



Microbial and Viral Genome and Proteome Nitrogen Demand Varies across Multiple Spatial Scales within a Marine Oxygen Minimum Zone

 Daniel Muratore,^{a,b,c}  Anthony D. Bertagnolli,^d  Laura A. Bristow,^e  Bo Thamdrup,^e  Joshua S. Weitz,^{b,f,g}  Frank J. Stewart^{d,h}

^aGeorgia Institute of Technology, Interdisciplinary Program in Quantitative Biosciences, School of Biological Sciences, Atlanta, Georgia, USA

^bGeorgia Institute of Technology, School of Biological Sciences, Atlanta, Georgia, USA

^cSanta Fe Institute, Santa Fe, New Mexico, USA

^dMontana State University, Department of Microbiology and Cell Biology, Bozeman, Montana, USA

^eDepartment of Marine Sciences, University of Gothenburg, Göteborg, Sweden

^fGeorgia Institute of Technology, School of Physics, Atlanta, Georgia, USA

^gÉcole Normale Supérieure, Institut de Biologie, Paris, France

^hGeorgia Institute of Technology, Center for Microbial Dynamics and Infection, Atlanta, Georgia, USA

ABSTRACT Nutrient availability can significantly influence microbial genomic and proteomic streamlining, for example, by selecting for lower nitrogen to carbon ratios. Oligotrophic open ocean microbes have streamlined genomic nitrogen requirements relative to those of their counterparts in nutrient-rich coastal waters. However, steep gradients in nutrient availability occur at meter-level, and even micron-level, spatial scales. It is unclear whether such gradients also structure genomic and proteomic stoichiometry. Focusing on the eastern tropical North Pacific oxygen minimum zone (OMZ), we use comparative metagenomics to examine how nitrogen availability shapes microbial and viral genome properties along the vertical gradient across the OMZ and between two size fractions, distinguishing free-living microbes versus particle-associated microbes. We find a substantial increase in the nitrogen content of encoded proteins in particle-associated over free-living bacteria and archaea across nitrogen availability regimes over depth. Within each size fraction, we find that bacterial and viral genomic nitrogen tends to increase with increasing nitrate concentrations with depth. In contrast to cellular genes, the nitrogen content of virus proteins does not differ between size fractions. We identified arginine as a key amino acid in the modulation of the C:N ratios of core genes for bacteria, archaea, and viruses. Functional analysis reveals that particle-associated bacterial metagenomes are enriched for genes that are involved in arginine metabolism and organic nitrogen compound catabolism. Our results are consistent with nitrogen streamlining in both cellular and viral genomes on spatial scales of meters to microns. These effects are similar in magnitude to those previously reported across scales of thousands of kilometers.

IMPORTANCE The genomes of marine microbes can be shaped by nutrient cycles, with ocean-scale gradients in nitrogen availability being known to influence microbial amino acid usage. It is unclear, however, how genomic properties are shaped by nutrient changes over much smaller spatial scales, for example, along the vertical transition into oxygen minimum zones (OMZs) or from the exterior to the interior of detrital particles. Here, we measure protein nitrogen usage by marine bacteria, archaea, and viruses by using metagenomes from the nitracline of the eastern tropical North Pacific OMZ, including both particle-associated and nonassociated biomass. Our results show higher genomic and proteomic nitrogen content in particle-associated microbes and at depths with higher nitrogen availability for cellular and viral genomes. This discovery suggests that stoichiometry influences microbial and viral evolution across multiple

Editor Jean-Baptiste Raina, University of Technology Sydney

Copyright © 2023 Muratore et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Daniel Muratore, dmuratore@santafe.edu.

The authors declare no conflict of interest.

Received 9 November 2022

Accepted 19 January 2023

Published 15 March 2023

scales, including the micrometer to millimeter scale associated with particle-associated versus free-living lifestyles.

KEYWORDS stoichiometry, comparative metagenomics, microbial genome evolution, viral genome evolution, marine particles

A diverse consortium of marine microorganisms drive global biogeochemistry (1) and form the foundation of the marine food web. Major efforts have been undertaken to characterize the diversity and biogeographical patterns of marine microorganisms (2, 3) as well as their potential and realized metabolic output (4, 5). Emergent differences in the bio-availability of nutrients that are shaped by microbial activity in turn shape the evolution of microbial genomes and metabolism. Nutrient availability is an important mechanism in the evolution of gene content (6, 7), and it has been implicated as the driver of observed patterns of genomic streamlining for bacterioplankton living in oligotrophic marine waters (8). Reduced nitrogen availability is thought to select for genomes with lower guanine-cytosine (GC) content as well as for proteins with lower nitrogen to carbon ratios (9). More generally, the study of “stochio-genomics” addresses how resource conditions can influence microbial genomic and proteomic elemental composition (7). For example, a survey of the North Pacific Subtropical Gyre at Station ALOHA (A long-term oligotrophic habitat assessment) found increasing genomic lengths, GC content, and nitrogen utilization in amino acid side chains, from the mixed layer to the mesopelagic, corresponding to increasing nitrogen availability (10). Global marine metagenomes exhibit increased nitrogen content in arginine synthesis genes when found along the coasts, as opposed to in the open ocean, where nitrogen is presumed to be more limiting (6). On a global scale, marine metagenomic nitrogen content appears to correlate to nitrate concentrations (11). Studies linking nutrient availability to genomic streamlining have tended to focus on macro-scale patterns (e.g., spanning kilometers of depth changes or tens to thousands of kilometers between the coast and open water) (7). A fertilization experiment that simulated a large change in nutrient availability also showed an increased average GC content and genome length of the associated microbial community (12).

Steep chemical gradients can occur along much smaller spatial scales. Micron-scale chemical heterogeneity imposes stark ecological transitions from the frame of reference of microorganisms (13). In marine environments, small particles, such as sinking dust and phytoplankton cells, constitute pockets of high availability and diversity of substrates for microbial metabolism (14). Taxonomic surveys of estuarine (15), coastal (16), and offshore (17) environments show that free-living microbial communities are taxonomically distinct from those on particles. Across these studies, particle-associated fractions are generally enriched in Gammaproteobacteria, such as *Vibrio* and *Pseudoalteromonas*, whereas free-living communities contain more Alphaproteobacteria, such as the ubiquitous SAR11 group (15, 17). A study at Station ALOHA found that metagenome-assembled genomes (MAGs) that were associated with sinking particles were longer, had higher GC content, and had predicted proteins with higher N-usage than those of MAGs from free-living communities at the same depth (18).

OMZs exhibit large differences in nutrient availability on small spatial scales, suggesting that they are model ecosystems for exploring genomic and proteomic streamlining. OMZs form in areas of enhanced nutrient loading, typically from upwelling (19). In these regions, nutrients fuel high surface primary production (20). The respiration of sinking organic matter by heterotrophs, coupled with the limited vertical mixing, drives oxygen depletion at midwater depths, often to levels below detection (a few nM dissolved O₂ [20, 21]). These dynamics result in stark vertical transitions between an oxic photic layer and a functionally anoxic, subphotoc OMZ layer. This transition is defined by several important and co-occurring chemical regime shifts. For example, organic carbon may become limiting for denitrification among heterotrophic bacteria in the suboxic zone (22). The transition from aerobic respiration to denitrification is also slightly less energy-yielding, changing the bioenergetics in this ecosystem. Critically, in

the suboxic layer, nitrite accumulates as nitrate becomes the dominant oxidant for microbial respiration (23). The OMZ environment is ideal for testing stochio-genomic questions, given that vertical gradients in bioavailable nitrogen and other key water chemistry parameters (e.g., total dissolved organic matter) are steeper in OMZ regions than in almost any other ocean environment (24, 25).

Despite evidence showing taxonomic and functional differences in particle-associated versus free-living marine microorganisms, such as the enrichment of copiotrophic gammaproteobacteria and the increased prevalence of carbohydrate-degrading enzymes (2, 18, 26), there are limited data on how such differences affect stoichiogenomic properties (18). Based on prior research focused on macroscale spatial gradients, we hypothesize that stoichiogenomic properties are also shaped by microscale resource variation between particle-associated versus nonassociated niches. Specifically, we hypothesize that, compared to particle-associated microbes, free-living microbes exhibit genome-wide reductions in the (i) GC content and (ii) nitrogen content of encoded amino acids. We also hypothesize that these stoichiogenomic differences occur regardless of depth in the water column. That is, the genomic effects of resource heterogeneity (in this case, nitrogen) between particle and nonparticle niches persist alongside changes in dissolved nitrate + nitrite concentrations with depth.

In this study, we analyzed 58 metagenomes sampled in 2013 and 2014 from the world's largest OMZ, which is located in the eastern tropical North Pacific (ETNP) (27). These metagenomes span the oxic surface, anoxic OMZ, and suboxic upper mesopelagic below the OMZ, as well as both particulate ($>1.6\ \mu\text{m}$) and free-living (0.2 to $1.6\ \mu\text{m}$) size fractions. We investigated drivers of the metagenomic GC content as well as the stoichiometric properties of bacterial, archaeal, and viral genes across depths and between size classes. We found trends in the metagenome-level and gene-level GC content and amino acid nitrogen contents that correspond to nitrate + nitrite concentrations, and these are similar to those seen in non-OMZ environments (10). We also found higher GC content and amino acid N contents in particle-associated metagenomes for bacterial and archaeal genes. Viral genes do not exhibit stoichiogenomic differences between size fractions, although viral genes increase in nitrogen content with depth, similarly to bacteria. We selected an example core gene for bacteria, archaea, and bacteriophages, and we found patterns in amino acid composition that drive differences in bulk genetic and proteomic C:N ratios. For bacteria, functional analysis identified the enrichment of genes involved in the synthesis of arginine, one of the amino acid drivers of bacterial core gene nitrogen content, on particle-associated samples in which nitrogen remineralization and availability may be high. These results suggest that heterogeneity in nitrogen availability on the microscopic scale in the vicinity of particles and on the meters scale of the oxycline in an OMZ can have as substantial an impact on microbial and viral genome evolution as does large-scale variability across ocean environments (e.g., pelagic versus coastal) or ocean basins.

RESULTS

Biogeochemical gradients in the ETNP OMZ. Metagenomic samples were taken from five stations in the region of the ETNP OMZ on two cruises in June 2013 and May 2014 (Fig. 1A). Samples were serially filtered onto particle-associated ($>1.6\ \mu\text{m}$) and free-living ($>0.2\ \mu\text{m}$) size fractions across depths ranging from the surface to 2,600 m (see Materials and Methods). Oxygen saturation levels (see Materials and Methods) were between 95 and 101% for all stations in the surface mixed layer. The oxic-anoxic interface increases with depth offshore, but it occurs in the top 100 m of the water column, with an average depth of approximately 80 m (Fig. 1B contains a representative profile). Dissolved nutrient profiles had canonical OMZ characteristics (23) (Fig. 1B and C). Using data from station 6 as an example (Fig. 1C), in the surface mixed layer, nitrite concentrations were undetected, and nitrate concentrations were below $1\ \mu\text{M}$. Between the base of the mixed layer (20 m) and the oxic-anoxic interface (80 m), nitrate concentrations increase to $20\ \mu\text{M}$. Nitrite concentrations decline after peaking at $6.06\ \mu\text{M}$ at 125 m, whereas nitrate concentrations continue to increase to above $40\ \mu\text{M}$ at 1,000 m.

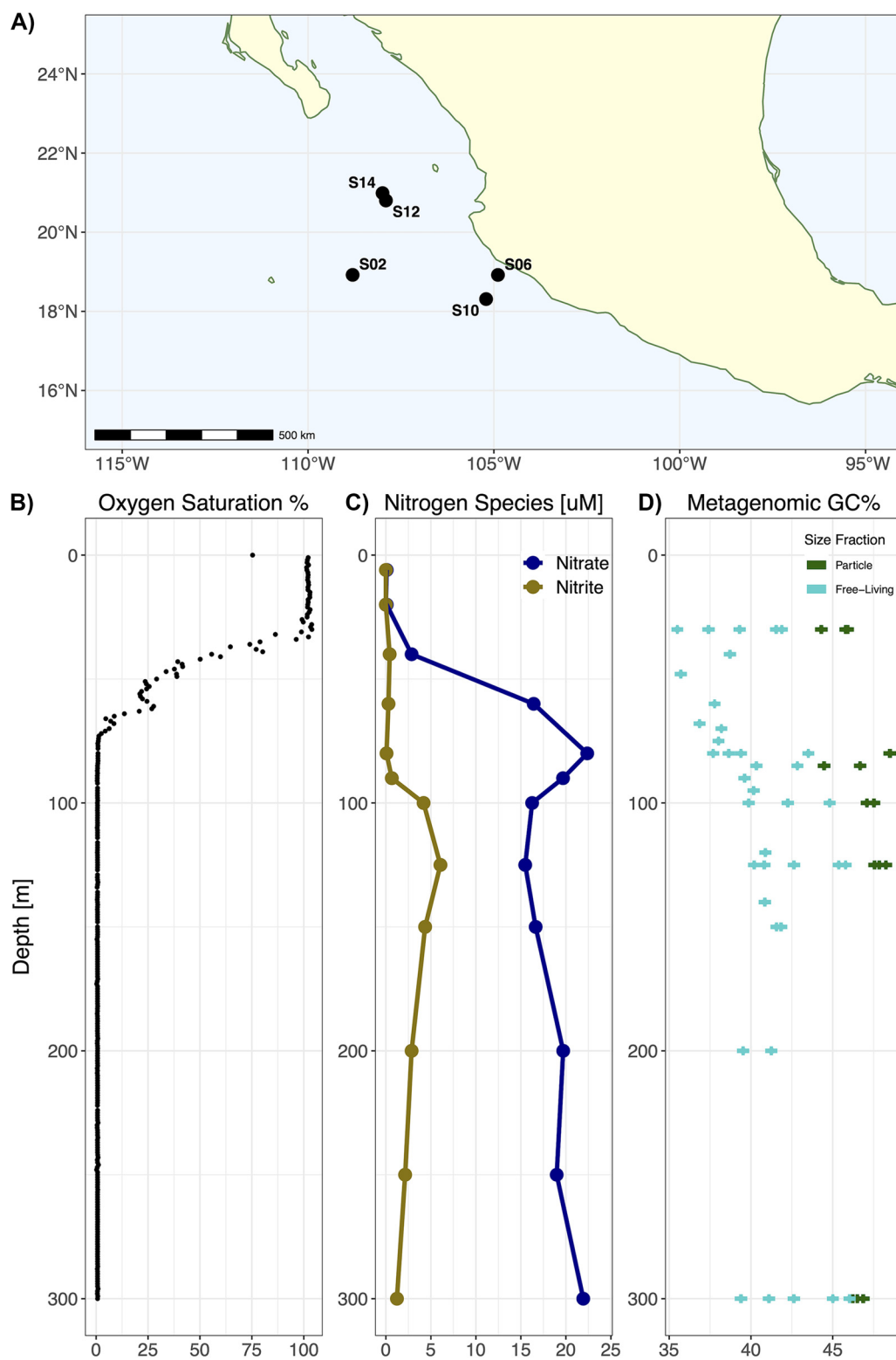


FIG 1 Environmental contextual data and sampling locations for this study. (A) Map showing sampling station locations for the 2013 and 2014 field collection efforts. (B) Representative oxygen saturation profile from Station 6 in the upper 300 m. (C) Dissolved nitrate and nitrite profile for Station 6, measured in 2013. (D) Average GC content of the metagenomic reads from all of the metagenomic samples that were taken from the upper 300 m, representing a $>1.6 \mu\text{m}$ particle-associated size fraction and a planktonic size fraction of $>0.2 \mu\text{m}$.

Genomic streamlining shaped by nitrogen availability at small spatial scales.

We analyzed the GC content of metagenomic reads to evaluate the potential for genomic streamlining, given the dissolved nitrate + nitrite differences across the water column and the putative differences in nitrogen availability between size fractions. For both size fractions and across all samples, the average GC content of the metagenomic reads (see Materials and Methods) is lowest in the oxycline, increases in alignment with nitrate concentrations in the anoxic OMZ core until approximately 150 m and remains roughly constant at lower depths (Fig. 1D). While this trend was consistent between size fractions, particle-associated metagenomes had consistently higher GC content at all depths, compared to free-living fraction metagenomes. Using a mixed effects model, controlling for sequencing depth and spatial correlation in depth, we identified a significant effect of a 5.32% ($\pm 0.587\%$) increase in metagenomic GC content between size fractions (see Materials and Methods; model parameters are described in Table S1).

We investigated 16S rRNA genes that were extracted from the metagenomes to explore the relationship between the changing GC content and the taxonomic composition (see Materials and Methods). The taxonomic patterns are summarized in Fig. S1 and match those reported in previous 16S-based surveys of the ETNP OMZ (2, 28). Key trends include a predominance of SAR11 Alphaproteobacteria at all depths throughout the water column, an abundance of nitrifying Thaumarchaeota along the oxycline, and an enrichment of Deltaproteobacteria and Gammaproteobacteria in particle-associated communities (Fig. S1). Notably, samples from the particle-associated fraction contained reads matching the mitochondria of picoeukaryotes, primarily obligate endosymbiotic dinoflagellates from the Syndiniales group. The eukaryotic taxa that we observed are consistent with 18S rRNA gene-based surveys in the ETNP OMZ (29).

Spatial structure in stoichiogenomic properties vary between bacteria, archaea, and viruses. Next, we investigated the stoichiometric properties of nucleotide and amino acid sequences of microbial and viral genes throughout the water column and between size fractions. Metagenomes were assembled, and we generated a gene catalogue of 718,947 gene clusters from the assembled contigs (see Materials and Methods). We annotated 713,019/718,947 (99.1%) of the gene clusters by using the KEGG Orthology database (30). Annotated genes were divided between bacterial, archaeal, and viral genes for further analysis, based on the KEGG Orthology search taxonomic assignments.

We hypothesized that the nitrogen content of the amino acid sequences of predicted genes would follow trends similar to those of the bulk metagenomic GC content. For each predicted gene, we calculated stoichiogenomic properties, including the GC content, the abundance of each amino acid in the encoded protein, the N:C ratios of the amino acid side chains in the sequence, and the codon usage bias (10). We generated “stoichiogenomic profiles” for bacterial, archaeal, and viral gene data sets via a coverage-weighted average of these properties over all of the genes within a sample. For each data set, we conducted a redundancy analysis (RDA), controlling for variable sequencing depth, to identify the “domain” (bacteria, archaea, or virus)-specific drivers of the covariation in the stoichiogenomic properties (Fig. 2). We modeled the relative explanatory effects of particle fraction and depth by using a permutational analysis of variance (see Materials and Methods).

The results of our ordination analysis are presented in Fig. 2. The bacterial stoichiogenomic profiles (Fig. 2A) split into two groups along the first RDA axis (64.46% of the total variance). Positive values are associated with communities from the surface mixed layer (yellow points), and negative values are associated with OMZ and mesopelagic samples (blue points). The second axis (20.77%) separated free-living samples (triangles) with high values from particle-associated samples (circles) with lower values. The estimated R^2 value for size fraction is 14.0% ($P < 0.001$) whereas the estimated R^2 for sample depth is 22.1% ($P < 0.001$). For archaeal profiles (Fig. 2B), the first axis (49.43% of the variance) separated the particle-associated samples at low values mostly from the mesopelagic free-living samples at high values. The second axis (22.91%) places the OMZ samples at higher values and the surface mixed layer samples at lower values. The estimated R^2 values for size fraction and depth are 14.1% ($P < 0.001$) and 13.1% ($P < 0.001$).

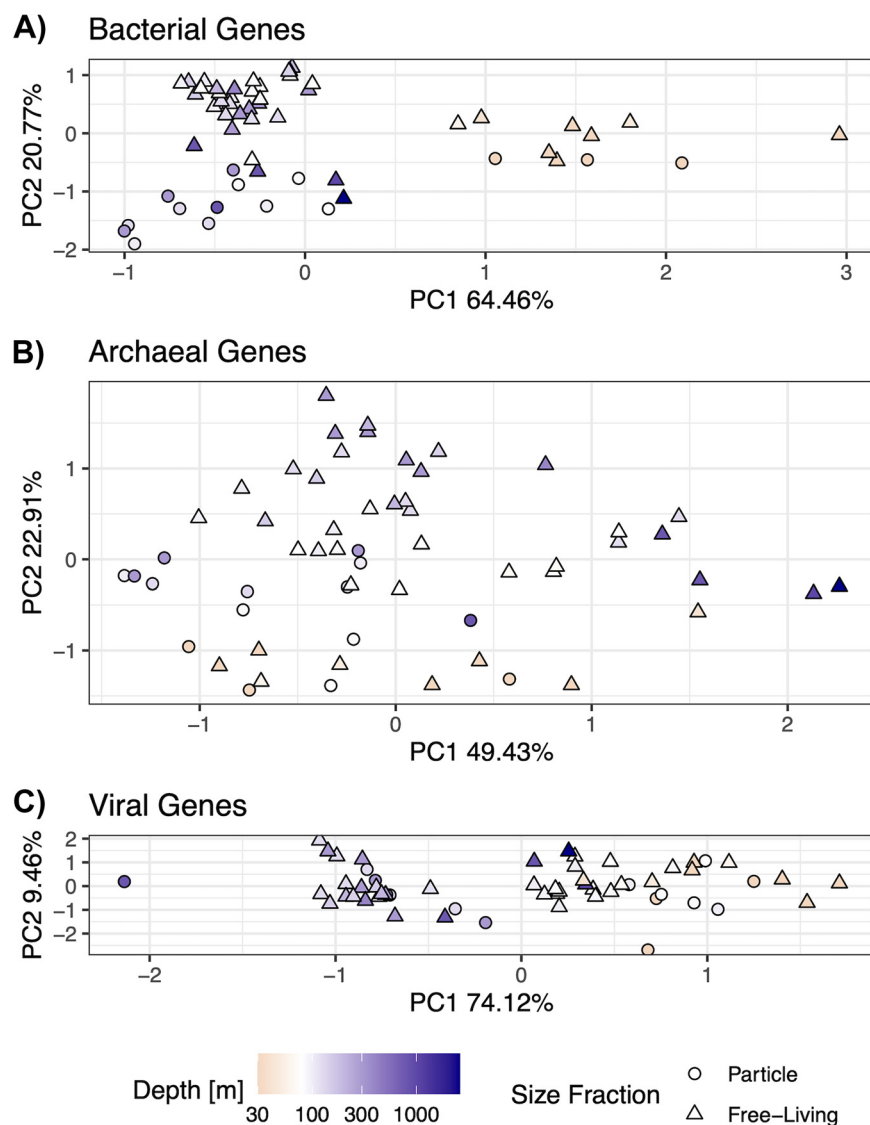


FIG 2 Stoichiogenomic profiles identify the size fraction and depth structure across domains. Redundancy analysis (RDA) ordination was constrained by the sequencing depth, and the first two axes uncorrelated with sequencing depth are presented. The panels are separated by the putative taxonomic assignment of the genes that were used to construct the ordinations: (A) bacterial genes, (B) archaeal genes, (C) viral genes. Point color indicates the sample depth, where yellow is above the average depth of the oxic-anoxic interface (80 m) and blue is below the interface. Point shape indicates particle-associated (circle) versus free-living (triangle) samples.

0.001), respectively. For the viral profiles (Fig. 2C), most of the variance is explained by the first axis (74.12%), which separates the surface mixed layer communities from the OMZ communities, with the mesopelagic samples (dark blue) being in the middle. For viruses, the R^2 for size fraction is 1.2% ($P = 0.363$), whereas the R^2 for depth is 29.5% ($P < 0.001$). While each domain has a unique underlying variance structure in its stoichiogenomic profiles, all three reveal patterns corresponding to the water column depth, and, in the cases of Bacteria and Archaea, the size fraction.

Particle fraction drives stoichiogenomic differences for Bacteria and Archaea.

We constructed linear mixed effects models of the GC content, amino acid N:C ratio, and number of nitrogen atoms in amino acid residue side chains (N-ARSC) to quantitatively test for differences between the size fraction for each domain (see Materials and Methods). Our model structure included depth as a random effect to account for spatial autocorrelation and nonlinearity in the water column chemistry (Fig. 1B and C). So, the depth effect

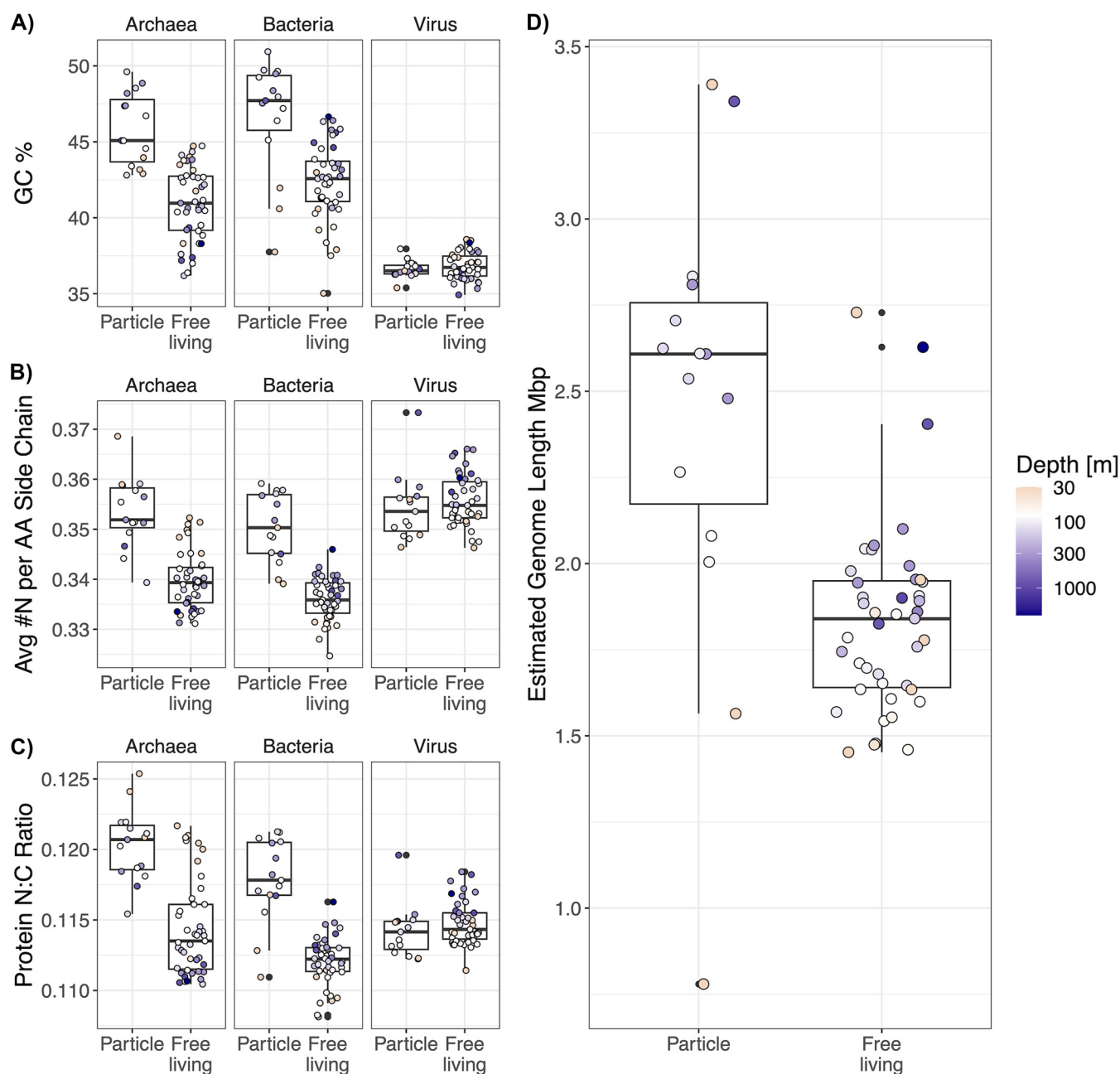


FIG 3 Microbial and viral genes exhibit stoichiogenomic structure across size fraction and depth. Each point represents a metagenome, where the value is the coverage-weighted average GC content (A), the nitrogen content of the amino acid side chains (B), and the total side chain N:C ratio (C) of all archaeal, bacterial, or viral annotated genes in that metagenome. Boxplots indicate the median, 25th, and 75th percentiles of the metagenomes within a size fraction. (D) The distribution of estimated average bacterial genome length between size fractions, as determined by the relative abundances of the single copy core genes in each metagenome.

cannot be expressed as a single parameter. Instead, we assessed the significance of depth via likelihood ratio tests against a null model without a depth effect (see Materials and Methods; estimated profiles in Fig. S2). For Bacteria, our models found significantly higher GC content ($3.82\% \pm 0.479\%$), NC ratios (0.00516 ± 0.000406), and N-ARSC values (0.0134 ± 0.00102) in particle-associated communities (Table S2; data plotted in Fig. 3A–C). We also found that depth had a significant effect for all three parameters and that the depth effects increase GC and N-ARSC until reaching a maximum near 150 m, after which they remain relatively constant with increasing depth, following the pattern of nitrate + nitrite (Fig. S2). For Bacteria, we also used the coverages of single copy core

genes to estimate the average genome length for each sample (see Materials and Methods), and we found a significant increase in genome length ($0.615 \text{ Mb} \pm 0.124 \text{ Mb}$, $P < 1E-5$) in the particle fraction samples (Fig. 3D).

We found significant increases in archaeal GC content ($4.58\% \pm 0.677\%$), NC ratios (0.00539 ± 0.000602), and N-ARSC values (0.0128 ± 0.00152) in the particle fraction samples (Table S2). We also found a significant depth effect for the N-ARSC values and the NC ratios for archaeal genes, but a significant effect was not observed for the gene GC content. These effects are highest at the surface and decrease with depth, in contrast to those of the bacteria (Fig. 3A). For the viruses, we did not find a significant increase in any stoichiometric parameter in the particle fraction. However, we did find a significant depth effect in the NC ratios and N-ARSC values for viral genes. The viral NC ratio and N-ARSC depth effects increased with depth, similarly to bacteria (Fig. 3B; Fig. S2).

Stoichiogenomic differences are characterized by unique amino acid frequencies. We chose one core functional gene for each domain to explore whether metagenome-wide differences in stoichiogenomic properties were detectable on the single gene level. Single gene-level analyses remove the potential effects that are associated with differences in the functional gene content between communities and focus on gene products that should have similar biochemical constraints to maintain the same function. We scanned the annotations of gene clusters for core functional genes that had many diverse representatives in the gene catalogue, had high coverage in the metagenomes, and were detected in all samples (see “Amino acid analysis” in Materials and Methods). There were 154 unique gene clusters that were annotated as bacterial *rpoZ*, the DNA-directed RNA polymerase omega subunit, 119 gene clusters that were annotated as archaeal *ftsZ*, the cell division protein, and 129 gene clusters that were annotated as phage structural protein Gp23. We selected Gp23 as a representative gene of the T4-like bacteriophages, an abundant and ecologically relevant group of marine viruses that all have dsDNA genomic structure (31). We used the coverage of the unique gene clusters in each sample to create a weighted average N:C ratio for *rpoZ*, *ftsZ*, and Gp23 for each sample. Then, we compared those weighted averages by using the same mixed effects model framework to look for significant effects of the particle size fraction on the gene-specific N:C ratios. For the bacterial *rpoZ*, we found an increase of 0.0304 (± 0.00891) in the N:C ratios for the particle-associated samples. For the archaeal *ftsZ*, we found an increase of 0.00373 (± 0.00153). For the viral Gp23, we found no significant difference in the N:C ratio between size fractions. These single gene-level patterns match those of our model of genome-wide stoichiogenomic profiles (Fig. S3; model information is presented in Table S3).

Then, we investigated the amino acid composition of the gene clusters to understand the mechanisms governing their N:C ratios. We used the amino acid frequencies of each gene cluster to create a weighted average amino acid distribution for *rpoZ*, *ftsZ*, and Gp23 for each sample. We specified multiresponse linear models with a lasso penalty to simultaneously identify the amino acids that were most predictive of the size fraction, protein N:C ratio, and N-ARSC (see Materials and Methods). The parameters for these models are shown in Fig. 4. Regularized regression allowed us to identify and rank the most important amino acids in predicting the size fraction, protein N:C ratio, and N-ARSC for each domain, excluding the amino acids that did not contribute much information. The top driver of increased N:C ratios for all three domains is arginine, which is the only amino acid with three additional N atoms on its side chain (Fig. 4). High bacterial *rpoZ* N:C ratios are also driven by asparagine, lysine, and histidine. Archaeal *ftsZ* N:C ratios are more strongly associated with histidine than is bacterial *rpoZ*, and are also associated with increased tryptophan, unlike *rpoZ*. High viral Gp23 N:C ratios are also associated with increased histidine and asparagine, but they are more strongly associated with glutamine than are *ftsZ* and *rpoZ*. Interestingly, methionine is also associated with an increased Gp23 N:C ratio, despite methionine having no N atoms on its side chain. This evidence shows that increased arginine is a common driver of stoichiogenomic variation across marker genes from all three of the domains in our data set.

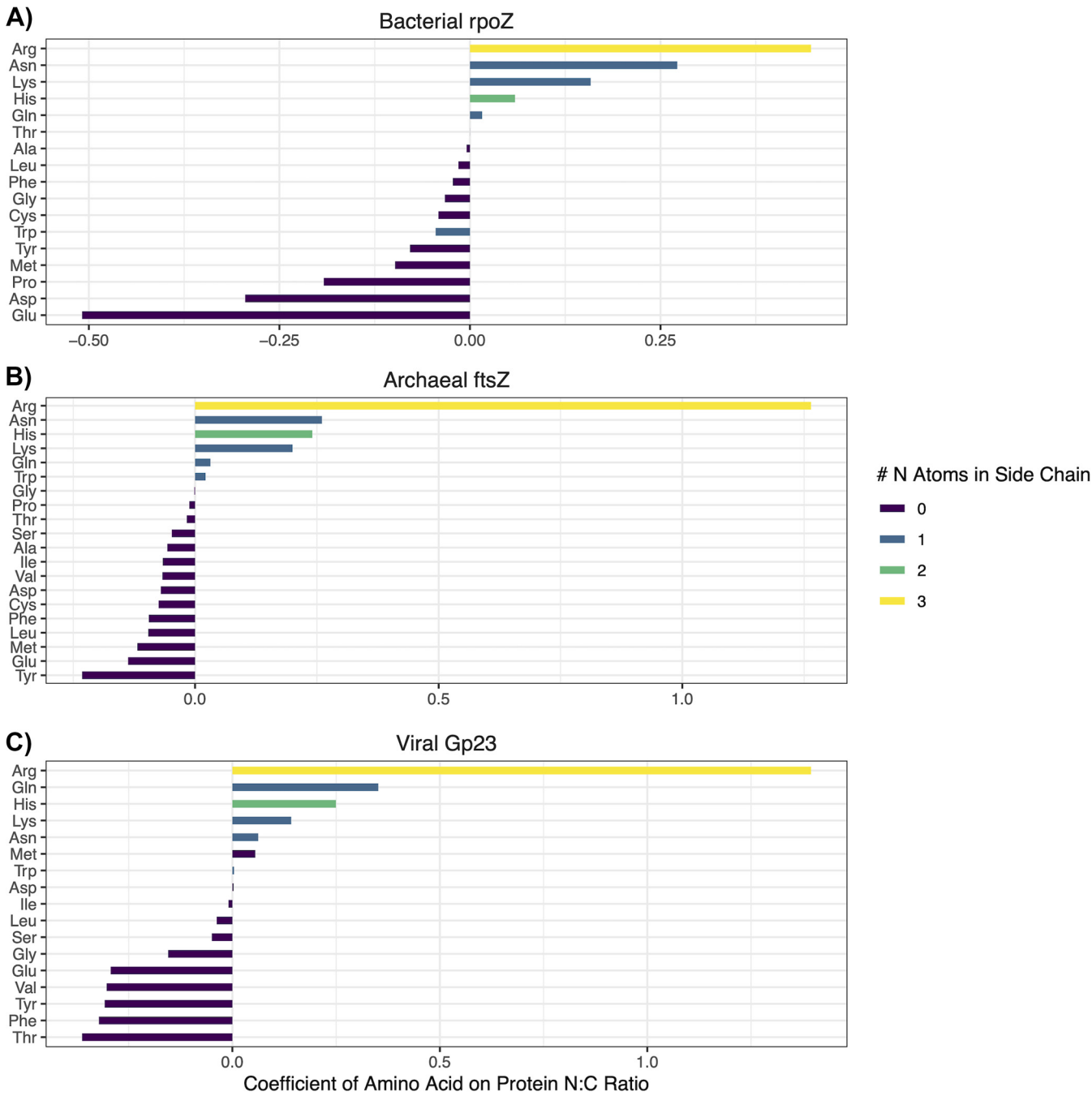


FIG 4 Domains have unique amino acid usage patterns to structure protein nitrogen content. The regularized amino acid regression parameters on the average N:C ratios of widely conserved functional proteins across samples are presented. Positive values indicate increases in the protein N:C ratio, and negative values indicate decreases. Amino acids are colored by the numbers of N atoms in their side chains. Note the differences in scale between domains. (A) Results from bacterial *rpoZ* sequences. (B) Results from archaeal *ftsZ* sequences. (C) Results from viral Gp23 sequences.

However, there is variability in which other amino acids contribute to increased protein nitrogen content.

Virus structural protein amino acid frequencies shift with oxygen concentration. Viral genes increased in GC, N-ARSC, and NC ratios with depth, similarly to bacterial genes, although there were no significant differences for viral genes between size fractions for these parameters (Fig. 3A–C; Table S2). While we did not find a significant difference between the size fractions in the stoichiometry of a phage structural protein (Table S4), we found that viral Gp23 and bacterial *rpoZ* N:C ratios both increase with increasing arginine, asparagine, histidine, lysine, and glutamine (Fig. 4C).

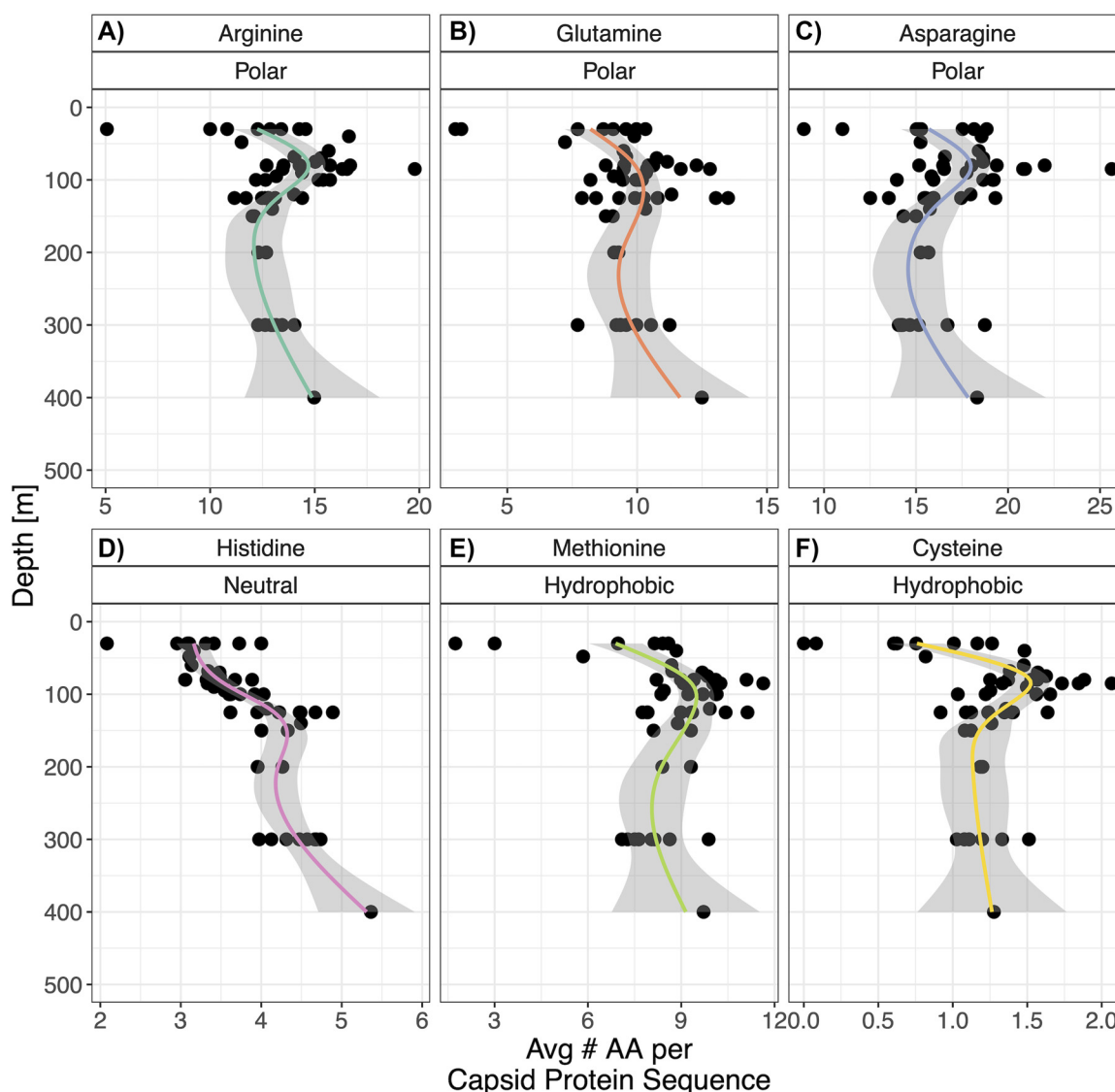


FIG 5 Viral Gp23 amino acid composition varies with depth. The x axis represents the coverage-weighted average frequency of each amino acid in a Gp23 sequence for each sample. Amino acids are labeled as polar (hydrophilic) (A to C), neutral (D), or hydrophobic (E and F). The smoothed lines, which are used to demonstrate overall trends, were fitted via general additive model smoothing, with shaded regions representing the 95% confidence intervals of the smoothing.

We plotted the profiles for the amino acids that were identified by the multiresponse regularized regression model (presented in Fig. 4C) to increase the average Gp23 N:C ratios (Fig. 5A–E). Most of these amino acids have a peak in abundance slightly above 100 m, which corresponds to the average oxic-anoxic interface depth across stations. Three of these amino acids had significant negative rank correlations with oxygen levels. Histidine ($\rho = -0.487$, $P < 1E-5$), an amino acid with two N atoms on its side chain, was negatively correlated with oxygen concentration, as was methionine ($\rho = -0.311$, $P = 0.0334$). From our previous ordination of viral stoichiogenomic properties (Fig. 2C), we recalled that the primary axis clustered samples from depths around the oxic-anoxic interface at negative values, whereas samples from the surface mixed layer as well as mesopelagic samples were clustered at positive values. Then, we looked for other amino acids that were inversely correlated with oxygen concentration, indicating enrichment in high nitrogen anoxic water, and we found cysteine ($\rho = -0.228$, $P = 0.0334$). Our amino acid-level analysis shows that the overall increase of viral N:C ratios with depth matches trends in bacteria and can be explained by a similar

set of amino acids. Methionine and cysteine, the two hydrophobic and sulfur-containing amino acids, are specifically associated with the anoxic core of the OMZ, and they may contribute to the separation of that depth layer from the mesopelagic waters in our earlier multivariate analysis of all viral genes.

Particles are enriched for functional genes utilizing organic nitrogen substrates. Our analysis of the stoichiometric properties of bacterial genomes indicates higher genomic and proteomic nitrogen content for particle-associated communities. This may be because some organic particles can act either directly as an organic nitrogen source or as a hot spot of nitrogen remineralization (32, 33). We compared the relative abundances of 412 bacterial KEGG orthologues between size fractions to identify which functional genes were enriched in particle fraction samples. We used a model based on the log-ratios of the relative abundances of each orthologue to a common “baseline” orthologue to mitigate the negative constrained covariance and detection biases of compositional metagenomic data (34). Setting the false discovery rate to 10%, we found 69 (16.8%) KEGG orthologues that were enriched in particle-associated bacterial communities (Fig. 6).

Our analysis included 12 transporters, 5 of which take up branched-chain amino acids or polyamines. The particle fraction was enriched in 4/12 transporters, including 3/5 transporters for organic nitrogen substrates. The particle fraction was also enriched in 9 genes related to amino acid degradation, including 4/13 studied genes that are involved in the degradation of branched-chain amino acids. The amino acid-level analysis identified arginine utilization as an important driver of protein-level N:C ratios for bacteria (Fig. 4A). The functional analysis also shows that 3/5 studied genes in the arginine biosynthesis pathway are enriched in particle fraction bacteria (Fig. 6E). We also found enrichment on the particle fractions for 3/4 protein excretion genes and 3/4 flagellum/pilus-related genes. For comparison, 8/30 genes involved in carbon metabolism were enriched in the particle fraction. Our results suggest that genes for the uptake and degradation of organic nitrogen molecules are more abundant within particle-associated bacterial communities, and this is consistent with the hypothesis that particles may offer usable, organic nitrogen-containing substrates.

DISCUSSION

Stoichiogenomic studies have identified genome streamlining-type evolutionary adaptations to nitrogen limitation across diverse marine microorganisms at spatial scales that span kilometers of depth to thousands of kilometers across the ocean surface (7). Here, we examined the principles of stoichiogenomic streamlining at smaller spatial scales, that is, across tens of meters surrounding the nitracline of the ETNP OMZ and across microns separating planktonic and particle-associated microbes. The difference in nitrate + nitrite at the surface and at the oxic-anoxic interface (20 μ M over 60 m) is equivalent to the difference in nitrate concentrations between the surface and 500 m, as found by a previous stoichiogenomic study in the North Pacific subtropical gyre (NPSG) (10), and it mirrors the average differences in nitrate + nitrite measurements between the surface and 350 to 400 m in the Station ALOHA NPSG climatology (<https://hahana.soest.hawaii.edu/hot/hot-dogs/>). 20 μ M is also similar to the total nitrogen differences in surface concentrations between open ocean regions and high nitrate coastal upwelling regions (35). Our observed differences of nitrate + nitrite concentrations between the surface mixed layer and the anoxic core are similar to the differences that were examined by previous stoichiogenomic studies at larger spatial scales.

Our analysis estimates that the bacterial genomic and proteomic nitrogen content is enriched on particles and increases with increasing depth (thus, increasing nitrate + nitrite). We found that arginine, asparagine, and lysine were associated with higher nitrogen-content protein sequences. There was a corresponding enrichment in arginine biosynthesis genes in particle-associated communities, along with other genes with functions associated with particle-associated bacterial lifestyles (2). Archaeal genes were also enriched in nitrogen in particle-associated samples, with protein sequences having increased arginine, asparagine, and lysine. However, archaeal genes decreased in nitrogen content with increasing depth. We cannot guarantee that changes in nitrogen availability

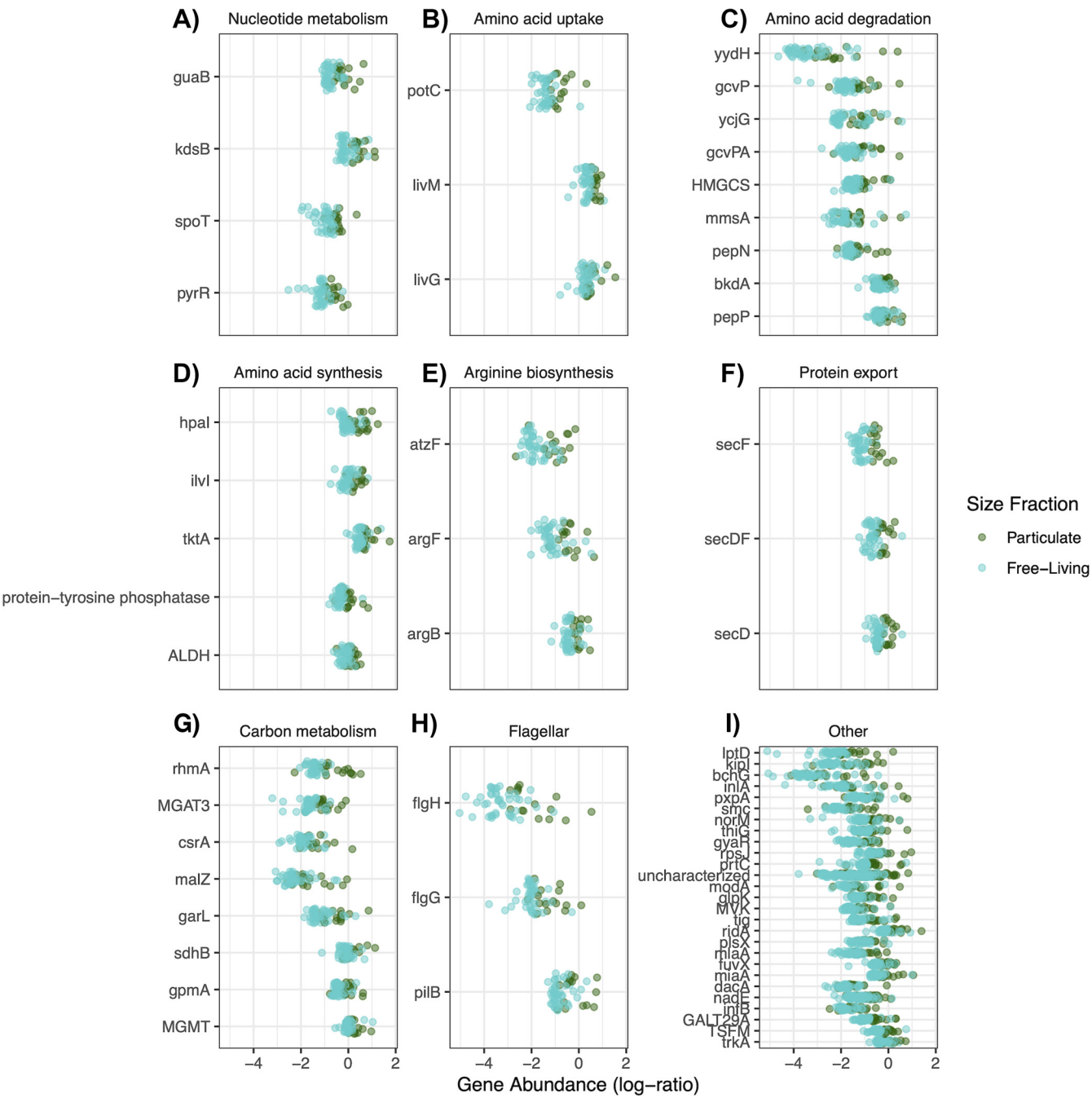


FIG 6 Bacterial functional gene content differences between size fractions. KEGG orthologues are grouped by metabolic pathways and are identified by KEGG gene nomenclature (y axis). The log ratios of the relative abundance of each gene to *sdhC*, a core functional gene that is present across all of the samples, are plotted on the x axis (see Materials and Methods). Point color indicates the size fraction.

drove these stoichiogenomic trends, as the OMZ environment has a complicated chemistry, in that changes in inorganic nitrogen availability correlate with changes in the dominant metabolic strategy (e.g., aerobic respiration versus denitrification) and the availability of organic carbon for heterotrophs. These changes may exert a pressure for genomic streamlining on the basis of reduced free energy being available for synthesizing receptors, gathering nutrients, and constructing macromolecules. We note that the decrease in the free energy yield between aerobic respiration and denitrification is relatively small. So, bioenergetic effects may not be as important as nutrient availability. Our results here suggest a role for the nitrate + nitrite supply in stoichiogenomic variation.

Beyond bacterial and archaeal populations, corresponding depth-related differences exist (although not differences between size fractions) in the nitrogen content of viral genes and protein sequences. Our sampling protocol was designed to collect biomass or particles greater than 0.2 μm in size. Therefore, it likely did not recover free virions. The viral sequences that were detected in our data set are therefore likely to be host-associated. Whether by viral attachment to host cell membranes, active infections, or potentially through lysogenic integration, these mechanisms would suggest that the viral genes in our study belong to ecologically active components of the viral community. Our analyses suggest stoichiogenomic adaptation to environmental nitrate + nitrite concentrations in viruses, adding to the body of stoichiogenomic literature that has thus far focused on cellular life. Economizing the viral elemental content to reflect the nitrogen limitation conditions of the host may be important for facilitating efficient infection (36, 37). Surveys of viral auxiliary metabolic genes in the eastern tropical South Pacific (ETSP) OMZ include viral genes for nitrogen cycling processes, such as denitrification and assimilatory nitrate reduction (38), suggesting that nitrogen metabolism influences viral adaptation in OMZ environments. We supplement this functional evidence by suggesting that nitrate + nitrite concentrations may also correlate to the nitrogen quota of viral genomes. We also found the stoichiogenomic parameters for viral genes in the anoxic core of the OMZ to be distinct from those at other depths via ordination analysis. This result contributes further evidence from an elemental composition and structural perspective to functional and taxonomic evidence that shows unique viral communities in the eastern tropical South Pacific (ETSP) OMZ (31, 38, 39). Our evidence suggests that the uniqueness of OMZ viruses is present not only in the taxonomy and functional gene content but also in the genomic and amino acid chemical composition of viral particles.

Our amino acid-level analysis of the variability of a phage structural protein also identified elevated methionine and cysteine as a unique signature that is specific to the anoxic core. These amino acids are both hydrophobic and sulfur-containing. The increase of sulfur-containing amino acids in this layer may be related to complex sulfur cycling dynamics in OMZs (40). Hydrophobic residues, specifically methionine, have been demonstrated to be important for phage capsid stability (41, 42). In the anoxic core of the ETNP OMZ, aerobic respiration rates are low (20), and less efficient terminal electron acceptors are used (2). Furthermore, in the ETNP anoxic core, cell numbers decrease, and the virus to microbe ratio increases (39). These factors suggest that encountering hosts with sufficient metabolic activity to support an infection may be more difficult at anoxic depths. Evidence exists for the long-term (annual scale) persistence of viruses in the absence of hosts in the Red Sea (43), suggesting that marine viruses may use persistence outside hosts as an ecological strategy. The speculation that low metabolic rates may influence viral evolutionary strategies in the ETSP OMZ anoxic core is supported by genomic evidence that viruses have adaptations for slow and intermittent replication (39). The replacement of structural protein amino acid residues with stabilizing hydrophobic residues, such as the cysteine and methionine enrichments we found, may be a novel mechanism that aids in viral structural stability and thereby allows for longer residence times between host encounters to overcome low host abundance and metabolic rates.

Despite finding viral genomic streamlining with depth, we were unable to detect a significant difference in the stoichiogenomic properties of viral genes between particle-associated and planktonic metagenomes. This leads us to at least two hypotheses: (i) selective forces for streamlining differ between planktonic viral and particle-associated viral communities and are potentially weaker for viruses that infect hosts that exhibit planktonic and particle-associated lifestyles; (ii) viruses detected on particles do not necessarily infect particle-associated hosts and instead passively aggregate onto particles. To the latter hypothesis, recent molecular studies of particulate metagenomic communities have recovered cyanophage DNA in abyssal depths (44), suggesting that viral DNA can potentially be vertically transported on particles and persist where hosts may not be

metabolically active. Our finding that the stoichiometric properties of planktonic and particle-associated viral genes are similar lends further evidence to the idea that viruses with planktonic-adapted hosts may passively aggregate to particles.

In summary, this study demonstrates that stoichiogenomic differences occur on spatial scales that are much smaller than the spatial scales that have previously been studied (6, 10). We show genomic differences in the GC content of 3 to 5% between particle-associated and free-living metagenomes from the same depth, similar to previously reported differences that were observed over ranges of hundreds of meters from Mende et al. (10). The increase in N-ARSC that was observed in particle-associated versus free-living metagenomes is within the range (4 to 7% difference) that was observed between coastal and open ocean proteins (6). Macromolecules with compositions that are susceptible to stoichiometric forcing (e.g., proteins and genomes) can comprise a substantial proportion of cell contents for small cells (45). Therefore, evolutionary flexibility in these traits may have important emergent biogeochemical effects at larger scales. Viral infection can also shape the ecology and biogeochemistry of microbial communities (37). The elemental composition of viruses can generate lysate with distinct elemental ratios, compared to the host (46). Here, we show evidence of viral streamlining occurring in parallel to hosts along nitrate + nitrite concentration gradients with depth, despite viruses lacking independent metabolism and nutrient uptake mechanisms. A better means of identification for the factors that shape the nutrient requirements of marine microbes and their viruses is vital to understanding the mechanisms that shape marine ecology and biogeochemistry. Likewise, determining the influence of viral infection can also shape the ecology and biogeochemistry of microbial communities (37). For example, the elemental composition of viruses can generate lysate with distinct elemental ratios, compared to the host (46). Modeling and predicting the influence of viral infection on the stoichiometry of marine detrital organic matter, and therefore carbon cycling, requires theory describing the evolutionary relationships between the host and viral elemental composition (47). Including stoichiogenomic effects, such as those described in this study, in future ecosystem models may offer an important route to our future comprehension of the roles of microbial adaptation and macromolecular contents in biogeochemical cycling.

MATERIALS AND METHODS

Metagenomic data. All of the metagenomic data from the ETNP were sequenced on the Illumina platform and are associated with BioProject ID [PRJNA632347](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA632347).

CTD and nutrient data. Sea-Bird conductivity, temperature, and depth (CTD) sensors were deployed with Niskin rosettes during sample collection on a package that included a fluorometer, a transmissometer, and a Sea-Bird SBE43 oxygen sensor (48). Samples for the nitrate and nitrite nutrient analyses were collected, and the year 2013 samples were processed as described in Ganesh et al. (2) and Glass et al. (48). Samples for the nitrate and nitrite analysis for the 2014 cruise were processed as described in Ganesh et al. (49).

For this study, down-casts of CTD profiles were analyzed using TEOS's Python Gibbs Seawater (gsw) toolbox v. 3.4.0, which was loaded into R via the reticulate library v. 3.38.0. Using latitude, longitude, pressure, temperature, and salinity readings, corrected height relative to sea level was calculated along with absolute salinity and conservative temperature. Absolute salinity and conservative temperature were used to derive density anomaly σ_{θ} with a reference pressure of 0 decibar. These properties were also used to calculate oxygen solubility, and oxygen saturation was derived from the oxygen solubility and oxygen readings from the CTD instrumentation package. The data were then smoothed via binned averaging to the nearest 1 m of height.

Sequence data quality control, taxonomic assessment, and assembly. Interleaved reads were deinterleaved using Nathan Haigh's deinterleaving script from github <https://gist.github.com/nathanhaigh/3521724>. Reads for all metagenomes were then adapter trimmed and quality filtered using TrimGalore https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/, using a Phred quality cutoff score of 25 and a minimum read length of 100 bp. After trimming and filtering, the reads were merged using FLASH (50), with the expected read length adjusted to 200 bp to minimize the exclusion of mostly overlapping reads.

Before the paired-end reads were merged, putative small subunit (SSU) ribosomal reads were identified using Metaxa2 (51). All of the identified 16S reads were then extracted from the forward and reverse reads for each metagenome, and the sequences were quality assessed and dereplicated using dada2, as implemented in R (52). Sequence variants were then assigned taxonomy using the SILVA v1.3.2 database (53) from dada2's assignTaxonomy function.

Parameters on the distributions of the merged paired-end reads (average length, mean and variance of GC content, total merged reads) were calculated for each metagenome using BioPython utilities. We modeled the average GC content of the metagenomic reads for each sample by using a mixed-effects linear model with the $\log_{10}$ total number of reads and size fraction as fixed effects and depth as a random effect, using an exponential correlation structure to account for the spatial autocorrelation in the water column.

Metagenomes were assembled using megahit (54) with a starting kmer length of 27 and the default parameters. Contig statistics were then calculated using custom Python and R scripts.

Gene catalogue. Using BioPython (55) utilities, the assemblies were filtered for all contigs with a length of $>1,000$. Then, genes were predicted for every metagenome using Prodigal v. 2.6.3 with the standard parameters on the metagenome setting (56). Partial genes were removed.

After combining all predicted genes from all metagenomes, a nonredundant catalog of genes was created using CD-HIT with amino acid sequences, clustering genes with an identity of greater than 95% and an alignment of 90% with the shorter sequence (57). To identify the COGs represented in the clusters, FetchMG's COG extraction script was used with the standard parameters (58).

Calculating gene characteristics. Using functionalities from scripts provided in Mende et al. (10) as well as additional BioPython utilities, the gene length, gene GC content, molecular weight of summed amino acids coded by each gene, codon usage (as defined in Mende et al., a ranking scheme for determining the diversity of codons used for a particular amino acid in a given gene), codon bias, number of nitrogen and carbon atoms in residue side chains for all coded amino acids (N/C-ARSC), NC-ratios of those side chains, and individual amino acid counts were calculated for each gene. Additionally, the bulk GC content of each metagenome was calculated for all merged reads using BioPython.

Read mapping for abundance and genome size estimation. The BWA-MEM algorithm was used to map individual metagenomes to the assembled gene catalog, using the standard parameters (59). After mapping, the results were filtered for only alignments with an identity of 95% or higher (calculated as $1 - \text{number of mismatches} / \text{alignment length}$) and an alignment length of greater than 60 bp, using a mix of samtools (60) and the pysam Python package (<https://www.osti.gov/servlets/purl/1559931>). The resulting alignments were then converted into the coverage for each gene, dividing the total number of base pairs mapped to each reference gene by the gene length. To estimate the average genome copy number, these coverages were normalized to the average coverage of 10 single-copy COGs (61), following the same method used in (10). The average numbers of genes per genome were then calculated by summing the copy numbers for all genes for each metagenome, using a custom script in R. The average weighted gene coverage was then calculated by dividing the coverages for each gene in a metagenome by the sum coverage of all of the genes in that metagenome so as to mitigate the differences of sequencing depth between samples.

To translate the gene properties to a metagenome-wide summary, the gene properties for each gene were then multiplied by their weighted coverages and summed for each metagenome. This created a "gene property profile" for each metagenome, which was then used in further statistical analyses.

Statistical analyses. (i) Ordination analysis. Because our samples were sequenced across a wide range of depths, and because Illumina sequencing has been shown to display molecular bias in sequencing (62, 63), we aimed to carefully undergo statistical analyses, keeping this effect in mind. Therefore, for the ordination analyses, we conducted RDA using sequencing depth as a covariate. We then removed that axis from the ordination to ensure that all of the patterns in the underlying data structure were uncorrelated with sequencing depth. For all of the regression models, sequencing depth was included as an independent variable so as to reduce the potential impact of confounding on parameter estimation.

To first explore the differences in the covariation between metagenomic gene properties, an RDA was constructed using the vegan package v.2.5-7 in R (64), based on all of the properties listed in "Calculating gene characteristics". Two axes were found to be associated with eigenvalues that were larger than the eigenvalue associated with sequencing depth. So, these were selected for presentation. *Post hoc* permutational analysis of variance tests were conducted to find the percent variance explained by size fraction and the asinh-transformed depth for each group, independently. We transformed the depth in order to account for the nonlinearity in the chemical structure of the water column, which changes more gradually with increasing depth below the anoxic core of the OMZ.

(ii) Modeling stoichiogenic parameters. We structured the comparison between the size fractions for the stoichiogenic parameters among bacteria, archaea, and viruses as separate linear mixed effects models. We specified depth as a random effect with an exponential correlation structure so as to represent the spatial autocorrelation in the water column chemistry and to capture nonlinear relationships between depth and stoichiogenic parameters. Sequencing depth was also used as a covariate in this model to control for the aforementioned biases. Model parameters were learned using maximum likelihood methods via R's nlme package v 3.1-155. For each model, to assess significance, another model was learned using only fixed effects, including depth. Model comparisons were conducted via likelihood ratio tests and comparisons of the Akaike information criterion (AIC) values so as to control for differences in model complexity. For models in which there was a significant likelihood gain, including depth, the random effect estimate that was learned for each depth was correlated with depth via Spearman's ρ .

(iii) Amino acid analysis. First, the coverages of all of the KEGG orthologues and the number of gene cluster representatives in the gene catalogue were assessed. For each domain, we selected the orthologue with the highest total coverage, coverage in all samples, the greatest number of gene cluster representatives in the gene catalogue, and a general function presumed to be universal to diverse taxa

within each domain. These genes, namely, *rpoZ* for bacteria, *ftsZ* for archaea, and Gp23 for viruses, were considered to be our “marker genes” for an amino acid frequency analysis. For the case of Bacteria, we chose *rpoZ* for its generic role in RNA polymerization. For Archaea, we chose *ftsZ* for its involvement in cell division. For viruses, we chose the structural protein for T4-like bacteriophages Gp23. Gp23 is a core structural gene for an abundant and (relatively) taxonomically cohesive group of viruses in the family Myoviridae. We specifically chose this structural gene as the virus representative so as to avoid the substantial differences in macromolecular and chemical composition between viruses with different genomic architectures (dsDNA versus ssDNA versus RNA) while still including a large, abundant, and likely ecologically relevant group. The amino acid counts for the representative sequence of each gene cluster for each marker gene were then used to construct a weighted average amino acid profile for each sample, using the coverage of each gene cluster in that sample as the weight.

The weighted amino acid averages for each domain were centered and scaled to account for the different dynamic ranges. We generated one model for each domain. The models used the scaled amino acid ranges as predictors and the sample size fraction (binarized as particle fraction = 1, planktonic = 0), the scaled weighted average N-ARSC for the domain protein, and the scaled weighted average N:C ratio of the domain protein as multiresponses. For example, for bacteria, we used the average number of alanine, cysteine, etc., for each *rpoZ* in each sample, weighted by the abundance of each *rpoZ* sequence, to simultaneously predict the size fraction, average N-ARSC of *rpoZ* sequences, and average N:C ratio of *rpoZ* sequences for that sample. We learned the parameters of a linear model via a maximum likelihood framework with an L_1 regularization to reduce redundancy among amino acids. 70% of the metagenomes were used for model learning and cross validation, whereas 30% were used for testing. The models were learned using the R glmnet package 4.13 “multigaussian” family of models. Cross validation using 10 folds was conducted to optimize the regularization parameter, and the model with the minimum average mean squared error across all 10 folds of the test data was selected. The model was fit and then assessed against the remaining 30% of test data (see Fig. S4 for the model evaluation).

(iv) Functional analysis. A log-ratio approach was used to compare the relative abundances of functional genes between samples. Log-ratios have been specifically recommended for use in metagenomic data sets to account for sequencing bias and to overcome negative constrained covariance structures surrounding compositional data (34). In order to construct the log-ratios, we selected *sdhC*, one of the components of succinate dehydrogenase from the TCA cycle, as the focal gene to be the denominator. This gene is part of a central metabolic pathway that is widely distributed across bacteria, and it had among the highest coverage in all of our samples. This left 412 remaining KEGG orthologues that had at least some reads mapping in all samples. For each orthologue, we summed the coverages of all gene clusters annotated as that orthologue for each sample, divided it by the summed coverage of *sdhC* in that sample, and then took the log to reduce the variance. Then, a linear model was constructed for each orthologue by using the log-ratio as the response, and the size fraction, the depth, and a size fraction/depth interaction term as predictors. The estimates for all three of these parameters were retained, and then an FDR threshold of 10% was applied to the P values of all of the parameters across all of the models to account for multiple testing.

Data availability. All of the metagenomic data from the ETNP were sequenced on an Illumina platform and are associated with BioProject ID [PRJNA632347](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA632347). The intermediate and final data products as well as the code are available in repository https://github.com/dmuratore/omz_stoichiometry with archival hosting at doi: <https://zenodo.org/badge/latestdoi/560085581>.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

FIG S1, PDF file, 0.01 MB.

FIG S2, PDF file, 0.02 MB.

FIG S3, PDF file, 0.02 MB.

FIG S4, PDF file, 0.01 MB.

TABLE S1, PDF file, 0.1 MB.

TABLE S2, PDF file, 0.3 MB.

TABLE S3, PDF file, 0.2 MB.

TABLE S4, PDF file, 0.1 MB.

ACKNOWLEDGMENTS

We thank the captain and crew of the R/V Oceanus for enabling the sample collection. This work was supported by the National Science Foundation grants 1151698, 1558916, 1564559, and 2130185 to F.J.S., Danmarks Frie Forskningsfond 9040-00327B to L.A.B., and Simons Foundation grant 721231 to J.S.W. The funding agencies had no role in conducting the research.

D.M. conceptualized the study, performed the bioinformatic analysis, performed the statistical analysis, and wrote the manuscript. A.D.B. and F.J.S. collected the samples, conceptualized the study, generated the sequence data, and advised in the analysis as well as the writing and editing of the manuscript. L.A.B. and B.T. carried out fieldwork,

processed the samples for the nutrient analysis, and advised in the writing and editing of the manuscript. J.S.W. contributed to the development of the analytic methods, analysis, and editing of the manuscript.

REFERENCES

- Falkowski PG, Fenchel T, Delong EF. 2008. The microbial engines that drive Earth's biogeochemical cycles. *Science* 320:1034–1039. <https://doi.org/10.1126/science.1153213>.
- Guidi L, Chaffron S, Bittner L, Eveillard D, Larhlami A, Roux S, Darzi Y, Audic S, Berline L, Brum JR, Coelho LP, Espinoza JCI, Malviya S, Sunagawa S, Dimier C, Kandels-Lewis S, Picheral M, Poulain J, Searson S, Stemmann L, Not F, Hingamp P, Speich S, Follows M, Karp-Boss L, Boss E, Ogata H, Pesant S, Weissenbach J, Wincker P, Acinas SG, Bork P, de Vargas C, Iudicone D, Sullivan MB, Raes J, Karsenti E, Bowler C, Gorsky G, Tara Oceans Consortium C. 2016. Plankton networks driving carbon export in the oligotrophic ocean. *Nature* 532:465–470. <https://doi.org/10.1038/nature16942>.
- Sunagawa S, Coelho LP, Chaffron S, Kultima JR, Labadie K, Salazar G, Djahanschiri B, Zeller G, Mende DR, Alberti A, Cornejo-Castillo FM, Costea PI, Cruaud C, d'Ovidio F, Engelen S, Ferrera I, Gasol JM, Guidi L, Hildebrand F, Kokoszka F, Lepoivre C, Lima-Mendez G, Poulain J, Poulos BT, Royo-Llonch M, Sarmiento H, Vieira-Silva S, Dimier C, Picheral M, Searson S, Kandels-Lewis S, Oceans C T, Bowler C, de Vargas C, Gorsky G, Grimsley N, Hingamp P, Iudicone D, Jaillon O, Not F, Ogata H, Pesant S, Speich S, Stemmann L, Sullivan MB, Weissenbach J, Wincker P, Karsenti E, Raes J, Acinas SG, Tara Oceans coordinators, et al. 2015. Ocean plankton. Structure and function of the global ocean microbiome. *Science* 348:1261359. <https://doi.org/10.1126/science.1261359>.
- Caputi L, Carradec Q, Eveillard D, Kirilovsky A, Pelletier E, Pierella Karlusich JJ, Rocha Jimenez Vieira F, Villar E, Chaffron S, Malviya S, Scalco E, Acinas SG, Alberti A, Aury JM, Benoiston AS, Bertrand A, Biard T, Bittner L, Boccara M, Brum JR, Brunet C, Busseni G, Carratalà A, Claustre H, Coelho LP, Colin S, D'Aniello S, Da Silva C, Del Core M, Doré H, Gasparini S, Kokoszka F, Jamet JL, Lejeune C, Lepoivre C, Lescot M, Lima-Mendez G, Lombard F, Lukeš J, Maillet N, Madoui MA, Martinez E, Mazzocchi MG, Néou MB, Paz-Yepes J, Poulain J, Ramondenc S, Romagnan JB, Roux S, Salvaggio Manta D, et al. 2019. Community-Level Responses to Iron Availability in Open Ocean Plankton Ecosystems. *Glob Biogeochem Cycles* 33:391–419. <https://doi.org/10.1029/2018GB006022>.
- Delmont TO, Quince C, Shaiber A, Esen OC, Lee STM, Rappe MS, McLellan SL, Lucker S, Eren AM. 2018. Nitrogen-fixing populations of Planctomycetes and Proteobacteria are abundant in surface ocean metagenomes. *Nat Microbiol* 3:804–813. <https://doi.org/10.1038/s41564-018-0176-9>.
- Dittbener H, Ohlmann N, Acquisti C. 2018. Stoichiometry-metagenomics of ocean waters: a molecular evolution approach to trace the dynamics of nitrogen conservation in natural communities. *Front Microbiol* 9:1590. <https://doi.org/10.3389/fmicb.2018.01590>.
- Elser JJ, Acquisti C, Kumar S. 2011. Stoichiogenomics: the evolutionary ecology of macromolecular elemental composition. *Trends Ecol Evol* 26:38–44. <https://doi.org/10.1016/j.tree.2010.10.006>.
- Morris JJ, Lenski RE, Zinser ER. 2012. The Black Queen Hypothesis: evolution of dependencies through adaptive gene loss. *mBio* 3:e00036-12. <https://doi.org/10.1128/mBio.00036-12>.
- Larsen ML, Wilhelm SW, Lennon JT. 2019. Nutrient stoichiometry shapes microbial coevolution. *Ecol Lett* 22:1009–1018. <https://doi.org/10.1111/ele.13252>.
- Mende DR, Bryant JA, Aylward FO, Eppley JM, Nielsen T, Karl DM, DeLong EF. 2017. Environmental drivers of a microbial genomic transition zone in the ocean's interior. *Nat Microbiol* 2:1367–1373. <https://doi.org/10.1038/s41564-017-0008-3>.
- Grzymalski JJ, Dussaq AM. 2012. The significance of nitrogen cost minimization in proteomes of marine microorganisms. *ISME J* 6:71–80. <https://doi.org/10.1038/ismej.2011.72>.
- Okie JG, Poret-Peterson AT, Lee ZM, Richter A, Alcaraz LD, Eguarte LE, Siefert JL, Souza V, Dupont CL, Elser JJ. 2020. Genomic adaptations in information processing underpin trophic strategy in a whole-ecosystem nutrient enrichment experiment. *Elife* 9:e49816. <https://doi.org/10.7554/eLife.49816>.
- Stocker R. 2012. Marine microbes see a sea of gradients. *Science* 338:628–633. <https://doi.org/10.1126/science.1208929>.
- Seymour JR, Amin SA, Raina JB, Stocker R. 2017. Zooming in on the phycosphere: the ecological interface for phytoplankton–bacteria relationships. *Nat Microbiol* 2:17065. <https://doi.org/10.1038/nmicrobiol.2017.65>.
- Crump BC, Armbrust EV, Baross JA. 1999. Phylogenetic Analysis of Particle-Attached and Free-Living Bacterial Communities in the Columbia River, Its Estuary, and the Adjacent Coastal Ocean. *Appl Environ Microbiol* 65:3192–3204. <https://doi.org/10.1128/AEM.65.7.3192-3204.1999>.
- Bachmann J, Heimbach T, Hassenrück C, Koppio GA, Iversen MH, Grossart HP, Gärdes A. 2018. Environmental Drivers of Free-Living vs. Particle-Attached Bacterial Community Composition in the Mauritania Upwelling System. *Front Microbiol* 9:2836. <https://doi.org/10.3389/fmicb.2018.02836>.
- DeLong EF, Franks DG, Alldredge AL. 1993. Phylogenetic diversity of aggregate-attached vs. free-living marine bacterial assemblages. *Limnol Oceanogr* 38:924–934. <https://doi.org/10.4319/lo.1993.38.5.0924>.
- Leu AO, Eppley JM, Burger A, DeLong EF. 2022. Diverse Genomic Traits Differentiate Sinking-Particle-Associated versus Free-Living Microbes throughout the Oligotrophic Open Ocean Water Column. *mBio* 13:e01569. <https://doi.org/10.1128/mbio.01569-22>.
- Helly JJ, Levin LA. 2004. Global distribution of naturally occurring marine hypoxia on continental margins. *Deep Sea Res Part I Oceanogr Res Pap: Oceanogr Res Pap* 51:1159–1168. <https://doi.org/10.1016/j.dsr.2004.03.009>.
- Tiano L, Garcia-Robledo E, Dalsgaard T, Devol AH, Ward BB, Ulloa O, Canfield DE, Revsbech NP. 2014. Oxygen distribution and aerobic respiration in the north and south eastern tropical Pacific oxygen minimum zones. *Deep Sea Res Part I Oceanogr Res Pap: Oceanogr Res Pap* 94:173–183. <https://doi.org/10.1016/j.dsr.2014.10.001>.
- Thamdrup B, Dalsgaard T, Revsbech NP. 2012. Widespread functional anoxia in the oxygen minimum zone of the Eastern South Pacific. *Deep Sea Res Part I Oceanogr Res Pap: Oceanogr Res Pap* 65:36–45. <https://doi.org/10.1016/j.dsr.2012.03.001>.
- Ward BB, Tuit CB, Jayakumar A, Rich JJ, Moffett J, Naqvi SWA. 2008. Organic carbon, and not copper, controls denitrification in oxygen minimum zones of the ocean. *Deep Sea Res Part I Oceanogr Res Pap: Oceanogr Res Pap* 55:1672–1683. <https://doi.org/10.1016/j.dsr.2008.07.005>.
- Ulloa O, Canfield DE, DeLong EF, Letelier RM, Stewart FJ. 2012. Microbial oceanography of anoxic oxygen minimum zones. *Proc Natl Acad Sci U S A* 109:15996–16003. <https://doi.org/10.1073/pnas.1205009109>.
- Kalvelage T, Lavik G, Lam P, Contreras S, Arteaga L, Löscher CR, Oschlies A, Paulmier A, Stramma L, Kuypers MM. 2013. Nitrogen cycling driven by organic matter export in the South Pacific oxygen minimum zone. *Nature Geosci* 6:228–234. <https://doi.org/10.1038/ngeo1739>.
- Logina AN, Thomsen S, Engel A. 2016. Chromophoric and fluorescent dissolved organic matter in and above the oxygen minimum zone off Peru. *J Geophys Res Oceans* 121:7973–7990. <https://doi.org/10.1002/2016JC011906>.
- Karner M, Herndl GJ. 1992. Extracellular enzymatic activity and secondary production in free-living and marine-snow-associated bacteria. *Mar Biol* 113:341–347. <https://doi.org/10.1007/BF00347289>.
- Paulmier A, Ruiz-Pino D. 2009. Oxygen minimum zones (OMZs) in the modern ocean. *Prog Oceanogr* 80:113–128. <https://doi.org/10.1016/j.pcean.2008.08.001>.
- Beman JM, Carolan MT. 2013. Deoxygenation alters bacterial diversity and community composition in the ocean's largest oxygen minimum zone. *Nat Commun* 4:1–11. <https://doi.org/10.1038/ncomms3705>.
- Parris DJ, Ganesh S, Edgcomb VP, DeLong EF, Stewart FJ. 2014. Microbial eukaryote diversity in the marine oxygen minimum zone off northern Chile. *Front Microbiol* 5:543. <https://doi.org/10.3389/fmicb.2014.00543>.
- Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, Mende DR, Letunic I, Rattei T, Jensen LJ, von Mering C, Bork P. 2019. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res* 47:D309–D314. <https://doi.org/10.1093/nar/gky1085>.
- Vik D, Gazitúa MC, Sun CL, Zayed AA, Aldunate M, Mulholland MR, Ulloa O, Sullivan MB. 2021. Genome-resolved viral ecology in a marine oxygen minimum zone. *Environ Microbiol* 23:2858–2874. <https://doi.org/10.1111/1462-2920.15313>.

32. Zehr JP, Ward BB. 2002. Nitrogen Cycling in the Ocean: new Perspectives on Processes and Paradigms. *Appl Environ Microbiol* 68:1015–1024. <https://doi.org/10.1128/AEM.68.3.1015-1024.2002>.
33. Wright JJ, Konwar KM, Hallam SJ. 2012. Microbial ecology of expanding oxygen minimum zones. *Nat Rev Microbiol* 10:381–394. <https://doi.org/10.1038/nrmicro2778>.
34. McLaren MR, Willis AD, Callahan BJ. 2019. Consistent and correctable bias in metagenomic sequencing experiments. *Elife* 8:e46923. <https://doi.org/10.7554/eLife.46923>.
35. Stock CA, Dunne JP, Fan S, Ginoux P, John J, Krasting JP, Laufkötter C, Paulot F, Zadeh N. 2020. Ocean Biogeochemistry in GFDL's Earth System Model 4.1 and Its Response to Increasing Atmospheric CO₂. *J Adv Model Earth Syst* 12:e2019MS002043. <https://doi.org/10.1029/2019MS002043>.
36. Waldbauer JR, Coleman ML, Rizzo AJ, Campbell KL, Lotus J, Zhang L. 2019. Nitrogen sourcing during viral infection of marine cyanobacteria. *Proc Natl Acad Sci U S A* 116:15590–15595. <https://doi.org/10.1073/pnas.1901856116>.
37. Zimmerman AE, Howard-Varona C, Needham DM, John SG, Worden AZ, Sullivan MB, Waldbauer JR, Coleman ML. 2020. Metabolic and biogeochemical consequences of viral infection in aquatic ecosystems. *Nat Rev Microbiol* 18:21–34. <https://doi.org/10.1038/s41579-019-0270-x>.
38. Gazitúa MC, Vik DR, Roux S, Gregory AC, Bolduc B, Widner B, Mulholland MR, Hallam SJ, Ulloa O, Sullivan MB. 2021. Potential virus-mediated nitrogen cycling in oxygen-depleted oceanic waters. *ISME J* 15:981–998. <https://doi.org/10.1038/s41396-020-00825-6>.
39. Jurgensen SK, Roux S, Schwenck SM, Stewart FJ, Sullivan MB, Brum JR. 2022. Viral community analysis in a marine oxygen minimum zone indicates increased potential for viral manipulation of microbial physiological state. *ISME J* 16:972–982. <https://doi.org/10.1038/s41396-021-01143-1>.
40. Canfield DE, Stewart FJ, Thamdrup B, De Brabandere L, Dalsgaard T, Delong EF, Revsbech NP, Ulloa O. 2010. A Cryptic Sulfur Cycle in Oxygen-Minimum-Zone Waters off the Chilean Coast. *Science* 330:1375–1378. <https://doi.org/10.1126/science.1196889>.
41. Foguel D, Teschke CM, Prevelige PEJ, Silva JL. 1995. Role of entropic interactions in viral capsids: single amino acid substitutions in P22 bacteriophage coat protein resulting in loss of capsid stability. *Biochemistry* 34:1120–1126. <https://doi.org/10.1021/bi00004a003>.
42. Lima SMB, Peabody DS, Silva JL, de Oliveira AC. 2004. Mutations in the hydrophobic core and in the protein–RNA interface affect the packing and stability of icosahedral viruses. *Eur J Biochem* 271:135–145.
43. Hevroni G, Flores-Urbe J, Béja O, Philofo A. 2020. Seasonal and diel patterns of abundance and activity of viruses in the Red Sea. *Proc Natl Acad Sci U S A* 117:29738–29747. <https://doi.org/10.1073/pnas.2010783117>.
44. Luo E, Leu AO, Eppley JM, Karl DM, DeLong EF. 2022. Diversity and origins of bacterial and archaeal viruses on sinking particles reaching the abyssal ocean. *ISME J* 16:1627–1635. <https://doi.org/10.1038/s41396-022-01202-1>.
45. Kempes CP, Wang L, Amend JP, Doyle J, Hoehler T. 2016. Evolutionary trade-offs in cellular composition across diverse bacteria. *ISME J* 10:2145–2157. <https://doi.org/10.1038/ismej.2016.21>.
46. Jover LF, Effler TC, Buchan A, Wilhelm SW, Weitz JS. 2014. The elemental composition of virus particles: implications for marine biogeochemical cycles. *Nat Rev Microbiol* 12:519–528. <https://doi.org/10.1038/nrmicro3289>.
47. Danovaro R, Corinaldesi C, Dell'anno A, Fuhrman JA, Middelburg JJ, Noble RT, Suttle CA. 2011. Marine viruses and global climate change. *FEMS Microbiol Rev* 35:993–1034. <https://doi.org/10.1111/j.1574-6976.2010.00258.x>.
48. Glass JB, Kretz CB, Ganesh S, Ranjan P, Seston SL, Buck KN, Landing WM, Morton PL, Moffett JW, Giovannoni SJ, Vergin KL, Stewart FJ. 2015. Metabolic signatures of microbial metal and nitrogen cycling in marine oxygen minimum zones. *Front Microbiol* 6:998. <https://doi.org/10.3389/fmicb.2015.00998>.
49. Ganesh S, Bertagnolli AD, Bristow LA, Padilla CC, Blackwood N, Aldunate M, Bourbonnais A, Altabet MA, Malmstrom RR, Woyke T, Ulloa O, Konstantinidis KT, Thamdrup B, Stewart FJ. 2018. Single cell genomic and transcriptomic evidence for the use of alternative nitrogen substrates by anammox bacteria. *ISME J* 12:2706–2722. <https://doi.org/10.1038/s41396-018-0223-9>.
50. Magoc T, Salzberg SL. 2011. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 27:2957–2963. <https://doi.org/10.1093/bioinformatics/btr507>.
51. Bengtsson-Palme J, Thorell K, Wurzbacher C, Sjöling A, Nilsson RH. 2016. Metaxa2 Diversity Tools: easing microbial community analysis with Metaxa2. *Ecol Informatics* 33:45–50. <https://doi.org/10.1016/j.ecoinf.2016.04.004>.
52. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. 2016. DADA2: high-resolution sample inference from Illumina amplicon data. *Nat Methods* 13:581–583. <https://doi.org/10.1038/nmeth.3869>.
53. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 41:D590–D596. <https://doi.org/10.1093/nar/gks1219>.
54. Li D, Liu CM, Luo R, Sadakane K, Lam TW. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31:1674–1676. <https://doi.org/10.1093/bioinformatics/btv033>.
55. Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, de Hoon MJL. 2009. BioPython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* 25:1422–1423. <https://doi.org/10.1093/bioinformatics/btp163>.
56. Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119–111. <https://doi.org/10.1186/1471-2105-11-119>.
57. Fu L, Niu B, Zhu Z, Wu S, Li W. 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28:3150–3152. <https://doi.org/10.1093/bioinformatics/bts565>.
58. Kultima JR, Sunagawa S, Li J, Chen W, Chen H, Mende DR, Arumugam M, Pan Q, Liu B, Qin J, Wang J, Bork P. 2012. MOCAT: a Metagenomics Assembly and Gene Prediction Toolkit. *PLoS One* 7:e47656. <https://doi.org/10.1371/journal.pone.0047656>.
59. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25:1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>.
60. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>.
61. Galperin MY, Makarova KS, Wolf YI, Koonin EV. 2015. Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic Acids Res* 43:D261–D269. <https://doi.org/10.1093/nar/gku1223>.
62. Benjamini Y, Speed TP. 2012. Summarizing and correcting the GC content bias in high-throughput sequencing. *Nucleic Acids Res* 40:e72. <https://doi.org/10.1093/nar/gks001>.
63. Dabney J, Meyer M. 2012. Length and GC-biases during sequencing library amplification: a comparison of various polymerase-buffer systems with ancient and modern DNA sequencing libraries. *Biotechniques* 52:87–94. <https://doi.org/10.2144/000113809>.
64. Dixon P. 2003. VEGAN, a package of R functions for community ecology. *J Veg Sci* 14:927–930. <https://doi.org/10.1111/j.1654-1103.2003.tb02228.x>.