Organic matter processing by microbial communities throughout the Atlantic water column as revealed by metaproteomics

Kristin Bergauer[a,1], Antonio Fernandez-Guerra[b,c], Juan A. L. Garcia[a], Richard R. Sprenger[d], Ramunas Stepanauskas[e], Maria G. Pachiadaki[e], Ole N. Jensen[d], and Gerhard J. Herndl[a,f,g]

[a]Department of Limnology and Bio-Oceanography, University of Vienna, A-1090 Vienna, Austria; [b]Microbial Genomics and Bioinformatics Research Group, Max Planck Institute for Marine Microbiology, D-28359 Bremen, Germany; [c]Oxford e-Research Centre, University of Oxford, Oxford OX1 3QG, United Kingdom; [d]Department of Biochemistry and Molecular Biology, VILLUM Center for Bioanalytical Sciences, University of Southern Denmark, DK-5230 Odense M, Denmark; [e]Bigelow Laboratory for Ocean Sciences, East Boothbay, ME 04544; [f]Department of Marine Microbiology and Biogeochemistry, Royal Netherlands Institute for Sea Research, Utrecht University, 1790 AB Den Burg, The Netherlands; and [g]Vienna Metabolomics Center, University of Vienna, A-1090 Vienna, Austria

Edited by David M. Karl, University of Hawaii, Honolulu, HI, and approved November 21, 2017 (received for review May 26, 2017)

The phylogenetic composition of the heterotrophic microbial community is depth stratified in the oceanic water column down to abyssopelagic layers. In the layers below the euphotic zone, it has been suggested that heterotrophic microbes rely largely on solubilized particulate organic matter as a carbon and energy source rather than on dissolved organic matter. To decipher whether changes in the phylogenetic composition with depth are reflected in changes in the bacterial and archaeal transporter proteins, we generated an extensive metaproteomic and metagenomic dataset of microbial communities collected from 100- to 5,000-m depth in the Atlantic Ocean. By identifying which compounds of the organic matter pool are absorbed, transported, and incorporated into microbial cells, intriguing insights into organic matter transformation in the deep ocean emerged. On average, solute transporters accounted for 23% of identified protein sequences in the lower euphotic and ∼39% in the bathypelagic layer, indicating the central role of heterotrophy in the dark ocean. In the bathypelagic layer, substrate affinities of expressed transporters suggest that, in addition to amino acids, peptides and carbohydrates, carboxylic acids and compatible solutes may be essential substrates for the microbial community. Key players with highest expression of solute transporters were Alphaproteobacteria, Gammaproteobacteria, and Deltaproteobacteria, accounting for 40%, 11%, and 10%, respectively, of relative protein abundances. The in situ expression of solute transporters indicates that the heterotrophic prokaryotic community is geared toward the utilization of similar organic compounds throughout the water column, with yet higher abundances of transporters targeting aromatic compounds in the bathypelagic realm.

transporter proteins | organic matter | deep sea | Atlantic Ocean | metaproteomics

The dark ocean is, together with the deep subsurface, the most extensive and least explored biome on Earth, characterized by high hydrostatic pressure, low temperature, and low metabolic activities (1, 2). The principal carbon and energy source for the dark ocean's biota is primary production in the sunlit surface waters of the ocean and the resulting sedimentation of particulate organic matter (POM) into the ocean's interior, known as the biological carbon pump (3, 4). Once exported into the mesopelagic waters, the vertical flux of POM attenuates with depth due to direct utilization by the biota and/or solubilization via extracellular enzymatic activity by heterotrophic microbes (5, 6). This solubilization process of POM to dissolved organic matter (DOM) and the subsequent uptake of cleavage products by heterotrophic microbes are considered the rate-limiting step in microbial metabolism in the deep sea. Based on microbial activity measurements and dissolved organic carbon profiles throughout the water column, it is estimated that direct DOM utilization accounts for only about 10% of the microbial carbon demand in the mesopelagic and bathypelagic waters (7, 8). Despite this apparent low contribution of DOM compared with POM in supporting heterotrophic microbial metabolism in the deep ocean, DOM quantity and quality decreases with depth (9). The decrease in the overall nutritional quality of the DOM with depth might reflect the generation of increasingly recalcitrant compounds by heterotrophic microbes (10) in mesopelagic and bathypelagic layers, as suggested by the microbial carbon pump hypothesis (11). Another likely mechanism, formulated in the selective preservation hypothesis, is that labile DOM compounds are preferentially used, leaving behind refractory molecules. Alternatively, it might be simply a consequence of diluting out the large number of DOM molecules, making their utilization inefficient (12). The low concentrations of organics and the extremely high diversity of DOM molecules in the dark ocean (12, 13) constitute a challenge for microbes to maintain a sufficiently high encounter rate with solutes and to minimize the energy requirements associated with compound assimilation (14). All of these factors might contribute to the overall rather inefficient

Significance

Circumstantial evidence indicates that especially deep-ocean heterotrophic microbes rely on particulate organic matter sinking through the oceanic water column and being solubilized to dissolved organic matter (DOM) prior to utilization rather than on direct uptake of the vast pool of DOM in the deep ocean. Comparative metaproteomics allowed us to elucidate the vertical distribution and abundance of microbially mediated transport processes and thus the uptake of solutes throughout the oceanic water column. Taken together, our data suggest that, while the phylogenetic composition of the microbial community is depth stratified, the composition and substrate specificities of transporters considered in this study are ubiquitous while their relative abundance changes with depth.

Author contributions: K.B., O.N.J., and G.J.H. designed research; K.B. and R.R.S. performed research; A.F.-G., R.S., and M.G.P. contributed new reagents/analytic tools; K.B., J.A.L.G., and R.R.S. analyzed data; and K.B., R.S., and G.J.H. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database [accession nos. SRP081826 and SRP081823 (genomic sequence data); KY241481–KY241660, KY194331–KY194691, KY193976–KY194214, KY081807–KY081876, KX426906–KX426937, KX427579–KX428016, KX426472–KX426524, and KX426391–KX426464 (assembled 16S rRNA sequences)]. A selection of Malaspina gene sequences is deposited in Pangaea (https://doi.org/10.1594/PANGAEA.883794).

[1]To whom correspondence should be addressed. Email: kristin.bergauer@univie.ac.at.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1708779115/-/DCSupplemental.

microbial utilization of the large stock of deep-sea DOM and the apparent preferential utilization of sedimenting POM in the ocean's interior (15).

In this study, we integrated metaproteomic, metagenomic, and single-cell genomic analyses to elucidate protein expression patterns of microbial communities from the euphotic zone (100 m) to the dark ocean (300–4,050 m) along a large latitudinal range (67°N to 49°S) in the Atlantic Ocean. We focused on the spatial distribution and abundance of expressed transporters as key mechanisms for the uptake of essential nutrients, including the primary active ATP-binding cassette (ABC) transporters, as well as the secondary active tripartite ATP-independent periplasmic transporters (TRAP-Ts), the tripartite tricarboxylate transporters (TTTs), and the TonB-dependent transporters (TBDTs). The prime determinant of selectivity in import systems are the extracytoplasmic substrate-binding proteins (SBPs) (16, 17), which are constituents of primary and secondary active transporters and capture the substrate with high affinity from the cells' surroundings (16, 18). Besides their primary function as substrate translocators, SBPs play an important role in signal transduction (19). All SBP-dependent ABC systems are importers and ubiquitous in Bacteria and Archaea (20). Knowledge on the variability of the transporter proteins and their phylogenetic origin provides insights into the dynamics in nutrient scavenging of the microbial community in response to the organic matter (OM) availability with depth (21–26).

In this study, we addressed the following specific research questions: (*i*) Are changes in the phylogenetic composition of the microbial community tightly or loosely linked to changes in the type and abundance of transporter proteins for OM throughout the water column? (*ii*) Does the distribution of transporter protein indicate major changes in the OM compound classes used as substrate by the microbial communities between surface and bathypelagic waters?

Results and Discussion

The Transporter Repertoire of Open-Ocean Microbes. As a first step toward establishing a proteomic inventory of marine microbial communities, we performed protein mass spectrometry (Fig. S1) at 14 sample sites (Table S1) in the Atlantic Ocean (Fig. S2). Additionally to protein annotations retrieved by homology searches against genomic databases (*Supporting Information* and Fig. S3), functional classifications were inferred using (*i*) clusters of orthologous groups (COGs) of proteins and (*ii*) the Kyoto Encyclopedia of Genes and Genomes (KEGG) database for higher-order cellular processes (Table 1). A combined set of 1,002 nonredundant transport-related membrane proteins (29% of total proteins identified; Datasets S1 and S2), subsequently referred to as TMPs, was identified. Approximately 71% of the TMPs were identified in at least two depth layers, whereas only 7% (75 proteins) were detected in all metaproteomes (Fig. 1*A*). COG categories, defined for the transporter protein repertoire, resulted in a conserved set of 96 COG families (Dataset S3), subsequently referred to as transporter COGs. The comparison between the major water layers, namely the euphotic, the mesopelagic (300–850 m depth), and upper (1,500–2,000 m depth) and lower (2,750–4,050 m) bathypelagic zones, enabled us to identify 29 COG families (~30%) common to all metaproteomes (i.e., involved in amino acid, inorganic ion, and carbohydrate transport; Fig. 1*B*). Among the 31 COG families present exclusively in the bathypelagic layer, we identified transporter COGs from the categories of "Carbohydrate transport and metabolism" (G; i.e., organic acids, xylose, maltose), "Inorganic ion transport and metabolism" (P), and "Poorly characterized" (R). The high number of unique transporter COGs recovered from deep ocean likely reflects the overall higher number of proteins recovered from these samples. Noteworthy, despite the greater diversity of differential transporter COGs, no additional substrate specificities were predicted from the protein-coding sequences. This observation points at the importance of suitable and "complete" genomic databases and to accurately annotate protein-coding sequences. These results highlight also the potentially broad substrate range encounter by marine microbes, which needs yet to be defined. To account for differences in protein recovery between the samples, we applied semiquantitative analyses of transporter COGs, introducing normalized area abundance factors (NAAFs) (*Supporting Information*). Resulting expression profiles of transporter COGs indicate that functionally similar transport processes are present throughout the water column yet differ in their relative abundances between sampling sites (Fig. 2).

Vertical DOM Uptake Patterns Based on Transporter Proteins. Similar to previous metaproteomic studies (14, 21, 24, 25, 27, 28), ABC transporters (662 proteins, 20% of total identified proteins)

Table 1. Transport-related membrane protein statistics summarizing unique COG and KO classifiers

Metaproteomics										Reference databases		
			TMP									
Layer	Depth, m	Sample ID	No.	%	COG, %	KO, %	Unique TMP, no.	Average TMP, %	Moca, Geotraces, %	Malaspina, %	SAG, %
EUPH	100	MED1_13	55	15	66	82	28	23	19	10	71
		GEO_5	63	25	77	63	16		24	26	49
		MOC_7	132	21	70	85	214		13	11	76
		MOC_6	150	31	72	73	58		26	18	56
MESO	300	MOC_5	87	26	71	59	7	32	23	47	30
	500	GEO_6	59	20	76	76	57		18	17	65
	850	MED1_16	169	51	81	70	21		25	32	43
BATHY	1,475	MED2_24	418	40	80	64	116	39	30	33	37
	1,750	MED2_12	594	38	83	68	68		34	31	35
	2,000	MED2_16	442	39	82	69	36		30	31	39
	2,750	MED2_8	472	40	83	65	2		35	36	29
	2,800	MED2_17	675	38	83	67	104		34	35	31
	3,200	MED2_5	227	44	79	63	10		29	43	28
	4,050	MED1_24	429	35	83	62	162		26	54	20

Detailed information on the reference databases is provided for all depth layers considered in this study. Number (no.) indicates the number of protein sequences identified, and percentages were calculated for the entire metaproteomic data/sample. COG, cluster of orthologous groups; EUPH, lower euphotic layer; KO, KEGG orthology; SAG, single amplified genome; TMP, transport-related membrane protein.

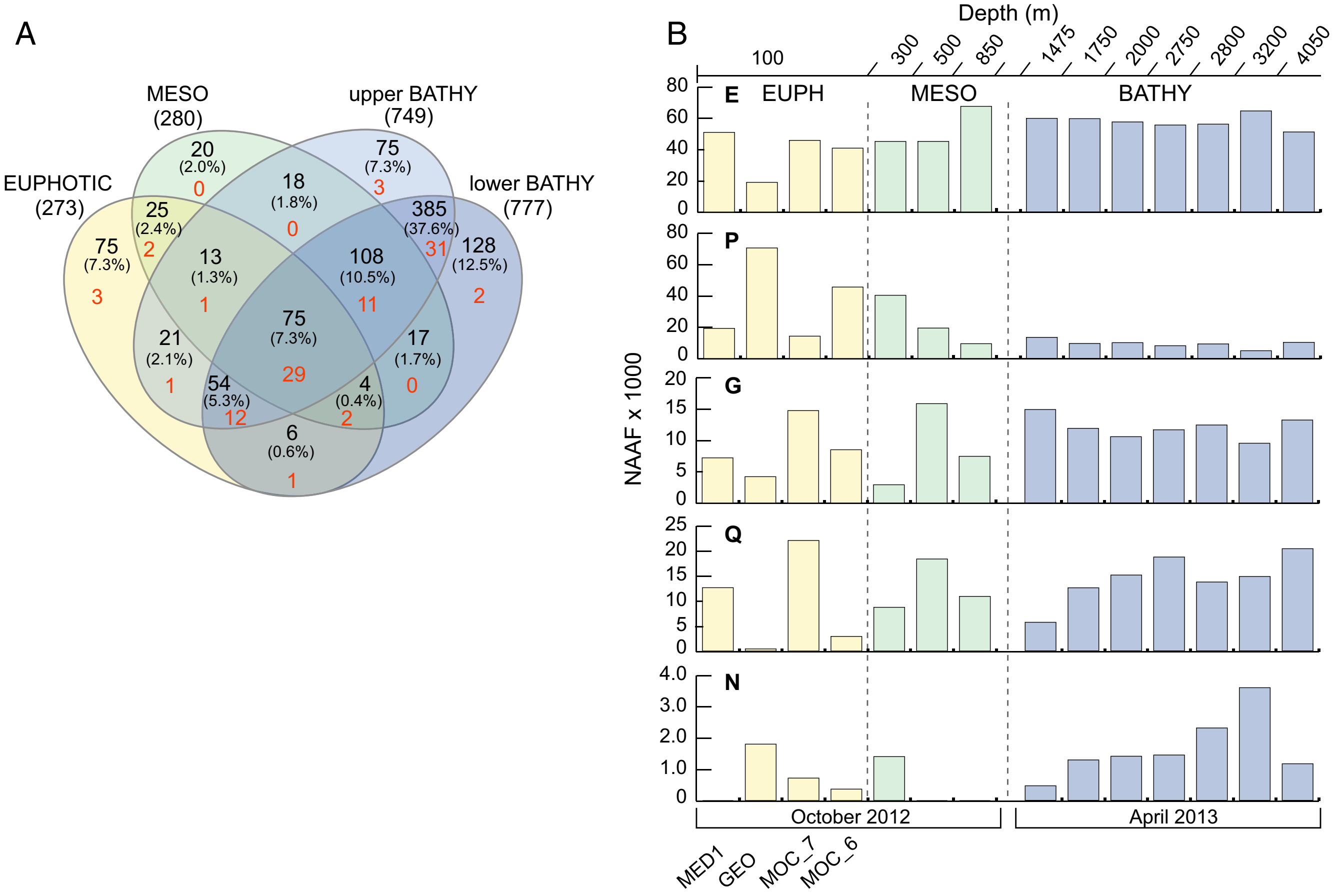

Fig. 1. Mascot and SEQUEST-HT search results were combined to create nonredundant lists of protein groups, and shared COGs as well as differences in transporter abundances between the samples are shown. (*A*) Venn diagram illustrating the number of COG families (red) shared between the 14 metaproteomes, grouped by the distinct water layers: EUPHotic (lower euphotic, 100 m), MESO (mesopelagic, 300–850 m), upper BATHY (bathypelagic, 1,475–1,973 m), and lower BATHY (2,750–4,050 m). Proteins not grouped into COGs are indicated in black letters. (*B*) Distribution and abundance of selected COG functional categories associated with transport functions and cell motility. COG categories: E, amino acid transport and metabolism; P, inorganic ion transport and metabolism; G, carbohydrate transport and metabolism; N, cell motility; Q, secondary metabolites biosynthesis, transport, and catabolism. Relative protein abundances are based on NAAF values.

comprised the most prevalent transporter system in our study. As reported from coastal surface waters off the Antarctic Peninsula (14), the SBPs were the most frequently encountered components of the ABC complex, whereas transmembrane domains were less often identified (4%; Dataset S1). Roughly 85% of ABC transporters were predicted periplasmic SBPs and 11% remained unassigned. TRAP transporters were among the 10 most expressed proteins in the metaproteomic repertoire (86 proteins; 9% of TMP and 2.5% of total identified proteins) and mostly comprised SBPs homologous to the DctP family. Only three SBPs were of the TRAP-associated extracytoplasmic immunity (TAXI) family (29, 30) (Dataset S1). Together with TRAP transporters (31), TTTs are considered as secondary active transporters found in Bacteria and Archaea that employ SBPs to capture their ligands (29, 30, 32, 33). TTTs accounted for the smallest fraction of prokaryotic transport systems (29 proteins; 3% of TMP and 1% of total identified proteins; Fig. 2), which might be explained by their narrow substrate range we know of until now (31, 34, 35), or due to the limited number of reference sequences available in public databases.

The third most abundant transporter system was the outer-membrane TBDT, which plays an important role in high–molecular-weight OM uptake (22). Metaproteomic (21), metagenomic (26), and metatranscriptomic (36) surveys indicate the ocean-wide (coastal to open ocean) presence of TBDTs in surface oceans and in deep-sea hydrothermal vents. In this study, we provide additional information on the vertical distribution and expression levels of TBDTs (107 proteins) in pelagic waters, accounting for 11% of TMP and 3% of total identified proteins.

The semiquantitative assessment of these transporter systems further confirmed the omnipresence and wealth of ABC transporters in the euphotic zone and the dark ocean (Fig. 2). Relative abundances of expressed ABC transporter proteins were on average six times higher than TRAP-Ts in the euphotic zone and 22 times higher than TBDTs in bathypelagic metaproteomes. TRAP-Ts and TTTs exhibited similar vertical expression patterns, with higher average abundances in the mesopelagic and bathypelagic zones, whereas TBDTs were expressed at higher values in the upper water column.

With the exception of TTT systems, transporter proteins mediating the uptake of both low– and high–molecular-weight OM were detected at all depths. However, relative expression levels of transporter systems varied with depths, suggesting depth-related quantitative changes in substrate uptake patterns while qualitative changes in the transporter systems were not apparent (Fig. S4). Importantly, the fraction of transporters increased from ~23% in the euphotic layers, to 32% in the mesopelagic and 39% in bathypelagic waters (Table 1). The increase in the fraction of transporter proteins from the euphotic to the bathypelagic zone as observed in this study (Table 1) might be interpreted as an

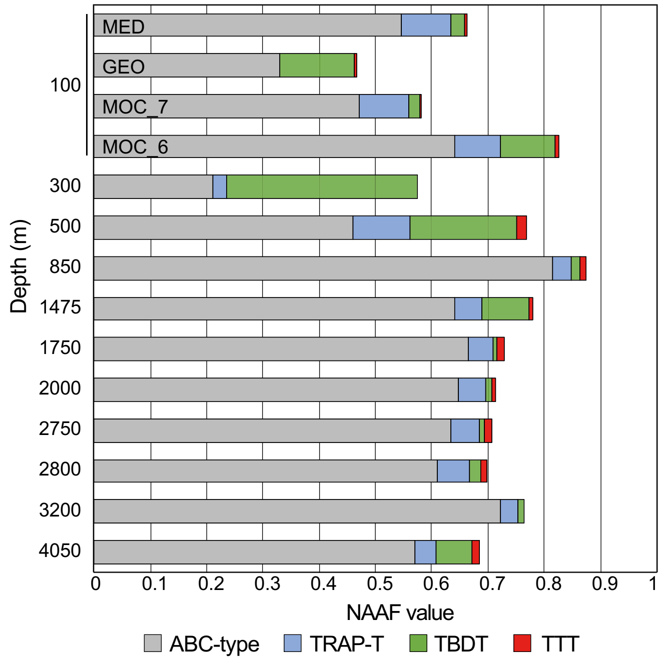

Fig. 2. Vertical distribution and relative abundance (NAAF) of ATP-binding cassette (ABC), tripartite ATP-independent (TRAP-T), tripartite tricarboxylate (TTT), and TonB-dependent (TBDT) transporters. The sample IDs for metaproteomes from the euphotic zone are indicated; the coordinates and the physical and chemical characteristics of the samples are given in Table S1.

adaptation of the deep-sea microbial community to the changes in quality and quantity of the OM in the water column. Prokaryotic abundances and leucine incorporation rates, used as a proxy for heterotrophic prokaryotic production, typically decrease from the sunlit to the abyssopelagic waters by roughly three orders of magnitude (37). Cell-specific leucine incorporation rates, however, analyzed along the cruise track of Geotraces and Medea (*Supporting Information*) remained fairly constant in bathypelagic and abyssopelagic waters (1,000–5,500 m; Table S2), with average rates of $2.1 \pm 2.2 \times 10^{-5}$ fmol Leu·cell^{-1}·d^{-1}, indicating a fairly constant heterotrophic activity below the mesopelagic zone.

The Substrate-Active Community. The taxonomic diversity derived from expressed transporter proteins, referred to as the substrate-active community, comprised mainly Bacteria (69%) and 2% Archaea (Fig. 3). Roughly 30% of TMPs (294 proteins) remained unclassified due to divergent assignments at the phylum level or insignificant hits (Dataset S1). Expression profiles of the substrate-active community revealed the following contribution of phyla: Proteobacteria (Alphaproteobacteria, 40% of NAAF values; Gammaproteobacteria, 10%; Deltaproteobacteria, 7%), Actinobacteria (5%), Bacteroidetes (5%), Thaumarchaeota (4%), Cyanobacteria (3%), and Euryarchaeota and Planctomycetes (each 1%). A depth-dependent stratification of substrate-responsive phyla, based on the expression of transporter-related proteins, was evident for Cyanobacteria, Actinobacteria, Bacteroidetes, Verrucomicrobia, Firmicutes, and the candidate phylum Marinimicrobia, accounting for higher relative abundances in euphotic and mesopelagic layers than in bathypelagic waters (Fig. 3). Conversely, transporters from Deferribacterales, Planctomycetes, Chloroflexi, and members of the Gammaproteobacteria and Deltaproteobacteria were relatively more abundant in the bathypelagic than in water masses above.

Concomitant metagenomes recovered between 1,809 and 65,228 predicted protein-coding genes (Table S3) and 32–438 near–full-length 16S rRNA gene sequences (*Supporting Information* and Table S4). The recovery of 16S rRNA genes from metagenomic libraries supports the findings on the diversity deduced from 16S rRNA gene PCR amplifications. Vertical distribution patterns based on variations in estimated 16S rRNA gene abundances revealed similar trends in the microbial community structure (*Supporting Information* and Fig. S5). Noteworthy, however, estimated 16S rRNA gene abundances indicated higher relative abundances of Thaumarchaeota throughout the water column than suggested by transporter protein abundances, contributing up to ∼20% of the microbial community in bathypelagic waters (Fig. S5). Thus, transporter proteins primarily selected as a proxy for transport and degradation of complex organic molecules, might underestimate the activity and abundance of marine Archaea and the relative contribution of chemolithoautotrophic metabolism versus heterotrophy. The genetic capabilities of Thaumarchaeota for the uptake or translocation of compounds are described in detail elsewhere (38). In our metaproteomes, predicted thaumarchaeal transporters included ammonium channel proteins

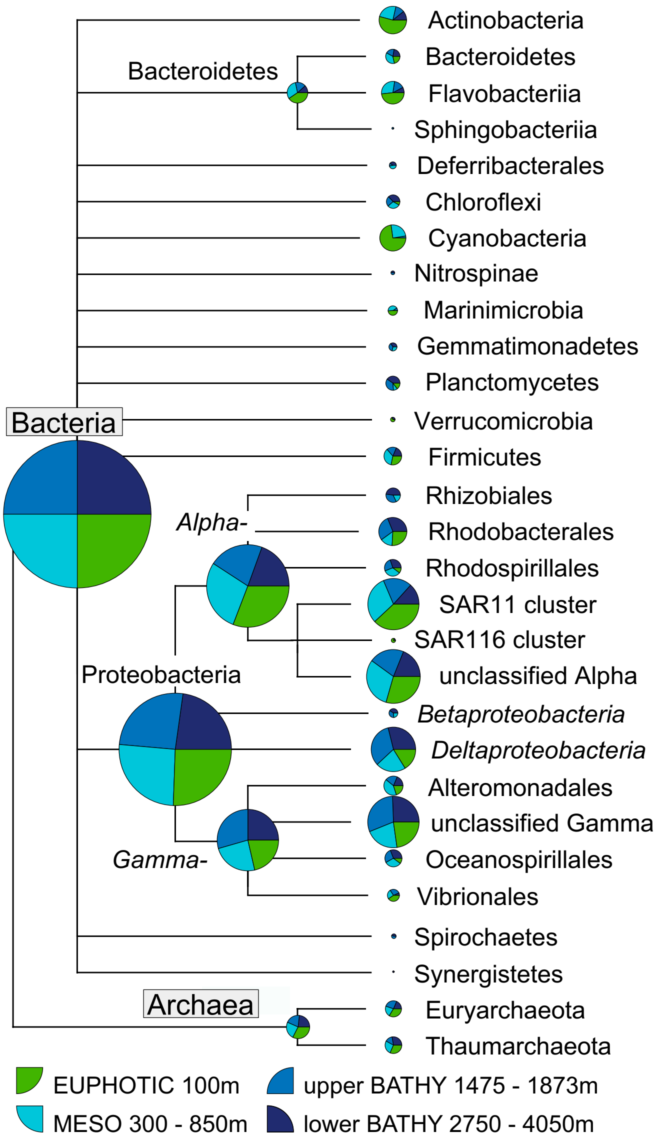

Fig. 3. Taxonomic distribution of *Bacteria* and *Archaea* assigned to transport-related membrane proteins (TMPs). Pie charts represent semi-quantitative abundance estimates based on averaged NAAF values for all TMPs at the designated water layers (see color key).

(Amt), urea, and SSS family transporters revealing, despite the comparatively low expression levels, the potential of the organism to utilize inorganic and organic energy sources (39).

Substrate Uptake Patterns at the Community Level. COG and KEGG Orthology (KO) classifiers were used to resolve the compound specificity of transport proteins (Fig. 4), and for prevalent microbial groups, average NAAFs, referred to as "protein abundance" from here on, were summarized for the euphotic, the mesopelagic, and the upper and lower bathypelagic zones (Fig. 5).

Predicted substrate affinities of expressed transporter proteins suggest that amino acids, including branched-chain amino acids, proline/glycine betaine, and di- and oligopeptides, as well as carbohydrates and carboxylic acids, might represent essential components of the OM pool in the oligotrophic open ocean.

ABC transporters exhibited the highest expression values in every depth layer attributed to their involvement in the uptake of amino acids (245 proteins; Fig. 4). Predicted ligands of the respective SBPs included proteinogenic amino acids such as glutamate, arginine and derivatives, histidine, proline, and glycine or more general branched-chain amino acid (leucine, isoleucine, and valine) and polar- and L-amino acids (Fig. S6 and Dataset S1). Various bacterial phyla and Euryarchaeota Marine Group II (MG-II) expressed amino acid transporters, albeit the proportions of active cells varied among groups (Dataset S1). In agreement with actual measurements of amino acid uptake rates (40), Alphaproteobacteria and Gammaproteobacteria accounted for high relative abundances of amino acid transporters, together contributing ~30% to the expression values in the lower euphotic, 83% in the mesopelagic, and 85% in the bathypelagic zone (Fig. 5). As previously described based on transcriptome data (41), members of the SAR11 clade invest substantially in the uptake of compatible solutes, for example, proline/glycine betaine (ProU operon), reportedly supporting cell growth (42). In sinking POM, amino acid-like material accounts for the largest component (40–50%) of particulate organic carbon (43). Interestingly, minimal changes in the chemical structure of the bulk POM composition were found throughout the water column (43). Together with amino acids and lipids, carbohydrates constitute a major group of biomolecules in the OM pool and comprise up to 30% (by mass) of sinking particles (43). We identified 109 proteins of ABC carbohydrate-binding transporters: (*i*) monosaccharide and (*ii*) disaccharide and oligosaccharide transporters, which are homologous to the family of oligopeptide and dipeptide transporters. Carbohydrate transporter abundances accounted for similarly high expression values as observed for peptides with only minor variations in relative abundances throughout the water column (Fig. 4). Predicted COG/KO functions indicate a broad range of potential target molecules (Fig. S6), with a remarkably high fraction of transporters involved in the uptake of glycerol-3-phosphate (G3P) (Dataset S4). Expression profiles of the carbohydrate-active community indicated a vertical stratification in the water column, with higher relative abundances of Actinobacteria, Firmicutes, and Gammaproteobacteria in the upper water column, whereas Deltaproteobacteria dominated in underlying waters (Fig. 5). SAR324 clade members showed high expression values of

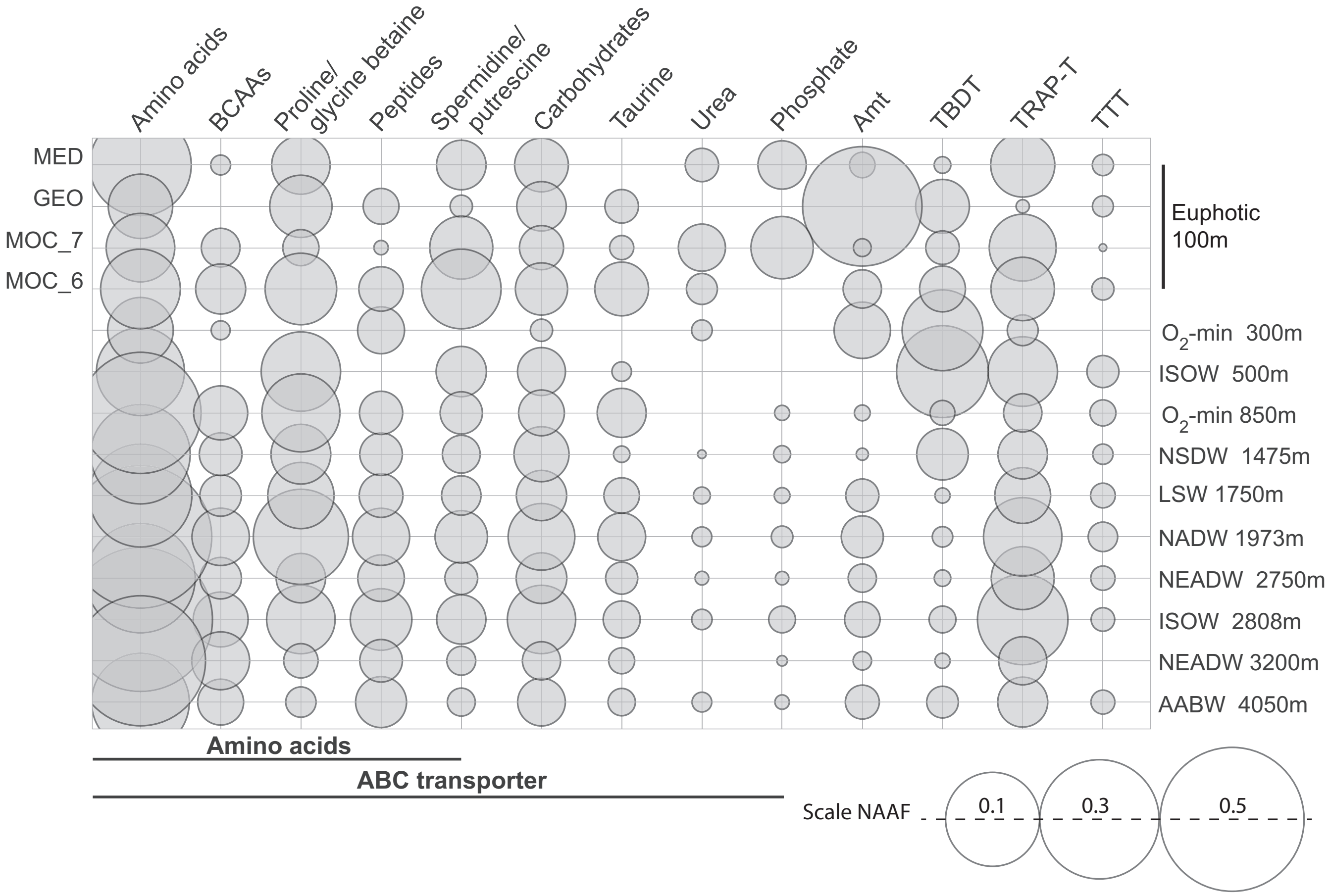

Fig. 4. Vertical expression profiles of selected transporters analyzed in a semiquantitative manner based on NAAF values. Transporter proteins were grouped by the predicted substrate specificity of the SBP. BCAAs, branched-chain amino acids.

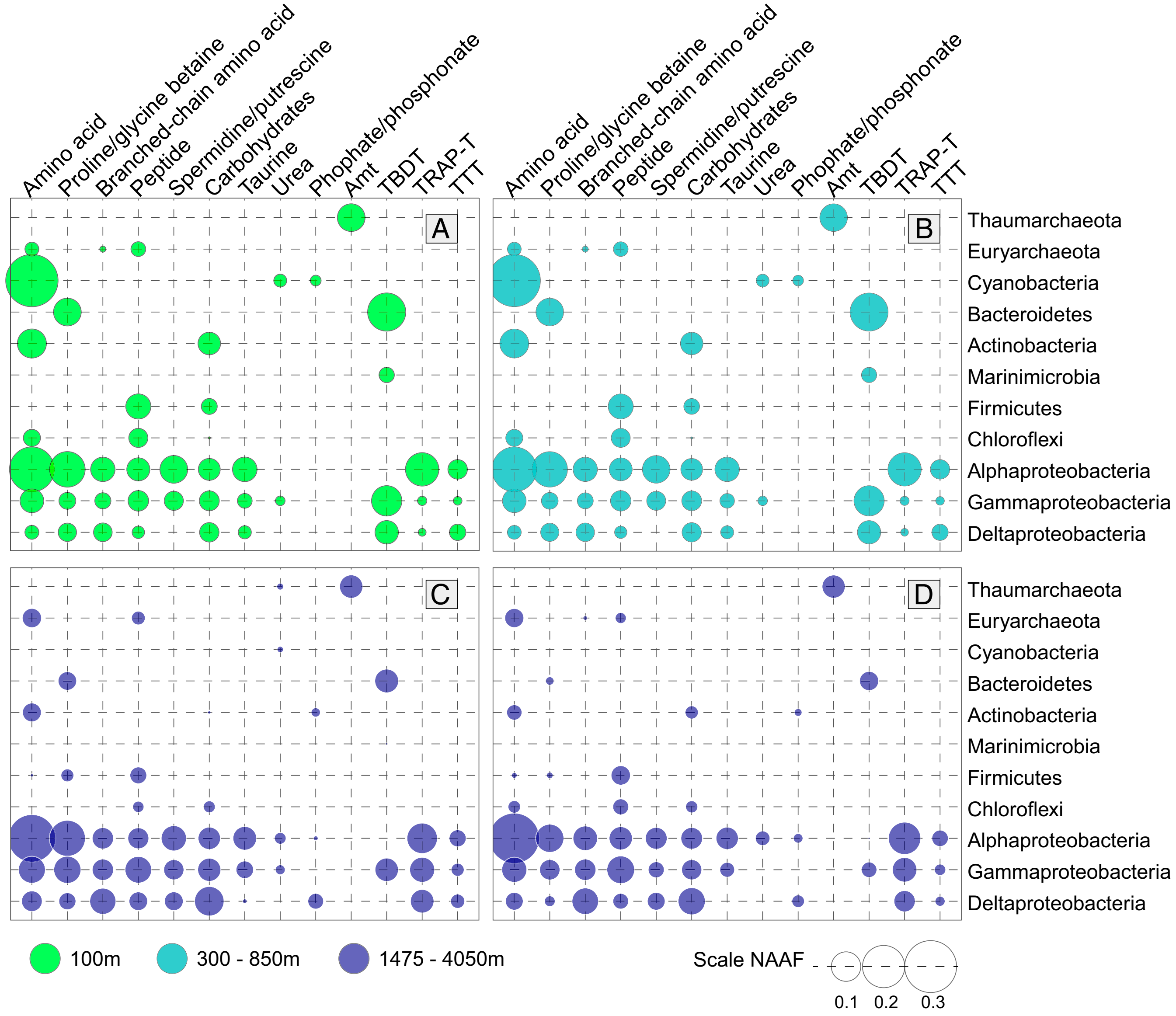

Fig. 5. Vertical expression profiles of transporter proteins of abundant taxa. Expression values were calculated in a semiquantitative manner and average abundances were plotted for selected members of the substrate active community residing in the (*A*) lower euphotic, (*B*) mesopelagic, and (*C*) upper and (*D*) lower bathypelagic water layers.

transporters specific to G3P, which might be utilized as a source of carbon or inorganic phosphate under nutrient-limiting conditions (44, 45). The expression of indicator genes for flagellar-related assembly (COG1344; Dataset S1) may indicate that at least some members of the SAR324 cluster possess a strong gliding or adhesion capability (46), suggesting the tendency to associate with particles (47).

Another omnipresent and abundantly expressed compound class was ABC peptide transporters, mediating the bulk uptake of dipeptides and oligopeptides (Dpp and Opp systems). We identified 115 transporter proteins involved in peptide utilization (Dataset S1). Independent of the sampling depth, relative abundances of dipeptide transporters were twice as high as those for oligopeptides (Fig. S6). While their primary function is the import of peptides as carbon and energy source, their role as messengers in virulence or signaling processes, for example, modulating chemotactic behavior or peptide export for biofilm formation, has to be taken into consideration. Thereby, peptide transporters exemplify that it is inevitable to perform biochemical analyses to verify substrate specificities and/or functions of the predicted transporter components. Various taxa demonstrated protein-processing capabilities, with the MG-II/III Euryarchaeota, Planctomycetes, and Gammaproteobacteria being the prime protein utilizers (Fig. 5). This is in agreement with high transcript levels of extracellular peptidases and carbohydrate-active enzymes (CAZymes) of MG-II and MG-III Euryarchaeota (48), suggesting the metabolic capacity of extracellular protein degradation and utilization of carbohydrates for heterotrophic growth (Table S5).

Another ubiquitous source of labile carbon and nitrogen in the ocean constitute polyamines like putrescine and spermidine. The occurrence of polyamine transporters has been studied in marine bacterial genomes (49), revealing the wide distribution of SBPs among the phyla of Actinobacteria, Chlamydiae/Verrucomicrobia, Cyanobacteria, Firmicutes, and Proteobacteria. Surveys of marine metagenomes revealed that up to 32% of surface ocean bacterioplankton potentially encode homologs implicated in the transport or degradation of polyamines, suggesting the important role of polyamines in carbon and nitrogen cycling (50). In our study, the active polyamine-transforming community comprised Alphaproteobacteria (25 proteins; including Rhodobacterales, Rhizobiales, and Pelagibacterales), Gammaproteobacteria (14 proteins), and Deltaproteobacteria (7 proteins), with SAR11 dominating in the euphotic zone (Fig. 5). Overall, relative abundances of polyamine transporter systems (PotD, PotF; 56 proteins) suggest a higher expression in the euphotic zones than underlying water masses (Fig. 4).

ABC transporters expressed at considerably lower abundances had predicted substrate specificities to urea, phosphate/phosphonate, or taurine (2-aminoethane sulfonate), sources of organic carbon, nitrogen, and/or sulfur moieties (Fig. 4). Taurine, in particular, constitutes a valuable food source for heterotrophic microbes due to its carbon, nitrogen, and sulfur moieties and is found at nanomolar concentrations in the marine environment (51). Key enzymes required to utilize these nutritional elements (52, 53) include the taurine-pyruvate aminotransferase (Tpa), alanine dehydrogenase (Ald), and sulfoacetaldehyde acetyltransferase (Xsc), facilitating the utilization of taurine as C source (52, 53). Thirteen proteins encoding the SBP (TauA) of the taurine transport system (TauABC) were detected in our metaproteomes at higher average abundances in the euphotic zone than in deeper layers (Fig. 4; NAAFs ranging between 0.003 and 0.06). Proteins encoding the Xsc (Dataset S1) were recovered at 500 m and 1,475- to 2,808-m depth, suggesting that taurine might serve as a favorable C source also in mesopelagic and bathypelagic waters. Function predictions of several proteobacterial single amplified genomes revealed genes or homologs for the TauD and/or Tpa, suggesting the genetic potential of C and S utilization. Our metagenomes clearly indicated an overall increase of TauD gene abundances below the euphotic zone, pointing at the importance of taurine as S source in the dark ocean (Dataset S5). Within the Alphaproteobacteria, SAR11 cluster members appeared as the prime utilizers of taurine in bathypelagic waters (Fig. 5). This is in agreement with a previous study, experimentally demonstrating the growth of SAR11 on taurine (54) and the ubiquity of SAR11 taurine transporters in metaproteomic and metatranscriptomic surveys in coastal surface waters (14, 28, 41).

Additionally to ABC transporters, diverse members of the microbial community expressed TRAP transporters at relatively high abundances, particularly in bathypelagic layers (Fig. 5). TRAP transporters are ATP independent, and thus less energy consuming than ABC transporters and, consequently, more advantageous under oligotrophic, deep-sea conditions. Known substrates recognized by the SBPs include a wide variety of compounds, all characterized by the presence of a carboxylate group, that is, organic acids (amino acids and acid sugars) and other carboxylate-containing small metabolites (29, 30, 32, 33, 55). Besides nutritional benefits, TRAP transporters support the thermostabilization of microbial cells by translocating osmoprotectants such as ectoine and 5-hydroxyectone (56). Predicted substrate affinities of TRAP SBPs identified in this study included mannitol/chloroaromatic compounds and aromatic acids (35) (COG4663, FcbT1; 45 proteins) and C4-dicarboxylates such as malate, fumarate, and succinate used as carbon and energy sources (COG1638, DctP; 36 proteins). Collectively, the phylogenetically diverse expression of these transporters and their ubiquity may indicate their ecological importance in dark ocean substrate turnover.

High proportional abundances of outer membrane TBDTs (107 proteins) were observed in euphotic and mesopelagic layers (NAAF euphotic, 0.0747 ± 0.047; meso, 0.1778 ± 0.117; bathy, 0.03 ± 0.03), similar to environmental metatranscriptomic (57) and metaproteomic (21) studies. TonB transporters were originally identified in the context of iron transport but are now reported to be involved in the uptake of a variety of compounds (58).

TonB-dependent transporter proteins encoded for cobalamine or vitamin B12 receptors (COG4206, BtuB), Ton box of ferric citrate (COG4772, FecA), outer membrane Fe transporters (COG1629, CirA), and outer membrane receptors for ferrienterochelin and colicins (COG4771, FepA). Also, we identified eight proteins encoding the SusC/RagA clade of TonB-linked outer membrane proteins, presumably utilizing large protein fragments (i.e., RagA) or carbohydrates (i.e., SusC) as organic compounds (Dataset S1). TonB-related proteins were also identified as among the most abundant transcripts assigned to DOM-responsive Idiomarinaceae and Alteromonadacea in a DOM-enriched marine microcosm (22). Both FepA and FecA facilitate the uptake of iron complexed to OM in diverse marine bacteria (59) in addition to other transport capacities. Fe is a critical trace element for bacterioplankton and impacts carbon and nitrogen fixation over broad regions of the ocean (60). However, to our knowledge, Fe(II) oxidation has not been observed in the open water column. In agreement with previous reports (26, 61), phylogenetic analyses of TBDTs strongly indicated that Gammaproteobacteria (34 proteins) and Bacteroidetes (including Cytophaga, Flavobacteria, and Sphingobacteria; 19 proteins) degrade polymeric matter throughout the water column. The taxonomic classification and relative abundance of TBDTs throughout the water column reveal a clear stratification of Flavobacteria, Alteromonadales, SAR86 cluster bacteria, and Marinimicrobia, dominating in the upper ocean. Conversely, TBDTs of Deferribacteres and Gemmatimonadetes were expressed at higher abundances in the dark ocean (Dataset S1). Taken together, these observations further highlight the significance of transporter processes throughout the water column and likely illustrate depth-related partitioning of marine bacteria according to environmental conditions.

Conclusion

Understanding the biogeochemistry of the OM pool remains a central challenge in microbial ecology (23, 62, 63), and thus, the analyses of expressed transporters in the open water column provide fundamental insights on this important topic. In concert, our data suggest that despite the commonly observed decreasing concentrations of OM with depth, transport proteins are expressed at high levels accounting for up to 39% of assigned COGs in bathypelagic zones (Table 1). Our study revealed expression patterns of prevalent transporter systems targeting various substrates, with no indications of major changes in the substrate affinity of transporter systems. The relative quantity of individual transporter systems, however, changed with depth such as TRAP transporters being particularly abundant among the bathypelagic microbial community (Fig. 4). Thus, while the phylogenetic composition of the microbial community is depth- and water mass-specific (Fig. S5), the composition of transporter proteins, indicative of the substrate quality, changes only marginally with depth. Using transporter components as proxy for OM translocation and utilization, we hypothesize that low substrate concentrations might be compensated by either modifying the abundance of the corresponding transporters by the individual microbes or by a shift in the microbial community. Despite multiple challenges imposed by "omics" tools, for example, peptide recovery and annotation accuracy (64, 65), or misleading annotations for protein families with diverse functions, transport proteins currently give the most subtle clue to assess the organic and inorganic compounds actually being used by microbes in the deep ocean. Overall, the distribution of expressed transporter proteins from the subsurface to the bathypelagic waters lends support to the major role of POM solubilization as a carbon and energy source for deep-ocean microbes rather than direct DOM utilization.

Materials and Methods

Sampling and Metadata Collection. Sampling was performed during the research cruises Geotraces III (March 2010), Moca (October 2010), and Medea I and II (October 2011, June/July 2012), spanning a latitudinal range in the Atlantic Ocean from 67°N to 49°S (Fig. S2). To investigate the vertical structure and activity of microbial communities, water samples (100–600 L) for metagenomic (100–5,002 m) and metaproteomic (100–4,050 m) analyses were collected from eight distinct water masses and 14 depth layers (Table S1).

Metagenomics. Preparation of genomic DNA was performed using a standard phenol extraction protocol (*Supporting Information*). Contigs were assembled using paired-end Illumina HiSeq reads, and the assembly and prediction of ORFs were performed using the software tools MetaVelvet and Prokka (Fig. S1). 16S rRNA gene sequences were assembled using EMIRGE (66) and identified by BLASTn searching against the Silva SSU 128 database (Fig. S5).

Metaproteomics. Protein extraction and spin filter-aided in-solution digestion (SF-ISD) was performed according to León et al. (67) with modifications to optimize protein recovery (*Supporting Information*). Equal peptide aliquots generated from each sample were analyzed by ultrahigh-pressure nanoLC coupled to an LTQ-Orbitrap Velos Mass Spectrometer (Thermo Scientific). Peptide separation was performed with a nanoAcquity UPLC system (Waters) fitted with a 2-cm, 180-μm ID Symmetry C18 trap column (Waters) and a 25-cm, 75-μm ID, BEH C18 (1.7-μm particles) analytical column (Waters). Peptides were trapped for 2 min at 10 μL/min with 0.1% TFA and separated at 350 nL/min using a gradient of 5–35% acetonitrile with 0.1% formic acid for 75 min.

Detailed information on the experimental procedures, functional assignments, and quantification methods can be found in *SI Materials and Methods*.

Data Availability. Genomic sequence data that support the findings of this study have been deposited in GenBank with accession codes SRP081826 and SRP081823 and Pangaea (https://doi.org/10.1594/PANGAEA.883794). Accession codes of assembled 16S rRNA sequences are as follows: KY241481–KY241660, KY194331–KY194691, KY193976–KY194214, KY081807–KY081876, KX426906–KX426937, KX427579–KX428016, KX426472–KX426524, and KX426391–KX426464. Peptide sequences generated and analyzed during this study are included in *Supporting Information*.

ACKNOWLEDGMENTS. We thank the participants of the Malaspina circumnavigation expedition for kindly providing metagenomic data, in particular Pablo F. Sanchez, Silvia G. Acinas, and Josep M. Gasol. Financial support was provided by Austrian Science Fund Projects I486-B09 and Z194-B17, the European Research Council under the European Community's Seventh Framework Program (FP7/2007-2013)/European Research Council Grant Agreement 268595 (MEDEA project) (to G.J.H.), and National Science Foundation Grants OCE-1232982 and OCE-1335810 (to R.S.).

1. Nagata T, et al. (2010) Emerging concepts on microbial processes in the bathypelagic ocean—ecology, biogeochemistry, and genomics. *Deep Sea Res Part II Top Stud Oceanogr* 57:1519–1536.
2. Aristegui J, Gasol JM, Duarte CM, Herndl GJ (2009) Microbial oceanography of the dark ocean's pelagic realm. *Limnol Oceanogr* 54:1501–1529.
3. Buesseler KO, Lampitt RS (2008) Introduction to "Understanding the ocean's biological pump: Results from VERTIGO"–preface. *Deep Sea Res Part II Top Stud Oceanogr* 55:1519–1521.
4. Ducklow HW, Steinberg DK, Buesseler KO (2001) Upper ocean carbon export and the biological pump. *Oceanography* 14:50–58.
5. Delong EF, Franks DG, Alldredge AL (1993) Phylogenetic diversity of aggregate-attached vs free-living marine bacterial assemblages. *Limnol Oceanogr* 38:924–934.
6. Herndl GJ, Reinthaler T (2013) Microbial control of the dark end of the biological pump. *Nat Geosci* 6:718–724.
7. Arístegui J, et al. (2002) Dissolved organic carbon support of respiration in the dark ocean. *Science* 298:1967.
8. Arístegui J, Del Giorgio PA, Williams PJlB (2005) Respiration in the mesopelagic and bathypelagic zones of the ocean. *Respiration in Aquatic Ecosystems* (Oxford Univ Press, Oxford), pp 181–205.
9. Baltar F, Aristegui J, Gasol JM, Sintes E, Herndl GJ (2009) Evidence of prokaryotic metabolism on suspended particulate organic matter in the dark waters of the subtropical North Atlantic. *Limnol Oceanogr* 54:182–193.
10. Brophy JE, Carlson DJ (1989) Production of biologically refractory dissolved organic-carbon by natural seawater microbial-populations. *Deep-Sea Res* 36:497–507.
11. Jiao N, et al. (2010) Microbial production of recalcitrant dissolved organic matter: Long-term carbon storage in the global ocean. *Nat Rev Microbiol* 8:593–599.
12. Arrieta JM, et al. (2015) Ocean chemistry. Dilution limits dissolved organic carbon utilization in the deep ocean. *Science* 348:331–333.
13. Zark M, Christoffers J, Dittmar T (2017) Molecular properties of deep-sea dissolved organic matter are predictable by the central limit theorem: Evidence from tandem FT-ICR-MS. *Mar Chem* 191:9–15.
14. Williams TJ, et al. (2012) A metaproteomic assessment of winter and summer bacterioplankton from Antarctic Peninsula coastal surface waters. *ISME J* 6:1883–1900.
15. Hansell DA, Carlson CA, Repeta DJ, Schlitzer R (2009) Dissolved organic matter in the ocean: A controversy stimulates new insights. *Oceanography* 22:202–211.
16. Berntsson RP, Smits SH, Schmitt L, Slotboom DJ, Poolman B (2010) A structural classification of substrate-binding proteins. *FEBS Lett* 584:2606–2617.
17. Davidson AL, Dassa E, Orelle C, Chen J (2008) Structure, function, and evolution of bacterial ATP-binding cassette systems. *Microbiol Mol Biol Rev* 72:317–364.
18. Lewinson O, Livnat-Levanon N (2017) Mechanism of action of ABC importers: Conservation, divergence, and physiological adaptations. *J Mol Biol* 429:606–619.
19. Piepenbreier H, Fritz G, Gebhard S (2017) Transporters as information processors in bacterial signalling pathways. *Mol Microbiol* 104:1–15.
20. Saier MH, Jr (1998) Molecular phylogeny as a basis for the classification of transport proteins from bacteria, archaea and eukarya. *Adv Microb Physiol* 40:81–136.
21. Morris RM, et al. (2010) Comparative metaproteomics reveals ocean-scale shifts in microbial nutrient utilization and energy transduction. *ISME J* 4:673–685.
22. McCarren J, et al. (2010) Microbial community transcriptomes reveal microbes and metabolic pathways associated with dissolved organic matter turnover in the sea. *Proc Natl Acad Sci USA* 107:16420–16427.
23. Poretsky RS, Sun S, Mou X, Moran MA (2010) Transporter genes expressed by coastal bacterioplankton in response to dissolved organic carbon. *Environ Microbiol* 12:616–627.
24. Sowell SM, et al. (2011) Environmental proteomics of microbial plankton in a highly productive coastal upwelling system. *ISME J* 5:856–865.
25. Sowell SM, et al. (2009) Transport functions dominate the SAR11 metaproteome at low-nutrient extremes in the Sargasso Sea. *ISME J* 3:93–105.
26. Tang K, Jiao N, Liu K, Zhang Y, 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.
27. Colatriano D, et al. (2015) Metaproteomics of aquatic microbial communities in a deep and stratified estuary. *Proteomics* 15:3566–3579.
28. Georges AA, El-Swais H, Craig SE, Li WKW, Walsh DA (2014) Metaproteomic analysis of a winter to spring succession in coastal northwest Atlantic Ocean microbial plankton. *ISME J* 8:1301–1313.
29. Mulligan C, Fischer M, Thomas GH (2011) Tripartite ATP-independent periplasmic (TRAP) transporters in bacteria and archaea. *FEMS Microbiol Rev* 35:68–86.
30. Kelly DJ, Thomas GH (2001) The tripartite ATP-independent periplasmic (TRAP) transporters of bacteria and archaea. *FEMS Microbiol Rev* 25:405–424.
31. Winnen B, Hvorup RN, Saier MH, Jr (2003) The tripartite tricarboxylate transporter (TTT) family. *Res Microbiol* 154:457–465.
32. Fischer M, Zhang QY, Hubbard RE, Thomas GH (2010) Caught in a TRAP: Substrate-binding proteins in secondary transport. *Trends Microbiol* 18:471–478.
33. Rabus R, Jack DL, Kelly DJ, Saier MH, Jr (1999) TRAP transporters: An ancient family of extracytoplasmic solute-receptor-dependent secondary active transporters. *Microbiology* 145:3431–3445.
34. Antoine R, et al. (2005) The periplasmic binding protein of a tripartite tricarboxylate transporter is involved in signal transduction. *J Mol Biol* 351:799–809.
35. Hosaka M, et al. (2013) Novel tripartite aromatic acid transporter essential for terephthalate uptake in *Comamonas* sp. strain E6. *Appl Environ Microbiol* 79:6148–6155.
36. Li M, et al. (2014) Microbial iron uptake as a mechanism for dispersing iron from deep-sea hydrothermal vents. *Nat Commun* 5:3192.
37. De Corte D, Sintes E, Yokokawa T, Reinthaler T, Herndl GJ (2012) Links between viruses and prokaryotes throughout the water column along a North Atlantic latitudinal transect. *ISME J* 6:1566–1577.
38. Offre P, Kerou M, Spang A, Schleper C (2014) Variability of the transporter gene complement in ammonia-oxidizing archaea. *Trends Microbiol* 22:665–675.
39. Kerou M, et al. (2016) Proteomics and comparative genomics of *Nitrososphaera viennensis* reveal the core genome and adaptations of archaeal ammonia oxidizers. *Proc Natl Acad Sci USA* 113:E7937–E7946.
40. Alonso C, Pernthaler J (2006) Concentration-dependent patterns of leucine incorporation by coastal picoplankton. *Appl Environ Microbiol* 72:2141–2147.
41. Gifford SM, Sharma S, Booth M, Moran MA (2013) Expression patterns reveal niche diversification in a marine microbial assemblage. *ISME J* 7:281–298.
42. Tripp HJ, et al. (2008) SAR11 marine bacteria require exogenous reduced sulphur for growth. *Nature* 452:741–744.
43. Hedges JI, et al. (2001) Evidence for non-selective preservation of organic matter in sinking marine particles. *Nature* 409:801–804.
44. Vetting MW, et al. (2015) Experimental strategies for functional annotation and metabolism discovery: Targeted screening of solute binding proteins and unbiased panning of metabolomes. *Biochemistry* 54:909–931.
45. Lackner G, Peters EE, Helfrich EJ, Piel J (2017) Insights into the lifestyle of uncultured bacterial natural product factories associated with marine sponges. *Proc Natl Acad Sci USA* 114:E347–E356.
46. Cao H, et al. (2016) Delta-proteobacterial SAR324 group in hydrothermal plumes on the South Mid-Atlantic Ridge. *Sci Rep* 6:22842.
47. Swan BK, et al. (2011) Potential for chemolithoautotrophy among ubiquitous bacteria lineages in the dark ocean. *Science* 333:1296–1300.
48. Li M, et al. (2015) Genomic and transcriptomic evidence for scavenging of diverse organic compounds by widespread deep-sea archaea. *Nat Commun* 6:8933.

49. Mou XZ, Sun SL, Rayapati P, Moran MA (2010) Genes for transport and metabolism of spermidine in *Ruegeria pomeroyi* DSS-3 and other marine bacteria. *Aquat Microb Ecol* 58:311–321, and erratum (2010) 59:102.
50. Howard EC, Sun S, Biers EJ, Moran MA (2008) Abundant and diverse bacteria involved in DMSP degradation in marine surface waters. *Environ Microbiol* 10:2397–2410.
51. Clifford EL, et al. (2017) Crustacean zooplankton release copious amounts of dissolved organic matter as taurine in the ocean. *Limnol Oceanogr* 62:2745–2758.
52. Chien CC, Leadbetter ER, Godchaux W (1999) *Rhodococcus* spp. utilize taurine (2-aminoethanesulfonate) as sole source of carbon, energy, nitrogen and sulfur for aerobic respiratory growth. *FEMS Microbiol Lett* 176:333–337.
53. Cook AM, Denger K, Smits THM (2006) Dissimilation of C3-sulfonates. *Arch Microbiol* 185:83–90.
54. Schwalbach MS, Tripp HJ, Steindler L, Smith DP, Giovannoni SJ (2010) The presence of the glycolysis operon in SAR11 genomes is positively correlated with ocean productivity. *Environ Microbiol* 12:490–500.
55. Mulligan C, Kelly DJ, Thomas GH (2007) Tripartite ATP-independent periplasmic transporters: Application of a relational database for genome-wide analysis of transporter gene frequency and organization. *J Mol Microbiol Biotechnol* 12:218–226.
56. Kuhlmann SI, Terwisscha van Scheltinga AC, Bienert R, Kunte HJ, Ziegler C (2008) 1.55 A structure of the ectoine binding protein TeaA of the osmoregulated TRAP-transporter TeaABC from *Halomonas elongata*. *Biochemistry* 47:9475–9485.
57. Ottesen EA, et al. (2011) Metatranscriptomic analysis of autonomously collected and preserved marine bacterioplankton. *ISME J* 5:1881–1895.
58. Schauer K, Rodionov DA, de Reuse H (2008) New substrates for TonB-dependent transport: Do we only see the “tip of the iceberg”? *Trends Biochem Sci* 33:330–338.
59. Chakraborty R, Storey E, van der Helm D (2007) Molecular mechanism of ferricsiderophore passage through the outer membrane receptor proteins of *Escherichia coli*. *Biometals* 20:263–274.
60. Martin JH, Gordon RM, Fitzwater SE (1990) Iron in Antarctic waters. *Nature* 345: 156–158.
61. Dupont CL, et al. (2012) Genomic insights to SAR86, an abundant and uncultivated marine bacterial lineage. *ISME J* 6:1186–1199.
62. Benner R (2002) *Biogeochemistry of Marine Dissolved Organic Matter*, eds Hansell DA, Carlson CA (Academic, London), pp 59–90.
63. Kirchman DL, Suzuki Y, Garside C, Ducklow HW (1991) High turnover rates of dissolved organic-carbon during a spring phytoplankton bloom. *Nature* 352:612–614.
64. May DH, et al. (2016) An alignment-free “metapeptide” strategy for metaproteomic characterization of microbiome samples using shotgun metagenomic sequencing. *J Proteome Res* 15:2697–2705.
65. Timmins-Schiffman E, et al. (2017) Critical decisions in metaproteomics: Achieving high confidence protein annotations in a sea of unknowns. *ISME J* 11:309–314.
66. Miller CS, et al. (2013) Short-read assembly of full-length 16S amplicons reveals bacterial diversity in subsurface sediments. *PLoS One* 8:e56018.
67. León IR, Schwämmle V, Jensen ON, Sprenger RR (2013) Quantitative assessment of in-solution digestion efficiency identifies optimal protocols for unbiased protein analysis. *Mol Cell Proteomics* 12:2992–3005.

Supporting Information

Bergauer et al. 10.1073/pnas.1708779115

SI Materials and Methods

Study Site and Sampling. To study the genomic and proteomic properties of microbial assemblages, water samples were collected from the following: the lower euphotic layer at 100-m depth, the oxygen minimum layer (O_2-min), the Labrador Seawater (LSW), the Antarctic Intermediate Water (AAIW), the North Atlantic Deep Water (NADW), the Iceland–Scotland Overflow Water (ISOW), the South Atlantic Central Water (SACW), and the Antarctic Bottom Water (AABW). Seawater was sequentially filtered through a 0.8-μm ATTP polycarbonate filter (142-mm filter diameter; Millipore) and a 0.2-μm Pall Supor polyethersulfone or polycarbonate filter (142-mm filter diameter; Millipore). Filtration of the samples was accomplished within 2.5–3 h after recovery of the conductivity, temperature, and depth (CTD) rosette sampler. Filters (0.8 and 0.2 μm) were immediately placed into liquid nitrogen and stored at −80 °C until nucleic acid and protein extractions were performed on the 0.2-μm filters. CTD and oxygen concentrations were measured using a Seabird CTD system mounted on the rosette sampler. Seawater was collected with 25-L Niskin bottles at 18 stations from the major water masses determined by their potential temperature and salinity. Apparent oxygen utilization (AOU) was calculated using Ocean Data View 4.6.2 (1). The oxygen minimum layer (O_2-min) represents the depth layer with the lowest dissolved oxygen concentration measured at the respective station. O_2 concentrations in the O_2-min layer ranged between 81.30 and 143.31 $\mu mol \cdot kg^{-1}$ (Table S1).

Metagenomic Setup: Nucleic Acid Extraction. Nucleic acid extraction was performed directly from the membrane filters kept on ice, using a standard phenol extraction protocol. Briefly, microbial biomass was treated with a 1% SDS extraction buffer and phenol–chloroform–isoamyl alcohol [25:24:1 (vol:vol)]. Mechanical disruption of cells was performed using Lysing Matrix E tubes (MP Biomedicals) in combination with the FastPrep instrument (speed 4 for 30 s). Precipitation of nucleic acids was achieved with 5 M NaCl, glycogen, and ethanol. Residual phenol was removed by mixing with an equal volume of chloroform–isoamyl alcohol (24:1), and pelleted nucleic acids were washed in ice-cold 70% ethanol and air-dried before resuspension.

Agarose Gel Analysis of Nucleic Acid Extracts. The quality of high–molecular-weight DNA was confirmed and quantified by gel electrophoresis and using a NanoDrop ND-1000 spectrophotometer. Yields between 0.7 and 14 μg of DNA were obtained in the total extracts from the various samples. Libraries were prepared from a minimum of 700 ng (MOC_3) and otherwise from 1 μg of DNA per sample using Illumina HiSeq 2000 sequencing facilities (Eurofins MWG Operon or www.sequencing.uio.no) (Table S1).

Quality Filtering of Metagenomic Sequences, Assembly, and Gene Annotation. A shotgun metagenomic approach was applied using paired-end pyro- and Illumina sequencing techniques. Paired-end sequencing reads were filtered using Prinseq-lite-0.20.3 (2) and reads shorter than 100 bp and/or with a quality mean <30 were removed from the datasets. Identical reads were removed using CD-HIT. Filtered reads with a minimum of 95% identity and 90% coverage over a viral or human reference genome were removed using Deconseq-0.4.3.

After quality filtering, reads were assembled into contigs by employing the de novo assembler software MetaVelvet-1.2.02 (3). The optimal k-mer sizes, calculated with VelvetOptimizer-2.2.5 (www.vicbioinformatics.com/software.velvetoptimiser.shtml), were 49 for the dataset of GEO_1 and 51 for the remaining metagenomic datasets (Table S3). The MetaVelvet parameters were as follows: ins_length, 220 (GEO); ins_length, 300 (MOCA); -cov_cutoff, 4; and min_contig_lgth, 1,000 (GEO and MOCA).

Functional annotation of metagenomic contigs was conducted using Prokka v1.11 software tool (4) with an e-value cutoff of 1 × 10^{-8}. To predict the location of ribosomal RNA genes on the contigs, Barrnap 0.6 (www.vicbioinformatics.com/software.barrnap.shtml) was implemented in the Prokka software package, and the SILVA-SSU database was used as data source for HMM models.

Further assignments into higher-resolution functional properties were based on protein domains, predicted and characterized with the database Pfam-A 28.0 (5), where each of the protein families is represented by multiple sequence alignments and hidden Markov models (HMMs). The program hmmscan in HMMER, version 3.1b2 (6), identified the protein domains using the amino acid sequences given by Prokka as input, and the collections of profile HMM in Pfam-A as reference database. The comparison was performed with an expected e-value cutoff of 1 × 10^{-3} in hmmscan.

Clusters of orthologous groups (COGs) were identified using the amino acid sequences annotated by Prokka as a query for the command BLASTp (7), version ncbi-blast-2.2.29+, and the database of orthologous groups and functional annotation, Eggnog 4.0 (8), as reference database. The BLASTp e-value cutoff was 1 × 10^{-5}.

The Kyoto Encyclopedia of Genes and Genomes (KEGG) Orthology (KO) database was used for functional assignments of the amino acid sequences given by Prokka, using the KEGG Automatic Annotation Server (KAAS) (9). The method used to identify orthologs was single-directional best hit with a representative set of 40 selected prokaryotes, the maximum allowed by KAAS, comprising 13 Archaea and 27 Bacteria.

Phylogenetic Analysis. We performed the phylogenetic classification of DNA contigs employing two independent methods. On the one hand, PhymmBL v4.0 (10), a hybrid classifier combining Phymm and BLAST results, was used together with the database RefSeq release 67. Here, only contigs with a confidence score threshold > 0.8 were binned. On the other hand, contigs were used to query the nt NCBI nonredundant nucleotide database (downloaded in December 2015) with the BLASTn tool (7), version ncbi-blast-2.2.29+, and an e-value cutoff of 1 × 10^{-8}. The 10 most significant hits of the BLASTn search were compared with PhymmBL results, and genes with a congruent taxonomic assignment were taken into account for subsequent analysis.

Reconstruction and Taxonomic Affiliation of Ribosomal Genes. To taxonomically characterize the microbial community and estimate taxon abundance, we reconstructed full-length 16S rRNA gene sequences from filtered reads using the iterative method EMIRGE (11). The insert size averages ± SDs were 220 ± 30 (Geotraces) and 300 ± 40 (Moca), respectively. The joined threshold used was 0.99. Chimeric sequences classified by DECIPHER, version 2.0 (12), were removed from the dataset. The remaining sequences were identified at a minimum of 80% identity using BLASTn against the Silva SSU 128 database (www.arb-silva.de/). Sequences that did not match at the minimum of 80% identity were discarded and were assumed to be either of poor quality or derived from unclassified microorganisms.

Metaproteomic Setup: Protein Extraction and Protein Identification. Protein extraction and spin filter-aided in-solution digestion (SF-ISD) was performed according to León et al. (13) with modifications to optimize protein extraction and recovery. Briefly, whole-cell protein extraction was performed after resuspension of cells in a disruption buffer by three freeze–thaw cycles, boiling at 85 °C for 10 min and three cycles of sonication at the following settings: 30-s intervals, 0.5-s pulse on/0.5-s pulse off, and cooling of the sample on ice. The disruption buffer contained 20 mM TEAB, 100 mM 1,4-dithio-D-threitol (DTT), 2% SDS, 100 mM EDTA, and a 1/4 tablet of Complete protease inhibitors (Roche Applied Science) to 1 mL of buffer. Samples were centrifuged at 14,000 × *g* to remove cell debris; the supernatant was collected, and protein concentrations were estimated using a Thermo Scientific BCA Protein Assay. The protein mixture was subjected to tryptic digestion in 50 µL of 0.5 µg/µL trypsin solution and subsequent incubation in a wet chamber at 37 °C overnight. After digestion, extracts were collected, and sodium deoxycholate removal was achieved by phase separation with ethylacetate after acidification with trifluoroacetic acid (TFA) as described elsewhere (14).

The LTQ Orbitrap Velos instrument (Thermo Scientific) operated in data-dependent acquisition mode to automatically select the 20 most intense precursor ions for fragmentation by collisionally induced dissociation. Survey MS scans were acquired from *m*/*z* 300 to *m*/*z* 1,650 with a resolution of 60,000 (at *m*/*z* 400). MS/MS spectra were acquired in the linear ion trap (target value of 10,000 ions) with a maximum fill time of 50 ms.

To identify peptide sequences, acquired MS/MS spectra were analyzed combining two database search engines, SEQUEST-HT (15) and Mascot (16) as well as the validation tool Percolator in Proteome Discoverer 2.1 (Thermo Fisher Scientific). MS/MS spectra were searched against a tailored database combining the (*i*) in-house constructed gene catalog "GeMo" comprising 226,000 protein-coding sequences from eight metagenomes; (*ii*) 759,480 protein-coding sequences from 455 single-amplified genomes (Dataset S6); and (*iii*) 1.070,281 protein-coding sequences from deep-ocean metagenomes from the Malaspina circumnavigation. Single-amplified genomes were selected from euphotic to bathypelagic zones as well as from axenic cultures to represent as much as possible. Sequences shorter than 15 aa were removed, and the emerging dataset "Combo_v02" contained a total number of 1.801,424 protein sequences. The database "Combo_v02" was appended with the common Repository of Adventitious Proteins (cRAP) contaminant database (The Global Proteome Machine, www.thegpm.org/cRAP/index.html). MS/MS spectra searches were performed as follows: precursor ion tolerance of 5.0 ppm; fragment ion tolerance of 0.5 Da; carbamidomethyl cysteine was specified as fixed modification, whereas oxidation (M), deamidation (N/Q), and N-terminal protein acetylation were set as variable modifications. Trypsin was specified as the proteolytic enzyme, allowing for two missed cleavages and a maximum of three variable PTMs (maximum equal modifications) per peptide. In Proteome Discoverer, the Percolator-based scoring was chosen to improve the discrimination between correct and incorrect spectrum identifications, learning from the results of a decoy and target database; settings were as follows: maximum delta Cn, 0.05; strict false-discovery rate of 0.01 and validation based on *q* values.

Protein matches were only accepted if they were identified by a minimum of two peptides/protein (unique and razor peptides) and with high confidence. We enabled the formation of "Protein Groups" and used the highest scoring protein in the group as representative protein (Master Protein).

Protein Quantification. Besides the qualitative assessment of protein frequencies (number of times a unique protein was identified), we conducted label-free quantitative mass spectrometry using chromatographic peak areas [normalized area abundance factor (NAAF)] (17) to estimate protein abundances from shotgun proteomics data. In our semiquantitative analyses of TMPs, unique and shared peptides between multiple proteins generated during the MS experiment are not distinguished:

$$\mathrm{NAAF}_k = \frac{(\mathrm{PA}/\mathrm{length})k}{\sum_{i=1}^{N}(\mathrm{PA}/\mathrm{length})i},$$

where the label-free abundance factor "peak area" (PA) is divided by the peptide length. The subscript *k* denotes a protein/protein group identity, and *N* is the total number of proteins *i* detected in an experiment. Comparative quantitative analyses of the 14 metaproteomes were accomplished by normalizing the PA to the set of transport-related proteins listed in Dataset S1.

Consensus Functional and Taxonomic Annotation of Transport-Related Proteins. The taxonomic and functional profiling of metaproteomic data provides extensive information about the composition and metabolic activities of microbial communities, and yet poses an enormous challenge (18). Here, the taxonomic interpretation of expressed transporter proteins was conservatively inferred from site-specific or selected marine (meta)genomic databases. Prediction and annotation of genomic ORFs were retrieved from the different functional annotation platforms or tools such as Prokka (4), IMG (19), and RAPSearch (20). KEGG pathway mapping and COG clustering were used to assist pathway curation. A "consensus" output was achieved by combining all tools. If annotations were controversial, the functional annotation was downgraded to the first concordant entry or categorized as "uncharacterized."

Prokaryotic Abundance. Samples for prokaryotic abundances were collected at 51 stations and 24 depth layers during the research cruises Geotraces-1 to -3 (21), and 47 stations and 7 depth layers during research cruises Medea-1 and -2. Prokaryotic abundances were determined using standard procedures with modifications. Briefly, 2-mL samples were fixed with glutaraldehyde (0.5% final concentration), shock-frozen in liquid N_2, and kept at −80 °C until analysis. Samples were thawed to room temperature, and 0.5-mL subsamples were stained with SYBR Green I (Molecular Probes; Invitrogen) in the dark for 10 min. The prokaryotes were enumerated on a FACSAria II flow cytometer (Becton Dickinson) by their signature in a plot of green fluorescence versus side scatter.

Leucine Incorporation. Samples to measure leucine incorporation into heterotrophic prokaryotes were collected at the above-mentioned stations and up to seven depth layers. Microbial heterotrophic production was measured as described elsewhere (22), by incubating 50 mL of seawater in triplicate with 10 nM [^{3}H]leucine (final concentration, specific activity, 5.809 × 10^6 MBq·mmol^{-1}; Amersham) in the dark at in situ temperature (±1 °C) for 4–8 h. Triplicate formaldehyde-killed blanks were treated in the same way as the samples. Incubations were terminated by adding formaldehyde (2% final concentration) to the samples. Samples and blanks were filtered through 0.2-µm polycarbonate filters (25-mm filter diameter; Millipore) supported by cellulose acetate filters (HAWP; 0.45-µm pore size; Millipore). Subsequently, the filters were rinsed with 5% ice-cold trichloroacetic acid and dried; 8 mL of scintillation mixture (FilterCount; Canberra-Packard) was added and, after about 18 h, counted on board in a liquid scintillation counter (LKB Wallac). The instrument was calibrated with internal and external standards.

Statistics. Hierarchical cluster analysis was used to examine multivariate similarities between the samples. The similarity matrix was based on the Bray–Curtis similarity index for which normalized area abundance data (NAAF) were used. Clusters for the stations were calculated by the group mean linkage method using the unweighted pair group method with arithmetic mean algorithm (UPGMA). Data were visualized in R Studio.

SI Results and Discussion

This material presents additional interesting data on transporter functions and the 16S rRNA gene-based community structure of the MOCA metagenomes that could not be discussed in the text because of space considerations.

Moreover, specific features of transporter proteins not discussed in the text are included here.

The Protein Repertoire of Open-Ocean Microbes. We captured 1.1 million MS/MS spectra and employed multiple search algorithms (Mascot and Sequest-HT) against tailored peptide databases: (*i*) metagenomes collected concurrently during the same cruise track (GeMo; Fig. S2); (*ii*) the Malaspinomics global ocean metagenome collection (MP) covering the subtropical Atlantic, Indian, and Pacific Ocean; (*iii*) single-amplified genomes (SAGs) from study sites and other marine locations (SAGs; Fig. S1). Search results were combined and resulted in 12,047 high-confidence proteins. Enabling protein grouping and the strict parsimony principle, we inferred a total 3,365 nonredundant proteins from 14 metaproteomes. The contributions of the different databases to protein inferences are summarized in Fig. S6. Briefly, the highest number of confidently scored peptide matches and protein inferences resulted from the SAG database (43%; 1,439 protein IDs), followed by the metagenomic databases MP (30%) and GeMo (27%). Database size (23) and the inclusion of 21 cyanobacterial SAGs (Dataset S6) had substantial impact on search sensitivity, and yielded higher spectrum matches, particularly in metaproteomes from the euphotic layer (49–76% of protein IDs). Furthermore, assembled metagenomes tend to lack sequencing reads that could not be reliably assembled into contigs and thus do not reflect the entire sequence diversity (18, 23, 24). In the aphotic zones, however, composite metagenomic databases augmented the number of protein identifications, emphasizing the importance of suitable reference databases (18).

Characterization of Microbial Communities Using 16S rRNA Genes Reconstructed from Short-Read Metagenomic Datasets. The reconstruction of SSU rRNA genes using EMIRGE analysis of shotgun metagenomic reads (11) and removal of chimeric sequences resulted in 404 full-length sequence clusters at 97% similarity (Table S4). As expected, estimated sequence abundances revealed an overall dominance of Bacteria (between 58–99% per metagenomic sample) with a varying fraction of Archaea (1–42%; Fig. S5).

Depth-related distribution patterns based on variations in estimated 16S rRNA gene abundances were observed for marine Alphaproteobacteria, decreasing from an average of 52% in the euphotic to 38% in the mesopelagic and 27% in the bathypelagic zones. Members of the SAR11 clade dominated the pool of alphaproteobacterial 16S rRNA gene sequences (83–86%) and affiliated with the SAR11 clades Deep 1 and Surface 1–4. The relative abundances of sequences affiliating with SAR11 clades Surface 1–4 were higher in surface waters (average 64% of Alphaproteobacteria) compared with mesopelagic and bathypelagic layers (37% and 40%, respectively). Conversely, SAR11 clade Deep 1 was more abundant in mesopelagic to bathypelagic layers. While Marinimicrobia accounted for only 2.5–10% of 16S sequences in the euphotic and mesopelagic, their relative abundance increased up to 18% in the bathypelagic realm (Fig. S5). Similarly, sequences of the Chloroflexi-related SAR202 cluster were particularly abundant at 2,745- and 5,001-m depth, where they comprised 13.5% and 7.5% of the deep-ocean microbial community composition. The relative abundance of Thaumarchaeota also demonstrated a clear depth-related trend, accounting for 9–12% of the microbial community in the euphotic and 21–38% in mesopelagic to abyssopelagic waters. In contrast, the relative abundance of Gammaproteobacteria and Deltaproteobacteria was similar in in surface and deep-water microbial communities. Other phyla, including Bacteroidetes, Actinobacteria, and Cyanobacteria, prevailed in the upper water column, albeit at low relative abundances.

The microbial community composition (Fig. S5), as revealed by shotgun metagenomics, was in agreement with previous reports, showing the partitioning of microbial populations with depth in the marine environment (25, 26).

As expected, a higher percentage of *Prochlorococcus*-like sequences was evident in the euphotic zone (3%, 8%, and 18% of total; Table S4), whereas Deltaproteobacteria-like sequences typically increased toward the mesopelagic and bathypelagic realm, as reported elsewhere (27). Also, the high relative abundance of SAR11 sequences in the euphotic and mesopelagic zone is consistent with previous reports from the North Atlantic (28, 29), the Sargasso Sea (30), and the coastal oligotrophic Mediterranean Sea (31).

We identified sequences of the recently named Candidate phylum "Marinimicrobia" (32) (formerly Marine Group A, SAR406), revealing a subtle increase in relative abundance at and below the mesopelagic zone (Fig. S5). Marinimicrobia have been ubiquitously encountered across deep-ocean basins and in oxygen minimum zones (29, 33), and are considered to contribute substantially to the deep-ocean microbial community.

In agreement with a previous study showing that members of the SAR202 cluster are highly abundant below the deep chlorophyll maximum in the Atlantic and Pacific Ocean and persist down to ~4,000-m depth (34), we detected highest abundances of the SAR202 clade in the lower bathypelagic zone (Table S4).

Abundance estimates of Thaumarchaeota 16S rRNA gene sequences clearly surpassed Euryarchaeota, as described previously for bathypelagic waters (33, 35), and were among the most abundant sequences in bathypelagic zones (Fig. S5). Despite their rather low relative abundances, particle-associated members of the Euryarchaeota MGII might play a role in the degradation of POM, and potentially outnumber their free-living counterparts under oligotrophic conditions.

Taken together, despite the low recovery of 16S rRNA gene sequences from shotgun metagenomic data, we were able to determine ecotypes of the SAR11 clade, supposedly indicating a separation by selection for specific traits. We note, however, that our metagenomic analyses were performed on different samples than metaproteomics, making it difficult to directly link the vertical stratification observed at the community level (16S rRNA) with substrate uptake patterns (metaproteomics).

1. Schlitzer R (2015) Ocean Data View, version 4.6.2. Available at https://odv.awi.de/software. Accessed June 27, 2014.
2. Schmieder R, Edwards R (2011) Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 27:863–864.
3. Namiki T, Hachiya T, Tanaka H, Sakakibara Y (2012) MetaVelvet: An extension of Velvet assembler to de novo metagenome assembly from short sequence reads. *Nucleic Acids Res* 40:e155.
4. Seemann T (2014) Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069.
5. Finn RD, et al. (2016) The Pfam protein families database: Towards a more sustainable future. *Nucleic Acids Res* 44:D279–D285.
6. Eddy SR (1998) Profile hidden Markov models. *Bioinformatics* 14:755–763.
7. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410.
8. Powell S, et al. (2014) eggNOG v4.0: Nested orthology inference across 3686 organisms. *Nucleic Acids Res* 42:D231–D239.
9. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M (2007) KAAS: An automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res* 35:W182–W185.
10. Brady A, Salzberg SL (2009) Phymm and PhymmBL: Metagenomic phylogenetic classification with interpolated Markov models. *Nat Methods* 6:673–676.
11. Miller CS, et al. (2013) Short-read assembly of full-length 16S amplicons reveals bacterial diversity in subsurface sediments. *PLoS One* 8:e56018.
12. Wright ES, Yilmaz LS, Noguera DR (2012) DECIPHER, a search-based approach to chimera identification for 16S rRNA sequences. *Appl Environ Microbiol* 78:717–725.

13. León IR, Schwämmle V, Jensen ON, Sprenger RR (2013) Quantitative assessment of in-solution digestion efficiency identifies optimal protocols for unbiased protein analysis. *Mol Cell Proteomics* 12:2992–3005.
14. Masuda T, Tomita M, Ishihama Y (2008) Phase transfer surfactant-aided trypsin digestion for membrane proteome analysis. *J Proteome Res* 7:731–740.
15. Eng JK, McCormack AL, Yates JR (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom* 5:976–989.
16. Perkins DN, Pappin DJC, Creasy DM, Cottrell JS (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 20:3551–3567.
17. Zhang Y, Wen Z, Washburn MP, Florens L (2015) Improving label-free quantitative proteomics strategies by distributing shared peptides and stabilizing variance. *Anal Chem* 87:4749–4756.
18. Timmins-Schiffman E, et al. (2017) Critical decisions in metaproteomics: Achieving high confidence protein annotations in a sea of unknowns. *ISME J* 11:309–314.
19. Markowitz VM, et al. (2012) IMG: The integrated microbial genomes database and comparative analysis system. *Nucleic Acids Res* 40:D115–D122.
20. Zhao Y, Tang H, Ye Y (2012) RAPSearch2: A fast and memory-efficient protein similarity search tool for next-generation sequencing data. *Bioinformatics* 28:125–126.
21. De Corte D, Sintes E, Yokokawa T, Reinthaler T, Herndl GJ (2012) Links between viruses and prokaryotes throughout the water column along a North Atlantic latitudinal transect. *ISME J* 6:1566–1577.
22. Reinthaler T, van Aken HM, Herndl GJ (2010) Major contribution of autotrophy to microbial carbon cycling in the deep North Atlantic's interior. *Deep Sea Res Part II Top Stud Oceanogr* 57:1572–1580.
23. May DH, et al. (2016) An alignment-free "metapeptide" strategy for metaproteomic characterization of icrobiome samples using shotgun metagenomic sequencing. *J Proteome Res* 15:2697–2705.
24. Cantarel BL, et al. (2011) Strategies for metagenomic-guided whole-community proteomics of complex microbial environments. *PLoS One* 6:e27173.
25. Gordon DA, Giovannoni SJ (1996) Detection of stratified microbial populations related to *Chlorobium* and *Fibrobacter* species in the Atlantic and Pacific oceans. *Appl Environ Microbiol* 62:1171–1177.
26. DeLong EF, et al. (2006) Community genomics among stratified microbial assemblages in the ocean's interior. *Science* 311:496–503.
27. Pham VD, Konstantinidis KT, Palden T, DeLong EF (2008) Phylogenetic analyses of ribosomal DNA-containing bacterioplankton genome fragments from a 4000 m vertical profile in the North Pacific Subtropical Gyre. *Environ Microbiol* 10:2313–2330.
28. Agogué H, Lamy D, Neal PR, Sogin ML, Herndl GJ (2011) Water mass-specificity of bacterial communities in the North Atlantic revealed by massively parallel sequencing. *Mol Ecol* 20:258–274.
29. Schattenhofer M, et al. (2009) Latitudinal distribution of prokaryotic picoplankton populations in the Atlantic Ocean. *Environ Microbiol* 11:2078–2093.
30. Morris RM, et al. (2002) SAR11 clade dominates ocean surface bacterioplankton communities. *Nature* 420:806–810.
31. Alonso-Sáez L, et al. (2007) Seasonality in bacterial diversity in north-west Mediterranean coastal waters: Assessment through clone libraries, fingerprinting and FISH. *FEMS Microbiol Ecol* 60:98–112.
32. Rinke C, et al. (2013) Insights into the phylogeny and coding potential of microbial dark matter. *Nature* 499:431–437.
33. Salazar G, et al. (2016) Global diversity and biogeography of deep-sea pelagic prokaryotes. *ISME J* 10:596–608.
34. Morris RM, Rappé MS, Urbach E, Connon SA, Giovannoni SJ (2004) Prevalence of the Chloroflexi-related SAR202 bacterioplankton cluster throughout the mesopelagic zone and deep ocean. *Appl Environ Microbiol* 70:2836–2842.
35. Herndl GJ, et al. (2005) Contribution of Archaea to total prokaryotic production in the deep Atlantic Ocean. *Appl Environ Microbiol* 71:2303–2309.

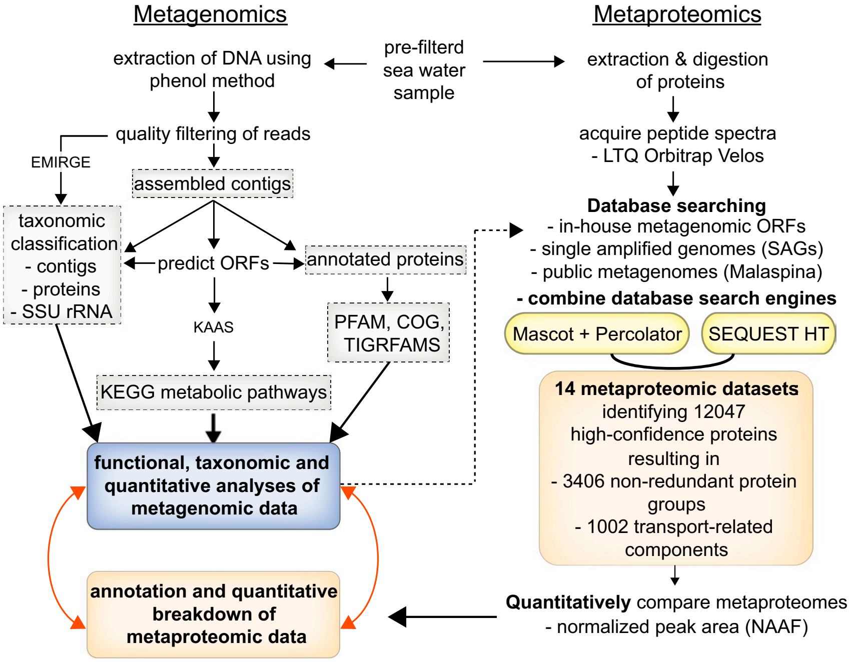
Metagenomics
Metaproteomics
pre-filterd sea water sample
extraction of DNA using phenol method
quality filtering of reads
EMIRGE
assembled contigs
taxonomic classification
- contigs
- proteins
- SSU rRNA
predict ORFs
annotated proteins
KAAS
PFAM, COG, TIGRFAMS
KEGG metabolic pathways
functional, taxonomic and quantitative analyses of metagenomic data
annotation and quantitative breakdown of metaproteomic data
extraction & digestion of proteins
acquire peptide spectra
- LTQ Orbitrap Velos
Database searching
- in-house metagenomic ORFs
- single amplified genomes (SAGs)
- public metagenomes (Malaspina)
- combine database search engines
Mascot + Percolator
SEQUEST HT
14 metaproteomic datasets
identifying 12047 high-confidence proteins
resulting in
- 3406 non-redundant protein groups
- 1002 transport-related components
Quantitatively compare metaproteomes
- normalized peak area (NAAF)

Fig. S1. Flowchart of the major steps in the metagenomic and metaproteomic analyses.

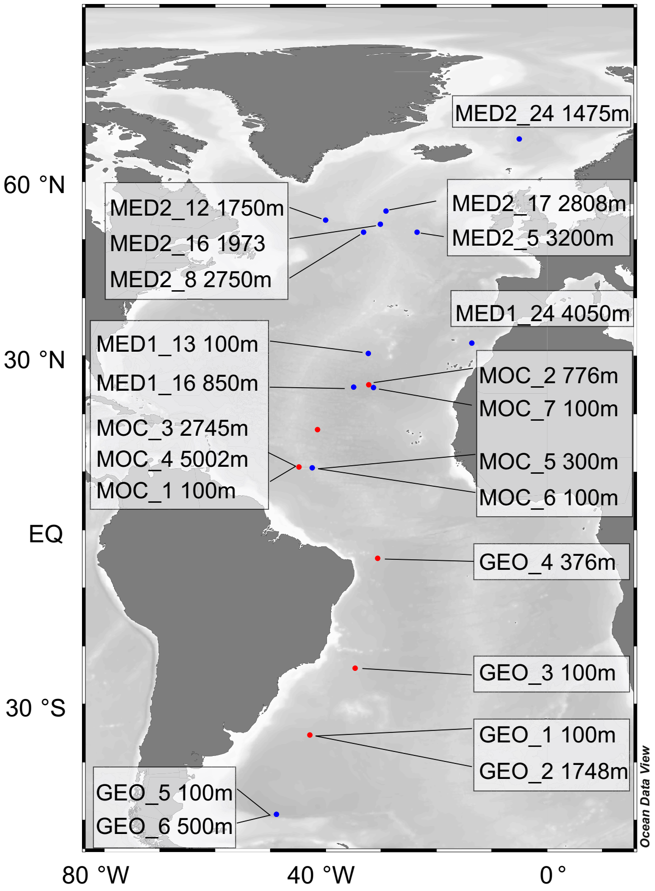

Fig. S2. Map showing the locations of metaproteomic and metagenomic sampling stations, occupied during the cruises: Geotraces (GEO), MOCA (MOC), MEDEA-I (MED1), and MEDEA-II (MED2). Color-coded symbols indicate metaproteomic (blue) or metagenomic (red) analyses.

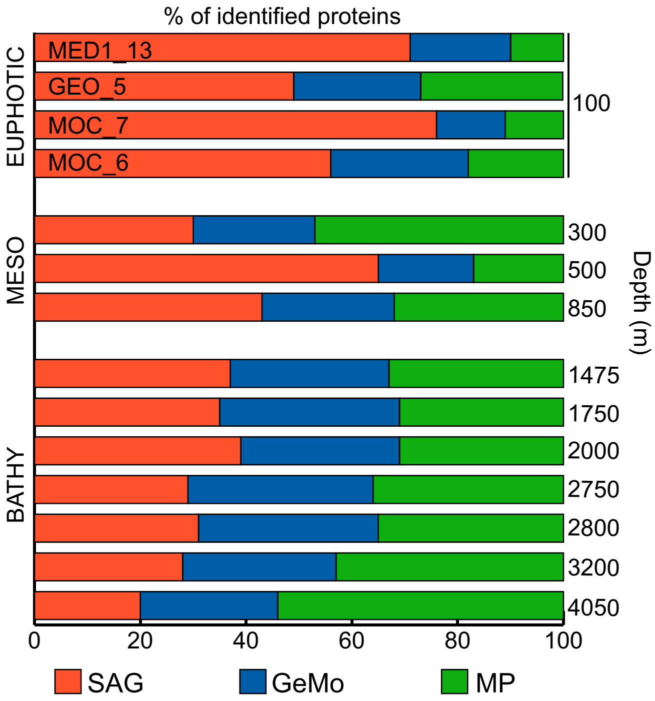

Fig. S3. Relative contribution of genomic databases used for protein identifications. Red, single-amplified genomes (SAG); blue, in-house established metagenomic libraries from Geotraces and Moca cruises (GeMo); green, Malaspina gene catalog (MP).

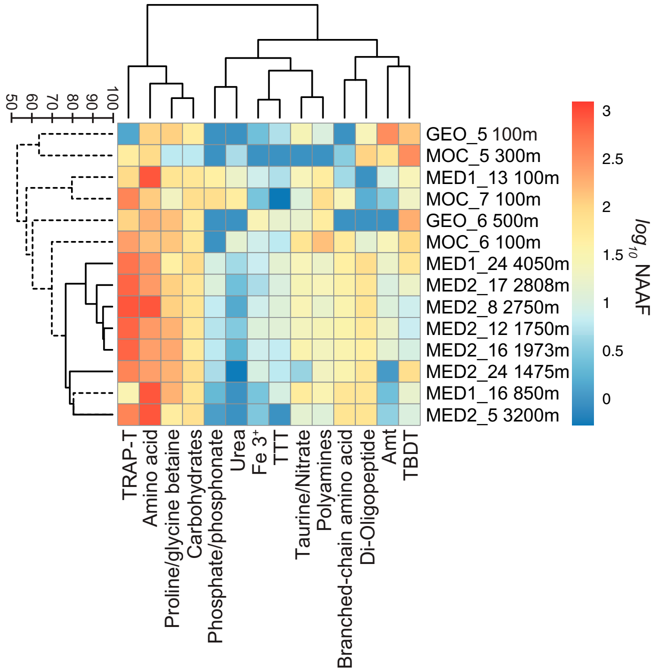

Fig. S4. Transporter proteins were combined by substrate affinity and semiquantitatively analyzed. The similarity matrix was based on the Bray–Curtis similarity index; relative abundance data (NAAF) were used. Clusters were built by the group mean linkage method using unweighted pair group method with arithmetic mean algorithm (UPGMA). Color intensity is proportional to $\log_{10}$-transformed NAAF values.

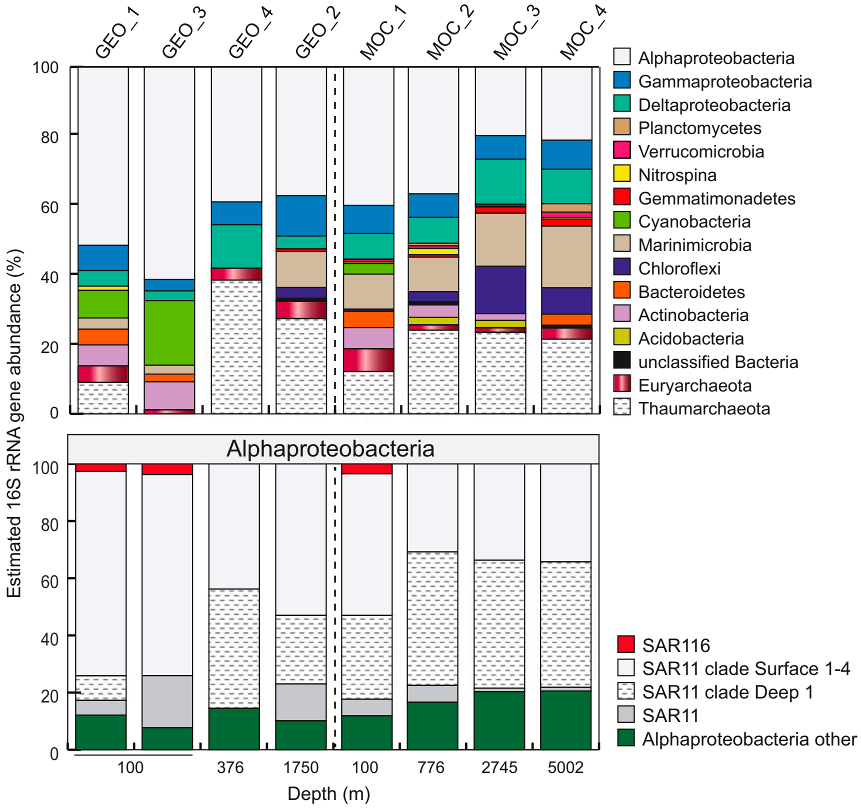

Fig. S5. 16S rRNA gene sequence-based microbial community composition derived from metagenomes, and relative abundance estimates reconstructed with the software tool EMIRGE. An individual EMIRGE sequence represents a consensus sequence in the sense that sequences are merged together at ≥97% sequence identity at each iteration (72). Metagenomes were obtained during Geotraces (GEO) and Moca (MOC) cruises and are sorted in order of increasing depth. Estimated relative abundances (*Supporting Information*) of taxonomic groups were summed at the phylum or class level (Table S4).

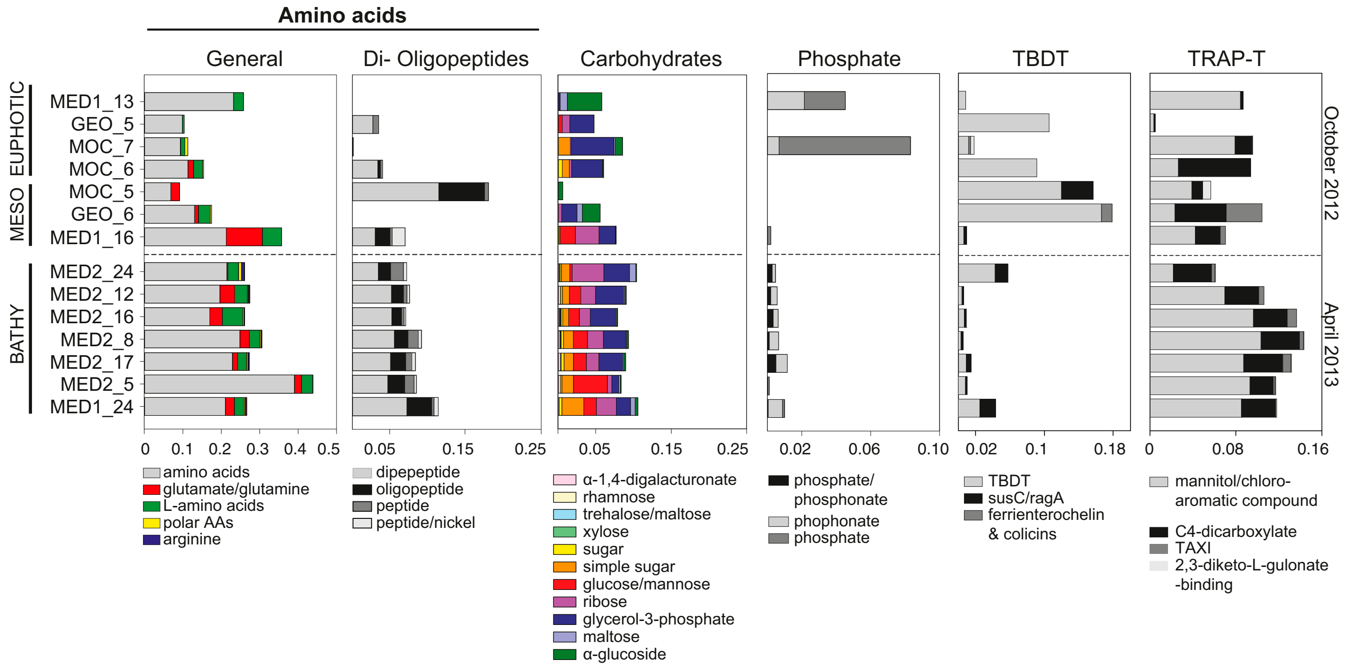

Fig. S6. High-resolution expression profiles of selected transporter systems and substrate affinities, based on NAAF values. Predicted substrate specificities of the SBPs are indicated.

Table S1. Summary of the physical and chemical characteristics of the samples used for metaproteomics and metagenomics in this study

Sample ID	Omic	Cruise	Depth, m	Date	Water mass	Longitude	Latitude	Temp., °C	Salinity	Oxygen, μmol/kg	PO_4, μmol/L	NO_3, μmol/L	Si, μmol/kg	AOU
GEO_1	G	Geotraces III	100	Mar-11	EUPHOTIC	41.13 W	37.83 S	15.44	35.67	230.00	0.34	3.15	1.59	36.92
GEO_2	G	Geotraces III	1,748	Mar-11	AAIW	41.13 W	37.83 S	2.86	34.64	184.90	2.17	31.94	61.83	177.37
GEO_3	G	Geotraces III	100	Mar-11	EUPHOTIC	34.26 W	26.09 S	21.33	36.64	212.62	0.08	0.00	0.77	98.12
GEO_4	G	Geotraces III	376	Mar-11	SACW/O_2-min	28.46 W	5.67 S	8.45	34.75	81.30	2.18	34.62	16.64	245.87
GEO_5	P	Geotraces III	100	Mar-11	EUPHOTIC	48.88 W	48.96 S	4.93	33.99	298.60	1.63	16.82	4.64	106.60
GEO_6	P	Geotraces III	500	Mar-11	ISOW	48.88 W	48.96 S	3.03	34.91	234.60	2.11	17.19	31.69	124.55
MED1_13	P	MEDEA I	100	Oct-11	EUPHOTIC	32.21 W	30.46 N	19.16	36.67	226.00	0.01	0.01	0.73	3.74
MED1_16	P	MEDEA I	850	Oct-11	O_2-min	34.94 W	24.67 N	9.23	35.33	143.31	1.56	24.75	13.50	142.47
MED1_24	P	MEDEA I	4,050	Oct-11	AABW	13.6 W	32.26 N	2.41	34.90	241.54	1.53	22.80	0.00	107.17
MED2_12	P	MEDEA II	1,750	Jul-12	LSW	40.07 W	53.38 N	3.36	34.92	274.36	1.12	17.18	11.67	48.15
MED2_16	P	MEDEA II	1,973	Jul-12	NADW/LSW	30.14 W	52.74 N	3.30	34.94	275.90	1.11	16.83	12.04	47.18
MED2_17	P	MEDEA II	2,808	Jul-12	ISOW	29.14 W	54.98 N	2.86	34.97	269.68	1.14	17.19	18.63	57.65
MED2_24	P	MEDEA II	1,475	Jul-12	NSDW	4.94 W	67.35 N	−0.73	34.91	284.86	1.02	15.26	8.79	66.82
MED2_5	P	MEDEA II	3,200	Jun-12	NEADW	23.46 W	51.27 N	2.86	34.95	263.10	1.18	17.73	19.82	64.47
MED2_8	P	MEDEA II	2,750	Jul-12	NEADW	33.13 W	51.34 N	2.98	34.93	274.45	1.10	16.82	14.06	50.43
MOC_1	G	MOCA	100	Oct-10	EUPHOTIC	44.67 W	10.91 N	18.36	36.18	112.84	0.84	14.28	4.68	204.25
MOC_2	G	MOCA	776	Oct-10	O_2-min	34.55 W	24.51 N	9.21	35.29	130.48	1.57	24.99	14.12	252.60
MOC_3	G	MOCA	2,745	Oct-10	NADW	40.96 W	20.79 N	3.00	34.95	228.66	1.47	21.86	32.69	298.22
MOC_4	G	MOCA	5,002	Oct-10	AABW	44.67 W	10.91 N	1.82	34.83	228.73	1.74	25.46	67.45	310.44
MOC_5	P	MOCA	300	Oct-10	O_2-min	42.4 W	10.76 N	9.40	34.81	89.90	2.30	35.98	14.33	246.79
MOC_6	P	MOCA	100	Oct-10	EUPHOTIC	42.4 W	10.76 N	17.21	36.11	108.88	1.00	16.57	5.62	209.28
MOC_7	P	MOCA	100	Oct-10	EUPHOTIC	31.36 W	24.54 N	22.80	37.31	203.51	0.01	0.00	0.58	185.24

AABW, Antarctic Bottom Water; AAIW, Antarctic Intermediate Water; AOU, apparent oxygen utilization; EUPHOTIC, lower euphotic layer; G, Metagenomics; ISOW, Iceland–Scotland Overflow Water; LSW, Labrador Sea Water; NADW, North Atlantic Deep Water; NEADW, Northeast Atlantic Deep Water; P, Metaproteomics; SACW, South Atlantic Central Water.

Table S2. Prokaryotic abundance and cell-specific leucine incorporation rate calculated for selected water masses occupied during Geotraces and Medea cruises

		EUPH		NACW	AAIW	O_2-min	LSW	MSOW	UCDW	NEADW	NADW	DSOW	LDW	AABW
Cruise	Water mass and depth, m	85–102	SACW 9–546	9–860	50–1701	248–981	701–2201	898–1201	999–2542	1248–3503	1248–5006	2148–4547	2952–4207	3002–5804
GEO_1	PA, ×10^5 mL^{-1}	1.33 ± 1.0					0.34 ± 0.22			0.34 ± 0.29	1.44 ± 1.05	0.21 ± 0.13		0.09 ± 0.01
	Cell-specific Leu, ×10^{-5} fmol·$cell^{-1}$·d^{-1}	55.8 ± 65.14					2.30 ± 1.67			3.48 ± 2.88	1.80 ± 2.23	2.33 ± 2.49		2.21 ± 0.01
GEO_2	PA, ×10^5 mL^{-1}		1.25 ± 1.00	0.98 ± 0.75	0.21 ± 0.07				0.19 ± 0.01		0.09 ± 0.02			0.09 ± 0.03
	Cell-specific Leu, ×10^{-5} fmol·$cell^{-1}$·d^{-1}		30.05 ± 38.6	57.67 ± 84.63	2.45 ± 0.98				1.49 ± 0.01		2.19 ± 1.66			1.62 ± 0.35
GEO_3	PA, ×10^5 mL^{-1}		2.18 ± 1.54		2.19 ± 2.12				0.42 ± 0.15		0.18 ± 0.07			0.21 ± 0.07
	Cell-specific Leu, ×10^{-5} fmol·$cell^{-1}$·d^{-1}		33.13 ± 36.36		17.13 ± 25.67				3.65 ± 5.97		2.34 ± 1.91			1.52 ± 0.97
MED_I	PA, ×10^5 mL^{-1}	2.9 ± 0.85			0.33 ± 0.26	0.53 ± 0.58	0.28 ± 0.13	0.54 ± 0.41		0.19 ± 0.16			0.26 ± 0.13	0.11 ± 0.05
	Cell-specific Leu, ×10^{-5} fmol·$cell^{-1}$·d^{-1}	27.94 ± 20.29			1.22 ± 0.72	2.18 ± 2.06	0.79 ± 0.23	1.56 ± 0.96		7.94 ± 0.44			0.88 ± 0.37	0.82 ± 0.43
MDE_22	PA, ×10^5 mL^{-1}	3.61 ± 1.05		1.72 ± 0.25	0.68 ± 0.07	1.42 ± 0.4	0.51 ± 0.1			0.43 ± 0.12			0.39 ± 0.05	0.22 ± 0.01
	Cell-specific Leu, ×10^{-5} fmol·$cell^{-1}$·d^{-1}	16.37 ± 13.16		4.63 ± 1.07	1.97 ± 0.72	4.4 ± 1.5	1.89 ± 0.59			2.10 ± 1.06			2.13 ± 0.39	1.02 ± 0.24

AABW, Antarctic Bottom Water; AAIW, Antarctic Intermediate Water; DSOW, Denmark Strain Overflow Water; EUPH, lower euphotic layer; LDW, Lower Deep Water; LSW, Labrador Seawater; MSOW, Mediterranean Sea Outflow Water; NACW, North Atlantic Central Water; NADW, North Atlantic Deep Water; NEADW, Northeast Atlantic Deep Water; O_2-min, oxygen minimum zone; PA, prokaryotic abundance; SACW, South Atlantic Central Water; UCDW, Upper Circumpolar Deep Water. Average and SD are indicated for all of the parameters.

Table S3. Illumina HiSeq metagenome library statistics

Metagenome ID	GEO_1	GEO_2	GEO_3	GEO_4	MOC_1	MOC_2	MOC_3	MOC_4
Water depth, m	100	1,748	100	376	100	776	2,745	5,002
Filter size, μm	0.22	0.22	0.22	0.22	0.22	0.22	0.22	0.22
k-mer size	49	51	51	51	51	51	51	51
No. of reads	113,198,592	106,867,548	98,909,012	82,284,654	425,280,050	423,903,158	401,600,508	415,710,740
Size of reads, bp	100 × 2	100 × 2	100 × 2	100 × 2	100 × 2	100 × 2	100 × 2	100 × 2
Insert size, bp	100	100	100	100	300	300	300	300
No. of filtered reads	66,400,342	62,632,906	64,068,762	46,532,000	257,089,524	224,068,850	136,302,196	190,114,406
No. of contigs	3,597	2,571	3,592	1,376	38,692	18,744	24,714	37,605
No. of protein coding genes	14,521	7,226	5,823	1,809	64,577	29,170	42,179	65,228
No. of genes with COG ID	11,621	6,520	3,473	1,407	44,788	21,341	32,406	49,960
No. of genes with KEGG ID	6,624	2,843	2,061	853	24,866	11,871	19,059	29,139
No. of genes with Pfam ID	20,862	11,760	6,039	2,380	79,678	39,085	62,228	93,995

GEO, Geotraces; MOC, MOCA.

Table S4. Abundance and relative estimated abundance (percentage) of SSU rRNA genes in metagenomic samples reconstructed with EMIRGE

		GEO_1		GEO _2		GEO _3		GEO _4		MOC_1		MOC_2		MOC_3		MOC_4	
		100 m		1,750 m		100 m		376 m		100 m		776 m		2,745 m		5,001 m	
Phylum	Order/Clade	No. SSU	EA %	No. SSU	EA %	No. SSU	EA %	No. SSU	EA %	No. SSU	EA %	No. SSU	EA %	No. SSU	EA %	No. SSU	EA %
Acidobacteria	Acidobacteria	—	—	—	—	—	—	—	—	—	—	14	2.1	8	3.4	—	—
Actinobacteria	Acidimicrobiia	3	6.0	1	3.6	4	8.0	—	—	11	6.1	5	4.2	1	0.6	5	1.3
Bacteroidetes	Bacteroidetes	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	0.1
	Cytophagia	1	0.6	—	—	—	—	—	—	5	1.6	—	—	—	—	1	0.4
	Flavobacteria	3	3.9	—	—	2	2.2	—	—	9	3.2	2	0.2	—	—	5	1.9
Chloroflexi	SAR202	—	—	1	1.1	—	—	—	—	6	0.7	14	2.9	31	13.5	17	7.5
Marinimicrobia		3	3.2	5	10.3	2	2.5	—	—	65	9.8	59	9.8	43	15.2	43	17.6
Cyanobacteria		3	7.9	—	—	4	18.5	—	—	2	3.0	1	0.03	—	—	—	—
Gemmatimonadetes		—	—	1	0.7	—	—	—	—	—	—	4	0.7	5	1.8	7	2.0
Planctomycetes		—	—	—	—	—	—	—	—	1	0.1	7	0.7	1	0.1	8	2.4
Proteobacteria	Alphaproteobacteria	22	51.9	23	37.7	53	61.6	11	39.5	191	41.1	156	37.1	35	20.5	55	21.8
	SAR116	1	1.4	—	—	2	2.3	—	—	5	1.5	—	—	—	—	—	—
	SAR11 clade Deep 1	2	4.5	4	9.0	—	—	3	16.5	75	11.8	79	17.4	14	9.2	19	9.6
	SAR11 clade Surface 1–4	13	37.1	12	20.0	32	43.3	3	17.3	73	20.0	42	11.4	10	6.9	19	7.5
	SAR11	1	2.7	2	4.9	13	11.2	—	—	20	2.3	12	2.2	2	0.3	3	0.3
	SAR 11 total	15	44.2	18	33.8	45	54.6	6	33.7	168	34.1	133	31.0	26	16.3	41	17.3
	Gammaproteobacteria	8	7.2	4	11.6	5	3.3	2	6.5	28	8.0	28	6.7	9	6.8	16	8.3
	SAR86	6	5.0	1	2.4	3	1.5	1	2.4	8	2.9	5	1.3	3	2.7	2	2.1
	SUP05	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2	0.9
	Deltaproteobacteria	1	4.5	2	3.7	3	2.8	4	12.4	15	7.3	18	7.4	10	12.8	16	9.8
	SAR324	1	4.5	2	3.7	2	2.5	4	12.4	11	6.9	16	7.3	8	12.3	12	9.2
Nitrospina		1	1.2	—	—	—	—	—	—	1	0.5	6	1.8	1	0.2	2	0.5
Verrucomicrobia		—	—	—	—	—	—	—	—	7	0.8	3	0.8	1	0.5	5	1.5
Bacteria	Unclassified	—	—	—	—	—	—	—	—	—	—	1	0.1	1	0.1	1	0.1
Thaumarchaeota		3	9.0	16	27.2	—	—	13	38.2	38	12.1	17	23.9	9	23.3	15	21.4
	Marine Group I	3	9.0	16	27.2	—	—	13	38.2	37	12.0	15	23.6	6	21.9	13	20.7
	Marine Benthic Group A	—	—	—	—	—	—	—	—	1	0.1	1	0.2	3	1.4	2	0.6
Euryarchaeota	Thermoplasmata	5	4.7	4	4.2	2	1.2	2	3.4	24	6.5	6	1.5	3	1.3	6	3.2
	Marine Group II	5	4.7	2	2.9	2	1.2	2	3.4	23	6.4	5	1.1	1	0.4	4	1.9
	Marine Group III	—	—	2	1.3	—	—	—	—	1	0.1	1	0.4	2	0.9	2	1.3
Total no. of SSU rRNA gene sequences		53		57		75		32		404		341		157		203	

EA %, estimated abundance in percent; GEO, Geotraces; MOC, MOCA; No. SSU, number of 16s rRNA sequences.

Table S5. Carbohydrate active enzymes (CAZyme) identified in the metagenomic libraries with relative gene abundances depicted for CAZyme families

CAZyme	Substrate family	Enzyme activity	EC no.	Metagenome	Relative abundance
GH1	Hemicellulose (xyloglucans)	Exo-β-1,4-glucanase	3.2.1.74	GEO_3	1.97
	Pectin (RGI)	β-Galactosidase	3.2.1.23	MOC_1	2.59
				MOC_3	0.6
GH2	Glycoproteins (mannans)	β-Mannosidase	3.2.1.25	MOC_1	0.19
		β-Glucuronidase	3.2.1.31	MOC_3	0.29
				MOC_4	2.42
GH3	Cellulose	β-Glucosidase	3.2.1.21	GEO_1	7.72
	Hemicellulose (xyloglucans)	Xylan 1,4-β-xylosidase	3.2.1.37	GEO_3	0.78
	Pectin	Exo-β-1,4-glucanase	3.2.1.58	MOC_1	5.71
				MOC_2	4.36
				MOC_3	15.33
				MOC_4	12.51
GH5	Cellulose	Endo-β-1,4-glucanase	3.2.1.4	GEO_1	8.04
	Hemicellulose (xylans)	β-1,4-Cellobiosidase	3.2.1.91	GEO_2	0.86
				GEO_3	0.53
				MOC_1	2.28
				MOC_2	1.91
				MOC_3	4.12
				MOC_4	1.02
GH13	Starch	αAmylase	3.2.1.1	GEO_1	186.62
				GEO_2	24.68
				MOC_1	0.58
				MOC_2	0.29
				MOC_3	6.46
				MOC_4	11.62
GH16	β-1,3-Glucans	β-1,3-Glucosidase	3.2.1.39	MOC_1	4.03
				MOC_2	1.07
				MOC_3	0.95
				MOC_4	3.57
GH20	Chitin	β-Hexosaminidase	3.2.1.52	GEO_2	9.61
				MOC_1	1.61
				MOC_2	0.51
				MOC_3	1.76
				MOC_4	4.46
GH43	Hemicellulose (xyloglucans)	β-Xylosidase	3.2.1.37	MOC_1	1.9
				MOC_4	0.17
GH65	Starch	Maltose phosphorylase	2.4.1.8	MOC_3	0.62
		Trehalose phosphorylase	2.4.1.64	MOC_4	0.47

GH, glycoside hydrolases.

Other Supporting Information Files

Dataset S1 (XLSX)
Dataset S2 (XLSX)
Dataset S3 (XLSX)
Dataset S4 (XLSX)
Dataset S5 (XLSX)
Dataset S6 (XLSX)