:3407–3418
21. Acosta-González A, Martirani-von Abercron S-M, Rosselló-Móra R, Wittich R-M, Marqués S (2015) The effect of oil spills on the bacterial diversity and catabolic function in coastal sediments: a case study on the Prestige oil spill Environ. Sci. Pollut. Res. 22: 15200–15214
22. Kimes NE, Callaghan AV, Suflita JM, Morris PJ (2014) Microbial transformation of the Deepwater Horizon oil spill—past, present, and future perspectives Front. Microbiol. 5:603
23. Kostka JE, Teske AP, Joye SB, Head IM (2014) The metabolic pathways and environmental controls of hydrocarbon biodegradation in marine ecosystems Front. Microbiol. 5:471
24. Guibert L, Loviso C, Marcos M, Commendatore M, Dionisi H, Lozada M (2012) Alkane biodegradation genes from chronically-polluted subantarctic coastal sediments and their shifts in response to oil exposure Microb. Ecol. 64:605–616
25. Guibert LM, Loviso CL, Borglin S, Jansson JK, Dionisi HM, Lozada M (2015) Diverse bacterial groups contribute to the alkane degradation potential of chronically polluted subantarctic coastal sediments Microb. Ecol. 71:100–112
26. Marcos MS, Lozada M, Di Marzio WD, Dionisi HM (2012) Abundance, dynamics, and biogeographic distribution of seven polycyclic aromatic hydrocarbon dioxygenase gene variants in coastal sediments of Patagonia Appl. Environ. Microbiol. 78: 1589–1592
27. Paisse S, Coulon F, Goñi-Urriza M, Peperzak L, McGenity TJ, Duran R (2008) Structure of bacterial communities along a hydrocarbon contamination gradient in a coastal sediment FEMS Microbiol. Ecol. 66:295–305
28. Paisse S, Duran R, Coulon F, Goñi-Urriza M (2011) Are alkane hydroxylase genes (alkB) relevant to assess petroleum bioremediation processes in chronically polluted coastal sediments? Appl. Microbiol. Biotechnol. 92:835–844
29. Paissé S, Goñi-Urriza M, Stadler T, Budzinski H, Duran R (2012) Ring-hydroxylating dioxygenase (RHD) expression in a microbial community during the early response to oil pollution FEMS Microbiol. Ecol. 80:77–86
30. Bargiela R, Mapelli F, Rojo D, Chouaia B, Tornés J, Borin S, et al. (2015) Bacterial population and biodegradation potential in chronically crude oil-contaminated marine sediments are strongly linked to temperature Sci Rep 5:11651
31. Mackelprang R, Waldrop MP, DeAngelis KM, David MM, Chavarria KL, Blazewicz SJ, et al. (2011) Metagenomic analysis of a permafrost microbial community reveals a rapid response to thaw Nature 480:368–371
32. Gilbert JA, Jansson JK, Knight R (2014) The Earth Microbiome project: successes and aspirations BMC Biol. 12:69
33. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. (2010) QIIME allows analysis of high-throughput community sequencing data Nat. Methods 7:335–336
34. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, et al. (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB Appl. Environ. Microbiol. 72:5069–5072

35. Matos MN, Lozada M, Anselmino LE, Musumeci MA, Henrissat B, Jansson JK, et al. (2016) Metagenomics unveils the attributes of the alginolytic guilds of sediments from four distant cold coastal environments Environ. Microbiol. 18:4471–4484
36. Markowitz VM, Chen I-MA, 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
37. Kanehisa M, Goto S (2000) KEGG: Kyoto encyclopedia of genes and genomes Nucleic Acids Res. 28:27–30
38. Parks DH, Beiko RG (2010) Identifying biologically relevant differences between metagenomic communities Bioinformatics 26: 715–721
39. Legendre P, Legendre LFJ (2012) Numerical ecology. Elsevier
40. Kuntze K, Shinoda Y, Moutakki H, McInerney MJ, Vogt C, Richnow H-H, Boll M (2008) 6-Oxocyclohex-1-ene-1-carbonyl-coenzyme a hydrolases from obligately anaerobic bacteria: characterization and identification of its gene as a functional marker for aromatic compounds degrading anaerobes Environ. Microbiol. 10: 1547–1556
41. Huson DH, Mitra S, Ruscheweyh H-J, Weber N, Schuster SC (2011) Integrative analysis of environmental sequences using MEGAN4 Genome Res. 21:1552–1560
42. Nie Y, Chi C-Q, Fang H, et al. (2014) Diverse alkane hydroxylase genes in microorganisms and environments Sci Rep 4:4968
43. Acosta-González A, Rosselló-Móra R, Marqués S (2013b) Diversity of benzylsuccinate synthase-like (bssA) genes in hydrocarbon-polluted marine sediments suggests substrate-dependent clustering Appl. Environ. Microbiol. 79:3667–3676
44. Berger SA, Krompass D, Stamatakis A (2011) Performance, accuracy, and web server for evolutionary placement of short sequence reads under maximum likelihood Syst. Biol. 60:291–302
45. Holte B, Dahle S, Gulliksen B, Næs K (1996) Some macrofaunal effects of local pollution and glacier-induced sedimentation, with indicative chemical analyses, in the sediments of two Arctic fjords Polar Biol. 16:549–557
46. Reimann S, Kallenborn R, Schmidbauer N (2009) Severe aromatic hydrocarbon pollution in the arctic town of Longyearbyen (Svalbard) caused by snowmobile emissions Env. Sci. Technol. 43:4791–4795
47. Dauner ALL, Hernández EA, MacCormack WP, Martins CC (2015) Molecular characterisation of anthropogenic sources of sedimentary organic matter from Potter Cove, King George Island, Antarctica Sci. Total Environ. 502:408–416
48. Korpinen S, Meski L, Andersen JH, Laamanen M (2012) Human pressures and their potential impact on the Baltic Sea ecosystem Ecol. Indic. 15:105–114
49. McLachlan A, Brown AC (2006) 14—Human impacts.The ecology of sandy shores, Second edn. Academic Press, Burlington, pp. 273–301
50. Rideout JR, He Y, Navas-Molina JA, Walters WA, Ursell LK, Gibbons SM, et al. (2014) Subsampled open-reference clustering creates consistent, comprehensive OTU definitions and scales to billions of sequences Peer J 2:e545
51. Wang Y, Sheng H-F, He Y, Wu J-Y, Jiang Y-X, Tam NF-Y, Zhou H-W (2012) Comparison of the levels of bacterial diversity in freshwater, intertidal wetland, and marine sediments by using millions of Illumina tags Appl. Environ. Microbiol. 78:8264–8271
52. Webster G, O'Sullivan LA, Meng Y, Williams AS, Sass AM, Watkins AJ, et al. (2015) Archaeal community diversity and abundance changes along a natural salinity gradient in estuarine sediments FEMS Microbiol. Ecol. 91:1–18
53. Purdy KJ, Nedwell DB, Embley TM (2003) Analysis of the sulfate-reducing bacterial and methanogenic archaeal populations in contrasting Antarctic sediments Appl. Environ. Microbiol. 69:3181–3191
54. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, et al. (2012) Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms ISME J. 6: 1621–1624
55. Apprill A, McNally S, Parsons R, Weber L (2015) Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton Aquat. Microb. Ecol. 75:129–137
56. Peter H, Beier S, Bertilsson S, Lindstrom ES, Langenheder S, Tranvik LJ (2011) Function-specific response to depletion of microbial diversity ISME J 5:351–361
57. Shade A, Peter H, Allison SD, Baho DL, Berga M, Bürgmann H, et al. (2012) Fundamentals of microbial community resistance and resilience Front. Microbiol. 3:417
58. Campbell BJ, Kirchman DL (2013) Bacterial diversity, community structure and potential growth rates along an estuarine salinity gradient ISME J. 7:210–220
59. Nobu MK, Dodsworth JA, Murugapiran SK, Rinke C, Gies EA, Webster G, et al. (2016) Phylogeny and physiology of candidate phylum "Atribacteria" (OP9/JS1) inferred from cultivation-independent genomics ISME J. 10:273–286
60. Youssef NH, Farag IF, Rinke C, Hallam SJ, Woyke T, Elshahed MS (2015) In silico analysis of the metabolic potential and niche specialization of candidate phylum "Latescibacteria" (ws3) PLOS ONE 10:e0127499
61. Gibbons SM, Caporaso JG, Pirrung M, Field D, Knight R, Gilbert JA (2013) Evidence for a persistent microbial seed bank throughout the global ocean Proc. Natl. Acad. Sci. 110:4651–4655
62. Zinger L, Boetius A, Ramette A (2014) Bacterial taxa–area and distance–decay relationships in marine environments Mol. Ecol. 23:954–964
63. Buttigieg PL, Ramette A (2014) A guide to statistical analysis in microbial ecology: a community-focused, living review of multivariate data analyses FEMS Microbiol. Ecol. 90:543–550
64. Ivanova EP, Flavier S, Christen R (2004) Phylogenetic relationships among marine Alteromonas-like proteobacteria: emended description of the family Alteromonadaceae and proposal of Pseudoalteromonadaceae fam. nov., Colwelliaceae fam. nov., Shewanellaceae fam. nov., Moritellaceae fam. nov., Ferrimonadaceae fam. nov., Idiomarinaceae fam. nov. and Psychromonadaceae fam. nov Int. J. Syst. Evol. Microbiol. 54: 1773–1788
65. Methé BA, Nelson KE, Deming JW, Momen B, Melamud E, Zhang X, et al. (2005) The psychrophilic lifestyle as revealed by the genome sequence of *Colwellia psychrerythraea* 34H through genomic and proteomic analyses Proc. Natl. Acad. Sci. U. S. A. 102: 10913–10918
66. Riley M, Staley JT, Danchin A, Wang TZ, Brettin TS, Hauser LJ, et al. (2008) Genomics of an extreme psychrophile, *Psychromonas ingrahamii* BMC Genomics 9:210
67. Dupont CL, Larsson J, Yooseph S, Ininbergs K, Goll J, Asplund-Samuelsson J, et al. (2014) Functional tradeoffs underpin salinity-driven divergence in microbial community composition PLoS One 9:e89549
68. 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
69. Oren A (2014) The family Rhodocyclaceae. In: Rosenberg E, DeLong EF, Lory S, Stackebrandt E, Thompson F (eds) The prokaryotes. Springer, Berlin, pp. 975–998
70. Muegge BD, Kuczynski J, Knights D, Clemente JC, González A, Fontana L, et al. (2011) Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans Science 332:970–974

71. Raes J, Letunic I, Yamada T, Jensen LJ, Bork P (2011) Toward molecular trait-based ecology through integration of biogeochemical, geographical and metagenomic data Mol. Syst. Biol. 7:473
72. Gianoulis TA, Raes J, Patel PV, Bjornson R, Korbel JO, Letunic I, et al. (2009) Quantifying environmental adaptation of metabolic pathways in metagenomics Proc. Natl. Acad. Sci. 106:1374–1379
73. Lozupone CA, Knight R (2007) Global patterns in bacterial diversity Proc Natl Acad Sci U S A 104:11436–11440
74. Sunagawa S, Coelho LP, Chaffron S, Kultima JR, Labadie K, Salazar G, et al. (2015) Structure and function of the global ocean microbiome Science 348:1261359
75. Hernández EA, Piquet AM-T, Lopez JL, Buma AGJ, Cormack WPM (2014) Marine archaeal community structure from Potter Cove, Antarctica: high temporal and spatial dominance of the phylum Thaumarchaeota Polar Biol. 38:117–130
76. Conley DJ, Björck S, Bonsdorff E, Carstensen J, Destouni G, Gustafsson BG, et al. (2009) Hypoxia-related processes in the Baltic Sea Environ. Sci. Technol. 43:3412–3420
77. Gil MN, Torres AI, Amin O, Esteves JL (2011) Assessment of recent sediment influence in an urban polluted subantarctic coastal ecosystem. Beagle Channel (southern Argentina) Mar. Pollut. Bull. 62:201–207
78. Estelmann S, Blank I, Feldmann A, Boll M (2015) Two distinct old yellow enzymes are involved in naphthyl ring reduction during anaerobic naphthalene degradation Mol. Microbiol. 95:162–172
79. Loviso C l, Lozada M, Guibert L m, Musumeci M a, Sarango Cardenas S, Kuin RV, et al. (2015) Metagenomics reveals the high polycyclic aromatic hydrocarbon-degradation potential of abundant uncultured bacteria from chronically polluted subantarctic and temperate coastal marine environments J. Appl. Microbiol. 119:411–424
80. Wöhlbrand L, Jacob JH, Kube M, Mussmann M, Jarling R, Beck A, et al. (2013) Complete genome, catabolic sub-proteomes and key-metabolites of *Desulfobacula toluolica* Tol2, a marine, aromatic compound-degrading, sulfate-reducing bacterium Environ. Microbiol. 15:1334–1355
81. Kniemeyer O, Musat F, Sievert SM, Knittel K, Wilkes H, Blumenberg M, et al. (2007) Anaerobic oxidation of short-chain hydrocarbons by marine sulphate-reducing bacteria Nature 449: 898–901
82. Jaekel U, Musat N, Adam B, Kuypers M, Grundmann O, Musat F (2013) Anaerobic degradation of propane and butane by sulfate-reducing bacteria enriched from marine hydrocarbon cold seeps ISME J. 7:885–895
83. Kleindienst S, Herbst F-A, Stagars M, von Netzer F, von Bergen M, Seifert J, et al. (2014) Diverse sulfate-reducing bacteria of the *Desulfosarcina*/*Desulfococcus* clade are the key alkane degraders at marine seeps ISME J. 8:2029–2044
84. Musat F (2015) The anaerobic degradation of gaseous, nonmethane alkanes—from in situ processes to microorganisms Comput. Struct. Biotechnol. J. 13:222–228
85. Beasley KK, Nanny MA (2012) Potential energy surface for anaerobic oxidation of methane via fumarate addition Environ. Sci. Technol. 46:8244–8252
86. Callaghan AV (2013) Enzymes involved in the anaerobic oxidation of n-alkanes: from methane to long-chain paraffins Front. Microbiol. 4:89
87. Mueller TJ, Grisewood MJ, Nazem-Bokaee H, Gopalakrishnan S, Ferry JG, Wood TK, Maranas CD (2015) Methane oxidation by anaerobic archaea for conversion to liquid fuels J. Ind. Microbiol. Biotechnol. 42:391–401
88. von Netzer F, Kuntze K, Vogt C, Richnow HH, Boll M, Lueders T (2016) Functional gene markers for fumarate-adding and dearomatizing key enzymes in anaerobic aromatic hydrocarbon degradation in terrestrial environments J. Mol. Microbiol. Biotechnol. 26:180–194
89. Rabus R, Boll M, Heider J, Meckenstock RU, Buckel W, Einsle O, et al. (2016) Anaerobic microbial degradation of hydrocarbons: from enzymatic reactions to the environment J. Mol. Microbiol. Biotechnol. 26:5–28