Diverse Genomic Traits Differentiate Sinking-Particle-Associated versus Free-Living Microbes throughout the Oligotrophic Open Ocean Water Column

Andy O. Leu,[a]* John M. Eppley,[a] Andrew Burger,[a] (ID) Edward F. DeLong[a]

[a]Daniel K. Inouye Center for Microbial Oceanography: Research and Education (C-MORE), University of Hawaii, Honolulu, Hawaii, USA

ABSTRACT Bacteria and archaea are central to the production, consumption, and remineralization of dissolved and particulate organic matter and contribute critically to carbon delivery, nutrient availability, and energy transformations in the deep ocean. To explore environmentally relevant genomic traits of sinking-particle-associated versus free-living microbes, we compared habitat-specific metagenome-assembled genomes recovered throughout the water column in the North Pacific Subtropical Gyre. The genomic traits of sinking-particle-associated versus free-living prokaryotes were compositionally, functionally, and phylogenetically distinct. Substrate-specific transporters and extracellular peptidases and carbohydrate-active enzymes were more enriched and diverse in particle-associated microbes at all depths than in free-living counterparts. These data indicate specific roles for particle-attached microbes in particle substrate hydrolysis, uptake, and remineralization. Shallow-water particle-associated microbes had elevated genomic GC content and proteome nitrogen content and reduced proteome carbon content in comparison to abyssal particle-associated microbes. An inverse trend was observed for their sympatric free-living counterparts. These different properties of attached microbes are postulated to arise in part due to elevated organic and inorganic nitrogen availability inside sinking particles. Particle-attached microbes also were enriched in genes for environmental sensing via two-component regulatory systems, and cell-cell interactions via extracellular secretion systems, reflecting their surface-adapted lifestyles. Finally, particle-attached bacteria had greater predicted maximal growth efficiencies than free-living bacterioplankton at all depths. All of these particle-associated specific genomic and proteomic features appear to be driven by microhabitat-specific elevated nutrient and energy availability as well as surface-associated competitive and synergistic ecological interactions. Although some of these characteristics have been previously postulated or observed individually, we report them together here in aggregate via direct comparisons of cooccurring free-living and sinking-particle-attached microbial genomes from the open ocean.

IMPORTANCE Particle-attached microbes play large roles in the ocean carbon cycle and help to sequester atmospheric CO_2 and to deliver nutrients and energy on sinking particles to the deep sea. Here, we report on the genomic traits of particle-attached versus free-living microbes throughout the ocean water column to better differentiate their specific metabolic and ecological roles in the sea. In general, the genomic properties and contents of particle-attached microbes reflected the physical and chemical compositions of their environment as well as their microhabitat-specific adaptive traits. In comparison to cooccurring free-living microbes, particle-attached microbes had larger genomes, greater capacity for extracellular polymer degradation, greater environmental sensing and response capacity, greater potential for motility and attachment, and higher growth efficiencies. Our results present an integrated new perspective on

sinking-particle-attached microbial adaptive traits that contribute to their critical ecological and biogeochemical roles and activities in the sea.

KEYWORDS carbon pump, deep sea, free-living, marine microbes, particle-attached

In the ocean, heterotrophic prokaryotes play a significant role in remineralizing and transforming large fractions of marine organic matter derived from phytoplankton biomass (1, 2). This microbially transformed carbon can be released as CO_2 into the atmosphere, transformed into dissolved organic carbon (which constitutes one of the largest actively cycled carbon reservoirs [3]), or incorporated into particulate organic carbon (POC) of living cells and decaying particulate organic matter (POM) (4). While most of this organic material is consumed in the photic zone, sinking particles that do reach abyssal depths comprise an important component of the biological pump and also provide essential nutrients and energy to the deep-sea ecosystem (5).

Particle-associated (PA) microbes may play crucial roles in POM degradation, in part due to their potential for producing extracellular enzymes that hydrolyze POC and dissolved organic matter (DOM) to lower-molecular-weight nutrients prior to cellular uptake (6, 7). These extracellular enzymes may also benefit other microbes in the PA community that lack extracellular enzymes by supplying them with lower-molecular-weight carbon sources (8, 9). Thus, sinking particles may provide a unique environmental condition with higher extracellular enzyme activity, biogeochemical transformation rates, bacterial cell densities, and overall microbial activity, in contrast to the free-living (FL) environments (10–13), leading to their contrasting microbial compositions (14, 15). However, PA communities are also highly diverse and heterogenous and may be strongly influenced by sampling regime and particle types (16–21), residence times (22), geographic location (23), depths, and seasonality (20, 24, 25). Consequently, the nature and composition of sinking particles remain to be fully characterized. Additionally, a deeper understanding of microbially driven organic carbon transformations is hampered by the complexity of processes involved in organic matter degradation as well as the challenges of linking specific microbial lineages involved with the relevant molecular processes.

Early studies contrasting coexisting PA and FL communities focused on 16S rRNA gene cloning and sequencing approaches to differentiate communities in these different habitats (14, 26). More recently, PCR-based amplicon sequencing (17, 21, 27–33), metagenomic analyses using gene-centric approaches (16, 18, 20, 28, 34–38), and examination of single amplified genomes (SAGs) have also been utilized (35). Different sampling approaches for assessing PA microbial communities have centered on collecting suspended, filter-fractionated particles (16, 28, 30, 32, 37–39) and direct collection in syringes *in situ* by divers (14) or via trapping of sinking particles by sediment traps (17, 18–21, 33–35). With respect to FL marine prokaryote communities, many metagenome-assembled genomes (MAGs) have also been recovered from the *Tara* Oceans and *Malaspina* sampling campaigns and other studies, spanning many samples and phyla (20, 40–42).

To further explore genomic differences that may exist between sinking-particle-associated (SPA) and FL microbes found throughout the water column in the oligotrophic Pacific Ocean, we leveraged time series metagenomic samples recovered at Station ALOHA in the North Pacific Subtropical Gyre, generating a total of 407 mid- to high-quality MAGs from both sample types. FL microbes (captured on 0.2-μm-pore-size filters) were collected from surface waters to a depth of 4,000 m at Station ALOHA, and metagenomes were generated, assembled, and analyzed (43–45). In parallel, 69 metagenomes collected from SPA microbes collected in sediment traps at 150 m and 4,000 m were similarly investigated. Comparative genome analysis of FL and SPA prokaryotes throughout the water column was conducted to determine microbial community composition, metabolic potential, and genomic characteristics. Our results reveal habitat-specific genomic traits that differentiate SPA from FL prokaryotes from surface waters to the abyss that relate to their organic matter utilization, metabolic lifestyles, and genomic and proteomic
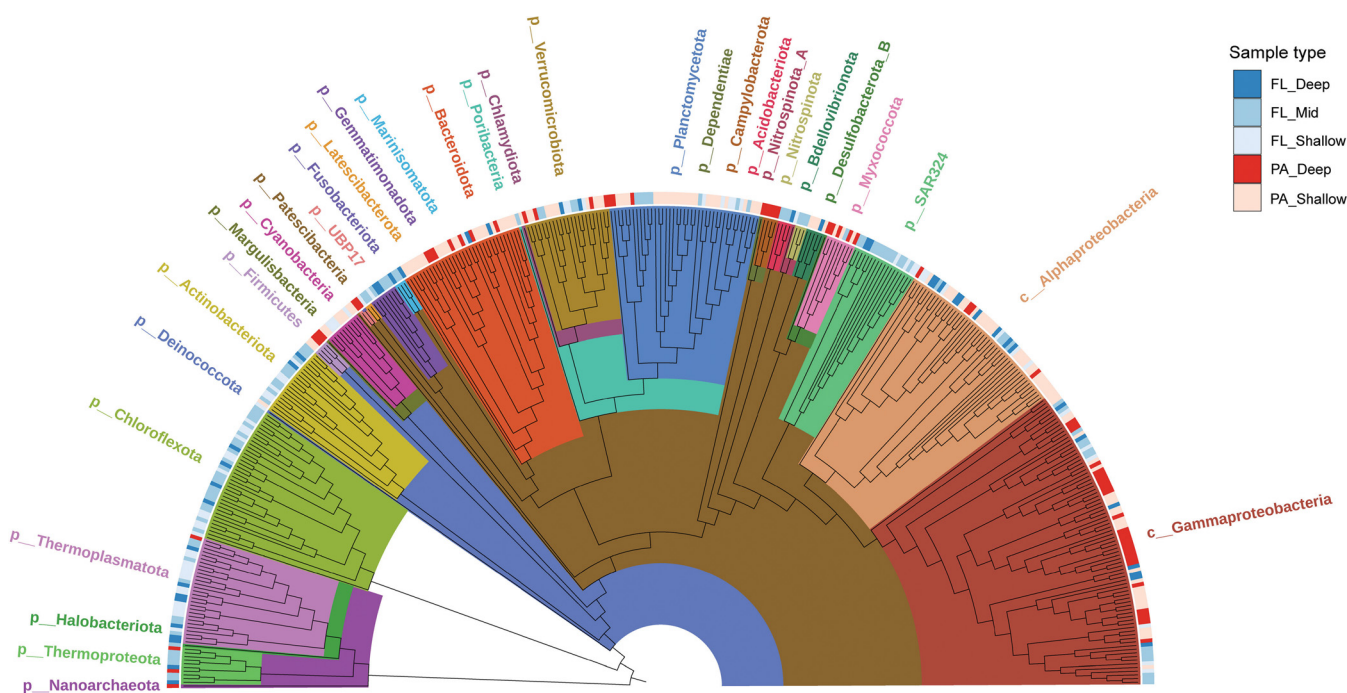
FIG 1 Sinking-particle-attached microbes are phylogenetically distinct from their free-living counterparts recovered from similar depths. Phylogenetic trees of the 407 dereplicated MAGs inferred using maximum likelihood analysis with a concatenated set of 122 archaeon-specific and 120 bacterium-specific marker genes. Phylum- and class-level taxonomic classifications were determined based on GTDB-Tk, GTDB v202. Each MAG was also annotated based on its sample type designation: FL, free-living; PA, particle-associated; FL_Shallow, 0 to 150 m; FL_Mid, >150 to 1,000 m; FL_Deep, 4,000 m; PA_Shallow, 150 m; PA_Deep, 4,000 m. See Table S2 in the supplemental material for more detailed taxonomic designations.

properties. The observed niche partitioning and habitat-specific genomic and proteomic trends identified differential adaptive characteristics and metabolic potential of SPA versus FL microbes and further illustrate their various roles in the biogeochemical transformation and organic carbon transformations in the ocean.

RESULTS

Recovery of MAGs from free-living and particle-associated bacteria and archaea. To expand the genomic data from sympatric SPA and FL prokaryote communities throughout the water column, metagenomic data sets were prepared from time series samples collected from the surface to 4,000 m during Station ALOHA time series efforts and expeditions (43–47). These included 63 metagenomic samples from a time series analysis of SPA microbes reaching 4,000 m at Station ALOHA from May 2014 to November 2016 (18, 20), six metagenomic samples collected from SPA microbes at 150 m at Station ALOHA from July to August in 2015, and Hawaii Ocean time series samples collected at depths between 5 m and 4,000 m from 2014 to 2017 (43–46) (see Table S1 in the supplemental material). A total of ~5.5 Tb of metagenomic sequence data was generated and used in the MAG assemblies. Each depth- and habitat-specific metagenome sample was assembled individually, and metagenomic binning was conducted to recover metagenome-assembled genomes (MAGs). In addition, quality control was performed to remove low-quality or redundant genomes, yielding a total of 407 medium- to high-quality MAGs (each with more than 70% completeness and less than 10% contamination) (see Materials and Methods) (Table S1).

The FL and PA MAGs were classified into five sample types (Fig. 1). FL MAGs were separated into FL_Shallow, FL_Mid, and FL_Deep based on their relative abundances throughout the water column. MAGs classified as FL_Shallow were most abundant at 150 m or above, those classified as FL_Mid were most abundant from 175 m to 1,000 m, and those classified as FL_Deep were most abundant at 4,000 m. SPA MAGs that were recovered from the 150-m sediment traps in August 2015 (47) were classified as PA_Shallow. SPA Mags recovered from the 4,000-m sediment traps were classified as

PA_Deep, except for those whose genomic and physiological features and distributions verify their surface water origins (20) (see Materials and Methods) (Table S2). The numbers of MAGs recovered from each of the different sample types were as follows: FL_Shallow, 64; FL_Mid, 105; FL_Deep, 43; PA_Shallow, 130; and PA_Deep, 65.

Phylogenetic classification of bacterial and archaeal MAGs. To resolve the taxonomic distribution of the FL and SPA MAGs, we evaluated their placement in a reference genome tree (48, 49) (see Materials and Methods). The MAGs encompassed 28 bacterial and 4 archaeal phyla (Fig. 1; Table S2). The five most abundant bacterial phyla in our data set were *Proteobacteria*, including *Alphaproteobacteria*, *Gammaproteobacteria*, *Planctomycetota*, *Chloroflexota*, *Bacteroidota*, and *Verrucomicrobiota*. Ten other bacterial phyla were each represented by only a single MAG. This included the poorly characterized phylum UBP17, formerly known as *Cloacimonetes* (WWE1), of which only a few MAGs have been included in the genome taxonomy database to date (49). Another MAG closely related at the genus level to the UBP17 MAG recovered was the NORP72 MAG (50), which shared an average amino identity of 72.97%. This previously reported MAG was recovered from a sediment basin sample collected at ~4,484 m in the Mid-Atlantic Ridge (50), consistent with the abyssal depth origin of the UBP17 MAG reported here, which was classified as PA_Deep.

MAGs from a total of four archaeal phyla were recovered in our Station ALOHA MAG data set, which included 29 from *Thermoplasmatota*, 10 from *Thermoproteota*, 1 from *Nanoarchaeota*, and 1 from *Halobacteriota*. In contrast to the FL samples, only four archaeal MAGs were recovered from the PA communities and were from the PA_Deep sample type. With respect to genera, these MAGs included close relatives of *Nitrosopelagicus*, *Nitrosopumilus*, *Thalassarchaeum*, and *Nanoarchaeia*. While the North Pacific Subtropical Gyre (NPSG) FL and PA MAGs shared some taxa at higher classification levels, only nine genera were found in both the FL and PA data sets. These included MAGs closely related to *Alcanivorax*, *Prochlorococcus*_A, *Thalassarchaeum*, *Henriciella*, *Idiomarina*, *Nitrosopumilus*, *Nitrosopelagicus*, *Roseibacillus*_B, and *Bythopirellula* (Table S2). These data support and extend previous observations suggesting that microorganisms found in PA versus FL communities are phylogenetically distinct (14, 16, 18, 20, 26, 35, 36, 51).

Functional genes differentiate environmental sample types. Metabolic reconstruction was conducted to identify functional differences between the PA and FL MAGs. Protein coding sequences were annotated with KEGG, generating 9,351 unique KEGG Orthology (KO) identifiers (IDs) and clustered into 49,146 orthologous protein families. Nonmetric multidimensional scaling (NMDS) profiles based on the presence of KO IDs (Fig. S1A) and orthologous protein families (Fig. S1B) present in each MAG were generated. Globally, both the KO and orthologous protein family NMDS profiles showed strong clustering of the MAGs according to their phylum-level composition.

To identify genes that were significantly associated with a given sample type, we used Scoary (52), a stringent, combined statistical test that leverages the presence/absence profiles of KO IDs and orthologous protein families (52). Overall, the PA_Shallow and PA_Deep samples showed greater numbers of genes that were significantly enriched than the FL sample types (false discovery rate-corrected P value < 0.05) (Fig. S2A and B and Tables S3A, S3B, and S4). The PA_Shallow and PA_Deep MAGs shared a large subset of significantly enriched genes in terms of both KO and orthologous protein family annotation (545 and 290, respectively) (Fig. S2A and B). For FL samples, only FL_Shallow and FL_Mid showed some gene overlap. No overlap of significantly enriched genes was found between FL and PA samples in terms of orthologous protein families. However, four KO IDs were shared between FL_Shallow and PA_Shallow that encoded zinc/manganese transport system ATP-binding protein as well as deoxyribodipyrimidine photolyase, bacteriorhodopsin, and phytoene desaturase (involved in carotenoid synthesis). Most of these genes are involved in light-driven energy generation or DNA damage repair, consistent with their surface origins. Two KO IDs were found shared between FL_Deep and PA_Deep that encoded the mercuric ion transport protein (merT) and

cobalt-zinc-cadmium efflux system protein (czcD), suggesting potentially higher concentrations or bioaccumulation of these cationic metals at greater depths.

The most enriched KO IDs identified by Scoary were significantly associated with the PA_Shallow MAGs (Table S3A and B). These genes encoded dihydrolipoamide dehydrogenase (DLD), i.e., an oxidoreductase linked to central metabolism, thioredoxin, which is involved in defending against oxidative stress, 1-acyl-sn-glycerol-3-phosphate acyltransferase, which is involved in glycerolipid metabolism, and most prevalent of all, a large family of proteins categorized as putative ABC transporters (Table S3A and B).

Substrate uptake capabilities of FL versus PA bacteria and archaea. Previous studies have shown that transporters of carbohydrates and energy sources are ubiquitous in heterotrophic prokaryotic communities throughout the open water column, from the surface to abyssal depths, and are crucial for organic matter transformation (53). To investigate essential nutrient uptake capabilities, transporters with compound specificity were examined in the different sample types (Fig. S3 and Table S4A). In general, MAGs in the SPA samples (both surface and deep) encoded a higher diversity of transporters than FL MAGs. Genes encoding transporters with substrate affinity to amino acids, aromatic compounds, carbohydrates, metal, nitrate/nitrite, peptide, phosphate, and spermidine/putrescine were more highly represented in SPA MAGs than FL MAGs. The PA_Shallow group had a relatively greater representation of MAGs (9.2%) that included genes for taurine transporters than did the PA_Deep group (1.5%) and the FL sample types (0 to 2.9%). A large percentage of MAGs in all sample types included genes for the assimilatory sulfate reduction pathway. Six percent of the PA_Shallow MAGs contained sulfate/thiosulfate transporter genes that were not identified in other sample types. The absence of the sulfate/thiosulfate transporters in some SPA microbes may indicate their reliance instead on reduced sulfur compounds, thereby alleviating energetic demands for transport and assimilatory reduction of sulfate (54). The PA_Deep group had the greatest relative proportion of MAGs encoding amino acids, aromatic compounds, di-oligopeptides, nitrate/nitrite, peptide, and spermidine/putrescine transporters in comparison to PA_Shallow and all FL MAG groups (Fig. S3; Table S4A), suggesting that deep-sea SPA microbes may be better adapted to incorporating these substrates at greater depths.

A greater proportion of MAGs encoding mannopine transporters were identified in the FL samples than in the PA samples (~6.2% versus ~1.2%) (Fig. S3; Table S4A). The substrate mannopine and other opine-like compounds are secondary amine derivatives that are utilized by many different bacteria as carbon, energy, and nitrogen sources (55). While opines are commonly found in crown gall tumor tissues in plants, other potential sources of opines have been isolated from marine invertebrates, such as mollusks, cnidarians, and sponges (56). This suggests that the mannopine-like transporters in FL_MAGs may acquire dissolved mannopine (or other opine-like substrates) derived from marine invertebrates or eukaryotic sources.

Polymer-degrading enzymes of PA versus FL bacteria and archaea. Given the higher prevalence and diversity of transporters in SPA MAGs, POC solubilization capabilities of both SPA and FL MAGs were investigated. Extracellular hydrolytic enzymes that include peptidases and carbohydrate-active enzyme (CAZymes), key components in particulate organic carbon hydrolysis, were examined in terms of their abundance, diversity, and function (Fig. 2; Table S4B and C).

The SPA MAGs possessed a significantly higher count of genes encoding peptidases and CAZymes per genome than the FL MAGs (Fig. 2A and B). A similar trend could also be observed for genes encoding secretory peptidases and CAZymes, with SPA MAGs showing higher average percentages of these secreted extracellular enzymes than FL MAGs (Fig. 2C and D). A significantly higher percentage of secretory peptidases were also observed in FL_Deep MAGs than the other FL sample types. In terms of functional diversity, the PA MAGs showed the highest overall diversity of peptidase and CAZyme classes per genome in comparison to the FL MAGs (Fig. 2E and F). Within the FL MAGs, the FL_Deep MAGs had a higher diversity of peptidase classes than MAGs of other depths and a CAZyme diversity similar to that of FL_Shallow MAGs.
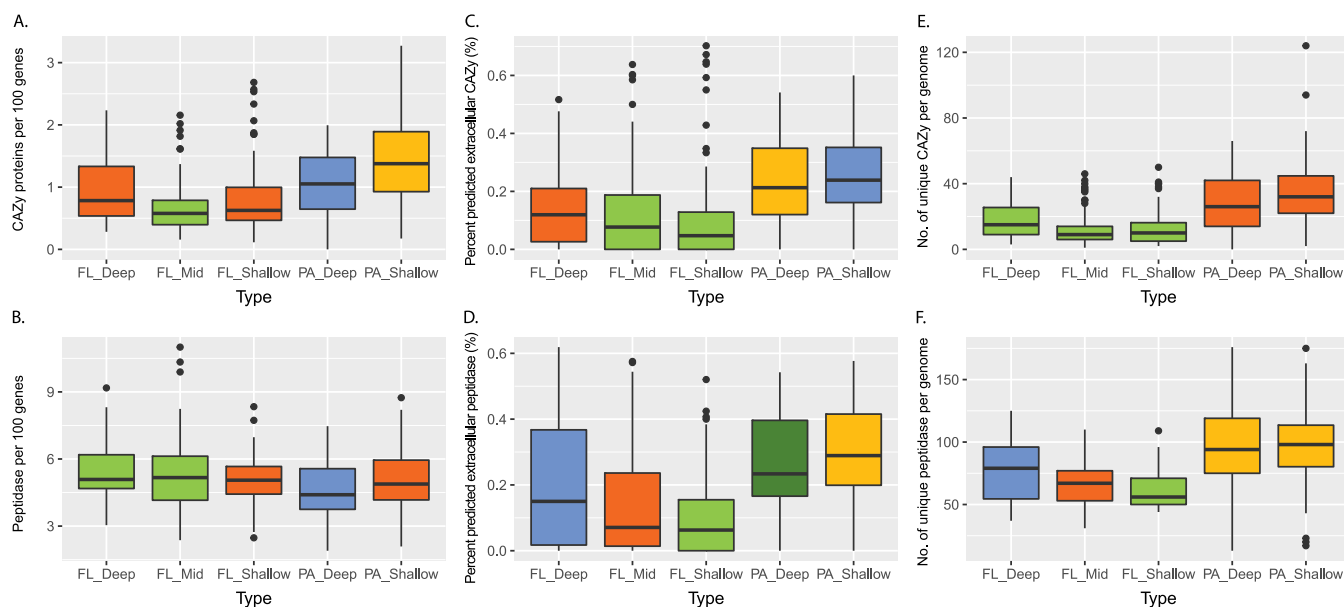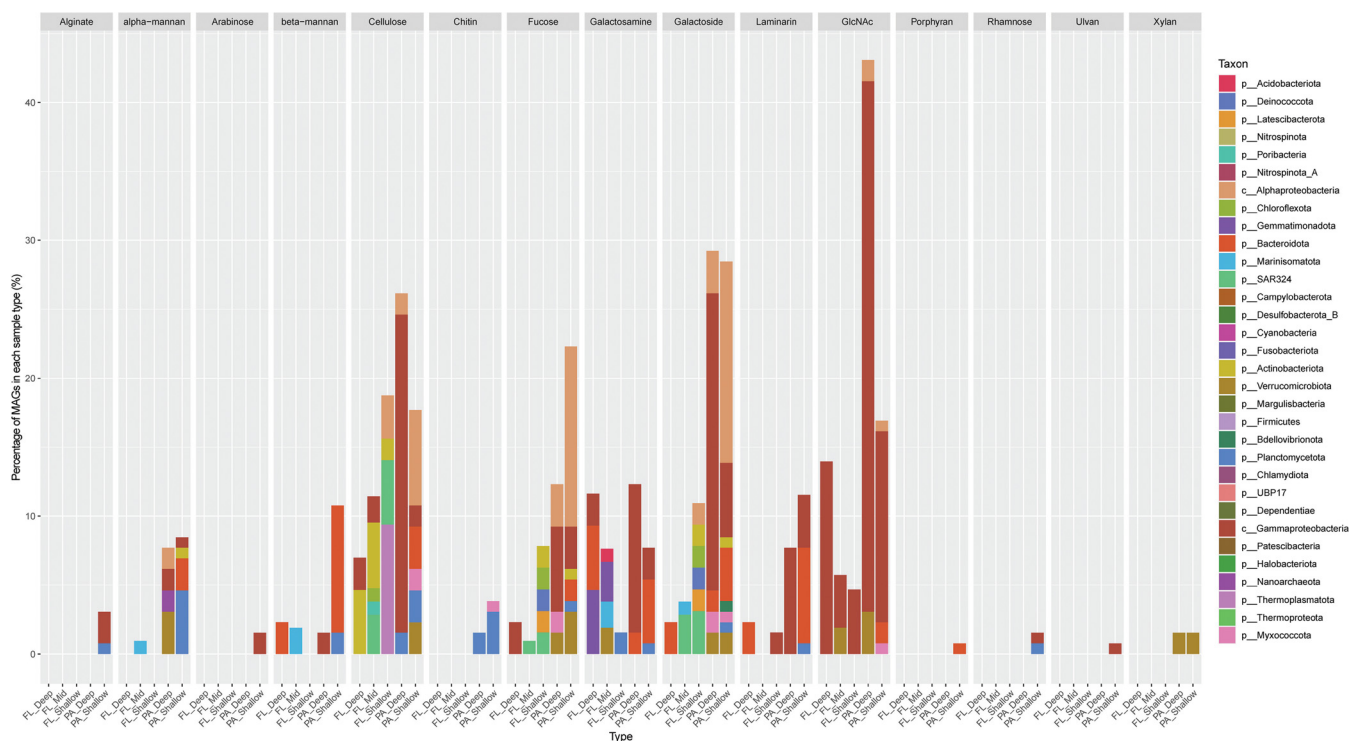
FIG 2 Sinking-particle-attached microbes contain relatively more secretory CAZyme and peptidase genes than do their free-living counterparts. (A and B) Counts of genes encoding CAZymes (A) and peptidases (B) per 100 genes in MAGs. (C and D) Percentages of genes encoding extracellular CAZymes and extracellular peptidases (D) in MAGs. (E and F) Numbers of unique CAZymes (E) and peptidases (F) per genome. Shared color denotes no significant difference. Box plots show medians and enclose 1st and 3rd quartiles. Whiskers denote 1.5 times the interquartile range (IQR) of the lower and upper quartiles. Statistics are based on the Tukey-Kramer test, and different colors denote significant differences ($P < 0.05$). PA, particle-associated.; FL, free-living; FL_Shallow, 0 to 150 m; FL_Mid, >150 to 1,000 m; FL_Deep, 4,000 m; PA_Shallow, 150 m; PA_Deep, 4,000 m.

Polysaccharide degradation potential in SPA versus FL bacteria and archaea.
Since polysaccharides constitute a large fraction of DOC (57) and POC (58), identifying genes encoding CAZymes and their substrate specificity in MAGs is crucial for determining the putative carbon source as well as the specific microbial members driving organic carbon transformation. Putative CAZymes with fully defined EC numbers (complete with four identifiers) were used to identify specific genes encoding glycoside hydrolase (GH) and polysaccharide lyase (PL) and the putative polysaccharides they target. In total, 45 EC numbers were identified that correspond to potential hydrolysis of 34 different putative substrates, which included common carbohydrates (e.g., starch) as well as marine relevant polysaccharides such as laminarin, xylan, and chitin (7, 59–61). PA_Shallow possessed greater diversity and a higher percentage of microbes capable of hydrolyzing a large variety of polysaccharides, followed by PA_Deep and the FL samples (Fig. 3; Fig. S4 and Table S4B).

Only one CAZyme, α-l-arabinofuranosidase (GH51), was found to be enriched in FL MAGs in comparison to PA MAGs (Table S4B). This enzyme is linked to the cleavage of α-arabinofuranoside from hemicellulose (62). In contrast, a large number of biochemically well characterized CAZymes (63–65) were present only in SPA MAGs, including β-xylosidases (GH52), ulvan lyase (PL24 and PL25), rhamnogalacturonan α-L-rhamnohydrolase (GH106), α-L-rhamnosidase (GH106), alginate lyase (PL6 and -14), endo-α-1,5-L-arabinanase (GH43), β-L-arabinofuranosidase (GH146), chitinase (GH18), and β-porphyranase (GH16). These CAZymes hydrolyze polysaccharides such as alginate, porphyran, ulvan, and xylan, which are structural components of algae, and chitin, a highly abundant polysaccharide in marine environments. These CAZymes also hydrolyze an assortment of polysaccharides, releasing sugars such as rhamnose, xylan, and arabinose, which are found enriched in spring phytoplankton blooms (66).

Other marine relevant CAZymes that are enriched in SPA MAGs also include α-mannosidase (GH38 and GH92), β-mannosidase (GH2), α-L-fucosidase (GH29 and GH151), β-glucosidase (GH1, -2, and -3), cellulase (GH5, -6, and -45), cellobionic acid phosphorylase (GH94), α-N-acetylgalactosaminidase (GH109 and -114), β-galactosidase (GH2 and -42), and α-galactosidase (GH4, -31, -36, and -97). The enrichment of these CAZymes in
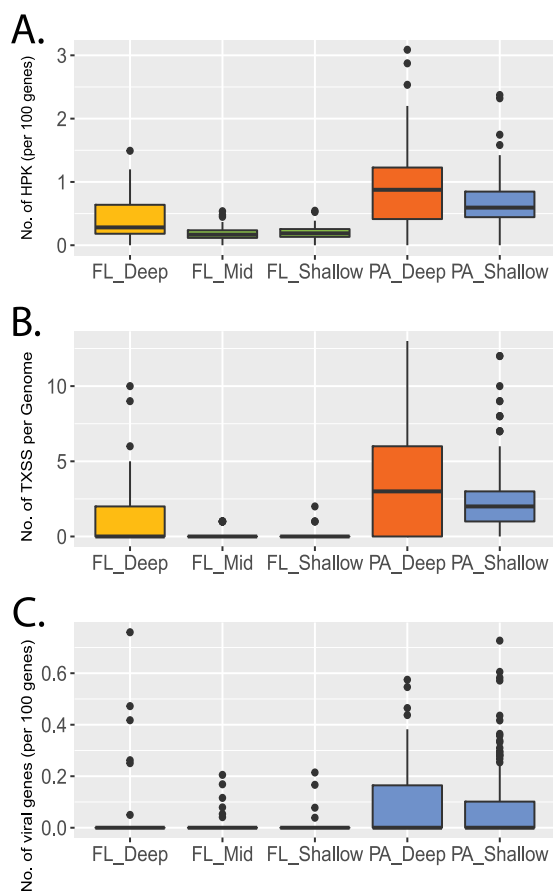
FIG 3 Sinking-particle-attached microbes contain relatively more polysaccharide-degrading genes than do their free-living counterparts. Percentages of genomes encoding substrate-specific CAZymes (see Table S4B in the supplemental material), separated by sample types. MAGs are colored based on their phylum- or class-level taxonomic classification. PA, particle-associated; FL, free-living; FL_Shallow, 0 to 150 m; FL_Mid, >150 to 1,000 m; FL_Deep, 4,000 m; PA_Shallow, 150 m; PA_Deep, 4,000 m.

SPA MAGs also correlates well with the spike in multiple sugars detected during summer phytoplankton blooms, such as mannan, fucose, glucose, galactosamine, and galactose (67).

The most enriched CAZymes identified in any sample type were in the PA_Deep sample, where 43.1% of all MAGs contained genes for β-N-acetylhexosaminidase (GH3, -20, -84), which is involved in the hydrolysis of N-acetylated oligo/polysaccharides such as chitooligosaccharides and chitin, which are prevalent and abundant in marine crustaceans. MAGs containing genes for β-N-acetylhexosaminidase were present only within the *Gammaproteobacteria* and include the orders *Arenicellales*, *Burkholderiales*, *Enterobacterales*, *Nitrosococcales*, *Pseudomonadales*, UBA11654, and *Xanthomonadales*. Lastly, genes encoding laminarin endo-1,3-β-D-glucosidase (GH16 and -81) were found to be enriched in 25 SPA MAGs, which degrades laminarin, a polysaccharide abundant in microalgae such as diatoms (61). These SPA MAGs belonged to multiple phyla, including *Proteobacteria*, *Planctomycetota*, *Verrucomicrobiota*, and *Bacteroidota*. Most of the *Bacteroidota* MAGs belonged to the *Flavobacteriaceae* family, which are well known degraders of polysaccharides and associated with phytoplankton blooms (67). The presence of these CAZymes in the PA_Shallow and PA_Deep MAGs suggests that many of these polysaccharides are likely more available as substrates in POC and accessible as polymeric substrates for SPA microorganisms. Our observations here are consistent with a recent study (68) that suggested that particle-associated enzymatic profiles had a greater representation of peptidases and a broader spectrum of polysaccharide hydrolases than did corresponding bulk seawater samples.

Enrichment of TCSs, extracellular secretion systems, and prophage in SPA bacteria and archaea. Two-component sensory systems (TCSs) enable rapid transcriptional responses to environmental variations, including light, temperature, and nutrient availability (69). TCSs are composed of a response regulator protein and a histidine kinase, which when activated by a molecule or by other physical stimuli induce a conformational change and interact with the response regulator protein. This typically stimulates

FIG 4 Sinking-particle-attached microbes contain more histidine kinase, extracellular secretion system, and prophage-associated genes than do their cooccurring free-living counterparts. Variance of histidine kinase (A), secretion system (B), and virus-associated (C) genes in MAGs from different sample types. Histidine kinases were identified based on hits to pfam02518, pfam13581, pfam13589, pfam14501, and pfam07536. Secretion system genes include genes encoding bacterial flagellum, T1SS, T2SS, T4P, T5aSS, T5bSS, T6SSi, T9SS, Tad(tight_adherence), and pT4SSt secretion systems. Prophage-associated gene markers included genes encoding phage CI repressor, phage capsid proteins, phage excisionase, and phage tail proteins. Box plots show medians and enclose 1st and 3rd quartiles. Whiskers denote 1.5 times the IQR of the lower and upper quartiles. Statistics are based on the Tukey-Kramer test, and different colors denote significant differences ($P < 0.05$). Shared color denotes no significant difference. PA, particle-associated; FL_Shallow, 0 to 150 m; FL_Mid, >150 to 1,000 m; FL_Deep, 4,000 m; PA_Shallow, 150 m; PA_Deep, 4,000 m.

binding of the response regulator to a DNA promoter region, allowing transcription of the genes in the downstream operon (70). Prior studies have also indicated that a lack of two-component regulatory systems in bacterial genomes may be a hallmark of oligotrophy for bacteria (69).

Examination of the SPA and FL MAGs for genes encoding TCSs showed significantly higher representation of histidine kinases (HPK; normalized to total gene content) in PA MAGs than in FL MAGs (Fig. 4A). In addition, PA_Deep MAGs possessed higher numbers of HPK than PA_Shallow MAGs, consistent with the higher organic carbon energy content previously documented for abyssal sinking particles (25). With respect to FL microbes, FL_Deep genomes had a higher HPK gene content than FL_Shallow and FL_Mid genomes. In conjunction with the higher percentage of secreted peptidases (Fig. 2C) found in FL_Deep MAGs, this suggests a more frequent association of deep-sea microbes with particles (36).

Extracellular protein secretion systems (ESSs) are widespread among bacteria and archaea and are central components in fimbria- and pilus-associated attachment and adhesion, nutrient transport, predation, cell-cell interactions, surface colonization, and pathogenicity in diverse environmental contexts and settings (71–74). The diversity of ESS

processes can be catalogued in part by the underlying mechanism of transport for any given ESS (75). Examination of the SPA and FL microbes showed that PA_Deep and PA_Shallow were significantly enriched in ESSs, compared to FL MAGs (Fig. 4B; Fig. S5 and Table S4D). In particular, SPA microbes had greater representation of genes coding for Tad systems, as well as ESS types I, II, III, IV, V, and IX, in their genomes than FL microbes at all depths. Two exceptions, however, included type 5cSS and type 4SSt, whose representations in FL and SPA microbial genomes were generally equivalent. SPA MAGs also had a higher representation of flagella, closely followed by a similar representation in FL_Deep microbes (Fig. S5). With respect to FL MAGs, FL_Shallow and FL_Mid genomes had a lower genomic representation of ESSs than their FL_Deep counterparts (Fig. 4B; Fig. S5 and Table S4D).

Prior work has suggested that deep-sea bacteria may have larger genomes than bacteria in the epipelagic zone (46, 76), and our observations here show that particle-attached bacteria on average have larger genomes than sympatric free-living counterparts. This provides potentially more genome "real estate" for expanded metabolic versatility and potentially greater capacity to accommodate mobile genetic elements. Examination of phage-associated genes in SPA versus FL MAGs revealed that SPA microbes (regardless of depth horizon) had a higher representation of phage genes, including both temperate phage markers and phage capsid structural genes (Fig. 4C). These data are consistent with the expectation that the larger genomes of particle-attached and deep-sea bacteria can accommodate greater numbers of mobile genetic elements.

Intrinsic genomic and proteomic traits differentiate SPA from FL bacteria and archaea. PA microbial communities are often viewed as "hot spots" for microbial activity in comparison to their FL counterparts, based on their larger cell size and cell densities, higher enzyme activities, and overall rates of heterotrophic microbial metabolism (10, 12, 13). This is a result of the aggregation of phytodetrital matter, fecal pellets, and high-molecular-weight nutrient-rich organic material such as proteins and polysaccharides, which are consumed by heterotrophic bacteria (77–79). Thus, nutrient-replete microhabitats on particles may influence the carbon and nitrogen budgets and core genomic properties (44, 80–83).

To further test whether observable genomic variability in MAGs was evident across multiple sample types and depths, we examined variability in a variety of genomic traits in SPA versus FL microbes throughout the water column. Analysis of these MAG genomic traits showed that SPA prokaryotes (from both deep and shallow water) had significantly larger genome sizes on average than FL microbes (from all depths) (Fig. 5A). In terms of GC content, there was a significant increase from FL_Shallow to FL_Deep MAGs, whereas PA_Deep MAGs had a lower GC content than PA_Shallow MAGs (Fig. 5B). In the context of predicted protein elemental composition, the average number of nitrogen atoms per amino acid residue side chain (N-ARSC) generally correlates with GC content variability in cultivar genomes (83). The results for the FL MAGs were consistent with recent gene-centric analyses showing increases in both GC content and N-ARSC values that tracked *in situ* inorganic nitrogen availability and depth for FL bacterioplankton at Station ALOHA (Fig. 5C) (44). Notably, the PA MAGs showed a depth trend opposite to that of FL MAGs, with PA_Deep MAGs having lower GC and N-ARSC values than PA_Shallow MAGs (Fig. 5C). As for the average number of carbon atoms per amino acid residue side chain (C-ARSC), only the PA_Deep MAGs possessed statistically higher values than all other sample types (Fig. 5D). These genomic trends in SPA bacteria may be driven by the elemental stoichiometry of organic carbon in shallow versus deep sinking particles, since shallow-water particles trend toward a higher organic nitrogen content, whereas sinking particles reaching the abyss are relatively depleted in organic nitrogen (25).

Growth efficiency estimates for PA versus FL bacteria and archaea. Recently, codon usage bias (CUB) metrics in genes encoding ribosomal proteins have been used to estimate the maximum growth rates (minimum doubling times) of microorganisms in both laboratory and environmental settings (84–87). We therefore tested whether CUB metrics might differentiate between the different SPA and FL sample types and
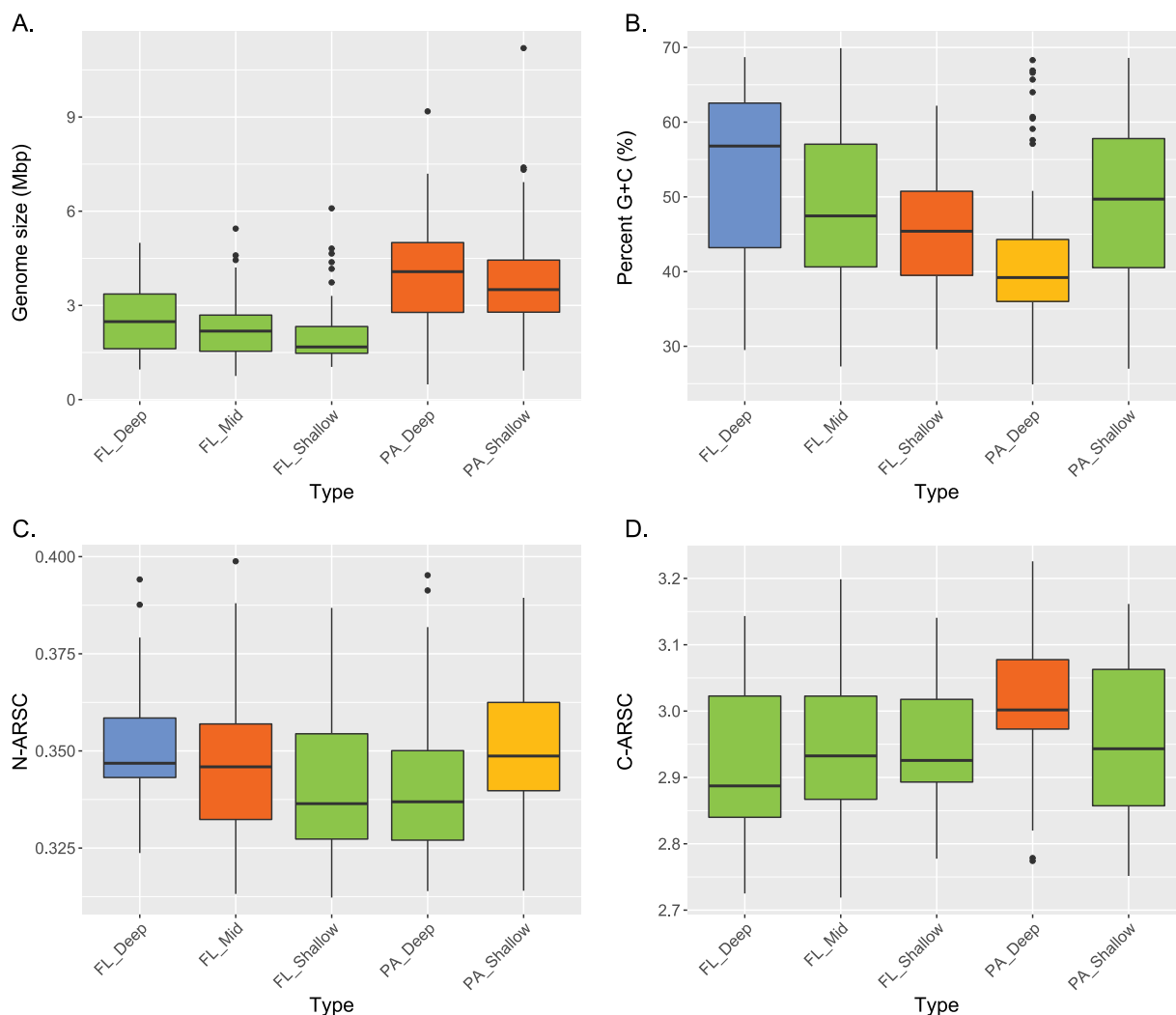
FIG 5 Sinking-particle-attached microbes have different genomic and proteomic compositions than their free-living counterparts. Variances of genome size (A), GC content (B), N-ARSC (C), and C-ARSC (D) in MAGs from the different sample types. Box plots show medians and enclose 1st and 3rd quartiles. Whiskers denote 1.5 times the IQR of the lower and upper quartiles. Statistics are based on the Tukey-Kramer test, and different colors denote significant differences ($P < 0.05$). Shared color denotes no significant difference. PA, particle-associated; FL_Shallow, 0 to 150 m; FL_Mid, >150 to 1,000 m; FL_Deep, 4,000 m; PA_Shallow, 150 m; PA_Deep, 4,000 m.

genomes (Fig. 6; Table S5). Using a Gaussian mixture model, the distribution of the SPA and FL MAGs separated into three clusters with predicted mean minimum doubling times of 4.7, 16.9, and 36.7 h, respectively (Fig. 6A). These three clusters did not correspond to depth of origin, even though they spanned a range of 4,000 m, with pressure differences of >400 atm and temperature differences of >22°C. Instead, the three clusters were strongly associated with PA_Shallow and PA_Deep in one cluster, and the FL samples split between two clusters. Specifically, the majority of PA_Deep and PA_Shallow MAGs had greater predicted growth efficiencies (as estimated by CUB-based theoretical maximum growth rates) than did the FL MAGs (Fig. 6B and C). While individual MAGs may differ in part due to phylogenetic trends, the overall higher predicted growth efficiencies among PA prokaryotes than among FL prokaryotes generally supports the postulated copiotrophic lifestyles of SPA microbes compared to free-living oligotrophs throughout the water column.

DISCUSSION

The application of phylogenetic, genomic, and metagenomic approaches to the study of particle-attached versus free-living microbes is now providing new insights
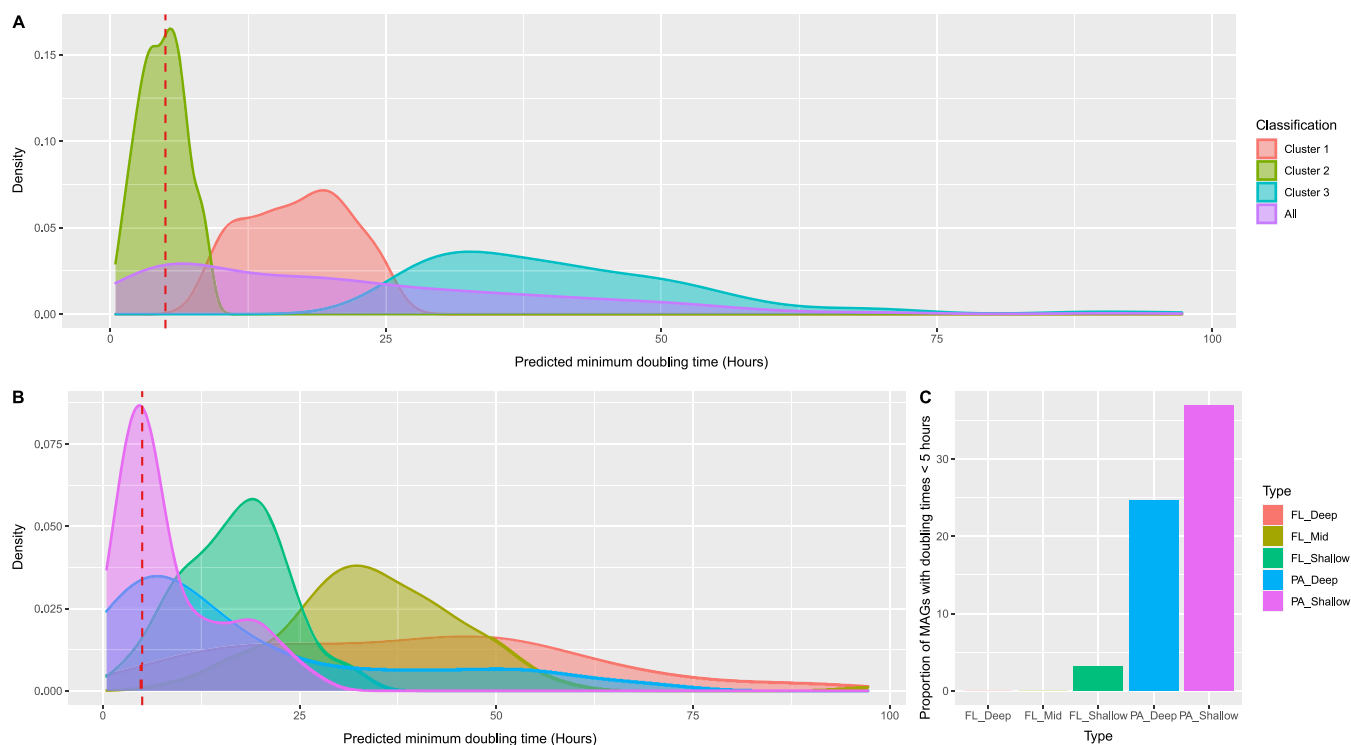
FIG 6 Predicted growth efficiencies of sinking-particle-attached microbes are greater than those of free-living microbes. (A) Predicted temperature-corrected maximum growth rates of all recovered MAGs, clustered based on a Gaussian mixture model. (B) Predicted temperature-corrected maximum growth rates of all recovered MAGs, clustered based on their sample types. (C) Proportion of MAGs in each sample type with temperature-corrected predicted doubling times of <5 h. A vertical dashed red line represents a doubling time of 5 h. FL, free-living; PA, particle-associated; FL_Shallow, 0 to 150 m; FL_Mid, >150 to 1,000 m; FL_Deep, 4,000 m; PA_Shallow, 150 m; PA_Deep, 4,000 m.

into microhabitat dynamics throughout the water column, as well as identifying habitat-specific adaptive characteristics. We report here microbial genomes associated with sediment trap-collected sinking particles versus filtered free-living microbes from the photic zone to the abyss, to identify any habitat-specific genomic traits. The majority of the SPA MAGs were found to be phylogenetically unrelated to FL MAGs. Only a few genera were captured in both the FL and PA samples and included phototrophic species like *Prochlorococcus* as well as motile heterotrophic copiotrophs like *Idiomarina*. These results further illustrate the distinctive microhabitat-specific phylogenetic, physiological, and ecological divergence between SPA and FL microbial communities (14, 16, 18, 20, 26, 35, 36, 51).

Rapidly sinking particles are rich in nutrients (including polysaccharides and proteins) and energy (7, 25, 61, 67). Consequently, we observed a higher percentage of secreted peptidases and CAZymes in SPA microbes than in FL microbes. This observation is consistent with the prediction that microbes inhabiting diffusion-limited environments (e.g., soils, sludges, sediments) encode more extracellular enzymes, in part because they diffuse away more slowly in structured (particles) versus unstructured (bulk seawater) habitats. SPA MAGs were also found to be enriched in CAZymes that degrade different polysaccharides known to be abundant in marine habitats. Notably, those microbes capable of hydrolyzing laminarin and known to be abundant in marine environments (7, 61, 67) were most highly represented in the SPA samples. With predictions of laminarin accounting for up to 50% of POC in sinking diatom-containing particles (61), the data suggest that SPA microbes may have a considerable influence on laminarin degradation and associated carbon export to the abyssal depths in the open ocean.

Previous metagenomic and metaproteomic studies of 0.2-μm-filtered samples throughout the water column reported that total peptidases and CAZymes decreased from epipelagic (0 to 200 m) to bathypelagic (>1,000 m) microbial communities, while the

percentage of extracellular enzymes and their diversity increased with depth (36). In our genomic survey, a decrease in total peptidase and CAZyme genes in genomes with depth was observed only in the SPA MAGs. However, the percentage of extracellular peptidases and CAZymes was not significantly higher with depth in SPA or FL sample types. Methodological differences in particle sampling methods (e.g., hand collection by SCUBA divers versus serial filtration versus sediment traps), along with differences in spatial resolution, analytes, and analytical strategies, may in part explain the differences between our results and prior studies (36). Our CAZyme, histidine kinase, and secretion system results do, however, support the postulate that filter-collected, suspended bathypelagic prokaryotes appear more adapted to particle colonization than their epipelagic counterparts (36). Additionally, the enrichment of total and secretory hydrolytic extracellular enzymes and their elevated diversity in SPA microbes suggest their greater metabolic potential for solubilization of a wide variety of POC substrates in sinking particles. Overall, the higher diversity of transporters in SPA MAGs also suggests that efficient solubilization may produce a larger variety of nutrients which are subsequently taken up by particle-associated communities.

Given their particle-associated lifestyles, we postulated that SPA microbial genomes are also enriched in genes that promote environmental sensing and response, motility, attachment, colonization, and cell-cell interactions. Consistent with this hypothesis, both the shallow- and deep-water SPA microbes contained greater proportions of genes associated with two-component regulatory systems, extracellular secretion systems, and flagella than did FL microbes from the same depth horizons. The genomic distributions of TCSs support prior observations with respect to general metabolic types (69) and suggest a greater potential for environmental sensing and response capacity in SPA microbes than in their sympatric oligotrophic FL counterparts. The diverse array and elevated levels of ESS genes in SPA microbes compared to FL microbes also support their postulated greater capacity for attachment, motility, and cell-cell interactions. Additionally, the elevated levels of phage and prophage marker genes in SPA versus FL microbial genomes are consistent with hypotheses suggesting that nutrient-replete, high-cell-density habitats may favor retention of greater numbers of lysogens and prophage (85). Taken together, the greater representation of TCS, ESS, and phage genes in SPA microbes likely contributes to their expanded behavioral, regulatory, and metabolic traits, reflecting adaptations to heterogeneous and ephemeral microhabitats.

Prior gene-centric studies have demonstrated that genome size, GC content, and genome and proteome nitrogen content tend to positively correlate with environmental inorganic nitrogen availability among FL bacterioplankton populations (44, 81, 82). The genome-based analyses reported here generally recapitulate these trends for FL bacterioplankton, with increased genome size, GC content, and N-ARSC being a hallmark for deeper-water FL genomes, in part driven by increased dissolved inorganic nitrogen availability.

SPA microbes, however, showed an opposite trend with respect to DNA GC content and proteome N-ARSC with increasing depth. Strikingly, shallow-water SPA microbes collected at 150 m had higher DNA GC and proteome nitrogen contents than did deeper, 4,000-m SPA microbes. The higher genome and proteome nitrogen content in shallow- versus deep-water SPA microbes is consistent with recent studies of the elemental composition and energy content (in Joules per unit mass of organic carbon) (25) of sinking particles. Specifically, sinking particles collected at 4,000 m had relatively higher organic carbon-to-nitrogen (C:N) contents (C:N, $\sim$10:1) than did sinking particles collected at 150 m (C:N, $\sim$6:1) (25). In total, our new data suggest that resource limitation may also drive genome and proteome elemental stoichiometry in SPA microbes. Specifically, more nitrogen-depleted particles found at greater depths appear to select for SPA microbes having lower DNA GC content and less nitrogen in their genomes and proteomes than their shallow-water, more nitrogen-enriched SPA counterparts. These trends suggest that shallow-water SPA microbes may rely more on

particle-derived organic nitrogen compounds than on more oxidized bulk inorganic nitrogen sources (NO_3^-, NO_2^-) to reduce the energetic costs of nitrogen acquisition. Our data further suggest that as particles sink and nitrogen depletion ensues, SPA microbial assemblages undergo dynamic compositional shifts in response to the more nutrient-depleted particle microenvironments found at greater depths.

Codon usage bias has been suggested to be capable of determining the *in situ* maximum growth rates (minimum doubling times) of prokaryote environmental settings (84–87). Accurate *in situ* growth rates cannot necessarily be determined by this approach, however, since temperature, pressure, nutrient availability, competition, mortality, and epigenetic phenomena are all expected to exert profound effects on actual *in situ* microbial growth rates in nature. Nonetheless, this approach may provide a useful metric for comparing relative growth efficiencies of different bacteria and archaea from similar habitats. We observed here that SPA and FL MAGs partitioned specifically into discrete categories based on codon usage bias estimates that generally reflected their copiotrophic versus oligotrophic lifestyles, respectively. This further reflects the major influence across phyla of nutrient availability as a central ecological and evolutionary driver of both SPA and FL gene acquisition, genome evolution, and metabolic capabilities.

In total, this genome-centric study demonstrates for the first time the aggregate genomic features and gene repertoires that differentiate sympatric SPA from FL bacteria and archaea in the open ocean and define their phylogenetic metabolic and behavioral characteristics. These SPA and FL genomic traits reflect both the physicochemical structure and chemical composition of their environment as well as their competitive and synergistic interactions throughout the water column. Future directed efforts exploring the biological properties and biogeochemical activities of SPA and FL microbes *in situ* at greater spatiotemporal resolution may help to provide a better understanding of the complex mechanisms of POM and DOM cycling, which in part drive the ocean's biological carbon pump.

MATERIALS AND METHODS

Sample collection. Metagenomic data for the PA_Shallow samples were generated from DNA extracted from bulk sinking particles collected in net traps at 150 m as described by Farnelid et al. (47). Briefly, particles collected in the traps were filtered through a 0.2-μm-pore-size Supor filter (Pall Corporation, New York, NY, USA), and DNA was subsequently extracted using a Qiagen plant minikit and a QIAcube (Qiagen, Valencia, CA, USA).

Metagenomic data from the PA_Deep samples were derived from the 63 samples collected from the deep-sea sequencing sediment traps at 4,000 m, as previously described (18, 20). Briefly, particles collected in preservative-containing trap cups were centrifuged to collect the sinking particulate material, and DNA was subsequently extracted using a Qiagen DNeasy PowerBiofilm kit (Qiagen, Valencia, CA, USA).

Metagenomic sequence data from FL samples (see Table S1 in the supplemental material) were derived from Station ALOHA time series samples collected between depths of 5 and 4,000 m on Hawaii Ocean Time series cruises. Samples were collected, processed, and DNA extracted as previously described (44–46). Briefly, 2 to 4 L of seawater was collected using conductivity-temperature-depth (CTD)-attached Niskin bottles and filtered (with no prefiltration) onto a 0.2-μm-pore-size, 25-mm Supor filter (VWR 28147-956). After filtration, the filters were removed and stored in 300 μL of RNAlater (Thermo Fisher Scientific, Waltham, MA) at $-80°C$. DNA extractions were performed by thawing filters on ice, removing the RNAlater, and adding 400 μL of sucrose lysis buffer (final concentrations of 40 mM EDTA, 50 mM Tris (pH 8.3), and 0.75 M sucrose). Cell homogenization was performed using a TissueLyser (Qiagen, Germantown, MD) programmed at 30 Hz for two rounds of 1 min each. This was followed by the addition of 0.5 mg mL^{-1} lysozyme (final concentration) at 37°C for 30 min. Subsequently, 50 μL of a proteinase K solution (0.8 mg mL^{-1}) was added, followed by the addition of 50 μL of 10% SDS. Samples were incubated at 55°C for 2 h. Final DNA purification was robotically performed using a Chemagen MSM I instrument with the CMG-1037 DNA saliva kit (Perkin Elmer, Waltham, MA).

Sequencing libraries were prepared and sequencing for all the above samples was performed as described previously (18, 45). Briefly, DNA was sheared to an average size of 350 bp using a Covaris M220 focused-ultrasonicator (catalog no. 4482277; Thermo Fisher Scientific, Waltham, MA) with Microtube-50 AFA fiber tubes in accordance with the manufacturer's instructions. Sequencing libraries were prepared using Illumina's TruSeq Nano LT library preparation kits (Illumina, San Diego, CA). Libraries were sequenced using a 150-bp paired-end NextSeq500/550 high-output v2 reagent kit (Illumina, San Diego, CA).

Data generation and processing. The PA_Shallow metagenomic paired-end read sets were trimmed and quality filtered using illumina-utils (88) with the iu-filter-quality-minoche script based on quality filtering suggestions made by Minoche et al. (89). The PA_Shallow metagenomic reads were individually assembled using metaSPAdes version 3.13.0 with the default parameters (90). The PA_Deep

metagenomic paired-end read sets were processed as previously reported (20). The FL metagenomic paired-end read sets were trimmed and quality filtered using the bbduk script from BBMap 38.22 (https://sourceforge.net/projects/bbmap/) in two passes. The first pass used parameters "ktrim=r k = 23 mink = 11 hdist = 1 tbo tpe tbo tpe" for Illumina sequencing adapters, and the second pass used parameters "k = 27 hdist = 1 qtrim=rl trimq = 17 cardinality=t mingc = 0.05 maxgc = 0.95" to remove phiX, low-quality bases, and sequences with unrealistically high or low GC. Additional low-quality bases and sequences were removed with Trimmomatic 0.38 (parameters: LEADING:10 TRAILING:10 MINLEN:100) (91). Unpaired reads were removed using the dropse command from seqtk 1.2 (https://github.com/lh3/seqtk). The cleaned reads from each sample were assembled using SPAdes 3.11.1 (parameters: -meta -k 21,33,55,77,99,127) (90).

Recovery of the MAGs. Mapping of quality reads was performed using CoverM v0.4.0 with default parameters (https://github.com/wwood/CoverM). For each assembled metagenome, metagenome-assembled genomes were recovered using MetaBAT1 v0.32.5 (92) with all the sensitivity settings, MetaBAT2 v2.12.1 (93), MaxBin2 v2.2.6 (94) using the 40 and 107 gene sets, and CONCOCT (95) under the default settings. The MAGs were filtered to remove contigs below 1,500 bp. The resulting MAGs from each assembly were assessed for quality and dereplicated using DAStool with default settings (96). Completeness and contamination rates of the MAGs were assessed using CheckM v1.0.13 (97) with the "lineage wf" command. MAGs from a single sample type were subsequently dereplicated using dRep v2.2.3 (98) using the dereplicate_wf at ≥97% average nucleotide identity over ≥70% alignment, and the representative MAGs were chosen based on genome completeness. The dereplicated MAGs were further refined by reassembling the mapped quality trimmed reads with SPAdes (99) using the –careful and –trusted-contigs setting. Additional scaffolding and resolving of ambiguous bases of the MAGs were performed using the "roundup" mode of FinishM v0.0.7 (https://github.com/wwood/finishm).

MAGs were classified into five sample types. FL MAGs were classified as FL_Shallow, FL_Mid, and FL_Deep based on their relative abundances at different depths revealed by read mapping. MAGs were classified as FL_Shallow when they were most abundant at 150 m or shallower, as FL_Mid at 175 m to 1,000 m, and as FL_Deep at 4,000 m. SPA MAGs that were recovered from the 150-m sediment traps (48) were classified as PA_Shallow. Since surface-water microbes were also exported on sinking particles that reached 4,000 m (20), MAGs recovered from the 4,000-m SPA microbes (PA_Deep) that contained surface-associated genes (including those encoding proteorhodopsins, chlorophyll/bacteriochlorophyll *a* synthase, and deoxyribodipyrimidine photolyase) were reclassified as PA_Shallow (see Methods) (Table S2), since these originated from the export of surface-water-derived sinking particles (20).

Sixty-one 4,000-m PA_Deep MAGs were reannotated as PA_Shallow based on their photic zone-dependent gene representation, well documented physiologies and lifestyles (e.g., oxygenic and anoxygenic photoautotrophs and photoheterotrophs), and/or depth profile read mapping densities. These included the following MAGs: DT-Alcanivorax-1, DT-Alteromonas-1, DT-Bdellovibrionales-2, DT-Bdellovibrionales-3, DT-Caenarcaniphilales-1, DT-Chromohalobacter-1, DT-Cognatishimia-1, DT-Coraliomargarita-1, DT-Dinoroseobacter-1, DT-Ekhidna-1, DT-Epibacterium_A-1, DT-Erythrobacter-1, DT-Erythrobacter-2, DT-Flavobacteriaceae-1, DT-Flavobacteriaceae-3, DT-Flavobacteriaceae-4, DT-Flavobacteriales-4, DT-Gilvibacter-1, DT-Halieaceae-1, DT-Halioglobus-1, DT-Halomonas-1, DT-Halomonas-3, DT-Henriciella-1, DT-Henriciella-2, DT-Henriciella-3, DT-Hyphomicrobiaceae-1, DT-Idiomarina-1, DT-Idiomarina-3, DT-Ilumatobacteraceae-1, DT-Legionellales-1, DT-Micavibrionaceae-1, DT-Mycoplasmatales-1, DT-Mycoplasmatales-2, DT-Oleispira-1, DT-Oligoflexales-1, DT-Parvularculaceae-1, DT-Phaeodactylibacter-1, DT-Phycisphaerales-1, DT-Pseudoalteromonas-1, DT-Pseudobacteriovorax-1, DT-Pseudomonadales-2, DT-Psychroserpens-1, DT-Psychrosphaera-1, DT-Rhizobiales-2, DT-Rhodobacteraceae-1, DT-Rhodobacteraceae-2, DT-Rhodobacteraceae-3, DT-Richelia-1, DT-Rickettsiaceae-1, DT-Rickettsiales-1, DT-Rivularia-1, DT-Salinicola-1, DT-Saprospiraceae-1, DT-Saprospiraceae-2, DT-Shewanella-1, DT-Simkaniaceae-1, DT-Verrucomicrobiales-1, DT-Vibrio-1, DT-Winogradskyella-1, DT-Winogradskyella-2, DT-Xanthomonadales-1, and DT-Crocosphaera.

Taxonomic inference of the MAGs. Classification of the MAGs was determined using GTDB-Tk (48) v.1.3.0 implementing the classify_wf command (https://github.com/Ecogenomics/GTDBTk). Briefly, marker genes were identified in each genome, aligned, concatenated, and classified with pplacer to identify the maximum-likelihood placement of each genome's concatenated protein alignment in the GTDB-Tk reference tree. GTDB-Tk classifies each genome based on its placement in the reference tree, its relative evolutionary distance, and FastANI distance.

Functional annotations. For all MAGs, genes were called and annotated using Prokka v.1.13 (100). Additional annotation was performed using the blastp "very sensitive" setting in Diamond v0.9.30.131 (101) (https://github.com/bbuchfink/diamond.git) against UniRef100 (accessed September 2019) (102), clusters of orthologous groups (COG) (103), and Pfam 31 (104) and TIGRfam 15.0 hidden Markov models (HMMs) (105). Genes were assigned KEGG Ontology (KO) IDs by hmmsearch against KOfam with predefined thresholds (106).

Calculation of encoded protein elemental composition. Amino acid sequences from the dereplicated genome set were used to calculate N-ARSC and C-ARSC values using custom scripts from Mende et al. (44) and available at https://github.com/JessAwBryant/gene-characteristics.

CAZy annotation. Genes were annotated against dbCAN (3.0) with default parameters (107) and filtered to retain hits with E values of <1e−102 and HMM coverage of ≥0.35. Filtered genes were further classified to their subfamily level, and their corresponding EC numbers were determined using eCAMI (107). SignalP v6.0 (108) was used to detect the presence of signal peptides in genes encoding putative CAZymes.

Peptidase annotation. Genes were annotated using the blastp "very sensitive" setting in Diamond v0.9.30.131 (100) against the MEROPS database v12.2 (109) and filtered to retain hits with E values

of $<1 \times 10^{-10}$. SignalP v5.0 (108) was used to detect the presence of signal peptides in genes encoding putative peptidases.

Histidine kinase annotation. Genes were annotated via hmmsearch using protein family (Pfam) domain annotations (104, 110). Histidine kinases were identified based on hits to pfam02518, pfam13581, pfam13589, pfam14501, and pfam07536. Putative histidine kinases with hits to DNA gyrase (pfam00204), HSP90 (pfam00183), and MutL (pfam13941) were filtered out. Genes encoding response regulator receivers were identified based on hits to pfam0072.

Extracellular secretion system annotations and analyses. Genes encoding the bacterial secretion systems were predicted using MacSysFinder (111) and the TXSScan reference database (v1.0rc1) (112).

Virus annotations and analyses. Putative prophages within the MAGs were identified using VirSorter2 v2.2.3 (113), and their quality was assessed based on checkV (114) v0.8.1. False-positive screening of prophages was performed based on the Sullivan lab's protocol (https://www.protocols.io/view/viral-sequence-identification-sop-with-virsorter2-5qpvoyqebg4o/v3). Prophage genes were annotated using DRAM-v.py v1.2.4 (115), and phage-associated marker genes encoding CI repressor, phage capsid proteins, phage excisionase, and phage tail proteins were identified.

Comparative growth efficiencies estimated via CUB-based maximal growth rate metrics. Maximal growth rates of MAGs based on codon usage bias metrics were predicted using the R package gRodon (84). Different temperature correction settings were applied to the MAGs as follows: PA_Deep and FL_Deep, 2°C; PA_Shallow and FL_Shallow, 24°C; FL-Mid, 10°C. The Gaussian mixture model shown in Fig. 6 was fitted using the mclust v5.4.9 package in R (116).

Identification of orthologous proteins. Homologous proteins across all MAGs were identified with OrthoFinder (117) v2.3.3 using default parameters.

Principal-coordinate analysis. The presence/absence of each KO ID or orthologous family in each genome was used as input to a nonmetric multidimensional scaling (NMDS) ordination analysis with the Jaccard distance matrix using the metaMDS function in the Vegan v2.6-2 package in R (118, 119).

Calculation of relative abundance. To calculate the relative abundance, reads from each metagenomic data set were mapped to the dereplicated MAGs using CoverM v0.3.1 with the "contig" command, a cutoff of 95% minimum identity, and a minimum aligned read length of 75% of each read. Coverage of each contig was calculated with the CoverM "trimmed_mean" option, and the coverage for each MAG was calculated as the average of all contig coverages, weighted by their length. The relative abundance of each MAG in each metagenomic data set was calculated as its coverage divided by total reads in the sample multiplied by 100,000,000.

Gene enrichment analysis. The presence/absence of each KO ID or orthologous family in each genome was used as input for the program Scoary (52) to identify significant correlations between gene presence/absence and sample types.

Statistical analysis. The mean statistical significance of different proteins of interests and genomic characteristic metrics between sample types were determined through one-way analysis of variance (ANOVA), followed by the Tukey test using the R package multicompview v0.1-8 with the Tukey honestly significant different (HSD) function. All P values that were affected by multiple testing were corrected for false discovery using the Benjamini-Hochberg procedure.

Calculation of amino acid identity. Average amino acid identity (AAI) between the genomes was calculated using orthologous genes identified through reciprocal best BLAST hits by use of compareM v0.0.5 (https://github.com/dparks1134/CompareM).

Data visualization. Figures were generated using pheatmap (120) and ggplot2 (121) packages in R (122). Trees were visualized in ARB (123) and presented using iTOL (124). Venn diagrams were generated using the Venn diagram tool (http://bioinformatics.psb.ugent.be/webtools/Venn/). Figures were further refined using Adobe Illustrator.

Data availability. All data were deposited in the NCBI SRA archive as follows. Deep 4,000-m-trap metagenomes and MAGs are deposited under NCBI BioProject no. PRJNA482655, and assembled MAGs are available under NCBI BioSample no. SAMN14675689 to SAMN14675809. Metagenomic reads produced from DNA extracted from shallow 150-m sediment trap samples (collected in 2015) (48) and PA_Shallow MAGs are deposited under NCBI BioProject no. PRJNA358725. For Station ALOHA 0.2-μm-pore-size filter-collected metagenomes/MAGs, read sequence data and FL MAGs are available at NCBI SRA under BioProject no. PRJNA352737, and assemblies can be found under BioSample no. SAMN12604809.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

FIG S1, PDF file, 1.1 MB.
FIG S2, PDF file, 0.5 MB.
FIG S3, PDF file, 0.1 MB.
FIG S4, PDF file, 0.1 MB.
FIG S5, PDF file, 0.1 MB.
TABLE S1, XLSX file, 0.03 MB.
TABLE S2, XLSX file, 0.04 MB.
TABLE S3, XLSX file, 0.4 MB.
TABLE S4, XLSX file, 0.1 MB.
TABLE S5, XLSX file, 0.03 MB.

ACKNOWLEDGMENTS

REFERENCES

1. Azam F. 1998. Microbial control of oceanic carbon flux: the plot thickens. Science 280:694–696. https://doi.org/10.1126/science.280.5364.694.

2. Ducklow H, Kirchman D, Quinby H, Carlson C, Dam H. 1993. Stocks and dynamics of bacterioplankton carbon during the spring bloom in the eastern North Atlantic Ocean. Deep Sea Res Part 2 Top Stud Oceanogr 40:245–263. https://doi.org/10.1016/0967-0645(93)90016-G.

3. Hansell DA. 2013. Recalcitrant dissolved organic carbon fractions. Annu Rev Mar Sci 5:421–445. https://doi.org/10.1146/annurev-marine-120710-100757.

4. Simon M, Grossart H-P, Schweitzer B, Ploug H. 2002. Microbial ecology of organic aggregates in aquatic ecosystems. Aquat Microb Ecol 28:175–211. https://doi.org/10.3354/ame028175.

5. Herndl GJ, Reinthaler T. 2013. Microbial control of the dark end of the biological pump. Nat Geosci 6:718–724. https://doi.org/10.1038/ngeo1921.

6. Arnosti C. 2011. Microbial extracellular enzymes and the marine carbon cycle. Annu Rev Mar Sci 3:401–425. https://doi.org/10.1146/annurev-marine-120709-142731.

7. Arnosti C, Wietz M, Brinkhoff T, Hehemann J-H, Probandt D, Zeugner L, Amann R. 2021. The biogeochemistry of marine polysaccharides: sources, inventories, and bacterial drivers of the carbohydrate cycle. Annu Rev Mar Sci 13:81–108. https://doi.org/10.1146/annurev-marine-032020-012810.

8. Vetter Y, Deming J. 1999. Growth rates of marine bacterial isolates on particulate organic substrates solubilized by freely released extracellular enzymes. Microb Ecol 37:86–94. https://doi.org/10.1007/s002489900133.

9. Reintjes G, Arnosti C, Fuchs B, Amann R. 2019. Selfish, sharing and scavenging bacteria in the Atlantic Ocean: a biogeographical study of bacterial substrate utilisation. ISME J 13:1119–1132. https://doi.org/10.1038/s41396-018-0326-3.

10. Caron DA, Davis PG, Madin LP, Sieburth JM. 1986. Enrichment of microbial populations in macroaggregates (marine snow) from surface waters of the North Atlantic. J Mar Res 44:543–565. https://doi.org/10.1357/002224086788403042.

11. 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.

12. Smith DC, Simon M, Alldredge AL, Azam F. 1992. Intense hydrolytic enzyme activity on marine aggregates and implications for rapid particle dissolution. Nature 359:139–142. https://doi.org/10.1038/359139a0.

13. Ziervogel K, Steen A, Arnosti C. 2010. Changes in the spectrum and rates of extracellular enzyme activities in seawater following aggregate formation. Biogeosciences 7:1007–1015. https://doi.org/10.5194/bg-7-1007-2010.

14. DeLong EF, Franks DG, Alldredge AL. 1993. Phylogenetic diversity of aggregate-attached versus free-living marine bacterial assemblages. Limnol Oceanogr 38:924–934. https://doi.org/10.4319/lo.1993.38.5.0924.

15. Allen AE, Allen LZ, McCrow JP. 2013. Lineage specific gene family enrichment at the microscale in marine systems. Curr Opin Microbiol 16:605–617. https://doi.org/10.1016/j.mib.2013.10.001.

16. Ganesh S, Parris DJ, DeLong EF, Stewart FJ. 2014. Metagenomic analysis of size-fractionated picoplankton in a marine oxygen minimum zone. ISME J 8:187–211. https://doi.org/10.1038/ismej.2013.144.

17. Duret MT, Lampitt RS, Lam P. 2019. Prokaryotic niche partitioning between suspended and sinking marine particles. Environ Microbiol Rep 11:386–400. https://doi.org/10.1111/1758-2229.12692.

18. Boeuf D, Edwards BR, Eppley JM, Hu SK, Poff KE, Romano AE, Caron DA, Karl DM, DeLong EF. 2019. Biological composition and microbial dynamics of sinking particulate organic matter at abyssal depths in the oligotrophic open ocean. Proc Natl Acad Sci U S A 116:11824–11832. https://doi.org/10.1073/pnas.1903080116.

19. Preston CM, Durkin C, Yamahara K. 2020. DNA metabarcoding reveals organisms contributing to particulate matter flux to abyssal depths in the North East Pacific ocean. Deep Sea Res Part 2 Top Stud Oceanogr 173:104708. https://doi.org/10.1016/j.dsr2.2019.104708.

20. Poff KE, Leu AO, Eppley JM, Karl DM, DeLong EF. 2021. Microbial dynamics of elevated carbon flux in the open ocean's abyss. Proc Natl Acad Sci U S A 118:e2018269118. https://doi.org/10.1073/pnas.2018269118.

21. Baumas CMJ, Le Moigne FAC, Garel M, Bhairy N, Guasco S, Riou V, Armougom F, Grossart HP, Tamburini C. 2021. Mesopelagic microbial carbon production correlates with diversity across different marine particle fractions. ISME J 15:1695–1708. https://doi.org/10.1038/s41396-020-00880-z.

22. Zhang R, Liu B, Lau SC, Ki J-S, Qian P-Y. 2007. Particle-attached and free-living bacterial communities in a contrasting marine environment: Victoria Harbor, Hong Kong. FEMS Microbiol Ecol 61:496–508. https://doi.org/10.1111/j.1574-6941.2007.00353.x.

23. Fuhrman JA, Steele JA, Hewson I, Schwalbach MS, Brown MV, Green JL, Brown JH. 2008. A latitudinal diversity gradient in planktonic marine bacteria. Proc Natl Acad Sci U S A 105:7774–7778. https://doi.org/10.1073/pnas.0803070105.

24. Karl DM, Church MJ, Dore JE, Letelier RM, Mahaffey C. 2012. Predictable and efficient carbon sequestration in the North Pacific Ocean supported by symbiotic nitrogen fixation. Proc Natl Acad Sci U S A 109:1842–1849. https://doi.org/10.1073/pnas.1120312109.

25. Grabowski E, Letelier RM, Laws EA, Karl DM. 2019. Coupling carbon and energy fluxes in the North Pacific Subtropical Gyre. Nat Commun 10:1895. https://doi.org/10.1038/s41467-019-09772-z.

26. 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.

27. Amacher J, Neuer S, Lomas M. 2013. DNA-based molecular fingerprinting of eukaryotic protists and cyanobacteria contributing to sinking particle flux at the Bermuda Atlantic time-series study. Deep Sea Res Part 2 Top Stud Oceanogr 93:71–83. https://doi.org/10.1016/j.dsr2.2013.01.001.

28. Smith MW, Zeigler Allen L, Allen AE, Herfort L, Simon HM. 2013. Contrasting genomic properties of free-living and particle-attached microbial assemblages within a coastal ecosystem. Front Microbiol 4:120. https://doi.org/10.3389/fmicb.2013.00120.

29. LeCleir GR, DeBruyn JM, Maas EW, Boyd PW, Wilhelm SW. 2014. Temporal changes in particle-associated microbial communities after interception by nonlethal sediment traps. FEMS Microbiol Ecol 87:153–163. https://doi.org/10.1111/1574-6941.12213.

30. Salazar G, Cornejo-Castillo FM, Borrull E, Diez-Vives C, Lara E, Vaque D, Arrieta JM, Duarte CM, Gasol JM, Acinas SG. 2015. Particle-association lifestyle is a phylogenetically conserved trait in bathypelagic prokaryotes. Mol Ecol 24:5692–5706. https://doi.org/10.1111/mec.13419.

31. Rieck A, Herlemann DP, Jurgens K, Grossart HP. 2015. Particle-associated differ from free-living bacteria in surface waters of the Baltic Sea. Front Microbiol 6:1297. https://doi.org/10.3389/fmicb.2015.01297.

32. Liu R, Wang L, Liu Q, Wang Z, Li Z, Fang J, Zhang L, Luo M. 2018. Depth-resolved distribution of particle-attached and free-living bacterial communities in the water column of the New Britain Trench. Front Microbiol 9:625. https://doi.org/10.3389/fmicb.2018.00625.

33. Valencia B, Stukel MR, Allen AE, McCrow JP, Rabines A, Palenik B, Landry MR. 2021. Relating sinking and suspended microbial communities in the California Current Ecosystem: digestion resistance and the contributions of phytoplankton taxa to export. Environ Microbiol 23:6734–6748. https://doi.org/10.1111/1462-2920.15736.

34. Fontanez KM, Eppley JM, Samo TJ, Karl DM, DeLong EF. 2015. Microbial community structure and function on sinking particles in the North Pacific Subtropical Gyre. Front Microbiol 6:469. https://doi.org/10.3389/fmicb.2015.00469.

35. Pelve EA, Fontanez KM, DeLong EF. 2017. Bacterial succession on sinking particles in the Ocean's interior. Front Microbiol 8:2269. https://doi.org/10.3389/fmicb.2017.02269.

36. Zhao Z, Baltar F, Herndl GJ. 2020. Linking extracellular enzymes to phylogeny indicates a predominantly particle-associated lifestyle of deep-sea prokaryotes. Sci Adv 6:eaaz4354. https://doi.org/10.1126/sciadv.aaz4354.

37. Acinas SG, Sanchez P, Salazar G, Cornejo-Castillo FM, Sebastian M, Logares R, Royo-Llonch M, Paoli L, Sunagawa S, Hingamp P, Ogata H, Lima-Mendez G, Roux S, Gonzalez JM, Arrieta JM, Alam IS, Kamau A, Bowler C, Raes J, Pesant S, Bork P, Agusti S, Gojobori T, Vaque D, Sullivan MB, Pedros-Alio C, Massana R, Duarte CM, Gasol JM. 2021. Deep ocean metagenomes provide insight into the metabolic architecture of bathypelagic microbial communities. Commun Biol 4:604. https://doi.org/10.1038/s42003-021-02112-2.

38. Mestre M, Ruiz-Gonzalez C, Logares R, Duarte CM, Gasol JM, Sala MM. 2018. Sinking particles promote vertical connectivity in the ocean microbiome. Proc Natl Acad Sci U S A 115:E6799–E6807. https://doi.org/10.1073/pnas.1802470115.

39. Ruiz-Gonzalez C, Mestre M, Estrada M, Sebastian M, Salazar G, Agusti S, Moreno-Ostos E, Reche I, Alvarez-Salgado XA, Moran XAG, Duarte CM, Sala MM, Gasol JM. 2020. Major imprint of surface plankton on deep ocean prokaryotic structure and activity. Mol Ecol 29:1820–1838. https://doi.org/10.1111/mec.15454.

40. Tully BJ, Sachdeva R, Graham ED, Heidelberg JF. 2017. 290 metagenome-assembled genomes from the Mediterranean Sea: a resource for marine microbiology. PeerJ 5:e3558. https://doi.org/10.7717/peerj.3558.

41. Tully BJ, Graham ED, Heidelberg JF. 2018. The reconstruction of 2,631 draft metagenome-assembled genomes from the global oceans. Sci Data 5:170203. https://doi.org/10.1038/sdata.2017.203.

42. Royo-Llonch M, Sánchez P, Ruiz-González C, Salazar G, Pedrós-Alió C, Sebastián M, Labadie K, Paoli L, M Ibarbalz F, Zinger L, Churcheward B, Chaffron S, Eveillard D, Karsenti E, Sunagawa S, Wincker P, Karp-Boss L, Bowler C, Acinas SG, Tara Oceans Coordinators. 2021. Compendium of 530 metagenome-assembled bacterial and archaeal genomes from the polar Arctic Ocean. Nat Microbiol 6:1561–1574. https://doi.org/10.1038/s41564-021-00979-9.

43. Luo E, Eppley JM, Romano AE, Mende DR, DeLong EF. 2020. Double-stranded DNA virioplankton dynamics and reproductive strategies in the oligotrophic open ocean water column. ISME J 14:1304–1315. https://doi.org/10.1038/s41396-020-0604-8.

44. 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.

45. Wilson ST, Aylward FO, Ribalet F, Barone B, Casey JR, Connell PE, Eppley JM, Ferron S, Fitzsimmons JN, Hayes CT, Romano AE, Turk-Kubo KA, Vislova A, Armbrust EV, Caron DA, Church MJ, Zehr JP, Karl DM, DeLong EF. 2017. Coordinated regulation of growth, activity and transcription in natural populations of the unicellular nitrogen-fixing cyanobacterium Crocosphaera. Nat Microbiol 2:17118. https://doi.org/10.1038/nmicrobiol.2017.118.

46. DeLong EF, Preston CM, Mincer T, Rich V, Hallam SJ, Frigaard NU, Martinez A, Sullivan MB, Edwards R, Brito BR, Chisholm SW, Karl DM. 2006. Community genomics among stratified microbial assemblages in the ocean's interior. Science 311:496–503. https://doi.org/10.1126/science.1120250.

47. Farnelid H, Turk-Kubo K, Ploug H, Ossolinski JE, Collins JR, Van Mooy BAS, Zehr JP. 2019. Diverse diazotrophs are present on sinking particles in the North Pacific Subtropical Gyre. ISME J 13:170–182. https://doi.org/10.1038/s41396-018-0259-x.

48. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. 2020. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Oxford University Press, Oxford, United Kingdom.

49. Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil P-A, Hugenholtz P. 2018. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nat Biotechnol 36:996–1004. https://doi.org/10.1038/nbt.4229.

50. Tully BJ, Wheat CG, Glazer BT, Huber JA. 2018. A dynamic microbial community with high functional redundancy inhabits the cold, oxic subseafloor aquifer. ISME J 12:1–16. https://doi.org/10.1038/ismej.2017.187.

51. Salazar G, Cornejo-Castillo FM, Benítez-Barrios V, Fraile-Nuez E, Álvarez-Salgado XA, Duarte CM, Gasol JM, Acinas SG. 2016. Global diversity and biogeography of deep-sea pelagic prokaryotes. ISME J 10:596–608. https://doi.org/10.1038/ismej.2015.137.

52. Brynildsrud O, Bohlin J, Scheffer L, Eldholm V. 2016. Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. Genome Biol 17:1–9. https://doi.org/10.1186/s13059-016-1108-8.

53. Bergauer K, Fernandez-Guerra A, Garcia JA, Sprenger RR, Stepanauskas R, Pachiadaki MG, Jensen ON, Herndl GJ. 2018. 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. https://doi.org/10.1073/pnas.1708779115.

54. Giordano M, Raven JA. 2014. Nitrogen and sulfur assimilation in plants and algae. Aquat Bot 118:45–61. https://doi.org/10.1016/j.aquabot.2014.06.012.

55. Moore LW, Chilton WS, Canfield ML. 1997. Diversity of opines and opine-catabolizing bacteria isolated from naturally occurring crown gall tumors. Appl Environ Microbiol 63:201–207. https://doi.org/10.1128/aem.63.1.201-207.1997.

56. Harcet M, Perina D, Pleše B. 2013. Opine dehydrogenases in marine invertebrates. Biochem Genet 51:666–676. https://doi.org/10.1007/s10528-013-9596-7.

57. Benner R, Pakulski JD, McCarthy M, Hedges JI, Hatcher PG. 1992. Bulk chemical characteristics of dissolved organic matter in the ocean. Science 255:1561–1564. https://doi.org/10.1126/science.255.5051.1561.

58. Hedges JI, Baldock JA, Gélinas Y, Lee C, Peterson M, Wakeham SG. 2001. Evidence for non-selective preservation of organic matter in sinking marine particles. Nature 409:801–804. https://doi.org/10.1038/35057247.

59. Mabeau S, Kloareg B. 1987. Isolation and analysis of the cell walls of brown algae: *Fucus spiralis*, *F. ceranoides*, *F. vesiculosus*, *F. serratus*, *Bifurcaria bifurcata* and *Laminaria digitata*. J Exp Bot 38:1573–1580. https://doi.org/10.1093/jxb/38.9.1573.

60. D'ambrosio L, Ziervogel K, MacGregor B, Teske A, Arnosti C. 2014. Composition and enzymatic function of particle-associated and free-living bacteria: a coastal/offshore comparison. ISME J 8:2167–2179. https://doi.org/10.1038/ismej.2014.67.

61. Becker S, Tebben J, Coffinet S, Wiltshire K, Iversen MH, Harder T, Hinrichs KU, Hehemann JH. 2020. Laminarin is a major molecule in the marine carbon cycle. Proc Natl Acad Sci U S A 117:6599–6607. https://doi.org/10.1073/pnas.1917001117.

62. Bouraoui H, Desrousseaux M-L, Ioannou E, Alvira P, Manaï M, Rémond C, Dumon C, Fernandez-Fuentes N, O'Donohue MJ. 2016. The GH51 alpha-l-arabinofuranosidase from Paenibacillus sp. THS1 is multifunctional, hydrolyzing main-chain and side-chain glycosidic bonds in heteroxylans. Biotechnol Biofuels 9:140. https://doi.org/10.1186/s13068-016-0550-x.

63. Ndeh D, Rogowski A, Cartmell A, Luis AS, Baslé A, Gray J, Venditto I, Briggs J, Zhang X, Labourel A, Terrapon N, Buffetto F, Nepogodiev S, Xiao Y, Field RA, Zhu Y, O'Neil MA, Urbanowicz BR, York WS, Davies GJ, Abbott DW, Ralet MC, Martens EC, Henrissat B, Gilbert HJ. 2017. Complex pectin metabolism by gut bacteria reveals novel catalytic functions. Nature 544:65–70. https://doi.org/10.1038/nature21725.

64. Nurizzo D, Turkenburg JP, Charnock SJ, Roberts SM, Dodson EJ, McKie VA, Taylor EJ, Gilbert HJ, Davies GJ. 2002. *Cellvibrio japonicus* α-L-arabinanase 43 has a novel five-blade β-propeller fold. Nat Struct Biol 9:665–668. https://doi.org/10.1038/nsb835.

65. Viladot JL, de Ramon E, Durany O, Planas A. 1998. Probing the mechanism of *Bacillus* 1,3–1,4-beta-D-glucan 4-glucanohydrolases by chemical rescue of inactive mutants at catalytically essential residues. Biochemistry 37:11332–11342. https://doi.org/10.1021/bi980586q.

66. Sperling M, Piontek J, Engel A, Wiltshire KH, Niggemann J, Gerdts G, Wichels A. 2017. Combined carbohydrates support rich communities of particle-associated marine bacterioplankton. Front Microbiol 8:65. https://doi.org/10.3389/fmicb.2017.00065.

67. Teeling H, Fuchs BM, Becher D, Klockow C, Gardebrecht A, Bennke CM, Kassabgy M, Huang S, Mann AJ, Waldmann J, Weber M, Klindworth A, Otto A, Lange J, Bernhardt J, Reinsch C, Hecker M, Peplies J, Bockelmann FD, Callies U, Gerdts G, Wichels A, Wiltshire KH, Glöckner FO, Schweder T, Amann R. 2012. Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton bloom. Science 336:608–611. https://doi.org/10.1126/science.1218344.

68. Balmonte JP, Simon M, Giebel H-A, Arnosti C. 2021. A sea change in microbial enzymes: heterogeneous latitudinal and depth-related gradients in bulk water and particle-associated enzymatic activities from 30 S to 59 N in the Pacific Ocean. Limnol Oceanogr 66:3489–3507. https://doi.org/10.1002/lno.11894.

69. Gao R, Mack TR, Stock AM. 2007. Bacterial response regulators: versatile regulatory strategies from common domains. Trends Biochem Sci 32:225–234. https://doi.org/10.1016/j.tibs.2007.03.002.

70. Held NA, McIlvin MR, Moran DM, Laub MT, Saito MA. 2019. Unique patterns and biogeochemical relevance of two-component sensing in marine bacteria. mSystems 4:e00317-18. https://doi.org/10.1128/mSystems.00317-18.

71. Tomich M, Planet PJ, Figurski DH. 2007. The tad locus: postcards from the widespread colonization island. Nat Rev Microbiol 5:363–375. https://doi.org/10.1038/nrmicro1636.

72. de Souza MP, Amini A, Dojka MA, Pickering IJ, Dawson SC, Pace NR, Terry N. 2001. Identification and characterization of bacteria in a selenium-contaminated hypersaline evaporation pond. Appl Environ Microbiol 67:3785–3794. https://doi.org/10.1128/AEM.67.9.3785-3794.2001.

73. Isaac A, Francis B, Amann RI, Amin SA. 2021. Tight adherence (Tad) pilus genes indicate putative niche differentiation in phytoplankton bloom associated Rhodobacterales. Front Microbiol 12:718297. https://doi.org/10.3389/fmicb.2021.718297.

74. Galan JE, Waksman G. 2018. Protein-injection machines in bacteria. Cell 172:1306–1318. https://doi.org/10.1016/j.cell.2018.01.034.

75. Abby SS, Cury J, Guglielmini J, Néron B, Touchon M, Rocha EP. 2016. Identification of protein secretion systems in bacterial genomes. Sci Rep 6:23080. https://doi.org/10.1038/srep23080.

76. Konstantinidis KT, Braff J, Karl DM, DeLong EF. 2009. Comparative metagenomic analysis of a microbial community residing at a depth of 4,000 meters at station ALOHA in the North Pacific Subtropical Gyre. Appl Environ Microbiol 75:5345–5355. https://doi.org/10.1128/AEM.00473-09.

77. Ziervogel K, Arnosti C. 2008. Polysaccharide hydrolysis in aggregates and free enzyme activity in aggregate-free seawater from the north-eastern Gulf of Mexico. Environ Microbiol 10:289–299. https://doi.org/10.1111/j.1462-2920.2007.01451.x.

78. Volkman JK, Tanoue E. 2002. Chemical and biological studies of particulate organic matter in the ocean. J Oceanogr 58:265–279. https://doi.org/10.1023/A:1015809708632.

79. Alldredge AL, Silver MW. 1988. Characteristics, dynamics and significance of marine snow. Prog Oceanogr 20:41–82. https://doi.org/10.1016/0079-6611(88)90053-5.

80. Engel A, Thoms S, Riebesell U, Rochelle-Newall E, Zondervan I. 2004. Polysaccharide aggregation as a potential sink of marine dissolved organic carbon. Nature 428:929–932. https://doi.org/10.1038/nature02453.

81. Grzymski 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.

82. Berube PM, Rasmussen A, Braakman R, Stepanauskas R, Chisholm SW. 2019. Emergence of trait variability through the lens of nitrogen assimilation in *Prochlorococcus*. Elife 8:e41043. https://doi.org/10.7554/eLife.41043.

83. Bragg JG, Hyder CL. 2004. Nitrogen versus carbon use in prokaryotic genomes and proteomes. Proc Biol Sci 271(Suppl 5):S374–S377. https://doi.org/10.1098/rsbl.2004.0193.

84. Weissman JL, Hou S, Fuhrman JA. 2021. Estimating maximal microbial growth rates from cultures, metagenomes, and single cells via codon usage patterns. Proc Nat Acad Sci U S A 118:e2016810118. https://doi.org/10.1073/pnas.2016810118.

85. Vieira-Silva S, Rocha EP. 2010. The systemic imprint of growth and its uses in ecological (meta) genomics. PLoS Genet 6:e1000808. https://doi.org/10.1371/journal.pgen.1000808.

86. Silveira CB, Luque A, Rohwer F. 2021. The landscape of lysogeny across microbial community density, diversity and energetics. Environ Microbiol 23:4098–4111. https://doi.org/10.1111/1462-2920.15640.

87. Kirchman DL. 2016. Growth rates of microbes in the oceans. Annu Rev Mar Sci 8:285–309. https://doi.org/10.1146/annurev-marine-122414-033938.

88. Eren AM, Vineis JH, Morrison HG, Sogin ML. 2013. A filtering method to generate high quality short reads using Illumina paired-end technology. PLoS One 8:e66643. https://doi.org/10.1371/journal.pone.0066643.

89. Minoche AE, Dohm JC, Himmelbauer H. 2011. Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and genome analyzer systems. Genome Biol 12:R112–15. https://doi.org/10.1186/gb-2011-12-11-r112.

90. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. 2017. metaSPAdes: a new versatile metagenomic assembler. Genome Res 27:824–834. https://doi.org/10.1101/gr.213959.116.

91. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

92. Kang DD, Froula J, Egan R, Wang Z. 2015. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. PeerJ 3:e1165. https://doi.org/10.7717/peerj.1165.

93. Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, Wang Z. 2019. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ 7:e7359. https://doi.org/10.7717/peerj.7359.

94. Wu YW, Simmons BA, Singer SW. 2016. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics 32:605–607. https://doi.org/10.1093/bioinformatics/btv638.

95. Alneberg J, Bjarnason BS, De Bruijn I, Schirmer M, Quick J, Ijaz UZ, Lahti L, Loman NJ, Andersson AF, Quince C. 2014. Binning metagenomic contigs by coverage and composition. Nat Methods 11:1144–1146. https://doi.org/10.1038/nmeth.3103.

96. Sieber CM, Probst AJ, Sharrar A, Thomas BC, Hess M, Tringe SG, Banfield JF. 2018. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. Nat Microbiol 3:836–843. https://doi.org/10.1038/s41564-018-0171-1.

97. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25:1043–1055. https://doi.org/10.1101/gr.186072.114.

98. Olm MR, Brown CT, Brooks B, Banfield JF. 2017. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. ISME J 11:2864–2868. https://doi.org/10.1038/ismej.2017.126.

99. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021.

100. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069. https://doi.org/10.1093/bioinformatics/btu153.

101. Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. Nat Methods 12:59–60. https://doi.org/10.1038/nmeth.3176.

102. Suzek BE, Huang H, McGarvey P, Mazumder R, Wu CH. 2007. UniRef: comprehensive and non-redundant UniProt reference clusters. Bioinformatics 23:1282–1288. https://doi.org/10.1093/bioinformatics/btm098.

103. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN, Rao BS, Smirnov S, Sverdlov AV, Vasudevan S, Wolf YI, Yin JJ, Natale DA. 2003. The COG

database: an updated version includes eukaryotes. BMC Bioinformatics 4:41. https://doi.org/10.1186/1471-2105-4-41.

104. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J, Bateman A. 2016. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res 44:D279–D285. https://doi.org/10.1093/nar/gkv1344.

105. Haft DH, Selengut JD, Richter RA, Harkins D, Basu MK, Beck E. 2013. TIGRFAMs and genome properties in 2013. Nucleic Acids Res 41: D387–D395. https://doi.org/10.1093/nar/gks1234.

106. Aramaki T, Blanc-Mathieu R, Endo H, Ohkubo K, Kanehisa M, Goto S, Ogata H. 2020. KofamKOALA: KEGG ortholog assignment based on profile HMM and adaptive score threshold. Bioinformatics 36:2251–2252. https://doi.org/10.1093/bioinformatics/btz859.

107. Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, Busk PK, Xu Y, Yin Y. 2018. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. Nucleic Acids Res 46:W95–W101. https://doi.org/10.1093/nar/gky418.

108. Armenteros JJA, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, von Heijne G, Nielsen H. 2019. SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat Biotechnol 37: 420–423. https://doi.org/10.1038/s41587-019-0036-z.

109. Rawlings ND, Barrett AJ, Thomas PD, Huang X, Bateman A, Finn RD. 2018. The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. Nucleic Acids Res 46:D624–D632. https://doi.org/10.1093/nar/gkx1134.

110. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer ELL, Tate J, Punta M. 2014. Pfam: the protein families database. Nucleic Acids Res 42:D222–D230. https://doi.org/10.1093/nar/gkt1223.

111. Abby SS, Neron B, Menager H, Touchon M, Rocha EP. 2014. MacSyFinder: a program to mine genomes for molecular systems with an application to CRISPR-Cas systems. PLoS One 9:e110726. https://doi.org/10.1371/journal.pone.0110726.

112. Abby SS, Rocha EPC. 2017. Identification of protein secretion systems in bacterial genomes using MacSyFinder. Methods Mol Biol 1615:1–21. https://doi.org/10.1007/978-1-4939-7033-9_1.

113. Guo J, Bolduc B, Zayed AA, Varsani A, Dominguez-Huerta G, Delmont TO, Pratama AA, Gazitua MC, Vik D, Sullivan MB, Roux S. 2021. VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. Microbiome 9:37. https://doi.org/10.1186/s40168-020-00990-y.

114. Nayfach S, Camargo AP, Schulz F, Eloe-Fadrosh E, Roux S, Kyrpides NC. 2021. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. Nat Biotechnol 39:578–585. https://doi.org/10.1038/s41587-020-00774-7.

115. Shaffer M, Borton MA, McGivern BB, Zayed AA, La Rosa SL, Solden LM, Liu P, Narrowe AB, Rodriguez-Ramos J, Bolduc B, Gazitua MC, Daly RA, Smith GJ, Vik DR, Pope PB, Sullivan MB, Roux S, Wrighton KC. 2020. DRAM for distilling microbial metabolism to automate the curation of microbiome function. Nucleic Acids Res 48:8883–8900. https://doi.org/10.1093/nar/gkaa621.

116. Fraley C, Raftery AE, Murphy TB, Scrucca L. 2012. mclust version 4 for R: normal mixture modeling for model-based clustering, classification, and density estimation. Technical report no. 597.

117. Emms DM, Kelly S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol 16:157. https://doi.org/10.1186/s13059-015-0721-2.

118. Dixon P. 2003. VEGAN, a package of R functions for community ecology. J. Veg. Sci. 14, 927.

119. Oksanen J, Kindt R, Legendre P, O'Hara B. 2017. The vegan package version 2.4.2. https://github.com/vegandevs/vegan/tags.

120. Kolde R, Kolde MR. 2015. Package 'pheatmap'. R Package 1. https://cran.r-project.org/web/packages/pheatmap/.

121. Wickham H. 2016. ggplot2: elegant graphics for data analysis. Springer, New York, NY.

122. R Core Team. 2013. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

123. Westram R, Bader K, Prüsse E, Kumar Y, Meier H, Gloeckner F-O, Ludwig W. 2011. ARB: a software environment for sequence data, p 399–406. In Handbook of molecular microbial ecology I: metagenomics and complementary approaches. Wiley-Blackwell, Hoboken, NJ.

124. Letunic I, Bork P. 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res 47:W256–W259. https://doi.org/10.1093/nar/gkz239.