



# Phytoplankton exudates and lysates support distinct microbial consortia with specialized metabolic and ecophysiological traits

Brandon Kieft<sup>a,1</sup> , Zhou Li<sup>b,c</sup>, Samuel Bryson<sup>a,d</sup>, Robert L. Hettich<sup>b</sup> , Chongle Pan<sup>b,c,e</sup> , Xavier Mayalf , and Ryan S. Mueller<sup>a,1</sup>

<sup>a</sup>Department of Microbiology, Oregon State University, Corvallis, OR 97331; <sup>b</sup>Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37830; <sup>c</sup>Graduate School of Genome Science and Technology, The University of Tennessee, Knoxville, TN 37996; <sup>d</sup>Department of Civil & Environmental Engineering, The University of Washington, Seattle, WA 98195; <sup>e</sup>Department of Microbiology and Plant Microbiology, University of Oklahoma, Norman, OK 73019; and <sup>f</sup>Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, CA 94550

Edited by Edward F. DeLong, University of Hawaii at Manoa, Honolulu, HI, and approved August 9, 2021 (received for review January 20, 2021)

**Blooms of marine phytoplankton fix complex pools of dissolved organic matter (DOM) that are thought to be partitioned among hundreds of heterotrophic microbes at the base of the food web. While the relationship between microbial consumers and phytoplankton DOM is a key component of marine carbon cycling, microbial loop metabolism is largely understood from model organisms and substrates. Here, we took an untargeted approach to measure and analyze partitioning of four distinct phytoplankton-derived DOM pools among heterotrophic populations in a natural microbial community using a combination of ecogenomics, stable isotope probing (SIP), and proteomics. Each <sup>13</sup>C-labeled exudate or lysate from a diatom or a picocyanobacterium was preferentially assimilated by different heterotrophic taxa with specialized metabolic and physiological adaptations. Bacteroidetes populations, with their unique high-molecular-weight transporters, were superior competitors for DOM derived from diatom cell lysis, rapidly increasing growth rates and ribosomal protein expression to produce new relatively high C:N biomass. Proteobacteria responses varied, with relatively low levels of assimilation by Gammaproteobacteria populations, while copiotrophic Alphaproteobacteria such as the Roseobacter clade, with their diverse array of ABC- and TRAP-type transporters to scavenge monomers and nitrogen-rich metabolites, accounted for nearly all cyanobacteria exudate assimilation and produced new relatively low C:N biomass. Carbon assimilation rates calculated from SIP data show that exudate and lysate from two common marine phytoplankton are being used by taxonomically distinct sets of heterotrophic populations with unique metabolic adaptations, providing a deeper mechanistic understanding of consumer succession and carbon use during marine bloom events.\***

proteomics SIP | dissolved organic matter | microbial loop | resource partitioning | phytoplankton bloom

Cyanobacteria and diatoms produce 65 to 85% of dissolved organic matter (DOM) in marine upwelling zones (1, 2), much of which is synthesized during short-term bloom events when carbon fixation exceeds respiration (3). The distinct ecophysiological features of these two phytoplankton lineages provide ecosystem services that are essential for the coastal ocean food web, such as seasonally offset bloom cycles (1, 4) and the production of DOM with unique compositions and stoichiometries that fuel successive trophic transfers of fixed carbon through initial assimilation by microbial osmotrophs (5–7).

Most of the DOM produced during diatom and cyanobacteria blooms is first channeled through the microbial loop (8), which assimilates or respire the complex pool of dissolved carbon and nutrients that is otherwise inaccessible to higher trophic levels (9). This metabolically coupled system of fixation and remineralization is a major determinant of marine food web stability and global DOM cycles (10). However, efforts to characterize how different microbial populations contribute to DOM turnover are

typically limited by the use of model organisms or model substrates to interpret emergent properties of the whole community (11–17).

In this study, we amended complex pools of diatom (*Thalassiosira pseudonana*) and cyanobacterium (*Synechococcus* sp. strain WH8102) DOM enriched with the <sup>13</sup>carbon (<sup>13</sup>C) stable isotope for proteomic stable isotope probing analysis (proteomic SIP) (17–19). Labeled cell lysate or exudate materials from each phytoplankton (four treatments) were separately amended to microcosms of coastal seawater and the <sup>13</sup>C isotopes of each DOM substrate pool were tracked into the newly synthesized proteinaceous biomass of the heterotrophic community with mass spectrometry. Our experiments use exudates to represent relatively low-molecular-weight DOM produced during the growth phase of a nonlimited phytoplankton bloom, while lysates represent relatively high-molecular-weight material produced by cell lysis during bloom senescence or from ongoing top-down processes, such as sloppy feeding and viral lysis (20–22). By combining proteomic SIP with metagenome-assembled genome (MAG) binning, we

## Significance

Marine dissolved organic matter, which originates from phytoplankton, holds as much carbon as Earth's atmosphere; yet, the biological processes governing its fate are primarily studied under idealized laboratory conditions or through indirect measures such as genome sequencing. In this work, we used isotope labeling to directly quantify uptake of complex carbon pools from the two primary sources of marine organic carbon (diatoms and cyanobacteria) by a natural microbial community. Our data show that carbon pools are partitioned into distinct microbial lineages whose physiological properties and resource acquisition strategies match the chemical nature of their preferred substrates. Our results provide ecological and functional insights into the patterns of microbial community structure changes that occur during marine phytoplankton blooms.

Author contributions: B.K., S.B., X.M., and R.S.M. designed research; B.K., Z.L., S.B., X.M., and R.S.M. performed research; Z.L., S.B., R.L.H., C.P., X.M., and R.S.M. contributed new reagents/analytic tools; B.K., Z.L., S.B., C.P., X.M., and R.S.M. analyzed data; and B.K., Z.L., S.B., R.L.H., C.P., X.M., and R.S.M. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

<sup>1</sup>To whom correspondence may be addressed. Email: kieft1bp@gmail.com or Ryan. Mueller@oregonstate.edu.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2101178118/-DCSupplemental>.

Published October 7, 2021.

\*See *SI Appendix, Fig. S8* for a visual abstract of this manuscript.

were able to measure how whole communities and individual populations assimilated DOM resources available during diatom and cyanobacteria blooms.

Patterns of assimilation of the four DOM substrates (diatom lysate [DLy], diatom exudate [DEx], cyanobacteria lysate [CLy], or cyanobacteria exudate [CEx]) into different heterotrophic population proteomes were used to 1) determine how communities partitioned phytoplankton DOM, 2) demonstrate how substrate specialization is linked to protein expression and ecophysiological adaptations, and 3) discuss how the activity of major microbial loop members impacts rates of freshly fixed carbon incorporation and net carbon and nutrient cycling. Our results suggest that the composition of the active heterotrophic community and their rates of processing phytoplankton DOM is controlled to a large extent by molecular properties of the DOM itself (e.g., here, cell lysates or exudates from different algal species); thus, changes in DOM source will have implications for organic matter cycling in the ocean and nutrient flow through food webs (23–25). By detailing the metabolic relationships between individual heterotrophic populations and phytoplankton-derived DOM constituents with methods such as proteomic SIP, we can provide empirical measurements for understanding population-level mechanisms of carbon use and developing community-level modeling of the fate of carbon and nutrients derived from marine primary productivity.

## Results

This work presents several lines of evidence from experimentally derived data to support one of the current models of the marine microbial loop, wherein distinct lineages of heterotrophic marine microbes are adapted to metabolize specific pools of phytoplankton-fixed DOM through resource partitioning. To investigate this model, we used a stable isotope tracing experiment in which  $^{13}\text{C}$ -labeled complex DOM pools were amended to mesocosms of coastal seawater containing natural microbial communities. Shotgun metagenomics and metaproteomic SIP data detailing the heterotrophic community before and after DOM inputs were used to determine community- and population-level changes in structure, biomass  $^{13}\text{C}$ -labeling, proteome functions with increased or decreased expression, the populations (and their proteins) enriched in  $^{13}\text{C}$ , and to ultimately calculate the contribution of specific populations to the uptake and potential turnover of common marine DOM pools.

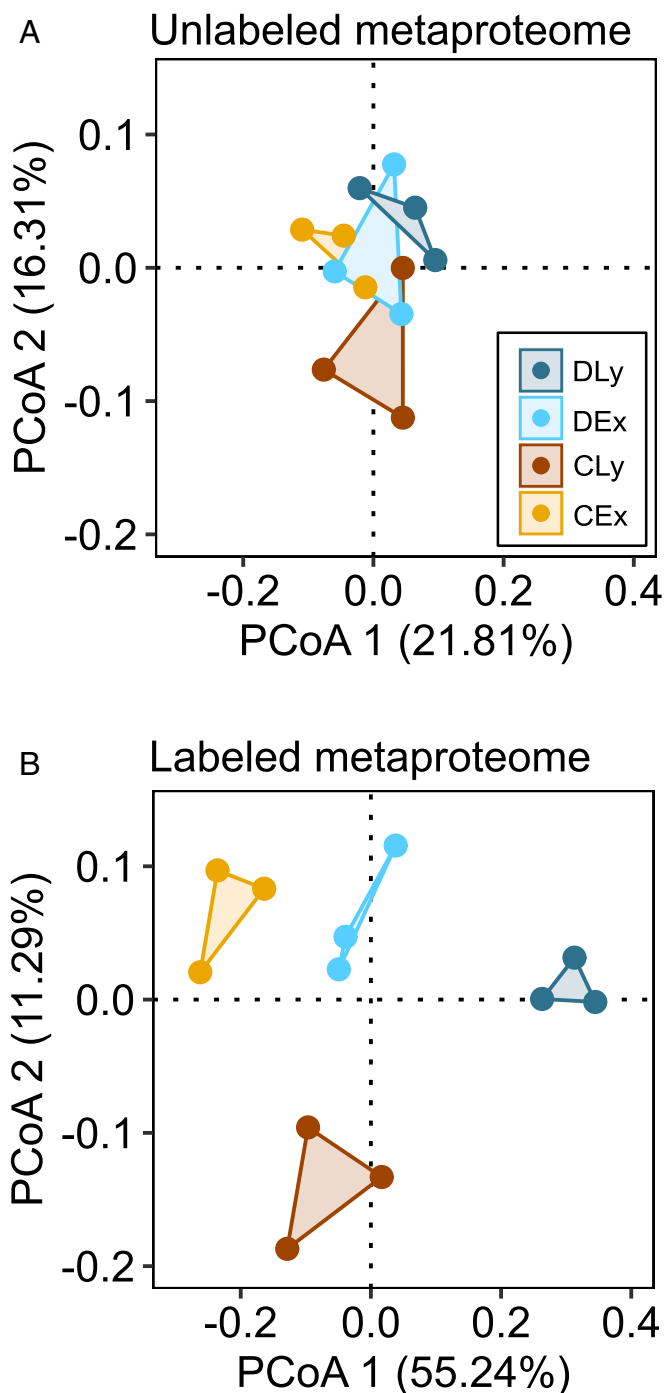
**Substrate-Dependent Labeling of the Community Metaproteome.** We observed significant  $^{13}\text{C}$  labeling of the coastal microbial community after only 15 h of 20 °C incubation with all four labeled substrates (*SI Appendix, Table S1*); however, the two phytoplankton species (diatom or cyanobacteria) and two cell fractions (lysate or exudate) did not elicit the same frequency or magnitude of  $^{13}\text{C}$  assimilation into heterotroph biomass. Community metaproteome label frequency (the percentage of a given set of proteins that is labeled with  $^{13}\text{C}$  above background) and average enrichment (the proportion of total C that is  $^{13}\text{C}$  in proteins enriched above background) were highest in the diatom lysate treatment ( $23.5 \pm 2.4\%$ ;  $86.2 \pm 1.6\%$ ) and lowest in the cyanobacteria exudate treatment ( $12.6 \pm 1.8\%$ ;  $54.6 \pm 8.6\%$ ) (*SI Appendix, Table S1*). Overall, label frequency and average enrichment were higher in diatom substrates compared to cyanobacteria and higher in lysate fraction compared to exudate from each phytoplankton species.

**Taxonomic Composition of Labeled and Unlabeled Metaproteomes.** The coastal microbial community metaproteome predominantly consisted of common marine prokaryotic taxa, including Actinobacteria, Archaea, Alpha-, Gamma-, and Betaproteobacteria, and Bacteroidetes (*SI Appendix, Fig. S1*). Taxonomic annotation of protein or peptide amino acid sequences can be challenging to confidently resolve to low taxonomic levels even with paired

metagenome and 16S ribosomal RNA (rRNA) subunit gene amplicon libraries; however, the majority of sequence annotations (>55%) could be confidently made to the genus level and >90% to the class level for well-represented marine lineages, and taxonomic structure gleaned from two DNA-based analyses showed similar patterns of microbial representation (*SI Appendix, Fig. S2*). The taxonomic composition of the unlabeled fraction of community metaproteomes was not significantly different between the four substrate treatments based on phytoplankton species, cell fraction, or the interaction of both (Fig. 1A), nor was the taxonomic composition of metagenomes or 16S rRNA gene sequences. Each substrate treatment did, however, elicit a significant shift from the no-substrate incubation ( $T_{15}$ ) and no-incubation ( $T_0$ ) controls (*SI Appendix, Fig. S3*). In comparison, the taxonomic composition of the labeled fraction of community metaproteomes was significantly different between substrate treatments and was most strongly differentiated by phytoplankton species, followed by cell fraction, but not on the interaction effect (permutational multivariate analysis of variance: phytoplankton species  $R^2 = 0.33$ ,  $F_{1,8} = 8.4$ ,  $P < 0.001$ , cell fraction  $R^2 = 0.26$ ,  $F_{1,8} = 6.6$ ,  $P < 0.001$ ; Fig. 1B).

Overall, 92.5% of labeled proteins with confident taxonomic annotation were assigned to Bacteroidetes, Alpha-, and Gammaproteobacteria populations. The taxonomic composition of labeled metaproteomes across treatments showed notable differences in representation of these three lineages (Fig. 2A). Proteins from populations of Alphaproteobacteria made up >75% of labeled proteins recovered from the cyanobacteria exudate treatment but <35% from diatom lysate, while Bacteroidetes proteins made up >50% of labeled proteins in the diatom lysate treatment and <10% from cyanobacteria exudate (Fig. 2A). Gammaproteobacteria populations made up a smaller proportion of the total labeled proteins across treatments (5 to 10%) but were proportionally higher in the exudate compared to lysate treatments from both phytoplankton. Within each lineage, active family-level populations generally followed a similar trend in relative labeled metaproteome abundance across treatments (*SI Appendix, Fig. S4*). For example, the abundant Alphaproteobacteria families, Rhodobacteraceae, Rhodospirillaceae, and Pelagibacteraceae, all comprised a higher proportion of labeled proteins from exudate treatments, particularly from the cyanobacterium, while both the abundant Bacteroidetes families, Flavobacteriaceae and Cryomorphaceae, were more represented in lysate treatments, particularly from the diatom (*SI Appendix, Fig. S4, Top*). These patterns in over- or under-representative protein abundance in a given treatment were not apparent when comparing bulk proteins from each treatment to bulk proteins from the control (*SI Appendix, Fig. S4, Bottom*), again highlighting the sensitivity of proteomic SIP to measure responses to substrate availability at the initial stages of resource partitioning.

In addition to differences in taxonomic representation of labeled proteomes across treatments labeled proteins expressed by the three dominant lineages also showed distinct patterns of  $^{13}\text{C}$  enrichment (Fig. 2B). Although labeled Alphaproteobacteria proteomes were not the most highly enriched on average among major lineages, their levels were consistent across all treatments, with 61.2 to 74.3%  $^{13}\text{C}$  enrichment in labeled proteins regardless of phytoplankton species or cell fraction (Fig. 2). Enrichment levels of labeled Gammaproteobacteria proteomes varied more based on treatment than Alphaproteobacteria and were much more labeled in diatom DOM treatments than cyanobacteria DOM (*SI Appendix, Table S2*), with particularly low enrichment in the cyanobacteria exudate treatment (24.5%). Bacteroidetes showed a similar pattern to Gammaproteobacteria but with a more extreme shift from very high average enrichment in the diatom lysate treatment (>95%) to the lowest of all three lineages in cyanobacteria exudate (21.6%). Label frequency of individual lineage proteomes followed a similar pattern to average enrichment



**Fig. 1.** PCoA plot depicting taxonomic structures of the unlabeled (A) and labeled (B) fractions of metaproteomes in the four DOM treatments. Proportion of variance explained in PCoA plots is shown along each axis in parentheses.

[e.g., with Alphaproteobacteria proteomes relatively consistent in label frequency across treatments (18.8 to 22.3%), while Bacteroidetes proteomes ranged from 48.2% in diatom lysate to just 5.2% in cyanobacteria exudate (*SI Appendix, Table S2*)].

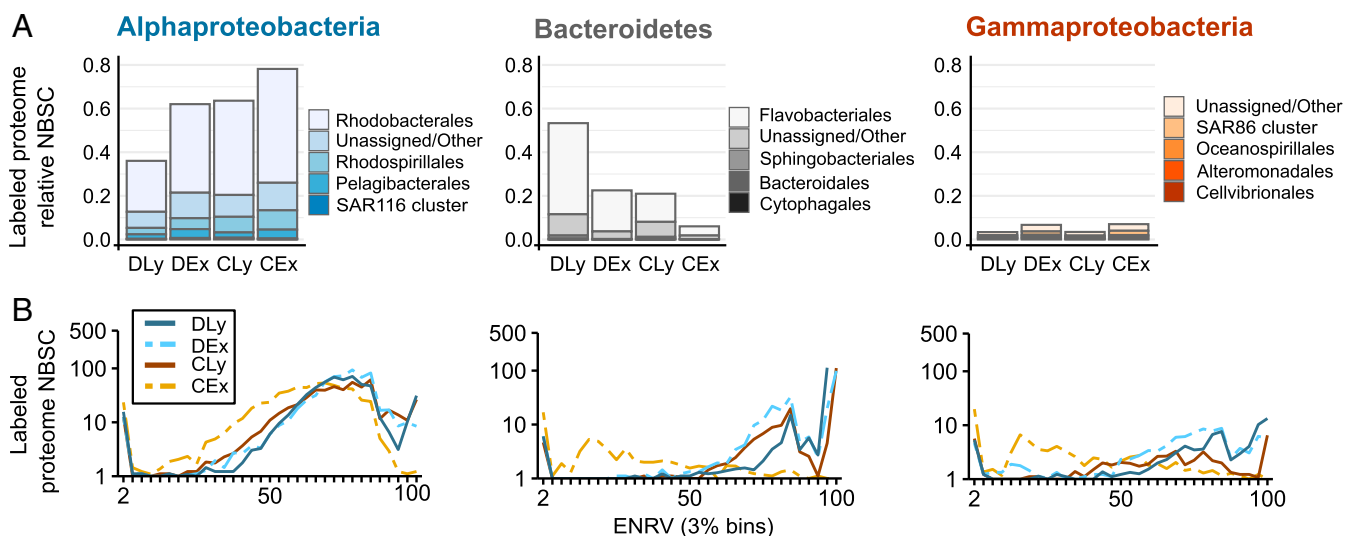
To determine if these enrichment patterns were consistent at more resolved taxonomic ranks, we performed nonparametric significance tests on  $^{13}\text{C}$ -enrichment values of all genus-level proteomes in our samples. Overall, >40 populations within Alphaproteobacteria, Gammaproteobacteria, and Bacteroidetes

were significantly enriched compared to the null expectation, and grouping these taxa by their enrichment values across treatments supported a general pattern of phylogenetic conservation in DOM assimilation (*SI Appendix, Fig. S5*). Given that derived traits (e.g., key metabolic genes and membrane uptake systems) are more likely to be conserved among closely related populations, we hypothesized that mechanisms underlying conserved patterns of assimilation among populations within the three dominant lineages but divergent between them would be evident in the functional compositions of their proteomes.

#### Functional Composition of Proteomes across Substrate Treatments and Taxa.

To explore this hypothesis and to identify any potential functional drivers of apparent niche differentiation, we examined how the abundance of 10 dominant protein functional groups (representing 72.1% of total annotated protein abundance) were distributed both within and between the proteomes of the three major bacterial lineages. Within a taxon, differences in the functional composition of the postincubation  $T_{15}$  control and labeled DOM treatment proteomes indicated a shared response across most populations of a shift from a physiological state of low activity, where transport proteins dominated the proteome, to one of growth and proliferation, where ribosomal proteins and transcription/translation proteins dominate new function synthesis (Fig. 3 and *SI Appendix, Table S3*). For example, the functional composition of the labeled Alphaproteobacteria proteomes in DOM treatments was considerably different from in the postincubation control, with significantly lower average proportion of ABC-type (−15.6%), TRAP-type (−5.5%), and polyamine-specific (−3.0%) transporters and higher average proportions of ribosomal protein (+15.9%), transcription/translation factor (+5.7%), and (to a lesser extent) polymerases (+0.5%). Similarly, in the Bacteroidetes proteomes, TonB-dependent (TBD) transporter proteins made up a lower average proportion (−8.1%) of labeled proteomes after DOM treatments than in the 15-h no-substrate control proteomes ( $T_{15}$ ), while ribosomal proteins (+17.9%), transcription/translation factors (+11.2%), and polymerases (+2.5%) increased in proteome relative abundance after DOM treatments (Fig. 3).

In addition to the general functional shift toward growth after substrate addition within each lineage proteome, there were notable changes to the taxonomic representation of these 10 most abundant metabolic functions in the community metaproteome. For example, the taxonomic distribution of proteins assigned to “growth and information processing” functions (ribosomal proteins, ATPases, transcription/translation factors, nucleotide polymerases, and nucleoside kinases) did not differ significantly between the initial ( $T_0$ ) and postincubation ( $T_{15}$ ) control communities and were distributed similarly to what might be expected from the community taxonomic composition (*SI Appendix, Fig. S6A*, functions near the ternary centers). Some functions, however, were clearly lineage specific (*SI Appendix, Fig. S6A*, functions near the ternary corners), as suggested by previous work (26–28). For example, the distribution of proteins annotated as TBD transporters and *sus*-like biopolymer transporters was skewed toward Bacteroidetes and Gammaproteobacteria, which collectively represented 94.3% of proteins with these functions expressed by the three major lineages. Conversely, the abundance of proteins annotated to ABC- and TRAP-type transporters and polyamine-specialized transporters was highly skewed (91.2%) toward Alphaproteobacteria populations. These stark differences in resource transport strategies between major lineages not only reflect expressed proteomes but also adaptations in genomes themselves. For example, nearly all TBD subunit genes from the metagenomes were annotated to Bacteroidetes populations, while ABC transport subunit genes were predominately annotated to Alphaproteobacteria, suggesting that past evolutionary selection, rather than regulatory



**Fig. 2.** Stacked bars represent the taxonomic composition of labeled metaproteomes in each DOM treatment, shown as relative NBSC abundance of order-level population-labeled proteomes in the three major microbial lineages. Community proteins not assigned to these three lineages are not shown but are included in the relative abundance calculations (A). Line plots depict absolute NBSC abundance of labeled proteins organized by average enrichment (ENRV histogram of 2 to 100%  $^{13}\text{C}$  in 3% bins) across treatments for the three major lineages (B).

mechanisms, strongly dictates the use of different resource acquisition systems by each respective lineage.

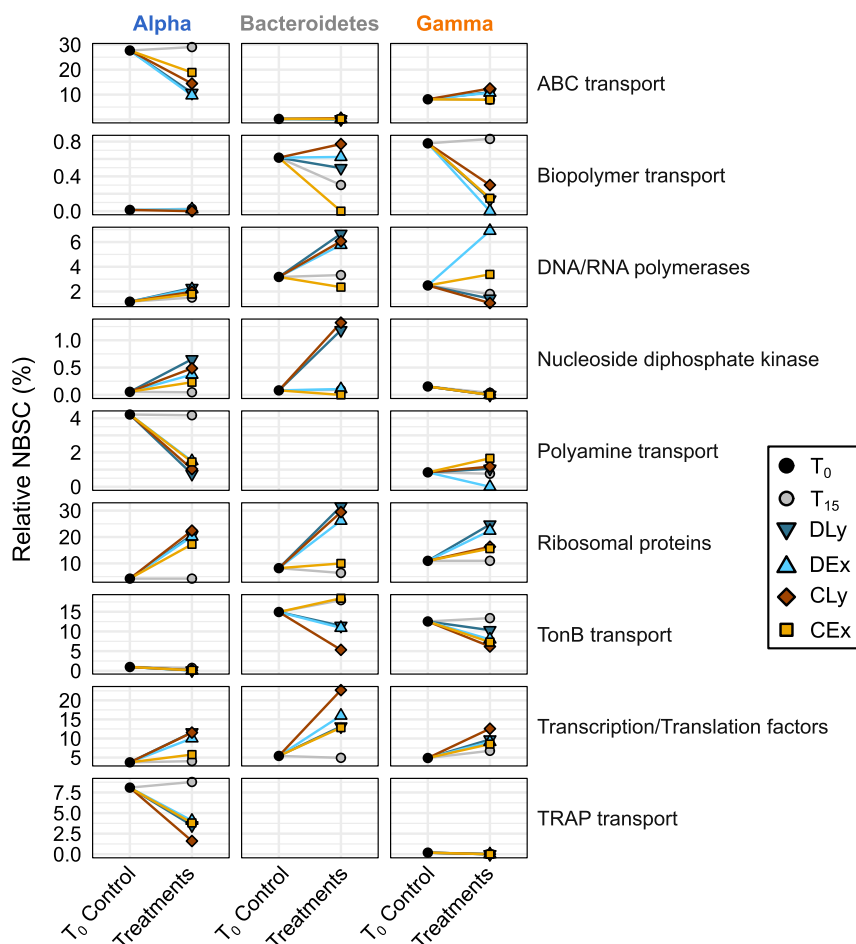
In DOM substrate treatments, labeled proteins from specialized functions (e.g., membrane transport systems) remained asymmetrical in taxonomic representation (*SI Appendix, Fig. S6B*), indicating that lineage-specific expression of these derived functions was maintained. However, labeled proteins annotated to the “growth and information processing” functional category shifted from relatively evenly distributed in the controls toward skewed, lineage-specific patterns in the substrate treatments (*SI Appendix, Fig. S6B*). In accordance with the  $^{13}\text{C}$ -enrichment data, substrate treatments with the most extreme shifts in taxonomic composition of these growth-related functional groups were the cyanobacteria exudate and the diatom lysate. In the former, all labeled functional groups except TBD and *sus*-like biopolymer transporters were strongly skewed toward Alphaproteobacteria representation. In the latter, Bacteroidetes populations synthesized most of the abundant proteins for all functions. While the percent of labeled proteins annotated to Gammaproteobacteria populations never reached levels observed for Alphaproteobacteria and Bacteroidetes, a significant increase in Gammaproteobacteria representation of the 10 functional categories was observed in both exudate treatments. For example, labeled Gammaproteobacteria proteins classified as ABC transporters, TBD transporters, ribosomal proteins, DNA/RNA polymerases, and transcription/translation factors all increased to 10 to 20% of total proteins identified in exudate treatments versus just 0 to 10% for lysate treatments.

**Relationships between Resource Preferences and Ecophysiology.** To determine whether protein  $^{13}\text{C}$  enrichment (i.e., de novo protein synthesis) and shifts to a proteome dominated by growth-related functions resulted in genome replication (especially after amendment of a preferred DOM fraction), we estimated growth rates for individual microbial populations by read-mapping to high-quality MAGs and estimated growth rates using resulting peak-to-trough coverage calculations (29). In support of the link between protein synthesis and cellular growth, average estimated growth rates for populations in the three dominant lineages were strongly correlated with their respective enrichment patterns across the four treatments (Pearson  $r = 0.76$ ; Fig. 4 A and B). Accordingly, Bacteroidetes populations exhibited the highest

estimated growth rates in the diatom lysate treatment and the lowest in the cyanobacteria exudate treatment, while Alphaproteobacteria had the highest growth estimate in the cyanobacteria exudate treatment and lowest in the diatom lysate treatment (Fig. 4B and *SI Appendix, Table S4*).

Evidence from these experimental data indicate that the four DOM components elicited bottom-up selection for a different set of active, specialized microbial populations. Based on previous work showing that the metabolomic profiles of phytoplankton DOM can vary significantly between species and cellular fractions (30, 31) and on the assumption that heterotrophic populations would show specialization on substrates with stoichiometries that match their biomass requirements, we sought to estimate whether C:N requirements of the most significantly enriched genera in each treatment varied according to substrate preferences. Indeed, the de novo-synthesized proteins of highly active populations in the three major heterotrophic lineages had distinct amino acid C:N, and the genes that encode those proteins also had distinct guanine+cytosine percent content (G+C%) (Fig. 4C and *SI Appendix, Table S5*). In general, Bacteroidetes populations had lower gene G+C% and higher protein C:N requirements compared to Alphaproteobacteria, with Gammaproteobacteria falling in between, corroborating our presumption of a link between stoichiometries of DOM for the four substrates and microbial biomass.

In order to estimate and compare the ecological outcomes of total DOM uptake by the three major lineages, labeled carbon assimilation rates were calculated in the four treatments using experimentally derived values: growth rates, proportional protein abundances,  $^{13}\text{C}$ -enrichment values, and label frequencies. Bacteroidetes populations exhibited the highest average daily substrate assimilation in diatom lysate treatments, at  $3,571.8 \text{ ng } ^{13}\text{C} \cdot \text{L}^{-1} \cdot \text{d}^{-1}$ , while their rate dropped to  $39.5 \text{ ng } ^{13}\text{C} \cdot \text{L}^{-1} \cdot \text{d}^{-1}$  in the cyanobacteria exudate treatments (Table 1 and *SI Appendix, Fig. S7*). The Gammaproteobacteria assimilation rate was highest in the diatom exudate treatments with an average of  $259.0 \text{ ng } ^{13}\text{C} \cdot \text{L}^{-1} \cdot \text{d}^{-1}$  and lowest in cyanobacteria exudate at  $60.9 \text{ ng } ^{13}\text{C} \cdot \text{L}^{-1} \cdot \text{d}^{-1}$ , while Alphaproteobacteria assimilation rates remained relatively stable across all DOM treatments ( $1,472.4$  to  $2,138.6 \text{ ng } ^{13}\text{C} \cdot \text{L}^{-1} \cdot \text{d}^{-1}$ ; Table 1).



**Fig. 3.** The functional structure of the initial no-incubation control community metaproteome ( $T_0$ ) compared to that of each postincubation treatment (the  $T_{15}$  no-substrate control plus the four DOM substrates). Data for the  $T_0$  and  $T_{15}$  control samples show the functional structure of total proteins recovered (all unlabeled because no  $^{13}\text{C}$  was amended), while data for the four DOM treatments show functional structures of only labeled proteins to highlight de novo protein synthesis responses by each lineage. The mean relative NBSC abundance of each of the 10 major functions on the right y-axis is faceted into three columns representing the proteomes of each of the three major lineages. Points are colored and shaped by treatment, and lines connecting  $T_0$  with each postincubation treatment are colored by treatment.

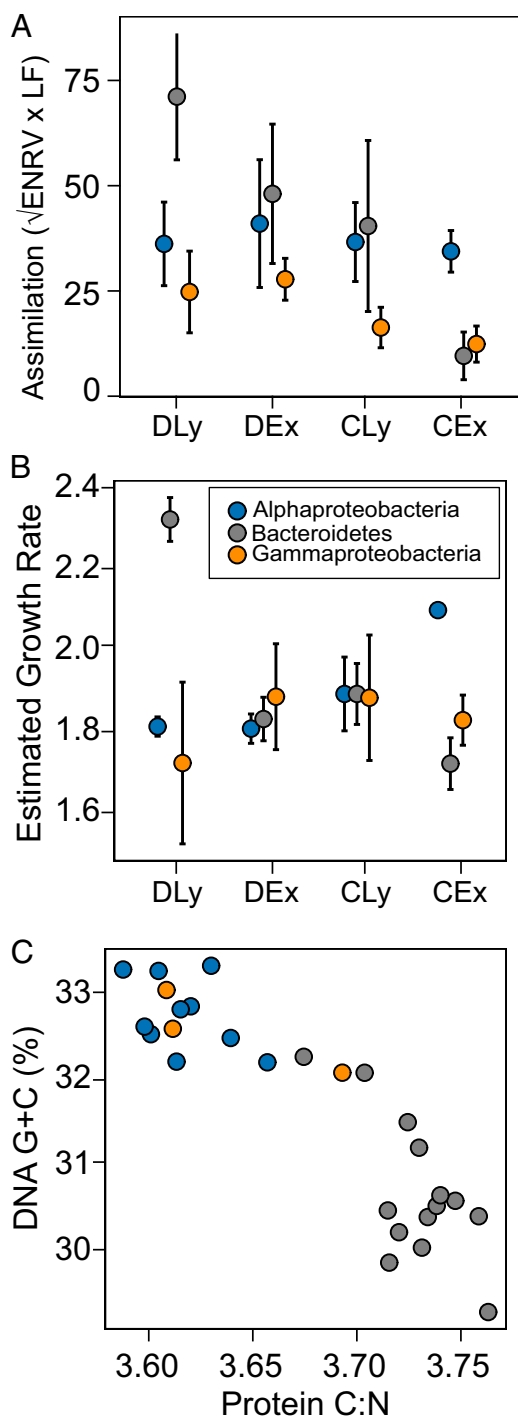
## Discussion

A significant fraction of newly fixed marine DOM flows through heterotrophic cells in the microbial loop, which control its subsequent release as  $\text{CO}_2$  through respiration or retention in food webs as biomass (32). The partitioning of the complex DOM pool among hundreds of co-occurring populations is a key mechanism controlling marine biogeochemical cycling (16–18, 26, 33–38). In this work, we combine the approach of proteomic SIP with paired metagenome sequencing and binning to measure partitioning of complex marine DOM into the population-specific biomass of a natural marine microbial community. Our results show that DOM composition has a strong bottom-up control in assembling the active subset of the total community, which is generally composed of populations with derived, specialized metabolic and ecophysiological traits. This implies that different lineages of abundant heterotrophic microbes can coexist in part due to differentiated resource use capabilities and that the distributed nature of metabolic strategies among active populations in the community facilitate the efficient turnover of the highly heterogeneous pool of DOM available throughout a phytoplankton bloom.

In accordance with the largely labile nature of fresh phytoplankton-derived DOM (20), all four  $^{13}\text{C}$ -labeled substrate pools were widely assimilated by the coastal heterotrophic community. Despite the fundamentally different character of the four substrates, the

15-h incubation periods led to high label frequency and  $^{13}\text{C}$ -enrichment of all treatment metaproteomes ( $12.1 \pm 3.0\%$ ;  $54.4 \pm 15.9\%$ ). The speed and extent of labeling observed here supports previous results using isotopically labeled model substrates (39) and phytoplankton-derived DOM (36). In our work, substrate addition did not significantly shift the taxonomic or functional compositions of metaproteomes recovered before versus after incubation. Rather, substrate-induced changes to the taxonomic composition of metaproteomes were only detected when comparing the initial metaproteome to the newly synthesized ( $^{13}\text{C}$ -labeled) metaproteome. These results highlight the sensitivity of the proteomic SIP method to detect rapid responses to environmental change prior to the large-scale shifts in biological activity that can be detected with unlabeled approaches (e.g., metagenomics), which are dependent on physiological responses such as genome replication and cell division. This observation has been noted elsewhere when measuring microbial responses to DOM inputs, where changes to protein expression (i.e., modulation of community function) occur before any detected changes in community taxonomic composition (40).

**Population-Level Substrate Preferences Are More Similar with Increasing Phylogenetic Relatedness.** Shifts in taxonomic composition of labeled metaproteomes in each treatment could be attributed to different



**Fig. 4.** Assimilation metrics (calculated as square root–transformed ENRV\*LF) of the three major lineages across substrate treatments (A). Average growth rate estimates of high-quality MAG populations in each lineage across substrate treatments (B). Bars in (A) and (B) depict the SE of each value from triplicate samples. The scatterplot of estimated carbon:nitrogen requirements of newly synthesized proteins in relation to the G+C content of DNA encoding those proteins for significantly enriched genera in each major lineage (C).

responses by common marine heterotrophic bacteria, confirming general patterns found in previous studies (41–44). The vast majority of peptides from labeled metaproteomes (>85%) were annotated to the marine and coastal microbial lineages Actinobacteria, Archaea, Alpha-, Gamma-, and Betaproteobacteria, and Bacteroidetes. All

four DOM pools were assimilated at some level by all of these lineages, but the magnitude and frequency of protein labeling in each lineage varied significantly across treatments. Importantly, resource assimilation patterns were similar among populations belonging to the same taxonomic class but distinct between classes, corroborating previous findings that responses to model DOM compounds are typically conserved at broad phylogenetic resolution (17, 22). This observation supports a model in which the strength of resource competition between heterotrophic populations is inversely related to their phylogenetic distance (45), implying that selection pressure exerted upon closely related taxa in the coastal ocean does not act on DOM utilization traits. Rather, as has been supported previously, allopatry and biogeography may be more important for generating species- or strain-level diversity in natural environments (46, 47), while higher-level diversity is maintained by divergent resource niches that help reduce competitive overlap.

Although the four substrate treatments appeared to strongly and reproducibly select for different sets of active taxa, individual population proteomes typically showed the same response, regardless of DOM treatment: proportional increases in synthesis of ribosomal proteins and decreases in substrate transporters after DOM amendment. This growth-prioritization response was exacerbated when the DOM was a preferred resource (i.e., when the overall proteome enrichment of the population was highest). From this pattern, we purport that the observed shift in metaproteome functional structure in each treatment is predominantly driven by a subset of “responder” populations that rapidly assimilate their preferred substrate into biomass when it becomes available, rather than being driven by a community-wide shift in which all populations alter their proteome in response to newly available resources.

#### Differences in DOM Assimilation between Highly Abundant Taxa.

Fitting our data into this theoretical framework, a general picture of how resources are distributed among taxa in the marine microbial loop emerges. Lineages of Bacteroidetes, represented predominantly by a few major Flavobacteriales taxa in our samples, preferentially assimilate DOM components derived from diatom lysates that are probably rapidly available upon cell death from sloppy feed, viral lysis, or mass senescence in the terminal phase of a bloom. On the other hand, lineages of Alphaproteobacteria, represented primarily by Pelagibacterales, Rhodobacterales, and Rhodospirillales taxa, assimilate mostly cyanobacteria- rather than diatom-derived substrates and also specialize on the exuded fraction of DOM that is more consistently available during nominal phytoplankton growth or in the early stages of a bloom cycle. Despite representing ~12% of unlabeled metaproteomes, Gammaproteobacteria lineages were not highly competitive in assimilating substrates in our experiments. For lineages such as Oceanospirillales and Cellvibrionales, which did demonstrate moderate labeling, their uptake patterns diverged from those of both the Alphaproteobacteria and Bacteroidetes lineages by exhibiting higher enrichment on diatom- compared to cyanobacteria-derived DOM and nearly twofold higher enrichment on exudates than lysates. Considering the magnitude of total assimilation accounted for by these three prevalent lineages, their strong preferences and competitive abilities for assimilating different resources are likely to be essential for the overall efficiency of DOM cycling in marine ecosystems (48), especially throughout the phases of a mixed-species phytoplankton blooms (6).

#### Protein Expression and Ecophysiological Evidence for Bacteroidetes Lysate DOM Resource Preference.

Combining metagenomics and metaproteomics with our SIP approach allowed for the exploration of how substrate preferences of active populations were related to their metabolic and ecophysiological characteristics. Bacteroidetes populations, which were most competitive in assimilating diatom lysate, almost exclusively expressed TBD membrane transport

**Table 1. Estimations of carbon assimilation rate (ng  $^{13}\text{C} \cdot \text{L}^{-1} \cdot \text{d}^{-1}$ ) for each major lineage across the substrate treatments**

Lineage	Treatment	Mean C assimilation rate (ng $^{13}\text{C} \cdot \text{L}^{-1} \cdot \text{d}^{-1}$ )	SE of C assimilation rate
Alphaproteobacteria	CEx	1,472.35232	20.7143259
Alphaproteobacteria	CLy	1,769.06243	68.2803903
Alphaproteobacteria	DEx	2,138.59708	3.93247522
Alphaproteobacteria	DLy	1,738.27875	76.1195034
Bacteroidetes	CEx	39.5413883	8.77430344
Bacteroidetes	CLy	669.653816	118.222284
Bacteroidetes	DEx	956.068627	6.35663931
Bacteroidetes	DLy	3,571.80057	118.436686
Gammaproteobacteria	CEx	60.8903329	5.05937951
Gammaproteobacteria	CLy	91.5045243	4.0882636
Gammaproteobacteria	DEx	259.002335	5.2456264
Gammaproteobacteria	DLy	203.802878	16.0257199

See *SI Appendix, Table S6* for the full table of empirically derived and cited values used to calculate these rates as well as replicate data used to calculate averages and SEs. See *SI Appendix, Fig. S7* for plotted rate values.

proteins among all identified substrate uptake systems in the metaproteome. These function by binding polymeric and carbon-rich molecules making up the structural and storage components of diatom cellular biomass (6, 30, 49–51). Not only were TBD transporters abundant within the genomes and proteomes of this lineage, but a significant proportion of newly synthesized proteins (i.e.,  $^{13}\text{C}$ -enriched proteins) from Flavobacteriales populations appear to function as part of the *sus*-like polysaccharide system, which acts to recognize and degrade glycans (52) and is involved in competitive uptake of phytoplankton carbohydrates (53, 54). These results suggest a metabolic program within Bacteroidetes taxa that specializes on carbon-rich high-molecular-weight DOM, resulting in these taxa being most competitive when this substrate type is present and poorly competitive when it is not (e.g., in cyanobacteria exudate). Physiological signatures of this specialization pattern on C-rich DOM was also borne out in the estimated carbon:nitrogen requirement of Bacteroidetes nucleotides and amino acids, which was higher than other active taxa in the community we sampled. Also of note and exclusive to the Bacteroidetes proteomes was the detection of peroxiredoxin proteins, which provide protection against oxidative stress. Both peroxiredoxin enzymes and biopolymer transporters assigned to the Bacteroidetes became enriched in all substrate treatments, except for the cyanobacteria exudate treatment, implying that cells of this lineage are particularly well-suited for growth in the phycosphere, where oxygen levels and mucilage will be highest.

The synthesis of these findings generally agrees with previous observations that Bacteroidetes populations increase in relative abundance and cell-specific activity in the middle and late phases of phytoplankton blooms and encode a large number of specialized enzymes allowing for the degradation and assimilation of polysaccharides (17, 26, 53, 55–58). Our data further suggest that lysate DOM, particularly those derived from diatoms, stimulate higher cellular growth rates in Bacteroidetes taxa relative to other co-occurring populations and other substrates we tested, allowing this lineage to translate their resource use strategy into growth and division in a suitable environment. This hypothesis was also supported by examining the shift in proportional representation of functions in the labeled metaproteomes from each treatment compared to the control incubations; when diatom lysate was amended, the majority of substrate-derived (i.e.,  $^{13}\text{C}$ -labeled) ribosomal proteins and transcription/translation factors synthesized by the community were annotated to Bacteroidetes populations. When contextualizing proteome enrichment and cell abundance of Bacteroidetes as DOM carbon uptake rates, the ecological role of this lineage was salient, with ~80-fold higher carbon uptake rates of diatom lysate compared to cyanobacteria exudate and over double

the assimilation rate of diatom lysate compared to the other two major lineages.

**Protein Expression and Ecophysiological Evidence for Alphaproteobacteria Exudate DOM Resource Preference.** In contrast to the Bacteroidetes, Alphaproteobacteria populations expressed mostly ABC and TRAP membrane transport proteins that target a wide range of polyamines, nucleosides, amino acids, short chain fatty acids, and other uncharacterized small molecules. Phytoplankton exudates are relatively rich in these compounds, and DOM exuded by cyanobacteria in particular can contain high concentrations of polyamines and organic osmolytes (31, 59, 60). Our observation that three-quarters of  $^{13}\text{C}$ -enriched peptides in cyanobacteria exudate treatments were annotated to Alphaproteobacteria suggests that their reliance on ABC and TRAP expression is a highly successful strategy for assimilation of this specific DOM fraction. The relatively low estimated carbon:nitrogen ratio requirements of Alphaproteobacteria nucleotides and proteins also suggests that cyanobacteria DOM (generally one-half the C:N ratio of diatoms) is more congruent with the elemental requirements of this lineage than the other substrates fractions tested (42, 43). As was the case in the Bacteroidetes lineage, the relatively higher estimated growth rates of Alphaproteobacteria taxa in the presence of their preferred resource pool suggests enhanced cellular growth and reproduction, and indeed, nearly 90% of newly synthesized ribosomal proteins and transcription/translation factors in the cyanobacteria exudate treatment were annotated to Alphaproteobacteria populations.

Although Alphaproteobacteria populations constituted the proportional majority of labeled proteins in the cyanobacteria exudate treatment (probably due to low assimilation by other taxa), the average enrichment values of this lineage were remarkably consistent across all four substrates. Most Alphaproteobacteria taxa have been shown to have relatively stable abundances and metabolic activities during phytoplankton bloom cycles compared with other heterotrophic bacteria (26) and have been noted to lack strong spatiotemporal gene expression patterns during active blooms (57, 61, 62). Similarly, populations of the most abundant Alphaproteobacteria family in our samples, Rhodobacteraceae, have been shown to retain high and consistent metabolic activity and growth rate despite marked fluctuations in organic matter and chlorophyll concentrations in a natural bloom (56). These results are seemingly at odds with some isolate and coculture studies which show substantial remodeling of gene expression among copiotrophic Alphaproteobacteria populations (11); however, observations during naturally occurring blooms or experiments that used complex DOM sources may be more representative of steady

metabolic activity in Alphaproteobacteria. Indeed, the Alphaproteobacteria as a whole showed little variability in carbon uptake rates across treatments and only dominated daily cyanobacteria exudate assimilation because Bacteroidetes and Gammaproteobacteria lineages appeared to have over an order of magnitude slower uptake rates. Taken together, our data suggest that populations in this lineage specialize on a subset of the low-molecular-weight and relatively N-rich molecules of marine DOM, which are probably present to some extent in all phytoplankton-derived resources but not accessed as readily by lineages of other phyla.

**Heterogeneity in DOM Responses between Alphaproteobacteria Populations.** Although aggregating Alphaproteobacteria populations at the class level was informative for community-scale interpretation, there was a notable pattern of heterogeneity in the representation of individual Alphaproteobacteria taxa in labeled and unlabeled treatment metaproteomes, implying more nuanced metabolic strategies than in other heterotrophic lineages. For example, Rhodobacteraceae and Rhodospirillaceae proteomes both had slightly (but not significantly) lower relative abundance in DOM treatments compared to the control, while Pelagibacteraceae proteomes were slightly more represented in the treatments (shown as deviations from the 1:1 relationship in *SI Appendix, Fig. S4, Bottom*); this may imply that the latter taxon increased protein expression more than the former taxa during substrate incubations. However, Pelagibacteraceae had very low representation in the labeled portion of treatment metaproteomes compared to total treatment metaproteomes (unlabeled + labeled), while Rhodobacteraceae and Rhodospirillaceae had higher representation in the labeled protein pools (shown as deviations from the 1:1 relationship in *SI Appendix, Fig. S4, Top*), implying a significant difference in the number of proteins synthesized from the amended  $^{13}\text{C}$ -labeled DOM by these populations.

We propose that this observation supports two possible conclusions. First, Pelagibacteraceae cells are synthesizing more proteins in the substrate-addition treatments compared to the no-substrate control, but they are not efficiently using  $^{13}\text{C}$ -labeled resources for anabolism and are instead relying on existing (unlabeled) substrates for biomass synthesis. Alternatively, since these results are based on proportional abundance, the relative increase in Pelagibacteraceae representation in substrate treatments could result from an offsetting loss of protein biomass assigned to other taxa, even if the Pelagibacteraceae remain largely unresponsive during the incubation period. While our data cannot definitively rule out this zero-sum-game scenario, our results do show that most other dominant taxa did exhibit high  $^{13}\text{C}$ -enrichment and activity, undermining the latter option that other abundant taxa were decreasing in biomass relative to the Pelagibacteraceae. Furthermore, there is precedence for a mechanism of metabolic partitioning that supports the former option describing a low carbon use efficiency model and that may be worth considering for future studies of Pelagibacteraceae physiology. For example, Guillemette et al. observed that freshwater bacterioplankton preferentially used autochthonous organic matter from phytoplankton for respiration, whereas allochthonous organic matter from terrestrial sources was used for biomass production (63). Applying this model to marine Pelagibacteraceae in our experiments would suggest that these cells preferentially use the amended freshly fixed, phytoplankton-derived ( $^{13}\text{C}$ -labeled) metabolites for respiration, and the preexisting pool of unlabeled DOM for biosynthesis. Evidence in partial support of this hypothesis comes from findings that key molecules of fresh DOM exuded from cyanobacteria are used primarily for ATP generation through respiration and only marginally for biomass production by *Pelagibacteria* spp. (64–66). Further experiments to trace labeled organic inputs into both anabolic and catabolic

products will be needed to address these more nuanced metabolic processes.

**Protein Expression and Ecophysiological Evidence for Mixed or Inconclusive DOM Resource Preferences of Other Taxonomic Lineages.** Gammaproteobacteria proteins comprised the third most abundant subset of all unlabeled metaproteomes and were highly represented in the compositional analysis of the metagenome assembly data. Gammaproteobacteria lineages appear to favor the assimilation of exudate rather than lysate DOM, with approximately double the number of labeled peptides being detected in both exudate treatments compared to lysate from the same phytoplankton. Overall, though, far fewer peptides could be confidently assigned to finer taxonomic levels compared to the proteomes of the Alphaproteobacteria and Bacteroidetes lineages, obscuring detailed analyses of most Gammaproteobacteria populations. For example, the relative abundances of the proteomes of assigned taxonomic families within the Gammaproteobacteria (i.e., Porticoccaceae, SAR86 and SUP05 clades) were generally comparable to those of other low-abundance taxa, such as the Marine Group II Euryarchaeota and Acidimicrobiaceae.

Low-abundance taxonomic groups showed similarly low levels of enrichment in substrate treatments on the time scale we used in our experiments. Given that we observed relatively little  $^{13}\text{C}$  assimilation by low-abundance groups, their contribution to ecosystem-scale carbon cycling appears to be proportional to their low representation. In a recent estimate of microbial population-level production compared to relative abundance in a coastal community, the most abundant bacteria accounted for a disproportionately high amount of biomass and activity, whereas rare taxa were proposed to not be able to grow fast enough in these competitive communities to add significantly to the total production (67). Importantly though, the low-abundance taxa may occupy niches not filled by the abundant populations in the initial communities, which implies that substrate turnover may be dependent on initial community structure in addition to DOM acting as a bottom-up mechanism of community assembly. This simultaneous and reciprocal relationship between resource composition and biological community structure is supported by a literature with many cases of microbe-dependent DOM turnover and DOM-dependent microbial community assembly (36, 50, 68).

## Conclusion

Contemporary oceans contain a highly heterogeneous pool of phytoplankton-derived DOM that varies substantially in quality and quantity over space and time (1). Variability in the type and timing of primary production is postulated to represent an important bottom-up driver of genotypic and phenotypic diversity in coexisting heterotrophic microorganisms (33), especially during phytoplankton blooms where substrate-controlled succession of taxa may hold integral control over the fate of fixed carbon (26). The results of our study support and extend our understanding of microbe–DOM interactions by demonstrating that separate pools of DOM from diatoms and cyanobacteria collected during simulated bloom phases of growth (exudates) and senescence (lysates) inordinately support the activities of taxonomically distinct subsets of the heterotrophic microbial community. Our approach using proteomic SIP allowed for the identification of newly synthesized proteins that represent rapid (<15 h) responses to DOM addition, revealing which taxonomic groups and metabolic functions were most active in assimilating the four distinct DOM pools before any observable shifts were recorded by unlabeled metaproteomic or metagenomic analyses. The strong signal of partitioned resource assimilation between phylogenetically distinct lineages allowed us to characterize the role of different heterotrophs in turning over each of the simulated bloom DOM pools. In a broad view, populations of Bacteroidetes are adapted to using relatively carbon-rich, high-molecular-weight DOM



compounds that are more common in the lysates from diatom cells, while Proteobacteria lineages show more variability in their resource specialization, with numerically abundant Alphaproteobacteria, such as SAR11 and Roseobacter-clade populations, assimilating a vast majority of bioavailable cyanobacteria exudate DOM but with different apparent carbon use efficiencies.

Based on these data we estimate the lineage-specific rates of carbon assimilation of these four DOM pools and conclude that these distributed resource preferences are crucial for efficient turnover (i.e., assimilation or respiration) of DOM during a bloom (26, 48). Therefore, changes to the origin and composition of DOM inputs under future ocean conditions, such as increased CO<sub>2</sub> concentrations and higher temperatures, may have predictable consequences for global carbon cycling and nutrient flux rates through the microbial loop (69, 70). For example, a Coupled Model Intercomparison Project (Phase 5) model suggests future surface oceans will be more stratified (71), indicating a potential for cyanobacterial blooms to become more prevalent due to their tolerance of higher temperatures and more oligotrophic conditions, while diatom blooms decrease due to reliance on cooler, more turbulent, and nitrate-rich waters. When considering this scenario in light of our results, the surface ocean could be expected to experience increased Alphaproteobacteria activity due to increased cyanobacterial exudation and a correspondingly lower contribution of Bacteroidetes activity due to the reduction of new diatom-derived DOM. Though our data does not provide direct insights into carbon use efficiencies of the taxa observed here, such a shift toward an Alphaproteobacteria-dominated community could depress carbon sequestration via the microbial carbon pump (72) by favoring organisms, such as *Pelagibacteria* spp., that have more tightly coupled and lower-efficiency carbon metabolisms (64).

Overall, the different metabolic strategies and adaptations used by coexisting surface ocean heterotrophic populations that consume phytoplankton DOM dictate carbon use efficiency at the community level. This distributed community metabolism not only allows taxa to coexist by decreasing direct resource competition but also indicates that trophic transfer and remineralization rates in the microbial loop are strongly coupled to DOM properties. Experiments that provide complex, naturally occurring substrates to resident microbial communities and track the fate of carbon and nutrients, as well as the activity and ecophysiological characteristics of microbial populations, will be crucial to the ongoing effort to mechanistically connect marine microbial community dynamics to global biogeochemical cycles.

## Materials and Methods

**Phytoplankton Growth Conditions and <sup>13</sup>C-Labeled Substrates.** Two phytoplankton species (*T. pseudonana* and *Synechococcus* sp. WH8102) were grown axenically in the laboratory under enriched Na<sub>2</sub><sup>13</sup>CO<sub>3</sub> conditions in artificial seawater F/2 media without unlabeled inorganic carbon. The F/2 media was sparged overnight with N<sub>2</sub>, autoclave sterilized, and amended with algal vitamins (40 mg thiamine-HCl, 20 mg biotin, and 20 mg cobalamin per 250 mL). A total of 1 mL stock culture of each phytoplankton was transferred to a 250-mL culture flask of medium with either Na<sub>2</sub><sup>13</sup>CO<sub>3</sub> or unlabeled Na<sub>2</sub>CO<sub>3</sub> (used to monitor growth) added to the F/2 medium via stirring to 3-mM final concentration. Diatom and cyanobacterial cells were grown at 22 °C with the bottle sealed from the atmosphere under continuous light until late exponential phase (optical density at 600 nm of 0.2 and 0.1 after 150 and 200 h, respectively), whereupon the <sup>13</sup>C-labeled cultures were centrifuged at 5,000 × g for 15 min. Supernatant was collected and sparged by bubbling with air, filter sterilized at 0.1 μm, and frozen at -80 °C (exudate). Cell pellets were rinsed and resuspended in sterile distilled and deionized water, freeze-thawed three times, sonicated, then centrifuged at 5,000 × g for 15 min. Supernatant was collected and frozen at -80 °C (lysate). Phytoplankton cells from the labeled cultures were confirmed to be >80% <sup>13</sup>C using Lawrence Livermore National Lab's nanoscale secondary ion mass spectrometry (nanoSIMS). In order to amend equal carbon mass to all treatments, DOM carbon content was measured by adding 250 to 750 μL exudate or lysate to 25-mm glass fiber filter membranes (baked for 4 h at

450 °C), allowing to dry overnight, then adding 250 μL 10% HCl to each filter and again drying overnight. Nutrient and elemental (CHN) analysis of DOM on filters was conducted by the Marine Science Institute (University of California, Santa Barbara). No measures of molecular weight or stoichiometric composition could be made on DOM, as the total exudate and lysate <sup>13</sup>C-biomass was required to amend to mesocosms.

**Sample Collection and Substrate Incubations.** Mesocosm water samples were collected on November 22, 2014 from the intake of the Hatfield Marine Science Center seawater distribution system at the mouth of Yaquina Bay (44.62°N, -124.04°W). A total of ~200 L (11 L/sample, 18 samples) surface seawater was prefiltered through 3-μm polyethersulfone (PES) membranes (Pall Corporation) and collected in separate 12-L acid-washed carboys. Samples were transported back to Oregon State University and placed in a dark 20 °C incubation room. Three samples were immediately filtered onto 0.2-μm PES membranes to collect microbial biomass for characterization of initial community DNA and protein (T<sub>0</sub>). The four substrate pools were amended to water samples in triplicate after being diluted in with 0.2-μm-filtered seawater from the sampling site to uniform concentrations based on C and N analysis (~40 μM C final concentration). The last set of triplicate carboys (T<sub>15</sub>) was amended with 800 mL 0.2-μm-filtered seawater from the sampling site as a no-substrate control. All 15 samples were incubated in the absence of light at 20 °C for ~15 h prior to biomass collection and processing.

**DNA and Protein Extraction.** Total community DNA was extracted from cells from 1 L water from each microcosm by filtering onto 0.2-μm PES filter membrane and performing the hexadecyltrimethylammonium bromide phenol-chloroform protocol (73). Total community protein was extracted from cells in 10 L as reported previously using sodium dodecyl sulfate cell lysis and protein precipitation with trichloroacetic acid (17). DNA and protein samples were stored at -80 °C until preparation for sequencing and mass spectrometry.

**Metagenome Sequencing, Assembly, and Annotation.** Metagenome libraries were prepared from whole-community DNA using the Illumina Nextera XT kit (74). DNA libraries were normalized by concentration and pooled based on quantification with Bioanalyzer HS-DNA Chips. Sequencing was performed on Illumina HiSeq 3000, generating 151-base pair, paired-end reads with >3 million reads per sample. All reads were trimmed and quality filtered using Sickle (75); bases with a phred score < 25 (across 15-base pair sliding windows) were trimmed, and only reads >100 bp were retained. This process yielded 4.8 to 7.5 Mbp per replicate. The software program IDBA-UD (76) was used for de novo metagenome assembly of combined replicate libraries to produce one assembly for each of the four substrate treatments and two controls.

Prokaryotic protein coding sequences (CDS) were predicted from assembled contigs using Prodigal (77). CDS were taxonomically annotated based on best hit to the RefSeq protein database (January 5, 2019) using the "diamond" program (78); annotation rates varied from >95% for domain level to 55% for genus level. CDS were functionally annotated to clusters of orthologous genes (COG) based on best rpsBLAST hits to the COG database (79); the annotation rate of protein sequences to COGs was 75.2%.

CDS read depth was measured by pairwise alignment of shotgun reads to the assembly of all samples using the "Bowtie2" program (80). CDS relative abundance across samples was then calculated by dividing read depth by a per-sample genome equivalence metric determined by the "MicrobeCensus" program (81). Relative abundance was reported as reads per kilobase of genome equivalents.

MaxBin2 was used to create population bins of contigs (82) that were manually curated and evaluated for overall quality using "CheckM" (83). Relative growth rates under different experimental conditions (reported as mean binned contig peak-to-trough ratio) were estimated for bins in each sample with >50% completion and <10% contamination using the "DEMIC" program (29). The main advantage of using metagenome-derived growth rate estimates in this study is that they can encompass a range of populations in an untargeted manner and can be performed with minimal disturbance to in situ processes.

**Peptide Mass Spectrometry, Identification, and Search Database Construction.** Extracted proteins were denatured with 6 M guanidine, reduced with 10 mM dithiothreitol, and digested with trypsin at room temperature at the enzyme:substrate ratio of 1:100 (w:w). A total of 25 μg peptide product per sample was loaded offline into a 150-μm inner diameter two-dimensional (2D) back column (Polymicro Technologies) packed with 3 cm C18 reverse-phase (RP)

resin (Luna; Phenomenex) and 3 cm strong cation exchange (SCX) resin (Luna; Phenomenex). Samples were analyzed using 2D liquid chromatography–tandem mass spectrometry (2D-LC-MS/MS) on an LTQ Orbitrap Elite mass spectrometer (Thermo Scientific) with previously described conditions (17, 18). Briefly, each mass spectrometric run was configured with 11 SCX fractionations using 5, 7, 10, 12, 15, 17, 20, 25, 35, 50, and 100% of solvent D (500 mM ammonium acetate dissolved in solvent A). Each SCX fraction was separated by a 110-min RP gradient from 100% solvent A to 50% solvent B. The LC eluent was directly nanosprayed (Proxeon) into an LTQ Orbitrap Elite mass spectrometer (Thermo Scientific). Both MS scans and collision-induced dissociation (CID) MS/MS scans were acquired in Orbitrap with a resolution of 30,000 and 15,000, respectively. After each MS scan, the eight most abundant precursor ions were selected under automated data-dependent acquisition for MS/MS analysis by CID.

Sipros version 1.3 (84) was used to search MS/MS spectra of each sample against a matched protein database constructed from its metagenome (85). The database searches used the following parameters: parent mass offsets of  $-3, -2, -1, 0, 1, 2, 3$  Da; mass tolerance of 0.05 and 0.02 Da for parent ions and fragment ions, respectively; up to two missed cleavages; and full enzyme specificity required. The database searching results were filtered to achieve 1% false discovery rate (FDR) at the peptide level. Proteins were inferred from the identified peptides using the parsimony rule (86). Indistinguishable proteins were combined as a protein group and a minimum of two peptides, at least one of which must be unique, were required for each inferred protein or protein group. An initial search was performed in the unlabeled mode against all metagenome CDS clustered at 100% identity using CD-hit (87) to identify the subset of proteins expressed by the community and create a smaller database for  $^{13}\text{C}$ -labeled spectra searches. Out of all metagenome CDS clusters, 11.8% (291,725) returned a positive peptide ID during the unlabeled search and were used to build a reduced database for Sipros version 1.3-labeled search mode, which identified both unlabeled ( $<2\%$   $^{13}\text{C}$ -enriched) and labeled ( $\geq 2\%$   $^{13}\text{C}$ -enriched) peptides from spectral matches. Data from the labeled search mode is presented, and the total numbers of spectra, peptides, and proteins identified are shown in *SI Appendix, Table S7*. Using the labeled search mode, two metrics were used to quantify  $^{13}\text{C}$  labeling of identified proteins: label frequency (LF) and average enrichment (ENRV). These metrics were calculated as previously described (17). Label frequency of a protein (or proteins aggregated by taxonomic level or treatment) is the percentage of its total peptides with a  $^{13}\text{C}$  content  $\geq 2\%$  (i.e., peptides enriched in isotope above the background of 1.1%). Average enrichment of a protein (or proteins aggregated by taxonomic level or treatment) is the percent of its total carbon atoms that are  $^{13}\text{C}$ .

Reverse sequences of CDS in the reduced database were added as decoys to calculate FDR of labeled searches. The FDRs of the labeled searches were independently estimated using the unamended control samples, which should have no enriched peptides. Only an average of 0.20% of peptide-spectrum matches were falsely identified to be enriched in the unamended control samples. This validated the identification confidence of the labeled searches. All CDS in this database were digested in silico, and those that were 7 to 60 residues long with a maximum of two missed tryptic cuts were considered in the final search database (17, 18). The conservative two-peptide threshold (i.e., one unique and one shared peptide or two unique peptides) was used to validate and filter protein hits after implementing a 1% peptide-level FDR. Normalized balanced spectral counts (NBSC) of detected peptides annotated to each identified protein were summed and used to quantify the relative abundance of a given protein in each sample.

**Statistical Analysis and Carbon Assimilation Calculations.** Principal component analysis (PCoA) was used to determine relationships between taxonomic

composition of proteins across samples and to calculate variance captured during dimension reduction. PCoAs were performed in R (version 3.4.0) using the function “*rca()*” from the package “*vegan*” (88), and resulting components were plotted with the package “*ggplot2*” (89).

To determine significant  $^{13}\text{C}$  enrichment in a given population, average enrichments for all genus-level populations were z-score normalized by sample to account for differences in the number of recovered proteins from each treatment. Compared to a null distribution created from 1,000 random subsamples of labeled spectra from all metaproteome peptides, a population had significantly high average enrichment if it was greater than or equal to 95% of average enrichment values from the null distribution (17). These tests were performed with custom python scripts using the “*scipy stats*” and “*skbio*” libraries. Heat maps of the statistics were made in R using the packages “*ggplot2*” and “*vegan*.”

The average DNA sequence G+C% of all search database CDS with protein hits, along with the carbon to nitrogen ratio (C:N) of protein amino acid sequences encoded by these CDS, were calculated using custom python scripts using the “*skbio*” library. Scatterplots of DNA G+C% and protein amino acid C:N per taxon were produced using the R package “*ggplot2*.”

Per-taxon carbon assimilation rates ( $\text{ng } ^{13}\text{C} \text{ assimilated } \text{L}^{-1} \cdot \text{d}^{-1}$ ) for each major taxon in each treatment were calculated with experimentally derived values using the following equation:

$$\frac{(N \times R) \times \text{ENRV} \times \text{LF} \times 10^{-5} \text{ ng}}{15 \text{ h}} \times \frac{1,000 \text{ mL}}{1 \text{ L}} \times \frac{24 \text{ h}}{1 \text{ d}}$$

where  $N$  is cell counts per treatment ( $\text{cells} \cdot \text{mL}^{-1}$ ),  $R$  is taxon-specific relative protein abundance per treatment (% NBSC), ENRV is taxon-specific average  $^{13}\text{C}$  enrichment value per treatment (%), LF is taxon-specific label frequency per treatment (%),  $10^{-5} \text{ ng}$  is a per-cell carbon mass factor [ $\text{ng C} \cdot \text{cell}^{-1}$  (90)], 15 h is the incubation period of our experiments, and the multiplication factors of 1,000 mL per 1 L and 24 h per day to allow for results to be reported in  $\text{ng } ^{13}\text{C} \text{ assimilated } \text{L}^{-1} \cdot \text{d}^{-1}$  (Table 1).

**Data Availability.** All R code and data tables for producing figures are available in GitHub at [https://github.com/kielt1bp-sys/Phytoplankton\\_DOM/Manuscript\\_code](https://github.com/kielt1bp-sys/Phytoplankton_DOM/Manuscript_code). Metagenome-assembled contig sequences are available from the National Center for Biotechnology Information (NCBI) database under BioSample [SAMN04917373](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA320136), BioProject [PRJNA320136](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA320136), and Short Read Archive [SRS1422236](https://www.ncbi.nlm.nih.gov/sra/SRS1422236), and unassembled paired-end metagenome reads are under archive [SRX1738728–SRX1738732](https://www.ncbi.nlm.nih.gov/sra/SRX1738728-SRX1738732). Proteomics data were deposited in the ProteomeXchange Consortium via the Proteomics Identification Database (PRIDE) partner repository with the dataset identifier [PXD008093](https://www.ebi.ac.uk/pride/archive/study/PXD008093).

**ACKNOWLEDGMENTS.** We would like to thank Dr. Kimberly Halsey for providing phytoplankton strains and assistance with media recipes and Dr. Steven Giovannoni for use of the Guava Flow Cytometer for cell counts and for personal communication during the revision process. We also thank Oregon State University’s Center for Genome Research and Biocomputing for biomolecule data generation and computing resources and the Hatfield Marine Science Center for resource access and laboratory use. This work was funded by the Gordon and Betty Moore Foundation Marine Microbiology Initiative (Grant No. GBMF3302). Part of this work was performed under the auspices of the US Department of Energy by Lawrence Livermore National Laboratory under Contract No. DE-AC52-07NA27344. Oak Ridge National Laboratory resources, including the Oak Ridge Leadership Computing Facility, were used in the research and are supported by the Office of Science of the US Department of Energy under Contract No. DE-AC05-00OR22725.

1. J. Uitz, H. Claustre, B. Gentili, D. Stramski, Phytoplankton class-specific primary production in the world’s oceans: Seasonal and interannual variability from satellite observations. *Global Biogeochem. Cycles* **24**, GB3016 (2010).
2. C. S. Rousseaux, W. W. Gregg, Interannual variation in phytoplankton primary production at a global scale. *Remote Sens.* **6**, 1–19 (2014).
3. J. Carstensen, R. Klais, J. E. Cloern, Phytoplankton blooms in estuarine and coastal waters: Seasonal patterns and key species. *Estuar. Coast. Shelf Sci.* **162**, 98–109 (2015).
4. H. Demarcq, G. Reygondeau, S. Alvain, V. Vantrepotte, Monitoring marine phytoplankton seasonality from space. *Remote Sens. Environ.* **117**, 211–222 (2012).
5. L. Aluwihare, D. Repeta, A comparison of the chemical characteristics of oceanic DOM and extracellular DOM produced by marine algae. *Mar. Ecol. Prog. Ser.* **186**, 105–117 (1999).
6. D. C. O. Thornton, Dissolved organic matter (DOM) release by phytoplankton in the contemporary and future ocean. *Eur. J. Phycol.* **49**, 20–46 (2014).
7. J. Grosse, A. Burson, M. Stomp, J. Huisman, H. T. S. Boschker, From ecological stoichiometry to biochemical composition: Variation in N and P supply alters key biosynthetic rates in marine phytoplankton. *Front. Microbiol.* **8**, 1299 (2017).

8. C. A. Vargas *et al.*, The relative importance of microbial and classical food webs in a highly productive coastal upwelling area. *Limnol. Oceanogr.* **52**, 1495–1510 (2007).
9. J. Cole, S. Findlay, M. Pace, Bacterial production in fresh and saltwater ecosystems: A cross-system overview. *Mar. Ecol. Prog. Ser.* **43**, 1–10 (1988).
10. G. J. Herndl, M. G. Weinbauer, “Marine microbial food web structure and function” in *Marine Science Frontiers for Europe*, G. Wefel, F. Lamy, F. Mantoura, Eds. (Springer Berlin Heidelberg, 2003), pp. 265–277.
11. M. Landa, A. S. Burns, S. J. Roth, M. A. Moran, Bacterial transcriptome remodeling during sequential co-culture with a marine dinoflagellate and diatom. *ISME J.* **11**, 2677–2690 (2017).
12. M. T. Cottrell, D. L. Kirchman, Natural assemblages of marine proteobacteria and members of the Cytophaga-Flavobacter cluster consuming low- and high-molecular-weight dissolved organic matter. *Appl. Environ. Microbiol.* **66**, 1692–1697 (2000).
13. G. Webster *et al.*, A comparison of stable-isotope probing of DNA and phospholipid fatty acids to study prokaryotic functional diversity in sulfate-reducing marine sediment enrichment slurries. *Environ. Microbiol.* **8**, 1575–1589 (2006).

14. E. Sintez, G. J. Herndl, Quantifying substrate uptake by individual cells of marine bacterioplankton by catalyzed reporter deposition fluorescence in situ hybridization combined with microautoradiography. *Appl. Environ. Microbiol.* **72**, 7022–7028 (2006).
15. S. DeLorenzo et al., Ubiquitous dissolved inorganic carbon assimilation by marine bacteria in the Pacific Northwest coastal ocean as determined by stable isotope probing. *PLoS One* **7**, e46695 (2012).
16. X. Mayali, P. K. Weber, S. Mabery, J. Pett-Ridge, Phylogenetic patterns in the microbial response to resource availability: Amino acid incorporation in San Francisco Bay. *PLoS One* **9**, e95842 (2014).
17. S. Bryson et al., Phylogenetically conserved resource partitioning in the coastal microbial loop. *ISME J.* **11**, 2781–2792 (2017).
18. S. Bryson et al., Proteomic stable isotope probing reveals taxonomically distinct patterns in amino acid assimilation by coastal marine bacterioplankton. *mSystems* **1**, e00027-15 (2016).
19. C. Pan et al., Quantitative tracking of isotope flows in proteomes of microbial communities. *Mol. Cell. Proteom.* **10**, M1110.006049 (2011).
20. B. Norrman, U. L. Zweifel, C. S. Hopkinson, F. Brian, Production and utilization of dissolved organic carbon during an experimental diatom bloom. *Limnol. Oceanogr.* **40**, 898–907 (1995).
21. B. Biddanda, R. Benner, Carbon, nitrogen, and carbohydrate fluxes during the production of particulate and dissolved organic matter by marine phytoplankton. *Limnol. Oceanogr.* **42**, 506–518 (1997).
22. C. Arnosti, G. Reintjes, R. Amann, A mechanistic microbial underpinning for the size-reactivity continuum of dissolved organic carbon degradation. *Mar. Chem.* **206**, 93–99 (2018).
23. J. Beardall, S. Stojkovic, S. Larsen, Living in a high CO<sub>2</sub> world: Impacts of global climate change on marine phytoplankton. *Plant Ecol. Divers.* **2**, 191–205 (2009).
24. Z. V. Finkel et al., Phytoplankton in a changing world: Cell size and elemental stoichiometry. *J. Plankton Res.* **32**, 119–137 (2010).
25. M. H. Schlüter, A. Kraberg, K. H. Wiltshire, Long-term changes in the seasonality of selected diatoms related to grazers and environmental conditions. *J. Sea Res.* **67**, 91–97 (2012).
26. H. Teeling et al., Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton bloom. *Science* **336**, 608–611 (2012).
27. K. Tang, N. Jiao, K. Liu, Y. Zhang, S. Li, Distribution and functions of TonB-dependent transporters in marine bacteria and environments: Implications for dissolved organic matter utilization. *PLoS One* **7**, e41204 (2012).
28. R. S. Poretsky, S. Sun, X. Mou, M. A. Moran, Transporter genes expressed by coastal bacterioplankton in response to dissolved organic carbon. *Environ. Microbiol.* **12**, 616–627 (2010).
29. Y. Gao, H. Li, Quantifying and comparing bacterial growth dynamics in multiple metagenomic samples. *Nat. Methods* **15**, 1041–1044 (2018).
30. A. Barofsky, C. Vidoudez, G. Pohnert, Metabolic profiling reveals growth stage variability in diatom exudates: Exometabolome of diatoms. *Limnol. Oceanogr. Methods* **7**, 382–390 (2009).
31. R. Baran et al., Functional genomics of novel secondary metabolites from diverse cyanobacteria using untargeted metabolomics. *Mar. Drugs* **11**, 3617–3631 (2013).
32. F. Azam, D. C. Smith, G. F. Steward, A. Hagström, Bacteria-organic matter coupling and its significance for oceanic carbon cycling. *Microb. Ecol.* **28**, 167–179 (1994).
33. M. F. Polz, D. E. Hunt, S. P. Preheim, D. M. Weinreich, Patterns and mechanisms of genetic and phenotypic differentiation in marine microbes. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **361**, 2009–2021 (2006).
34. L. Gómez-Cansarnau, M. V. Lindh, J. M. Gasol, J. Pinhassi, Structuring of bacterioplankton communities by specific dissolved organic carbon compounds. *Environ. Microbiol.* **14**, 2361–2378 (2012).
35. H. Sarmento et al., Phytoplankton species-specific release of dissolved free amino acids and their selective consumption by bacteria. *Limnol. Oceanogr.* **58**, 1123–1135 (2013).
36. H. Sarmento, J. M. Gasol, Use of phytoplankton-derived dissolved organic carbon by different types of bacterioplankton. *Environ. Microbiol.* **14**, 2348–2360 (2012).
37. S. Muthusamy et al., Comparative proteomics reveals signature metabolisms of exponentially growing and stationary phase marine bacteria. *Environ. Microbiol.* **19**, 2301–2319 (2017).
38. M. T. Duret, R. S. Lampitt, P. Lam, Prokaryotic niche partitioning between suspended and sinking marine particles. *Environ. Microbiol. Rep.* **11**, 386–400 (2018).
39. P. A. del Giorgio, J. M. Gasol, “Physiological structure and single-cell activity in marine bacterioplankton” in *Microbial Ecology of the Oceans*, D. L. Kirchman, Ed. (John Wiley & Sons, Ltd, 2008), pp. 243–298.
40. M. P. Mikan et al., Metaproteomics reveal that rapid perturbations in organic matter prioritize functional restructuring over taxonomy in western Arctic Ocean microbiomes. *ISME J.* **14**, 39–52 (2020).
41. S. Burkhardt, I. Zondervan, U. Riebesell, Effect of CO<sub>2</sub> concentration on C:N:P ratio in marine phytoplankton: A species comparison. *Limnol. Oceanogr.* **44**, 683–690 (1999).
42. D. Stramski, A. Sciandra, H. Claustre, Effects of temperature, nitrogen, and light limitation on the optical properties of the marine diatom *Thalassiosira pseudonana*. *Limnol. Oceanogr.* **47**, 392–403 (2002).
43. S. Bertilsson, O. Berglund, D. M. Karl, S. W. Chisholm, Elemental composition of marine *Prochlorococcus* and *Synechococcus* cells: Implications for the ecological stoichiometry of the sea. *Limnol. Oceanogr.* **48**, 1721–1731 (2003).
44. C. Six, J. Thomas, B. Brahmasha, Y. Lemoine, F. Partensky, Photophysiology of the marine cyanobacterium *Synechococcus* sp. WH8102, a new model organism. *Aquat. Microb. Ecol.* **35**, 17–29 (2004).
45. R. Levy, E. Borenstein, Metabolic modeling of species interaction in the human microbiome elucidates community-level assembly rules. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 12804–12809 (2013).
46. D. T. Truong, A. Tett, E. Pasolli, C. Huttenhower, N. Segata, Microbial strain-level population structure and genetic diversity from metagenomes. *Genome Res.* **27**, 626–638 (2017).
47. Y. Yan, L. H. Nguyen, E. A. Franzosa, C. Huttenhower, Strain-level epidemiology of microbial communities and the human microbiome. *Genome Med.* **12**, 71 (2020).
48. J. McCarren et al., Microbial community transcriptomes reveal microbes and metabolic pathways associated with dissolved organic matter turnover in the sea. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 16420–16427 (2010).
49. R. Urbani, E. Magaletti, P. Sist, A. M. Cicero, Extracellular carbohydrates released by the marine diatoms *Cylindrotheca closterium*, *Thalassiosira pseudonana* and *Skelletonema costatum*: Effect of P-depletion and growth status. *Sci. Total Environ.* **353**, 300–306 (2005).
50. M. Landa et al., Phylogenetic and structural response of heterotrophic bacteria to dissolved organic matter of different chemical composition in a continuous culture study. *Environ. Microbiol.* **16**, 1668–1681 (2014).
51. K. Longnecker, M. C. Kido Soule, E. B. Kujawinski, Dissolved organic matter produced by *Thalassiosira pseudonana*. *Mar. Chem.* **168**, 114–123 (2015).
52. D. N. Bolam, N. M. Koropatkin, Glycan recognition by the Bacteroidetes Sus-like systems. *Curr. Opin. Struct. Biol.* **22**, 563–569 (2012).
53. G. Reintjes, C. Arnosti, B. M. Fuchs, R. Amann, An alternative polysaccharide uptake mechanism of marine bacteria. *ISME J.* **11**, 1640–1650 (2017).
54. G. Reintjes, C. Arnosti, B. Fuchs, R. Amann, Selfish, sharing and scavenging bacteria in the Atlantic Ocean: A biogeographical study of bacterial substrate utilisation. *ISME J.* **13**, 1119–1132 (2018).
55. H. Elifantz, R. R. Malmstrom, M. T. Cottrell, D. L. Kirchman, Assimilation of polysaccharides and glucose by major bacterial groups in the Delaware estuary. *AEM* **71**, 7799–7805 (2005).
56. Y. Tada et al., Differing growth responses of major phylogenetic groups of marine bacteria to natural phytoplankton blooms in the western North Pacific Ocean. *Appl. Environ. Microbiol.* **77**, 4055–4065 (2011).
57. A. Klindworth et al., Diversity and activity of marine bacterioplankton during a diatom bloom in the North Sea assessed by total RNA and pyrotag sequencing. *Mar. Genomics* **18**, 185–192 (2014).
58. G. Reintjes et al., Short-term changes in polysaccharide utilization mechanisms of marine bacterioplankton during a spring phytoplankton bloom. *bioRxiv* [Preprint] (2020). <https://doi.org/2020.01.08.899153> (Accessed 9 January 2020).
59. R. Baran et al., Metabolite identification in *Synechococcus* sp. PCC 7002 using untargeted stable isotope assisted metabolite profiling. *Anal. Chem.* **82**, 9034–9042 (2010).
60. C. L. Fiore, K. Longnecker, M. C. Kido Soule, E. B. Kujawinski, Release of ecologically relevant metabolites by the cyanobacterium *Synechococcus elongatus* CCMP 1631. *Environ. Microbiol.* **17**, 3949–3963 (2015).
61. N. J. West, I. Obernosterer, O. Zemb, P. Lebaron, Major differences of bacterial diversity and activity inside and outside of a natural iron-fertilized phytoplankton bloom in the Southern Ocean. *Environ. Microbiol.* **10**, 738–756 (2008).
62. H.-A. Giebel et al., Distribution of Roseobacter RCA and SAR11 lineages in the North Sea and characteristics of an abundant RCA isolate. *ISME J.* **5**, 8–19 (2011).
63. F. Guillemette, S. Leigh McCallister, P. A. Del Giorgio, Selective consumption and metabolic allocation of terrestrial and algal carbon determine allochthony in lake bacteria. *ISME J.* **10**, 1373–1382 (2016).
64. S. J. Giovannoni, J. Cameron Thrash, B. Temperton, Implications of streamlining theory for microbial ecology. *ISME J.* **8**, 1553–1565 (2014).
65. K. H. Halsey et al., Biological cycling of volatile organic carbon by phytoplankton and bacterioplankton. *Limnol. Oceanogr.* **62**, 2650–2661 (2017).
66. J. Sun et al., The abundant marine bacterium *Pelagibacter* simultaneously catabolizes dimethylsulfoniopropionate to the gases dimethyl sulfide and methanethiol. *Nat. Microbiol.* **1**, 1–5 (2016).
67. D. L. Kirchman, Growth rates of microbes in the oceans. *Annu. Rev. Mar. Sci.* **8**, 285–309 (2016).
68. S. Langenheder, E. S. Lindström, Factors influencing aquatic and terrestrial bacterial community assembly. *Environ. Microbiol. Rep.* **11**, 306–315 (2019).
69. T. Trombetta, F. Vidussi, C. Roques, M. Scotti, B. Mostajir, Marine microbial food web networks during phytoplankton bloom and non-bloom periods: Warming favors smaller organism interactions and intensifies trophic cascade. *Front. Microbiol.* **11**, 502336 (2020).
70. C. J. Gobler, Climate change and harmful algal blooms: Insights and perspective. *Harmful Algae* **91**, 101731 (2020).
71. W. Fu, J. T. Randerson, J. K. Moore, Climate change impacts on net primary production (NPP) and export production (EP) regulated by increasing stratification and phytoplankton community structure in the CMIP5 models. *Biogeosciences* **13**, 5151–5170 (2016).
72. N. Jiao et al., The microbial carbon pump and the oceanic recalcitrant dissolved organic matter pool. *Nat. Rev. Microbiol.* **9**, 555 (2011).
73. B. Kieft et al., Microbial community structure-function relationships in Yaquina Bay Estuary reveal spatially distinct carbon and nitrogen cycling capacities. *Front. Microbiol.* **9**, 1282 (2018).
74. B. Kieft et al., Yaquina Bay Estuary communities. National Center for Biotechnology Information (NCBI). <https://www.ncbi.nlm.nih.gov/biosample/SAMN04917373>. Deposited 29 April 2016.
75. N. Joshi, J. Fass, Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files. Version 1.33. <https://github.com/najoshi/sickle>. Accessed 1 February 2018.

76. Y. Peng *et al.*, IDBA-tran: A more robust de novo de Bruijn graph assembler for transcriptomes with uneven expression levels. *Bioinformatics* **29**, i326–i334 (2013).
77. D. Hyatt, P. F. LoCasio, L. J. Hauser, E. C. Uberbacher, Gene and translation initiation site prediction in metagenomic sequences. *Bioinformatics* **28**, 2223–2230 (2012).
78. B. Buchfink, C. Xie, D. H. Huson, Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59–60 (2015).
79. J. Huerta-Cepas *et al.*, eggNOG 4.5: A hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res.* **44**, D286–D293 (2016).
80. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
81. S. Nayfach, K. S. Pollard, Average genome size estimation improves comparative metagenomics and sheds light on the functional ecology of the human microbiome. *Genome Biol.* **16**, 51 (2015).
82. Y. W. Wu, B. A. Simmons, S. W. Singer, MaxBin 2.0: An automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* **32**, 605–607 (2016).
83. D. H. Parks, M. Imelfort, C. T. Skennerton, P. Hugenholtz, G. W. Tyson, CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043–1055 (2015).
84. X. Guo *et al.*, Sipros Ensemble improves database searching and filtering for complex metaproteomics. *Bioinformatics* **34**, 795–802 (2018).
85. B. Kieft *et al.*, Proteomics for Newport Estuary samples. Proteomics Identification Database (PRIDE). <http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PXD008093>. Deposited 4 July 2018.
86. A. I. Nesvizhskii, R. Aebersold, Interpretation of shotgun proteomic data: The protein inference problem. *Mol. Cell. Proteomics* **4**, 1419–1440 (2005).
87. W. Li, A. Godzik, Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659 (2006).
88. J. Oksanen *et al.*, vegan: Community ecology package. R package version 2.4-3. <https://CRAN.R-project.org/package=vegan>. Accessed 21 September 2021.
89. H. Wickham, *ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag, 2009).
90. R. Fukuda, H. Ogawa, T. Nagata, I. Koike, Direct determination of carbon and nitrogen contents of natural bacterial assemblages in marine environments. *Appl. Environ. Microbiol.* **64**, 3352–3358 (1998).