UC Santa Barbara
UC Santa Barbara Previously Published Works

Title

Microbial functional diversity across biogeochemical provinces in the central Pacific Ocean.

Permalink

https://escholarship.org/uc/item/1962f12b

Journal

Proceedings of the National Academy of Sciences of USA, 119(37)

Authors

Saunders, Jaclyn

McIlvin, Matthew

Dupont, Chris

et al.

Publication Date

2022-09-13

DOI

10.1073/pnas.2200014119

Peer reviewed

Microbial functional diversity across biogeochemical provinces in the central Pacific Ocean

Jaclyn K. Saunders[a,1,2] , Matthew R. McIlvin[a], Chris L. Dupont[b,c,d], Drishti Kaul[b] , Dawn M. Moran[a], Tristan Horner[a], Sarah M. Laperriere[e],
Eric A. Webb[f,g] , Tanja Bosak[h], Alyson E. Santoro[e] , and Mak A. Saito[a,2]

Enzymes catalyze key reactions within Earth's life-sustaining biogeochemical cycles. Here, we use metaproteomics to examine the enzymatic capabilities of the microbial community (0.2 to 3 μm) along a 5,000-km-long, 1-km-deep transect in the central Pacific Ocean. Eighty-five percent of total protein abundance was of bacterial origin, with Archaea contributing 1.6%. Over 2,000 functional KEGG Ontology (KO) groups were identified, yet only 25 KO groups contributed over half of the protein abundance, simultaneously indicating abundant key functions and a long tail of diverse functions. Vertical attenuation of individual proteins displayed stratification of nutrient transport, carbon utilization, and environmental stress. The microbial community also varied along horizontal scales, shaped by environmental features specific to the oligotrophic North Pacific Subtropical Gyre, the oxygen-depleted Eastern Tropical North Pacific, and nutrient-rich equatorial upwelling. Some of the most abundant proteins were associated with nitrification and C1 metabolisms, with observed interactions between these pathways. The oxidoreductases nitrite oxidoreductase (NxrAB), nitrite reductase (NirK), ammonia monooxygenase (AmoABC), manganese oxidase (MnxG), formate dehydrogenase (FdoGH and FDH), and carbon monoxide dehydrogenase (CoxLM) displayed distributions indicative of biogeochemical status such as oxidative or nutritional stress, with the potential to be more sensitive than chemical sensors. Enzymes that mediate transformations of atmospheric gases like CO, CO_2, NO, methanethiol, and methylamines were most abundant in the upwelling region. We identified hot spots of biochemical transformation in the central Pacific Ocean, highlighted previously understudied metabolic pathways in the environment, and provided rich empirical data for biogeochemical models critical for forecasting ecosystem response to climate change.

marine microbial ecology | metaproteomics | mesopelagic | nitrification | methylotrophy

Earth's elemental cycles are heavily influenced by biological systems, where enzymatic catalysts contribute to large-scale biochemical transformations of carbon, nitrogen, sulfur, and trace metals. Proteins are necessary for life through their role in key cellular functions including transporting essential molecules into and out of the cell, forming cellular structures, storing key micronutrients, regulating cellular responses to environmental perturbations, and catalyzing biochemical transformations. The latter of these functions, enzymatic catalysts, are not only essential in cellular metabolism within organisms but can also directly influence global biogeochemical cycles and ecosystems (1). The recently developed ability to directly measure proteins in the ocean (2–10) has enabled the study of biogeochemically active proteins and their physiochemical drivers. Scaling measurements of adaptive response proteins over large geographic regions provides the opportunity to use microbes as geochemical sensors of environmental conditions (7, 11). Together, the direct analysis of proteins on an oceanic scale provides a rich empirical dataset to constrain models of biochemical transformations in the ocean (12–15), including those that impact major elemental cycles like carbon, nitrogen, oxygen, and sulfur.

Here, we applied high-resolution sampling and metaproteomic analyses to determine how geophysical forcings and microbial community dynamics shape spatial distributions of pelagic marine microorganisms and the biogeochemical cycles they mediate across biomes in the central Pacific Ocean. Ninety-eight microbial community protein samples (0.2 to 3 μm) and geochemical samples were collected at 11 locations through the euphotic and mesopelagic zones (depth range of 20 to 1,250 m) along a 5,000-km transect from Hawai'i to Tahiti (Fig. 1A) aboard the research vessel (R/V) *Falkor* in January–February of 2016, during a strong El Niño period. Metaproteomic samples were analyzed by high-resolution two-dimensional active modulation nanospray mass spectrometry (16) and mapped to a corresponding metagenome dataset yielding 107,579 unique peptides and 56,543 protein groups (88,251 unique proteins). Protein abundance data are represented as corrected spectral counts per liter of seawater (sc_{corr}/L_{sw}) by scaling spectral counts to

Significance

Enzymatic proteins are the engines of life, generating energy, fixing CO_2 into organic matter, and building biomass. At scale, these biological catalysts have the power to drive global elemental cycles. These microbial machines were responsible for the oxygenation of the atmosphere, can mediate the formation and destruction of greenhouse gases, and potentially influence cloud formation. Here, we used metaproteomics across the central Pacific Ocean to investigate how microbial protein distributions change across vertical and horizontal scales, identify interactions among microbial consortia, and characterize metabolic hot spots in line with major geochemical features. These observations provide insights into microbial community dynamics and can act as empirical constraints for biogeochemical ecosystem models to improve understanding of microbial interactions on a changing planet.

[1]Present address: Department of Marine Sciences, University of Georgia, Athens, GA 30609.

[2]To whom correspondence may be addressed. Email: msaito@whoi.edu or jaci@uga.edu.
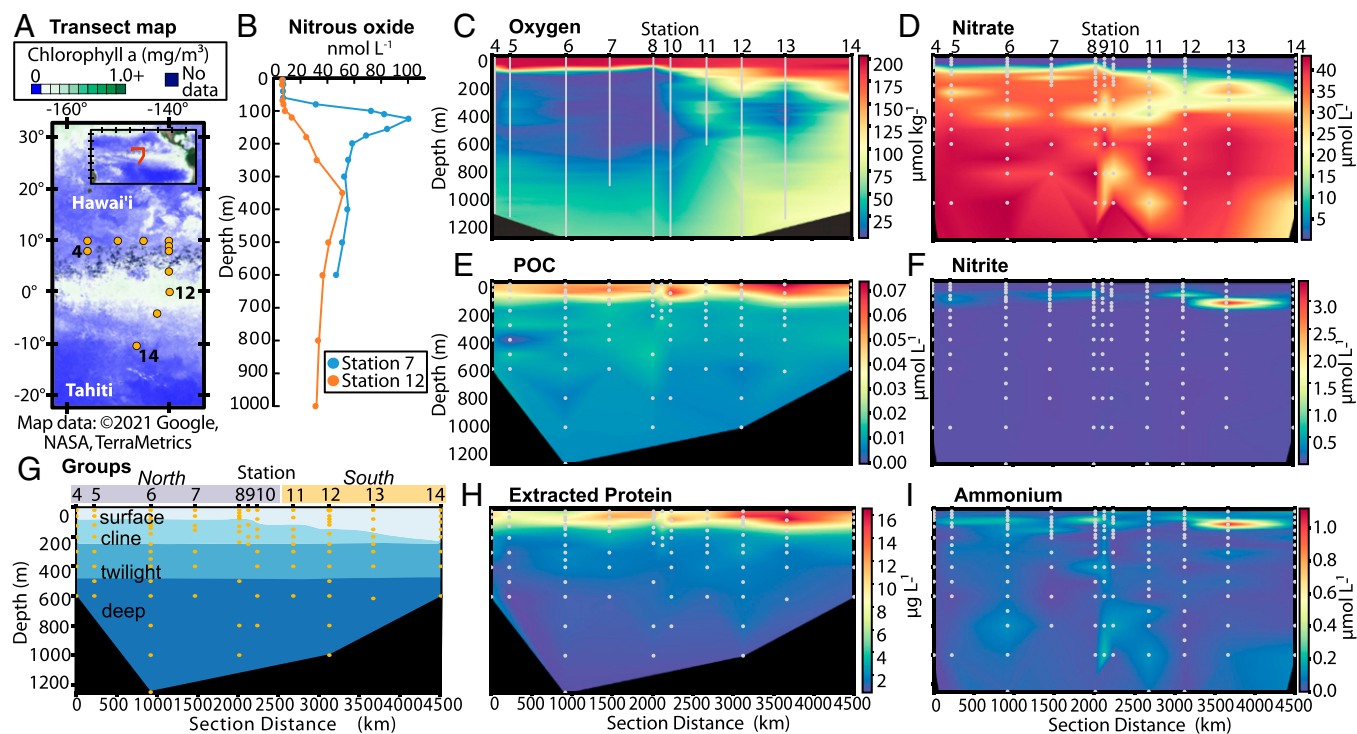
Fig. 1. Station locations, geochemical features, and sample groupings along vertical and horizontal scales. (*A*) Cruise track for the ProteOMZ research expedition across the central Pacific Ocean aboard the R/V *Falkor* in early 2016 overlaid on a composite image of estimated chlorophyll-*a* from the moderate resolution imaging spectroradiometer (MODIS) Aqua satellite data during the expedition with an additional month bounded on both sides. The more productive waters associated with equatorial upwelling are visible in a band centered on station 12 ranging from stations 11 to 13. Dark blue represents lack of satellite data due to dense cloud cover. (*B*) Depth profiles of nitrous oxide concentrations from stations 7 and 12. (*C*) Oxygen concentrations across the transect as observed from a conductivity, temperature, and depth instrument with reduced oxygen regions associated with the extremities of the Eastern Tropical North and South Pacific ODZs evident near stations 7 and 13, respectively. (*D*, *F*, and *I*) Concentrations of nitrate, nitrite, and ammonium, respectively, as determined from water samples collected by Niskin bottle. (*E*) POC per volume of seawater sampled determined from GF/F filters attached to McLane in situ pumps collected concurrently with proteomics samples, peaking in abundance near the region of equatorial upwelling in the south. (*G*) Samples, represented by individual dots, grouped by region (north and south) as well as by depth (surface, cline, twilight, and deep) as identified by machine learning clustering analyses of geochemical and hydrographic data. While station 11 is in the Northern Hemisphere and station 12 is at the equator, based upon environmental data they clustered together with the southernmost stations. (*H*) Total protein extracted per volume of seawater sampled from the 0.2- to 3.0-μm filters.

the abundance of protein sampled from a volume of seawater while correcting for instrument technical ionization bias (*Materials and Methods*). This study paper focuses on only some proteins across this transect, but all of the proteins in this large dataset are available for exploration and visualization in the Ocean Protein Portal (17).

The central Pacific Ocean is a region of strong physicochemical gradients due to upwelling of nutrient-rich waters on the equator surrounded by regions of depleted nutrients in the subtropical gyres to the north and south (Fig. 1*A*) with low-oxygen interior waters throughout (Fig. 1*C*). Patterns of microbial vertical stratification and regional variability within the metaproteomic data were consistent with varying physical and geochemical contexts (Fig. 1 *B–E*, *G*, and *H* and *SI Appendix*, Fig. S1) (18–20). We studied the vertical stratification of pelagic free-living microbial communities (not particle-associated) as they transitioned from the euphotic zone—a region of photosynthetic carbon fixation—to the mesopelagic with subsequent removal of labile dissolved organic carbon with depth (21). Protein distributions were not only analyzed in relation to these geochemical and physical features but were also investigated as they related to one another revealing complex interdependencies between various actors in the microbial communities across this transect.

Results and Discussion

To interrogate this large dataset, we utilized multiple statistical and modeling approaches to identify the major patterns in protein distributions (*Materials and Methods*). Machine learning clustering analyses of hydrographic and geochemical data resulted in partitioning of microbial communities along the transect into cohesive depth bins along a broad regional scale (Fig. 1*G*). Four major depth groups resulted from *k*-means clustering: surface, cline, twilight, and deep depth groups (*SI Appendix*, Fig. S2). Hierarchical clustering also divided stations into two regions: North Pacific (stations 4 to 10) and South Pacific regions (stations 11 to 14; *SI Appendix*, Fig. S3). These two clusters corresponded to multiple biogeochemical provinces: North Pacific stations captured the low-nutrient surface waters of the North Pacific Subtropical Gyre (NPSG; stations 4 and 5) and the western flank of the Eastern Tropical North Pacific (ETNP) oxygen deficient zone (ODZ) within the cline and twilight depths. In contrast, South Pacific stations captured the highly productive and relatively nutrient-replete area of equatorial upwelling, with corresponding enhanced particulate organic carbon (POC), NO_2^-, and NH_4^+ in the surface and cline (Fig. 1 *E*, *F*, and *I*). To better understand vertical stratification of proteins, we performed a community attenuation analysis by fitting a power law model to protein abundance through the water column. Microbial protein attenuation (*c*) is an indicator of the importance of specific proteins to microbial communities with increasing depth through the water column (Figs. 1 *E* and *H* and 2*A* and *SI Appendix*, Fig. S4). For example, the average *c* for all proteins is −1.25. Very negative values of *c* indicate proteins that are predominantly found in the surface relative to communities at depth. Less negative values
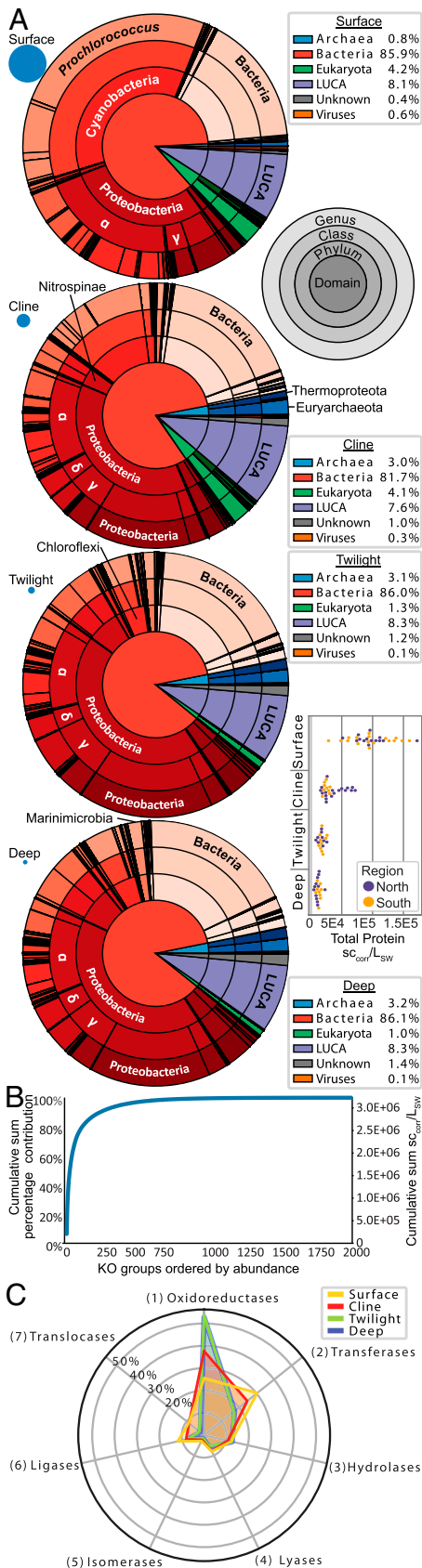
Fig. 2. Taxonomic and functional diversity within the ProteOMZ metaproteome. (*A*) The large concentric circles represent the relative abundance of peptides attributed to taxonomic groups according to an LCA analysis of peptides per depth group. The legend for each depth group also displays the percentage contribution of each major domain to the taxonomic profile with LUCA assigned to peptides which are highly conserved and thus found in multiple domains. Unknown indicates peptides which have no taxonomic

of *c* indicate proteins that are more important to microbial communities in deeper depths as they have a less pronounced decline in abundance with depth through the water column, with positive values indicating the few proteins that increase in abundance with depth even though total biomass dramatically decreases through the water column. Finally, an artificial neural network analysis provided insight across these biogeochemical provinces by highlighting protein distributions characteristic to specific samples, generating sample-specific "fingerprints," while simultaneously providing an exploration of microbial consortia interactions among proteins of varying functions and taxonomic origin.

The contribution of different taxa to the microbial community proteome varied along both vertical and regional dimensions. A least common ancestor (LCA) analysis of individual peptide constituents was conducted, with the relative contributions from each taxonomic group reported along depth groups (Fig. 2*A*) (22). The overall microbial proteome was dominated by marine Bacteria, with 85.1% of peptides assigned to bacterial groups in the surface, twilight, and deep depth groups. Bacteria had the lowest relative contribution in the cline at 81.7%. Cyanobacteria were the primary source of peptides in the surface (36.1%) with *Prochlorococcus* as the dominant taxon (25.8%). *Pelagibacter* was the primary contributor among the Alphaproteobacteria, the second-largest source of peptides in the surface (20.8%). Both the Cyanobacteria and Alphaproteobacteria peaked in abundance in the South Pacific near the region of equatorial upwelling. This coincided with the peak in POC (Pearson correlation coefficient $r \geq 0.94$; Fig. 1 *E* and *H*). The dominance of these two groups inferred from metaproteomics is consistent with taxonomic distributions from metagenomic data collected from similar pelagic regions (23–27). Integrating over all depth groups, Proteobacteria were the single largest source of peptides, contributing 32.6% overall. Eukaryotes, and Nitrospinae bacteria were the next largest groups, contributing 3.8 and 2.1% of the total peptides, respectively. Eukaryotic peptides were primarily found within the euphotic zone and were associated with the picoeukaryotic phytoplankton *Pelagomonas* and diatoms because the size fraction analyzed here (0.2 to 3.0 μm) precluded most eukaryotes and large particle-associated organisms (9). Archaeal peptides were most abundant below the surface depth group, ranging from 3.0 to 3.2%. About 2/3 of archaeal peptides were from Euryarchaeota in which the dominant peptides were associated with L-amino acid transport and metabolism, and 1/3 from Thermoproteota (Thaumarchaeota, Aigarchaeota, Crenarchaeota, and Korarchaeota), in which the bulk of the peptides were associated with the uptake and metabolism of nitrogenous compounds, in line with prior findings from studies of Archaea (28). Interestingly, Archaea are known to be highly abundant by cell number in the mesopelagic (29, 30), but their small cell size and slow metabolism appeared to result in a small contribution to overall microbial protein abundance. The same pattern was also recently observed in global marine metatranscriptomes (31). A relatively small contribution to the overall proteome was

homology to known organisms. The size of blue circles under the depth group names represents the relative contribution of peptides from that depth group to the overall metaproteome. (*Inset*) The swamp plot displays the abundance of peptides in sc_{corr}/L_{SW} per depth group, colored by region. (*B*) A cumulative summation plot categorically displaying KO groups in rank order of abundance along the *x* axis and the relative contribution of each KO group to the total peptides in the KO-identifiable metaproteome along the *y* axis. (*C*) The relative abundance of peptides assigned to the major Enzyme Commission categories according to depth group.

made by peptides that lacked association with any known taxonomic group (0.6%) demonstrating the capacity for dark environmental DNA to be translated into "dark protein."

The functional capabilities of the metaproteome were investigated by assigning traits using KEGG Orthology (KO) and Enzyme Commission (E.C.) identifiers. Seventy percent of the normalized spectral counts were assigned functional traits with KOs. Notably, out of 2,037 functional KO groups identified (Movie S1), over 51.6% of the peptide abundance was associated with only 25 KO groups (Fig. 2B and SI Appendix, Fig. S5). The chaperone protein GroEL was the single most abundant functionally characterized protein across the dataset (Fig. 3 and SI Appendix, Fig. S6), accounting for 6.4% of the KO-identifiable normalized spectral counts. The sheer abundance of the GroEL protein has not been previously observed using metagenomic or metatranscriptomic approaches highlighting the differences between transcription and protein abundance and the universal importance of protein folding in marine microbes.

A photosynthesis-centered taxonomic community characterized the surface depth group, while a more diverse community was present at depth responding to more variable and diffuse sources of energy (Fig. 2A). The functional focus of the microbial community also changed in the cline, twilight, and deep depth groups. E.C. numbers classify the chemical reactions carried out by enzymes into seven broad categories (Fig. 2C). Evaluation of these broad reaction categories revealed that the microbial catalytic focus shifted from that associated with rapid growth and reproduction in the surface, characterized by abundant transferases (E.C. class 2), to a focus on energy harvesting dominated by oxidoreductases (E.C. class 1) at depth (Fig. 2C). The transferase DNA-directed RNA polymerase (2.7.7.6) was the most abundant enzyme in the surface and throughout the entire dataset, consistent with microbial growth and associated transcription (32). In accordance with the biomass production and growth in the surface, peptides from lyase class (E.C. class 4) including ribulose-1,5-bisphosphate carboxylase/oxygenase (RubisCO, RbcLS), the major CO_2 fixation enzyme, and the HCO_3^- consuming phosphoenolpyruvate carboxylase (PEPC), which participates in the anaplerotic synthesis of oxaloacetate in Cyanobacteria (33, 34), were most abundant in the surface (Fig. 3). The transferases were the most abundant enzyme group in the photosynthesis-dominated surface layer where the majority of proteinaceous biomass occurred. The relative contribution of oxidoreductase enzymes increased with depth through the mesopelagic where the microbial focus shifted to maintaining life-sustaining energetic demands and were the most abundant enzyme class throughout the entire water column, composing 30% of total enzyme abundance.

Transport proteins were abundant, contributing at least 28% of the total proteome, similar to the findings of microbial metaproteomes from the pelagic Atlantic Ocean (2, 5). Transport protein distributions in the surface ocean corroborated prior findings of nitrogen stress in Cyanobacteria from the oligotrophic NPSG (7) while expanding upon the biomarker catalog (Fig. 3). Proteobacterial iron transporters were vertically stratified, with the iron(III) transport system substrate-binding protein (AfuA) significantly more abundant in the surface and the TonB-dependent siderophore receptor protein (TC.FEV.OM) deeper in the euphotic zone. The community attenuation coefficients (c) of -4.13 and -1.28 for AfuA and the TonB-dependent transporter, respectively, indicated the greater importance of the TonB-dependent iron transporter at depth because it attenuated more slowly through the water column than AfuA at a level more similar to the attenuation for all proteins ($c = -1.25$; Fig. 3).

This stratification hints at the differing microbial strategies for uptake of limiting nutrients, suggesting ligand-bound mechanisms are preferentially utilized by free-living microbes deeper in the water column. Some organic carbon transport proteins also displayed vertical and regional-scale variations. The simple sugar transport substrate-binding protein (ABC.SS.S: $c = -4.36$) attenuated more quickly than the multiple sugar transport substrate-binding protein (ABC.MS.S: $c = -1.17$), suggesting a shift in the DOC pool with depth (Fig. 3). Notably, the inositol-phosphate transport substrate-binding protein of Proteobacteria increased in abundance in the deep, indicative of the importance of this protein to community success in the mesopelagic (InoE: $c = +0.26$; Figs. 3 and 4A). Some transport proteins displayed variation across regional scales, suggestive of variation in DOC utilization along horizontal gradients; for example, the fructose transport substrate-binding protein (FcrB), primarily produced by Alphaproteobacteria, was significantly more abundant in the surface of the North Pacific (Fig. 3). Variations in transport protein abundance can act as biological indicators (or biomarkers) of scarce nutrient distributions and stress in microbial communities (5, 7, 35).

The patterns in the distribution and abundances of oxidoreductase enzymes indicated carbon stress, oxidative stress, and use of alternative respiratory pathways. The oxidoreductase class of enzymes catalyze redox reactions and often utilize transition metals at their catalytic sites, most often using Fe cofactors, as well as Mo, Cu, and W, among others (36). Oxidoreductases are critical components of respiration and include the CO_2-evolving enzymes pyruvate dehydrogenase (PdhA) and isocitrate dehydrogenase (IDH), both enriched in the surface (Fig. 3) in conjunction with POC and O_2 (Fig. 1 C and E). Many biogeochemically relevant oxidases support chemolithoautotrophy by mediating the oxidation of reduced substrates. Distributions of oxidoreductase enzymes associated with nitrification were linked with major biogeochemical features along this transect, specifically within the ETNP ODZ and equatorial upwelling. The Mo- and Fe-containing enzyme nitrite oxidoreductase (NxrAB) is responsible for the final step in nitrification through the oxidation of NO_2^- to NO_3^- (SI Appendix, Fig. 10). NxrAB was the most abundant oxidoreductase protein along this transect (Figs. 3 and 4B) and in the mesopelagic central Pacific Ocean, in general, at over 60 billion molecules per liter (8). While NxrAB was extraordinarily abundant in the mesopelagic with low attenuation coefficients ($c = -0.38$ and -0.55 for NxrA and NxrB, respectively), it peaked in abundance in the oxygenated surface region of equatorial upwelling at station 12 (Fig. 4B). The abundance of NxrAB here may be supported through elevated production of NO_2^- via remineralization of photosynthetically derived organic matter and aerobic ammonia oxidation.

The Cu-containing archaeal ammonia monooxygenase (including subunits AmoABC) is responsible for the first step of nitrification through the oxidation of NH_4^+ (SI Appendix, Figs. S11–S13). AmoABC peptides were associated with Thermoproteota and peaked in abundance at the interface of the surface and cline depth groups (Figs. 3 and 4C). The Cu-utilizing nitrite reductase (NirK), which reduces NO_2^- to NO, displayed variable distribution patterns related to taxonomic origin. Bacterial NirK was primarily found in low-oxygen waters (Figs. 1B, 4D, and 5) with the explicit functional role of NirK in these free-living communities still uncertain. Similarly, the function of archaeal NirK is also uncertain but has been suggested as a source of NO, an essential intermediate of ammonia oxidation (37, 38). Archaeal NirK occurred at the top of the ETNP ODZ at an abundance roughly similar to that of bacterial NirK; however,
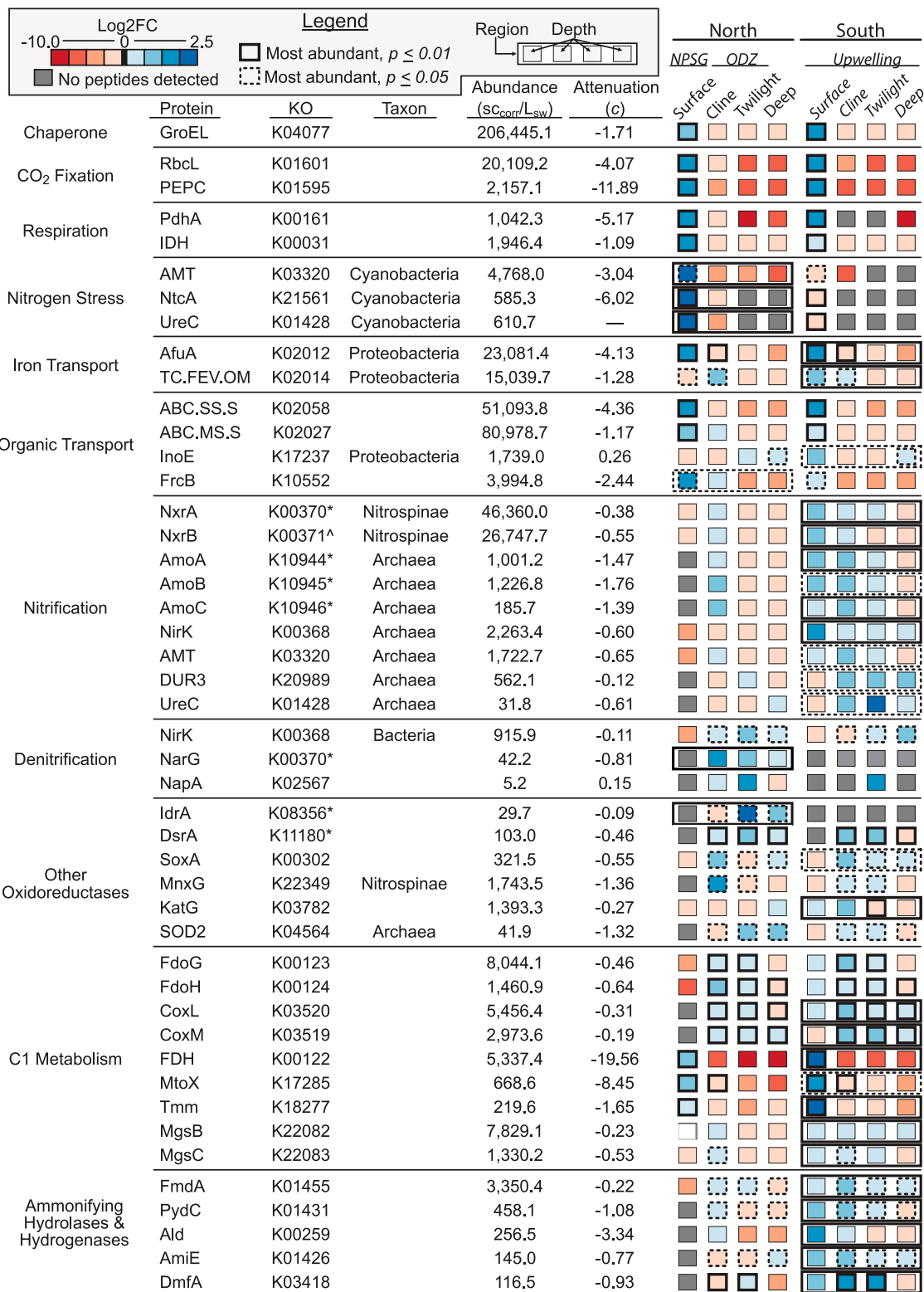
Fig. 3. Summary table of protein abundance and attenuation across the transect. Data in the table include the gene name, Kegg Ontology identifier (KO), specific taxonomic group determined by LCA analysis of peptides (blank taxa indicate all taxa for that protein presented), total abundance (sc_{corr}/L_{sw}: spectral counts per liter of seawater), community attenuation (c) through the water column (dash indicates lack of data for calculating; *Materials and Methods*). The heat map represents an aggregate of the individual depths (surface, cline, twilight, and deep) across the regions (north and south). The colors in the heat map represent the log2 fold change of the average abundance for each depth/region combination compared to the overall average abundance for all samples. Lines around groups represent where a protein in a particular location is significantly more abundant than other locations, with a dashed line indicating $P \leq 0.05$ and a thick solid line indicating $P \leq 0.01$. Lines around an entire region indicate where a protein is significantly more abundant than the other region as assessed by a Mann–Whitney u test. Lines around individual depths indicate where a protein is significantly more abundant in one or more depths when compared to the other depths as assessed by a Kruskal–Wallis H test with post hoc Dunn's tests. Asterisk indicates KO groups that contain multiple different functional proteins, such as K00370, which contains both NarG and NxrA proteins (*Materials and Methods* and *SI Appendix*, Figs. S10–S15). Caret indicates peptides identified through parsimony analysis from protein group inference in the software package Scaffold as opposed to LCA analysis which was utilized elsewhere due to the high conservation of the peptides among taxonomic groups (*Materials and Methods*). Section profiles of the distribution of these proteins can be found in Fig. 4 or *SI Appendix*, Figs. S6–S9.
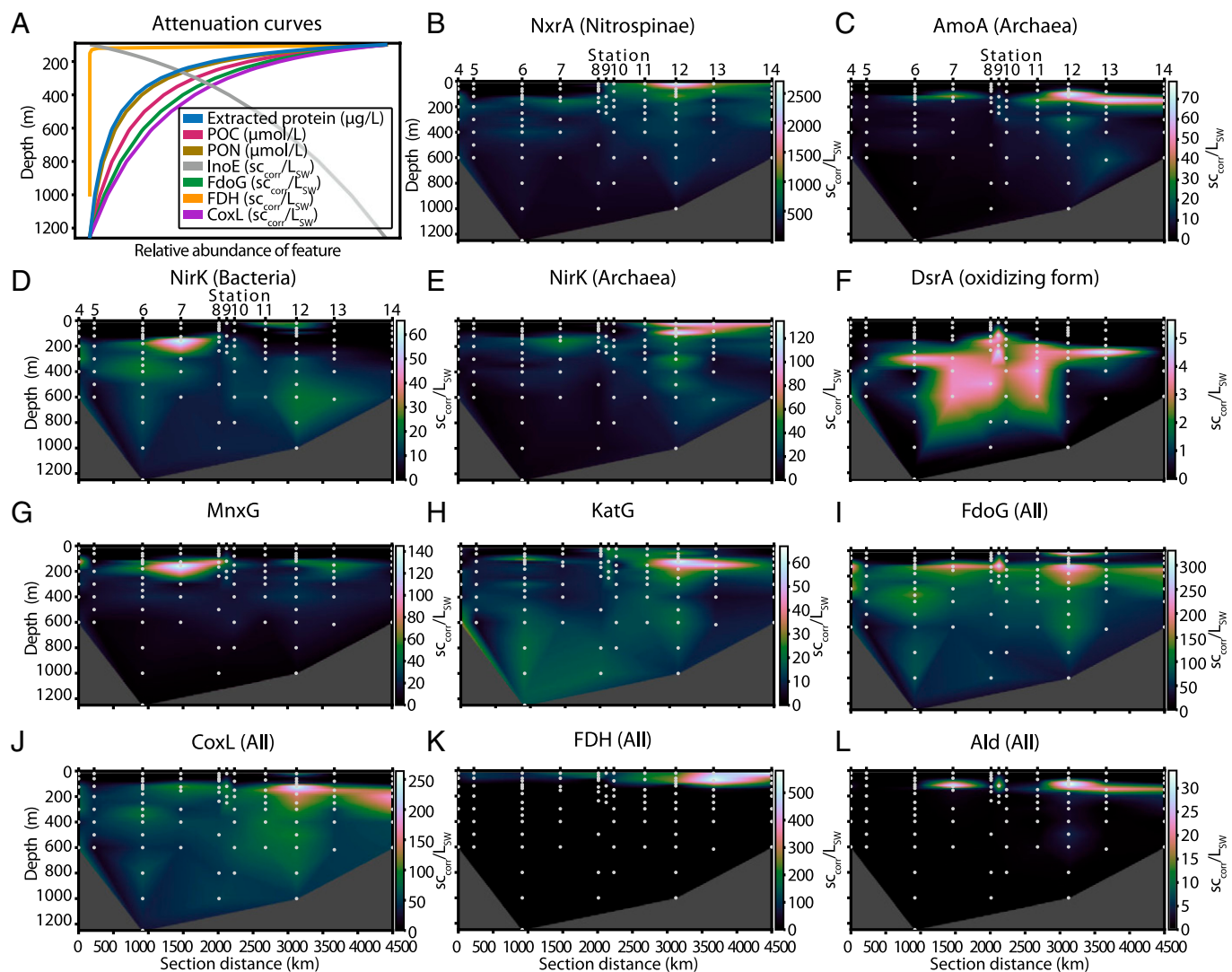
Fig. 4. Attenuation profiles and distributions of the most abundant oxidoreductases in the mesopelagic. (*A*) Attenuation lines calculated by fitting a power law through abundance data of total extracted protein ($c = -1.15$), POC ($c = -0.7$), PON ($c = -1.05$), transport, and oxidoreductase enzymes that dominate the mesopelagic. The attenuation for all proteins combined was $c = -1.25$; more negative attenuations mean that proteins are more abundant in the surface and are reduced more quickly from communities at depth. Proteins with relatively slow attenuation rates, like the low-oxygen formate dehydrogenase (FdoG; $c = -0.46$), which oxidizes the C1 compound formic acid to CO_2, and carbon monoxide dehydrogenase (CoxL; $c = -0.31$), which oxidizes carbon monoxide to CO_2, are shown. Also presented is the aerobic formate dehydrogenase (FDH; $c = -19.56$), which is associated with methylotrophy and attenuates rapidly as it is far more abundant in the surface than in the mesopelagic. The inositol-phosphate transport system substrate-binding protein (InoE; $c = +0.26$), which increases in abundance with depth, is also shown. (*B*) Abundance of the most abundant oxidoreductase enzyme, nitrite oxidoreductase (NxrA), across the transect. This enzyme is abundant in the mesopelagic as well as in the surface waters near the equatorial upwelling region in the south. (*C*) Ammonium monooxygenase alpha subunit (AmoA) of Thermoproteota is most abundant at the interface of the surface and cline and is also more abundant in the south. (*D*) Bacterial nitrite reductase (NirK) peaks in abundance in the low-oxygen waters above the ETNP ODZ. (*E*) Archaeal nitrite reductase (NirK) is present in the waters above the ETNP ODZ but peaks in abundance in the surface waters associated with equatorial upwelling. (*F*) The oxidizing form of dissimilatory sulfite reductase (DsrA) is found within the mesopelagic waters. (*G*) The copper monooxygenase from Nitrospinae, putatively a Mn oxidase (MnxG), is most abundant at the top of the ETNP ODZ and is also found at similar depths in more oxygenated waters in the South Pacific. (*H*) MnxG shows a similar distribution pattern in the South Pacific to the bacterial catalase (KatG), which is indicative of a region characterized by oxidative stress. (*I*) The third most abundant oxidoreductase protein in the mesopelagic, formate dehydrogenase (FdoG), is most abundant along oxic transitional regions. (*J*) The abundant protein carbon monoxide oxidoreductase (CoxL) is also found throughout the mesopelagic, peaking in abundance in the South Pacific. (*K*) The formate dehydrogenase (FDH), however, is most abundant in the South Pacific surface near high POC. (*L*) The distribution of multiple ammonifying catabolic enzymes like alanine dehydrogenase (Ald) is tightly correlated with ammonia monooxygenase.

archaeal NirK was significantly more abundant in the South Pacific near the equatorial upwelling region of station 12 (Fig. 4*E*). Archaeal NirK is located near NxrAB from *Nitrospina* in the neural net (39) (Fig. 5*C*) and is closely correlated ($r = 0.78$ for NxrA and archaeal NirK) showing that these enzymes have similar distribution patterns indicative of related geochemical niches linked to NO_2^-. The high abundance of NO_2^- utilizing enzymes in conjunction with the typically lower standing NO_2^- concentrations in the surface ocean above the primary nitrite maximum suggests active chemical cycling where the available

NO_2^- was consumed as quickly as it was produced in these regions (Figs. 1*E* and 4 *B* and *E*). Notably, higher concentrations of N_2O were observed at the top of the ETNP ODZ at station 7 (Fig. 1*B*), where archaeal NirK and AmoABC peaked in the North Pacific. This cooccurrence of N_2O and archaeal ammonia oxidation proteins in the ETNP ODZ may be due to the activity of archaeal ammonia oxidizers, whose N_2O yield from ammonia oxidation increases at low oxygen concentrations (40). Purified archaeal NirK has been shown to catalyze both the formation of N_2O from hydroxylamine under aerobic conditions, and the
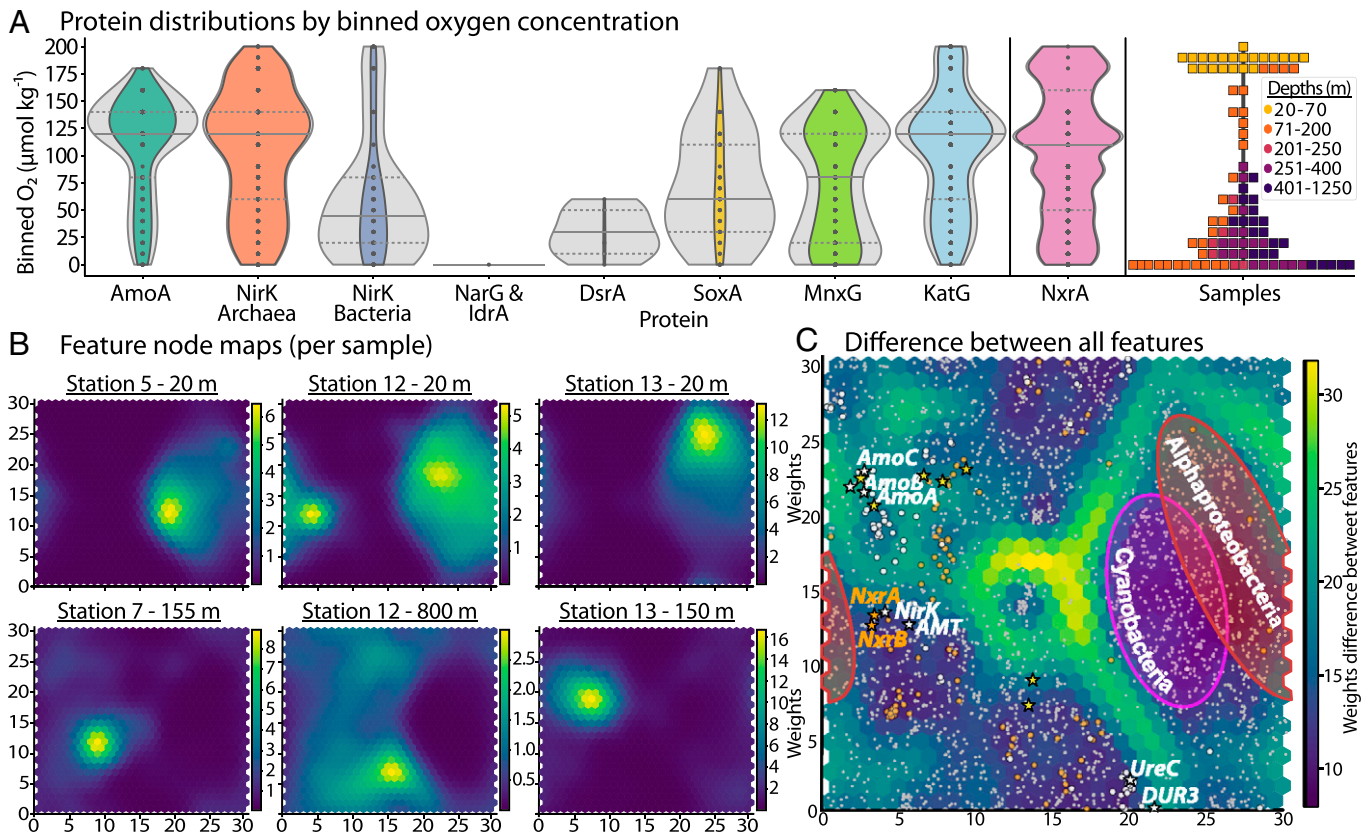
Fig. 5. Distribution of select oxidoreductases across the transect, according to oxygen concentrations, and all proteins as analyzed by an artificial neural network. (*A*) Distributions of proteins binned by oxygen concentrations (10 μmol/kg) and normalized by the number of samples per bin. Gray outlines of protein distributions show the shape of the distribution for each individual enzyme. The colored interiors of AmoA, Archaeal and Bacterial NirK, DsrA, SoxA, MnxG, and KatG are distributions scaled to the relative abundance of Archaeal NirK, the most abundant of these proteins. NxrA is presented separately as it is significantly more abundant than the other proteins (20× more than Archaeal NirK). NarG and IdrA are presented together as they both have the same distribution pattern and only occur in the lowest oxygen bin. The samples column displays the discrete number of samples in each oxygen bin by depth illustrating the variability in sampling across oxygen concentrations. (*B*) Neural network feature maps of individual samples showing the unique fingerprints of each sample, highlighting the major underlying protein distributions associated with each sample. Note how the fingerprints change with depth as well as along regional scales. The neural net is composed of 900 individual nodes (a 30 × 30 matrix) using periodic boundary conditions. (*C*) The integrated differences of weights across all samples in the neural net are displayed in the background. Overlaid on top are points which represent the location of individual proteins according to nodes in the neural net (best matching unit). Individual points within a single node are offset slightly to show density of points. Nodes closer to each other with lower weight differences between them are more similar to each other.

reduction of NO_2^- with hydroxylamine to form N_2O under anaerobic conditions (41). Further, production of labeled N_2O from NH_4^+ has been demonstrated over a range of oceanic oxygen concentrations where archaea are the only ammonia oxidizers (42, 43). However, we cannot conclusively rule out that this N_2O was instead advected from a remote zone of production further east in the ETNP and not locally produced.

The three nitrification enzymes (NxrAB, Archaeal NirK, and AmoABC) also cooccurred within the low-oxygen transition into the ETNP ODZ in the North Pacific (Fig. 1*C*), where anaerobic respiratory pathways can occur in free-living communities within the pelagic water column (44). The high oxygen affinities of archaeal ammonia oxidation and *Nitrospina*-mediated nitrite oxidation support these metabolic processes in the ETNP ODZ (45) and the Eastern Tropical South Pacific (ETSP) ODZ (46). The first step of heterotrophic denitrification is the dissimilatory reduction of nitrate. While particles can be hot spots of denitrification (44), the size fraction analyzed here precludes most particle-associated microorganisms. However, some peptides from nitrate reductase were identified (NarG and NapA) within the ODZs but only in samples from locations with oxygen concentrations less than 5 μmol kg^{-1} (Figs. 3 and 5*A*). Aside from bacterial NirK, other signatures of bacterial denitrification such as nitrite reductase (NirS), nitrous oxide reductase (NosZ)

which reduces N_2O to N_2, or nitric oxide reductase (NorBC) which reduces NO to N_2O were not found in the free-living community using our two-dimensional (2D) metaproteomic analysis (16), although representative sequences were present in the metagenome, indicating the proteins either were not present or were rare and below detection. Additionally, the marker protein for anammox, hydrazine dehydrogenase (Hdh/Hzo), was not identified in the metaproteome or the metagenome. These proteins may have been primarily associated with particles and thus were not identified in the free-living pelagic community, or an exact representative may not have been available in the corresponding environmental metagenome used for peptide identification (47). Additional anaerobic respiratory enzymes were identified in association with the ETNP ODZ. The distribution of heterotrophic dissimilatory iodate reductase (IdrA, formerly called AioA-Like; *SI Appendix*, Fig. S14) (48) followed the pattern of oxygen depletion (Figs. 1*C*, 3, and 5*A*). IdrA was identified in similar abundance to the oxidizing form of dissimilatory sulfite reductase, DsrA (*SI Appendix*, Fig. S15), similar to prior observations of transcripts in Pacific ODZs (49), implying a reliance on iodate metabolism comparable to that of sulfur oxidation in ODZs. Within the ETNP ODZ samples, these two proteins had similar abundances (sum from stations 6 to 8, 29.7 and 37.0 sc$_{corr}$/L$_{sw}$ for IdrA and DsrA,

respectively). Proteins associated with sulfur oxidation (SoxA and the oxidizing form of DsrA) were significantly more abundant below the surface group but did not display a clear relationship with the ODZ regions (Figs. 3 and 4F), although they were negatively correlated with O_2 concentrations (r = −0.32 and −0.53 for SoxA and DsrA, respectively; Fig. 5A).

Some nonrespiratory enzymes also displayed noteworthy relationships with oxygen concentrations across regional scales. The second most abundant oxidoreductase from Nitrospinae, manganese oxidase (MnxG), peaked in abundance in the cline of the ETNP ODZ (Figs. 4G and 5A) and was negatively correlated with O_2 concentrations (r = −0.47). While the Mn-oxidizing function of this specific multicopper oxidase in Nitrospinae has not been experimentally confirmed, the homology of this sequence with known Mn-oxidizing proteins suggests this as a likely function (50), warranting further investigation of this environmentally important enzyme. In the oxygen-depleted north region, Nitrospinae MnxG cooccurred with *Nitrospina* NxrA ($r = 0.64$); however, in the oxygenated south, MnxG peptides did not correlate with NxrA but instead correlated with catalase (KatG; $r = 0.64$). The abundance of KatG in these samples (Fig. 4H) was indicative of a region of high oxidative stress, likely driven by the generation of reactive oxygen species as by-products of ammonia oxidation (51). In keeping with this, archaeal Fe-Mn superoxide dismutase (SOD2) that converts superoxide radicals into H_2O_2, AmoAB, and archaeal NirK all peaked in abundance at station 12, 100 m. While no catalase or peroxiredoxin associated with Thermoproteota were identified, KatG displayed a similar distribution to other ammonia oxidation proteins in the south ($r = 0.73$ with AmoA; Fig. 5A). Aerobic ammonium oxidation may be enhancing oxidative stress in the South Pacific which is then moderated by catalases produced by Bacteria, a commensal process previously demonstrated in laboratory cocultures of Thermoproteota and heterotrophic bacteria (51). Additionally, the abundance of MnxG suggests a coupling of N and Mn cycles catalyzed by Nitrospinae. This coupling may manage oxidative stress (50) by oxidizing reduced Mn and removing it from the water column through Mn-oxide particle formation (52).

Across the entire transect, the most abundant oxidoreductases in the dark mesopelagic—in the twilight and deep groups— were the Fe- and Mo-containing formate dehydrogenase, FdoGH (Figs. 3A and 4I), and Cu- and Mo-containing carbon monoxide oxidoreductase CoxLM (Figs. 3 and 4J). Both FdoGH and CoxLM help sustain microbial energy demands during periods of nutritional or oxidative stress (53, 54). The large Mo-containing subunits FdoG and CoxL were the third and fourth most abundant oxidoreductase proteins in the mesopelagic. The importance of these enzymes to microbial communities at depth was evidenced by their slow attenuations through the water column ($c = -0.46$ and -0.31, for FdoG and CoxL, respectively; Fig. 4A). CoxLM likely supports mixotrophic growth under organic carbon stress at depth: the vast majority of CoxLM-harboring environmental microorganisms are carboxydovores capable of scavenging CO at subatmospheric levels for electrons to support aerobic respiration producing CO_2 when organic carbon is limiting (53). CoxLM proteins, primarily from Bacteria—including from Bacteroidetes, Actinobacteria, Chloroflexi, Alpha- and Gammaproteobacteria, among others—were found throughout the dark mesopelagic and were significantly more abundant in deeper depths compared to the surface (Fig. 3). The importance of CoxLM to mesopelagic communities is consistent with a high genomic capacity for CO oxidation in free-living microbial communities in the bathypelagic (55). Regional variations in

the abundance of CoxLM was observed, where CoxLM was significantly more abundant in the South Pacific (Figs. 3 and 4G). Formate dehydrogenases oxidize another C1 compound, formic acid, to CO_2. Formate dehydrogenases are classified into two families, both of which were observed in the proteome: one uses FeS catalytic subunits, such as FdoGH, and the other, FDH, utilizes NAD(P)+ as electron acceptors (56). FdoGH, primarily associated with Bacteria—including from Alpha- and Gammaproteobacteria, candidate division NC10, *Candidatus* Tectomicrobia, and Actinobacteira among others—peaks in abundance in the cline; however, it does not show a significant regional bias overall. Laboratory studies have shown that FdoGH assists in oxic/anoxic transitions, supporting substrate-level bioenergetic conservation in anaerobic chemoorganotrophic microbial respiration with NO_3^- or NO_2^- (54). In culture, the nitrifier *Nitrospira* was also shown to increase FdoGH protein in response to the onset of oxygen limitation to support cellular energetic requirements (57). In contrast, the NAD+ dependent formate dehydrogenase, FDH, functions optimally in aerobic conditions and has a very different distribution than FdoGH across this dataset as this protein is primarily found in the more productive South Pacific surface group, correlating with total extracted protein ($r = 0.78$; Figs. 1 E and H, 3, and 4I). FDH supports methylotrophy as the final catabolic step in conversion of C1 compounds to CO_2 and can account for 10 to 15% of cellular protein content in methylotrophs (56). Given that the methylotrophy-related FDH is significantly more abundant in the surface upwelling region (Figs. 3 and 4K), the variability in the distributions of FDH and CoxLM demonstrates the use of metaproteomics for constraining microbial production and consumption of gases such as CO that are generally sparsely sampled at depth (58).

Other oxidoreductase enzymes that can support methylotrophic growth also showed a similar regional distribution to FDH. Peptides from the Cu-containing methanethiol oxidase (MtoX), primarily produced by Proteobacteria, were significantly more abundant in the surface and cline of the South Pacific (Fig. 3). The gas methanethiol can be generated through the degradation of sulfur-containing amino acids (59) and is also an intermediate in the biotic degradation of the phytoplankton metabolite dimethylsulfoniopropionate (DMSP) and the volatile dimethylsulfide (DMS), a source of sulfur to the atmosphere and hypothesized contributor to cloud formation (Fig. 3) (60–62). Notably, MtoX peptides were found to positively correlate with eukaryotic RbcL ($r = 0.75$), suggesting a tight relationship between MtoX, which can be used by methylotrophs, and the likely source of the enzyme's substrate: picoeukaryotic phytoplankton. Trimethylamine monooxygenase (Tmm), which catalyzes the oxidation of trimethylamine (TMA) to trimethylamine *N*-oxide (TMAO), can also support methylotrophy (Fig. 3) (63, 64). The Tmm peptides predominantly originated from *Pelagibacter* and were also significantly more abundant within the surface of the South Pacific. Peptides of the methylamine–glutamate *N*-methyltransferase enzyme associated with methylamine oxidation (MgsBC) also peaked in abundance in the South Pacific; however, these were found deeper in the water column ($c = -0.23$ and -0.53 for MgsB and MgsC, respectively) and were also abundant above the ETNP ODZ. Methylamines make up a significant portion of both the volatile and dissolved C and N pools with oxidation of these compounds able to provide an exogenous source of ammonia (65) that can cross feed to other organisms (64).

Nitrification and C1 metabolisms described above are dependent on the activities of other enzymes and members of the

microbial consortium, including enzymes involved in the catabolism of organic matter. Numerous catabolic hydrolases (E.C. class 3) that participate in ammonification had distribution patterns linked to major biogeochemical features and showed distribution patterns where these degradative enzymes incidentally support other members of the microbial consortia, namely, Thermoproteota. The hydrolase formamidase (FmdA), which produces formate and ammonia as by-products, displayed a similar distribution pattern to the formate dehydrogenase FdoGH ($r = 0.71$ with FdoG). Other bacterial ammonia-producing hydrolases like beta-ureidopropionase (PydC) primarily from Actinobacteria and Proteobacteria, amidase (AmiE) primarily from Alphaproteobacteria, and N,N-dimethylformamidase (DmfA) from Alpha- and Gammaproteobacteria were significantly more abundant in the South Pacific near equatorial upwelling. These enzymes were positively correlated with archaeal AmoA ($r = 0.78$, 0.60, and 0.62) and were located close to AmoABC in the neural net (Fig. 5C). Additionally, the ammonifying catabolic enzyme alanine dehydrogenase (Ald; Fig. 4L) from Proteobacteria was one of the most tightly correlated with AmoA ($r = 0.87$). The cooccurrence of these ammonifying enzymes with AmoABC suggests a consortial syntrophic relationship between exogenous ammonia production and archaeal ammonia oxidation (Fig. 5C). Ammonifying hydrogenases also can support ammonia oxidation within the organism that produces them. For example, the hydrolytic Ni-containing urease enzyme (UreC) that releases ammonia from urea produced by Thermoproteota was significantly more abundant in the South Pacific and peaked in abundance in the twilight depth here (Fig. 3). This enzyme supports ammonia oxidation in Thermoproteota and displayed noteworthy taxonomic variability as the UreC of Cyanobacteria, used when Cyanobacteria are nutrient stressed, was significantly more abundant in the surface of the North Pacific associated with the oligotrophic waters of the NPSG. Notably, the Thermoproteotal UreC, which was colocated with the urea symporter (DUR3) in the neural net (Fig. 5C), did not have as strong of a correlation with AmoA ($r = 0.28$ and 0.20 for UreC and DUR3, respectively) as the ammonium transporter AMT ($r = 0.66$), suggesting archaeal use of urea when free ammonia is scarce. Ammonifying hydrolases, in addition to methylamine oxidation, likely contribute to the significant abundance of nitrification enzymes in the surface upwelling region of the South Pacific.

Conclusions

The distribution of microbial proteins across large transects spanning major ecosystems enables comparisons of microbial processes in various biogeochemical provinces, providing a holistic view to investigate microbial function across large geospatial scales. We identified how protein diversity of the microbial community, from both taxonomic and functional capacities, was dominated by a few major protein groups and a multiplicity of lower-abundance groups, while highlighting the interconnectedness among microbial consortia and the biogeochemical cycles of nitrogen, carbon, oxygen, and sulfur. Notably, many of the critical oxidoreductase enzymes are metalloenzymes or utilize metals as substrates (like MnxG), thus also impacting global trace metal cycling. Community and enzyme shifts along vertical scales attest to the rapid growth and reproduction in the euphotic zone and the need to meet energetic demands at depth. Community shifts were also observed along regional horizontal scales associated with varying oxygen concentrations and nutrient availability. The array of detected and quantified enzymes reflects carbon utilization and nitrification pathways including the production and

consumption of volatiles like CO, CO_2, NO, methanethiol, and methylamines. Chemical interactions between members of the microbial consortia were also observed in the neural net, for example, between the nitrite-oxidizing *Nitrospina* and ammonia-oxidizing Thermoproteota, connections between photosynthetic production by Cyanobacteria and methyltrophy, and the catabolism of organics by heterotrophs supporting chemoautotrophic ammonia oxidation by Thermoproteota. Direct measurements of these critical microbial enzymes as the engines of biochemical transformations can provide high-resolution empirical data to refine complex global biogeochemical models (12, 58) and improve our understanding of the ocean in response to a changing climate.

Materials and Methods

Sample Collection. The ProteOMZ research expedition through the central Pacific Ocean occurred in January–February of 2016 aboard the R/V *Falkor* (FK160115; chief scientist M. Saito). Dissolved macronutrient data were collected by discrete water sampling using Niskin bottles on a trace metal rosette at all proteomic sampling stations (Fig. 1G) roughly every 20 m in the surface and cline depth groups and every 100 m from the twilight and deep depth groups. Samples were passed through 0.2-μm filters and then frozen and analyzed as previously described (66) at the Oregon State University Nutrient Autoanalysis Facility (20). Nitrous oxide (N₂O) concentrations were measured at sea by headspace equilibration followed by analysis with a greenhouse gas monitoring gas chromatograph (SRI Instruments) equipped with an electron capture detector, dual HayeSep D packed columns, and a 1-mL sample loop, as previously described (19, 67). Nutrient and hydrographic data are available at the Biological and Chemical Data Management Office (BCO-DMO) repository (https://www.bco-dmo.org/; project no. 685696, datasets 730912 and 775849). POC and particulate organic nitrogen (PON) were collected onto borosilicate glass microfiber filters (Whatman grade GF/F) and processed at the Woods Hole Oceanographic Institution Nutrient Cost Center. Particulate organic matter on filters was combusted between 900 and 1,000 °C, with carbon converted to CO_2 and nitrogen into N_2 gases which column separated and measured on an Elemental Microanalysis Flash EA 1112 in a manner as described in ref. 68. Protein biomass was collected on 142-mm 0.2-μm Supor filters (Pall Corporation) after prefiltration through a 3.0-μm filter. The volume of seawater that passed through the filters was measured via a flow gauge integrated into the pump.

Hydrographic Clustering. Nutrient data did not always line up with protein sampling depths due to sampling circumstances and the McLane pumps being deployed by hanging at discrete depths from a line. Values for parameters at missing depths were extrapolated by conducting 10-m linear interpolations of nutrient data through the water column at each station (Dataset S1). Hydrographic and nutrient parameters (temperature, salinity, oxygen, nitrate, POC, PON, silicate, nitrite, and ammonium) were transformed using a min-max scaling function so that no feature was weighted greater than another. K-means clustering analysis of depth groups was conducted in R. While the major partitioning of depths was among two clusters–euphotic vs. aphotic zones, a well-established pattern in oceanographic domains–in order to gain an understanding of the finer-scale community shifts through a water column, we selected for the next best explanatory cluster number of four depth groups (SI Appendix, Fig. S2). Stations were clustered using a hierarchical clustering analysis in python using SciPy hierarchical clustering (69) with a Euclidian distance matrix and Ward variance minimization (SI Appendix, Fig. S3).

Proteomics: Extraction and Mass Spectrometry. Proteins were extracted from quarter sections of the 142-mm 0.2-μm filters using a modified SP3 magnetic bead method (70) following extraction, purification, and digestion methodology described in ref. 8. Briefly, this method involved use of filters stored at −80 °C, placed into an SDS detergent buffer, and heated at 95 °C for 10 min to lyse cells and solubilize proteinaceous material. Magnetic beads (SpeedBeads, GE Healthcare) were used to purify away detergent and after alkylation and reduction. Purified protein extract was then digested with trypsin (Promega). Protein was quantified after the extraction step and again after the purification

step. An aliquot of 2 μL was used for protein quantification steps, in duplicate, using the bicinchoninic acid method (Thermo Scientific Micro BCA protein assay kit). Absorbance was measured on a Nanodrop ND-1000 spectrophotometer (Thermo Scientific) and compared against a standard curve generated with an albumin standard (Thermo Scientific). An aliquot of 5 μL, at a concentration of 1 μg μL^{-1}, of purified and digested protein extract was injected into an online nanoflow 2D active modulation (2D-AM) liquid chromatography separation following the methods described in ref. 16. Briefly, this method involved a first column separation used a PLRP-S column (200 μm × 159 mm, 3-μm bead size, 300-Å pore size; NanoLCMS Solutions) with a nonlinear 8-h (pH = 10) gradient (10 mM ammonium formate in water and 10 mM ammonium formate in 90% acetonitrile). The eluent then flowed inline onto dual alternative (30 min) column traps (100 μm × 150 mm, 3-μm bead size, 120-Å pore size, C18 Reprosil Gold, Dr. Maisch, packed in a New Objective PicoFrit column), with nonlinear 30-min gradients (0.1% formic acid in water and 0.1% formic acid in 99.9% acetonitrile). Eluent flowed inline into a Thermo Flex ion source attached to the Thermo Fusion quadrupole-Orbitrap mass spectrometer (Thermo Scientific). Scans were set to 240,000 resolution and a 380 to 1,580 mass-to-charge ratio (m/z) window for MS1 scans in the Orbitrap. MS2 scans used a 1.6-m/z window with 50-ms maximum injection times using higher-energy C trap dissociation activation and 5-s dynamic exclusion in the ion trap. Each 2D-AM run took 8 h and resulted in 98 files (1 file per sample) of mass spectra (71).

Proteomics: Informatics. Peptide to spectrum matching (PSM) was conducted with SEQUESTHT within Thermo Proteome Discoverer v. 2.1 software (parent ion tolerance of 10 ppm and fragment tolerance of 0.6 Da). Following best practices in environmental metaproteomics (47, 72), PSM matching was conducted against a database of predicted proteins assembled from a metagenome collected from the ProteOMZ sampling region in 2011 aboard the METZYME expedition (73, 74). Additional proteomics identifications and assignments were conducted with Scaffold v. 4.8.7. Identification was conducted with decoy false discovery rates (FDRs) with a threshold of 95% minimum for peptides (FDR = 0.1%) and a threshold of 99% (1 peptide minimum) for proteins (FDR = 1.6%) (47). Protein level inference for parsimony-based assignments of specific proteins was conducted using experiment-wide grouping with binary peptide–protein weights in Scaffold. The mass spectrometry proteomics data, including sequences of identified proteins, have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the Proteomics Identification Database (PRIDE) partner repository (75) with the dataset identifier PXD030684 and DOI: 10.6019/PXD030684 (71). Peptide and protein abundance data in spectral counts (PSMs), scaled units (sc$_{corr}$/L$_{sw}$), and annotations are available at the BCO-DMO repository under project no. 685696, dataset no. 3868030 (18). Additionally, the complete protein dataset can be explored interactively through the Ocean Protein Portal (https://www.oceanproteinportal.org/) under expedition FK160115. All proteins in Fig. 3 were identified by ≥3 unique peptides (SI Appendix, Table S1) with an example of identified peptides mapped onto a multiple sequence alignment of open reading frames associated with IDH (K00031; SI Appendix, Fig. S16). Metagenome quality analysis and annotations were described in ref. 73, with assembly conducted using metaSPAdes (76). Metagenome sequencing can be found at the Joint Genome Institute (JGI)'s Integrated Microbial Genomes (IMG) database (https://img.jgi.doe.gov/) under JGI Sequencing Project IDs Gp0055157, Gp0055156, Gp0055136, Gp0055135, Gp0055134, and Gp0055133 (74). The metagenome assemblies are available in the IMG database under the Analysis Project IDs Ga0209753, Ga0209752, Ga0209433, Ga0209432, Ga0209228, and Ga0209019. Kegg Orthology classifications and Enzyme Commission numbers were annotated for sequences identified in the proteome using GhostKOALA (77). KOs K00370, K10944, K10945, K10946, K11180, and K08356, which have proteins of multiple different functions in one KO identification, were additionally checked for functional assignment with phylogenetic analyses (SI Appendix, Figs. S10–S15).

LCA analysis of peptides was performed using METATRYP v. 2.0 (22). For LCA analysis, each peptide was assessed separately to identify if the peptide could be found in proteins from multiple taxonomic origins. Any peptides associated with multiple KO functional assignments were also identified and annotated. Peptides in the conserved regions of highly conserved proteins are often shared among multiple taxonomic groups, and thus, a confident taxonomic assignment cannot be made, but rather an LCA taxonomic assignment is provided to the

shared ancestral group for all the possible proteins from which the peptide may have originated. For the most conserved of peptides, the LCA assignment of last universal common ancestor (LUCA) was provided when a peptide was found in protein sequences from multiple taxonomic domains.

Protein abundance data were presented in relative units across this dataset of corrected spectral counts per liter of seawater (sc$_{corr}$/L$_{sw}$). This unit was calculated with the following equation:

$$\frac{\text{Spectral Counts}}{\text{μg protein}_{\text{injected}}} \times \text{Scaling Factor} \times \frac{\text{μg protein}_{\text{total}}}{\text{L seawater filtered}} = \frac{\text{Spectral Counts}}{\text{L}_{\text{seawater}}}.$$

Spectral counts are the total number of exclusive PSMs identified for a peptide sequence. Protein injected is the amount of purified digested protein extract that was injected onto the inline LC setup. The scaling factor was used to reduce the bias in peptide detection between samples (SI Appendix, Fig. S17). The scaling factor is calculated by normalizing the total PSMs per sample in a manner similar to a normalized spectral abundance factor calculation where the sum of PSMs per sample are normalized relative to one another, making the assumption that the same amount of purified protein will yield equivalent numbers of PSMs across the samples. This was done by taking the inverse of the sum of PSMs per sample over the maximum sum of PSMs per sample. The total amount of protein is the abundance of protein extracted from the Supor filter over the volume of seawater that was filtered through the McLane pump head onto the filters.

Statistical analyses of protein abundance for proteins across regions (north vs. south) were performed by conducting Mann–Whitney u tests with the SciPy package. Statistically significant distributions in protein abundance varying by depth group were calculated with a Kruskal–Wallis H test followed by post hoc Dunn's tests, also using SciPy (69). Significance values depicted in Fig. 3 for depth groups were according to the Dunn's tests of comparisons of specific depth groups.

The protein attenuation model was inspired by the Martin curve for decay of POC with depth (78); however, the data and interpretation differ as protein abundance is from the in situ community sampled at discrete depths as opposed to exported sinking material sampled via sediment traps in the Martin et al. (78) method. Attenuation of protein abundance through the microbial communities with depth was calculated by fitting a power law model to protein abundance using the following equation with NumPy (79):

$$f_{\text{protein}}(z) = az^c.$$

The abundance of a protein at a particular depth (z) is calculated as the product of depth (z) and a protein-specific constant (a) to the power of the attenuation rate (c). Only protein data from depths of 100 m or more were considered in the calculation, and proteins must have been identified in at least three separate depths. Positive values of c indicate where the abundance of a protein increases with depth. The more negative the value of c, the faster that protein attenuates with depth.

The unsupervised neural net, a self-organizing map or Kohonen map (80), was conducted by summing abundance of proteins as groups using the taxonomic and KO annotations from the inferred protein group assigned by Scaffold. This protein abundance data were then min-max scaled to evaluate the relative distribution patterns of all proteins compared to one another without more abundant proteins being weighted more heavily than low-abundance proteins. All protein groups analyzed in the neural net were identified by a minimum of 20 PSM matches to remove stochastic noise from very low abundance proteins. Proteins were separated by phylum level, except for Proteobacteria, where Alphaproteobacteria and Gammaproteobacteria were separated as these groups were abundant and display different distribution patterns. The neural net analysis was conducted using the package SimpSOM v. 1.3.4 (81), using 30 × 30 node grid, periodic boundary conditions, PCA initialization of weights, and a 0.05 learning rate for 7,500 epochs (optimized epochs were selected by running incrementally from 10 to 100,000 epochs by 10× and selecting for optimal separation and coherent clustering of proteins) (39) .

Data Availability. Mass spectrometry files have been deposited in ProteomeXchange (PRIDE) (http://proteomecentral.proteomexchange.org, dataset ID PXD030684) (71), and oceanographic nutrient and hydrographic data modeling output have been deposited in BCO-DMO (https://www.bco-dmo.org/, project no. 685696, datasets 730912, 775849, and 868030) (18–20). Results and code for SOM analyses have been deposited in Zenodo (https://zenodo.org/, record no. 7005414) (39).

Author affiliations: [a]Marine Chemistry and Geochemistry Department, Woods Hole Oceanographic Institution, Woods Hole, MA 02543; [b]Genomic Medicine Group, J. C. Venter Institute, La Jolla, CA 92037; [c]Environment and Sustainability Group, J. C. Venter Institute, La Jolla, CA 92037; [d]Synthetic Biology Group, J. C. Venter Institute, La Jolla, CA 92037; [e]Department of Ecology, Evolution and Marine Biology, University of California, Santa Barbara, CA 93106; [f]Department of Biological Sciences, University of Southern California, Los Angeles, CA 90089; [g]Department of Environmental Studies, University of Southern California, Los Angeles, CA 90089; and [h]Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139

1. P. G. Falkowski, T. Fenchel, E. F. Delong, The microbial engines that drive Earth's biogeochemical cycles. *Science* **320**, 1034–1039 (2008).
2. K. Bergauer *et al.*, Organic matter processing by microbial communities throughout the Atlantic water column as revealed by metaproteomics. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E400–E408 (2018).
3. M. P. Mikan *et al.*, Metaproteomics reveal that rapid perturbations in organic matter prioritize functional restructuring over taxonomy in western Arctic Ocean microbiomes. *ISME J.* **14**, 39–52 (2020).
4. J. S. P. McCain *et al.*, Cellular costs underpin micronutrient limitation in phytoplankton. *Sci. Adv.* **7**, eabg6501 (2021).
5. R. M. Morris *et al.*, Comparative metaproteomics reveals ocean-scale shifts in microbial nutrient utilization and energy transduction. *ISME J.* **4**, 673–685 (2010).
6. E. K. Moore, B. L. Nunn, D. R. Goodlett, H. R. Harvey, Identifying and tracking proteins through the marine water column: Insights into the inputs and preservation mechanisms of protein in sediments. *Geochim. Cosmochim. Acta* **83**, 324–359 (2012).
7. M. A. Saito *et al.*, Multiple nutrient stresses at intersecting Pacific Ocean biomes detected by protein biomarkers. *Science* **345**, 1173–1177 (2014).
8. M. A. Saito *et al.*, Abundant nitrite-oxidizing metalloenzymes in the mesopelagic zone of the tropical Pacific Ocean. *Nat. Geosci.* **13**, 355–362 (2020).
9. N. R. Cohen *et al.*, Dinoflagellates alter their carbon and nutrient metabolic strategies across environmental gradients in the central Pacific Ocean. *Nat. Microbiol.* **6**, 173–186 (2021).
10. A. K. Hawley, H. M. Brewer, A. D. Norbeck, V. Paša-Tolić, S. J. Hallam, Metaproteomics reveals differential modes of metabolic coupling among ubiquitous oxygen minimum zone microbes. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 11395–11400 (2014).
11. N. G. Walworth *et al.*, Why environmental biomarkers work: Transcriptome-proteome correlations and modeling of multistressor experiments in the marine bacterium *Trichodesmium*. *J. Proteome Res.* **21**, 77–89 (2022).
12. E. J. Zakem, M. F. Polz, M. J. Follows, Redox-informed models of global biogeochemical cycles. *Nat. Commun.* **11**, 5680 (2020).
13. T. S. Weber, C. Deutsch, Ocean nutrient ratios governed by plankton biogeography. *Nature* **467**, 550–554 (2010).
14. M. J. Follows, S. Dutkiewicz, S. Grant, S. W. Chisholm, Emergent biogeography of microbial communities in a model ocean. *Science* **315**, 1843–1846 (2007).
15. S. A. Henson, B. B. Cael, S. R. Allen, S. Dutkiewicz, Future phytoplankton diversity in a changing climate. *Nat. Commun.* **12**, 5372 (2021).
16. M. R. McIlvin, M. A. Saito, Online nanoflow two-dimension comprehensive active modulation reversed phase-reversed phase liquid chromatography high-resolution mass spectrometry for metaproteomics of environmental and microbiome samples. *J. Proteome Res.* **20**, 4589–4597 (2021).
17. M. A. Saito *et al.*, Development of an ocean protein portal for interactive discovery and education. *J. Proteome Res.* **20**, 326–336 (2020).
18. J. K. Saunders, M. R. McIlvin, M. A. Saito, ProteOMZ exclusive peptide level spectral counts. Biological and Chemical Oceanography Data Management Office. https://doi.org/10.26008/1912/bco-dmo.868030.1. Deposited 13 January 2022.
19. S. M. Laperriere, M. A. Saito, A. E. Santoro, ProteOMZ nitrous oxide data. Biological and Chemical Oceanography Data Management Office. https://doi.org/10.26008/1912/bco-dmo.775849.1. Deposited 8 June 2022.
20. A. E. Santoro, M. A. Saito, ProteOMZ nutrient, CTD, and oxygen data. Biological and Chemical Oceanography Data Management Office. https://www.bco-dmo.org/dataset/730912. Deposited 19 November 2018.
21. D. A. Hansell, C. A. Carlson, R. Schlitzer, Net removal of major marine dissolved organic carbon fractions in the subsurface ocean. *Global Biogeochem. Cycles* **26**, 1016 (2012).
22. J. K. Saunders *et al.*, METATRYP v 2.0: Metaproteomic least common ancestor analysis for taxonomic inference using specialized sequence assemblies-standalone software and web servers for marine microorganisms and coronaviruses. *J. Proteome Res.* **19**, 4718–4729 (2020).
23. D. B. Rusch *et al.*, The Sorcerer II global ocean sampling expedition: Northwest Atlantic through eastern tropical Pacific. *PLoS Biol.* **5**, e77 (2007).
24. S. Pesant *et al.*; Tara Oceans Consortium Coordinators, Open science resources for the discovery and analysis of Tara Oceans data. *Sci. Data* **2**, 150023 (2015).
25. S. Sunagawa *et al.*; Tara Oceans Consortium Coordinators, Ocean plankton. Structure and function of the global ocean microbiome. *Science* **348**, 1261359 (2015).
26. D. M. Karl, M. J. Church, Microbial oceanography and the Hawaii Ocean Time-series programme. *Nat. Rev. Microbiol.* **12**, 699–713 (2014).
27. F. O. Aylward *et al.*, Microbial community transcriptional networks are conserved in three domains at ocean basin scales. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 5443–5448 (2015).
28. A. E. Santoro *et al.*, Genomic and proteomic characterization of "Candidatus Nitrosopelagicus brevis": An ammonia-oxidizing archaeon from the open ocean. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 1173–1178 (2015).
29. M. B. Karner, E. F. DeLong, D. M. Karl, Archaeal dominance in the mesopelagic zone of the Pacific Ocean. *Nature* **409**, 507–510 (2001).
30. C. A. Francis, K. J. Roberts, J. M. Beman, A. E. Santoro, B. B. Oakley, Ubiquity and diversity of ammonia-oxidizing archaea in water columns and sediments of the ocean. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 14683–14688 (2005).
31. G. Salazar *et al.*; Tara Oceans Coordinators, Gene expression changes and community turnover differentially shape the global ocean metatranscriptome. *Cell* **179**, 1068–1083.e21 (2019).
32. M. G. Mazzotta *et al.*, Characterization of the metalloproteome of Pseudoalteromonas (BB2-AT2): Biogeochemical underpinnings for zinc, manganese, cobalt, and nickel cycling in a ubiquitous marine heterotroph. *Metallomics* **13**, mfab060 (2021).
33. K. Izui, H. Matsumura, T. Furumoto, Y. Kai, Phosphoenolpyruvate carboxylase: A new era of structural biology. *Annu. Rev. Plant Biol.* **55**, 69–84 (2004).
34. C. Yang, Q. Hua, K. Shimizu, Metabolic flux analysis in Synechocystis using isotope distribution from 13C-labeled glucose. *Metab. Eng.* **4**, 202–216 (2002).
35. S. M. Sowell *et al.*, Transport functions dominate the SAR11 metaproteome at low-nutrient extremes in the Sargasso Sea. *ISME J.* **3**, 93–105 (2009).
36. E. K. Moore, B. I. Jelen, D. Giovannelli, H. Raanan, P. G. Falkowski, Metal availability and the expanding network of microbial metabolisms in the Archaean eon. *Nat. Geosci.* **10**, 629–636 (2017).
37. W. Martens-Habbena *et al.*, The production of nitric oxide by marine ammonia-oxidizing archaea and inhibition of archaeal ammonia oxidation by a nitric oxide scavenger. *Environ. Microbiol.* **17**, 2261–2274 (2015).
38. C. B. Walker *et al.*, Nitrosopumilus maritimus genome reveals unique mechanisms for nitrification and autotrophy in globally distributed marine crenarchaea. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 8818–8823 (2010).
39. J. K. Saunders, M. A. Saito, SOM results for Saunders et al., 2022 "Microbial functional diversity across biogeochemical provinces in the central Pacific Ocean." Zenodo. https://doi.org/10.5281/zenodo.7005414. Deposited 18 August 2022.
40. W. Qin *et al.*, Influence of oxygen availability on the activities of ammonia-oxidizing archaea. *Environ. Microbiol. Rep.* **9**, 250–256 (2017).
41. S. Kobayashi *et al.*, Nitric oxide production from nitrite reduction and hydroxylamine oxidation by copper-containing dissimilatory nitrite reductase (NirK) from the aerobic ammonia-oxidizing archaeon, Nitrososphaera viennensis. *Microbes Environ.* **33**, 428–434 (2018).
42. Q. Ji, A. R. Babbin, A. Jayakumar, S. Oleynik, B. B. Ward, Nitrous oxide production by nitrification and denitrification in the Eastern Tropical South Pacific oxygen minimum zone. *Geophys. Res. Lett.* **42**, 10755–10764 (2015).
43. A. E. Santoro *et al.*, Nitrification and nitrous oxide production in the offshore waters of the eastern tropical South Pacific. *Global Biogeochem. Cycles* **35**, e2020GB006716 (2021).
44. J. J. Wright, K. M. Konwar, S. J. Hallam, Microbial ecology of expanding oxygen minimum zones. *Nat. Rev. Microbiol.* **10**, 381–394 (2012).
45. J. M. Beman *et al.*, Substantial oxygen consumption by aerobic nitrite oxidation in oceanic oxygen minimum zones. *Nat. Commun.* **12**, 7043 (2021).
46. L. A. Bristow *et al.*, Ammonium and nitrite oxidation at nanomolar oxygen concentrations in oxygen minimum zone waters. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 10601–10606 (2016).
47. M. A. Saito *et al.*, Progress and challenges in ocean metaproteomics and proposed best practices for data sharing. *J. Proteome Res.* **18**, 1461–1476 (2019).
48. C. Yamazaki *et al.*, A novel dimethylsulfoxide reductase family of molybdenum enzyme, Idr, is involved in iodate respiration by Pseudomonas sp. SCT. *Environ. Microbiol.* **22**, 2196–2212 (2020).
49. J. K. Saunders, C. A. Fuchsman, C. McKay, G. Rocap, Complete arsenic-based respiratory cycle in the marine microbial communities of pelagic oxygen-deficient zones. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 9925–9930 (2019).
50. S. Luecker, B. Nowka, T. Rattei, E. Spieck, H. Daims, The genome of Nitrospina gracilis illuminates the metabolism and evolution of the major marine nitrite oxidizer. *Front Microbiol.* **4**, 27 (2013).
51. J.-G. Kim *et al.*, Hydrogen peroxide detoxification is a key mechanism for growth of ammonia-oxidizing archaea. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 7888–7893 (2016).
52. C. N. Butterfield, A. V. Soldatova, S.-W. Lee, T. G. Spiro, B. M. Tebo, Mn(II,III) oxidation and MnO$_2$ mineralization by an expressed bacterial multicopper oxidase. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 11731–11735 (2013).
53. P. R. F. Cordero *et al.*, Atmospheric carbon monoxide oxidation is a widespread mechanism supporting microbial survival. *ISME J.* **13**, 2868–2881 (2019).
54. H. Abaibou, J. Pommier, S. Benoit, G. Giordano, M. A. Mandrand-Berthelot, Expression and characterization of the Escherichia coli fdo locus and a possible physiological role for aerobic formate dehydrogenase. *J. Bacteriol.* **177**, 7141–7149 (1995).
55. S. G. Acinas *et al.*, Deep ocean metagenomes provide insight into the metabolic architecture of bathypelagic microbial communities. *Commun. Biol.* **4**, 604 (2021).
56. V. O. Popov, V. S. Lamzin, NAD(+)-dependent formate dehydrogenase. *Biochem. J.* **301**, 625–643 (1994).
57. B. Bayer *et al.*, Metabolic versatility of the nitrite-oxidizing bacterium Nitrospira marina and its proteomic response to oxygen-limited conditions. *ISME J.* **15**, 1025–1039 (2020).

58. L. Conte, S. Szopa, R. Séférian, L. Bopp, The oceanic cycle of carbon monoxide and its emissions to the atmosphere. *Biogeosciences* **16**, 881–902 (2019).

59. Ö. Eyice *et al.*, Bacterial SBP56 identified as a Cu-dependent methanethiol oxidase widely distributed in the biosphere. *ISME J.* **12**, 145–160 (2018).

60. M. Rinaldi *et al.*, Primary and secondary organic marine aerosol and oceanic biological activity: Recent results and new perspectives for future studies. *Adv. Meteorol.* **2010**, 310682 (2010).

61. P. K. Quinn, T. S. Bates, The case against climate regulation via oceanic phytoplankton sulphur emissions. *Nature* **480**, 51–56 (2011).

62. K. J. Mayer *et al.*, Secondary marine aerosol plays a dominant role over primary sea spray aerosol in cloud formation. *ACS Cent. Sci.* **6**, 2259–2266 (2020).

63. Y. Chen, N. A. Patel, A. Crombie, J. H. Scrivens, J. C. Murrell, Bacterial flavin-containing monooxygenase is trimethylamine monooxygenase. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 17791–17796 (2011).

64. I. D. E. A. Lidbury, J. C. Murrell, Y. Chen, Trimethylamine and trimethylamine N-oxide are supplementary energy sources for a marine heterotrophic bacterium: Implications for marine carbon and nitrogen cycling. *ISME J.* **9**, 760–769 (2015).

65. D. D. Nayak, C. J. Marx, Methylamine utilization via the *N*-methylglutamate pathway in Methylobacterium extorquens PA1 involves a novel flow of carbon through C_1 assimilation and dissimilation pathways. *J. Bacteriol.* **196**, 4130–4139 (2014).

66. A. E. Noble *et al.*, Basin-scale inputs of cobalt, iron, and manganese from the Benguela-Angola front to the South Atlantic Ocean. *Limnol. Oceanogr.* **57**, 989–1010 (2012).

67. S. M. Laperriere *et al.*, Nitrification and nitrous oxide dynamics in the Southern California Bight. *Limnol. Oceanogr.* **66**, 1099–1112 (2021).

68. C. F. Zimmerman, C. W. Keefe, J. Bashe, "Method 440.0 determination of carbon and nitrogen in sediments and particulates of estuarine/coastal waters using elemental analysis" (Document EPA/600/R-15/009, US Environmental Protection Agency, Washington, DC, 1997).

69. P. Virtanen *et al.*; SciPy 1.0 Contributors, SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).

70. C. S. Hughes *et al.*, Ultrasensitive proteome analysis using paramagnetic bead technology. *Mol. Syst. Biol.* **10**, 757 (2014).

71. J. K. Saunders, M. R. McIlvin, M. A. Saito, The ProteOMZ Expedition: Investigating life without oxygen in the Pacific Ocean. Proteomics Identifications Database. http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PXD030684. Deposited 31 December 2021.

72. E. Timmins-Schiffman *et al.*, Critical decisions in metaproteomics: Achieving high confidence protein annotations in a sea of unknowns. *ISME J.* **11**, 309–314 (2017).

73. C. L. Dupont *et al.*, Genomes and gene expression across light and productivity gradients in eastern subtropical Pacific microbial communities. *ISME J.* **9**, 1076–1092 (2015).

74. A. E. Santoro *et al.*, Thaumarchaeal ecotype distributions across the equatorial Pacific Ocean and their potential roles in nitrification and sinking flux attenuation. *Limnol. Oceanogr.* **62**, 1984–2003 (2017).

75. Y. Perez-Riverol *et al.*, The PRIDE database and related tools and resources in 2019: Improving support for quantification data. *Nucleic Acids Res.* **47** (D1), D442–D450 (2019).

76. S. Nurk, D. Meleshko, A. Korobeynikov, P. A. Pevzner, metaSPAdes: A new versatile metagenomic assembler. *Genome Res.* **27**, 824–834 (2017).

77. M. Kanehisa, Y. Sato, K. Morishima, BlastKOALA and GhostKOALA: KEGG Tools for functional characterization of genome and metagenome sequences. *J. Mol. Biol.* **428**, 726–731 (2016).

78. J. H. Martin, G. A. Knauer, D. M. Karl, W. W. Broenkow, VERTEX: Carbon cycling in the northeast Pacific. *Deep-Sea Res. A, Oceanogr. Res. Pap.* **34**, 267–285 (1987).

79. C. R. Harris *et al.*, Array programming with NumPy. *Nature* **585**, 357–362 (2020).

80. T. Kohonen, The self-organizing map. *Proc. IEEE* **78**, 1464–1480 (1990).

81. F. Comitani, SimpSOM. Zenodo. https://doi.org/10.5281/zenodo.2621559. Accessed 7 December 2021.