Predictable functional biogeography of marine microbial heterotrophs

Emily J. Zakem[1*], Jesse McNichol[2†], J.L. Weissman[3,4†], Yubin Raut[4], Liang Xu[1], Elisa R. Halewood[5], Craig A. Carlson[5], Stephanie Dutkiewicz[6,7], Jed A. Fuhrman[4], Naomi M. Levine[4]

[1]Department of Global Ecology, Carnegie Institution for Science, Stanford, CA, USA
[2]Department of Biology, St. Francis Xavier University, Antigonish, Nova Scotia, Canada
[3]Department of Biology, The City College of New York, New York, NY, USA
[4]Department of Biological Sciences, University of Southern California, Los Angeles, CA, USA
[5]Department of Ecology, Evolution, and Marine Biology, Marine Science Institute, University of California, Santa Barbara, CA, USA
[6]Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA
[7]Center for Global Change Science, Massachusetts Institute of Technology, Cambridge, MA, USA

[*]Correspondence: ezakem@carnegiescience.edu
[†]These authors contributed equally to this work.

Abstract

Microbial heterotrophs ('picoheterotrophs') drive global carbon cycling, but how to quantitatively organize their functional complexity remains unclear. Here, we generate a global-scale, mechanistic understanding of marine picoheterotrophic functional biogeography with a novel model-data synthesis. We build picoheterotrophic diversity into a trait-based marine ecosystem model along two axes: substrate lability and optimization for growth rate (copiotrophy) vs. substrate affinity (oligotrophy). Using genetic sequences along an Alaska-to-Antarctica Pacific Ocean transect, we compile 21 picoheterotrophic guilds and estimate their degree of copiotrophy. Data and model agreement suggests that gradients in predation and substrate lability predominantly set biogeographical patterns, and identifies 'slow copiotrophs' subsisting at depth. Results demonstrate the predictability of the marine microbiome and connect ecological dynamics with carbon storage, crucial for projecting changes in a warming ocean.

Introduction

The global carbon cycle plays a central role in regulating the Earth's climate by setting the amount of carbon dioxide (CO_2) in the atmosphere. Photosynthesis removes CO_2 from the atmosphere and converts it to organic carbon, which rapidly cycles through the biosphere. Ultimately, most organic carbon is consumed by heterotrophic organisms and respired back into CO_2. In the ocean, both photosynthesis and respiration are driven predominantly by microorganisms[1], and their decoupling in time and space sequesters carbon in the deep ocean[2].

Accurately predicting the Earth's climate requires understanding how microorganisms mediate ocean carbon storage. However, in ocean biogeochemical models used for climate projections, the activities of heterotrophic Bacteria and Archaea ('picoheterotrophs') that respire organic carbon are typically represented implicitly and simplistically[3]. This prevents climate models from capturing microbial feedbacks[4], and may explain in part why models cannot agree on the sign of the change in the organic carbon exported to the deep ocean in a warming climate[5].

Before we can build a dynamic picoheterotrophic response into climate models, we must develop a predictive understanding of how picoheterotrophic communities are currently structured and how this structure relates to their biogeochemical function. This is a challenging task because picoheterotrophic communities are immensely complex, consisting of thousands of mostly uncultivated populations consuming thousands of organic substrates, many of which are uncharacterized[6–9]. What are the broad classifications appropriate for representing picoheterotrophic diversity quantitatively at large scales? Rapidly expanding global-scale sequencing efforts have the potential to address some of these unknowns and to provide the data needed to inform and analyze ecosystem models[10–12]. However, how to fully harness these datasets and connect them with modeling frameworks remains unclear.

A tractable approach for untangling picoheterotrophic community complexity is to synthesize established organizing principles. First, picoheterotrophs have long been classified as "copiotrophs",

with cellular machinery optimized for rapid growth when substrate is abundant, or "oligotrophs", optimized for competitive uptake of limited substrates (i.e., high affinity)[13–16]. This tradeoff between growth rate and substrate affinity reflects a cell's finite proteome, and thus the amount of cellular machinery that can be allocated to a given function[17]. This classification has been used to explain biogeographical patterns in the ocean surface: oligotrophs dominate the microbial community in subtropical 'ocean deserts' while copiotrophs become more prevalent in high-latitude, nutrient-rich waters[18–20]. Second, with respect to the organic carbon substrates consumed by picoheterotrophs, previous work has organized these compounds along an axis of "lability," correlating with turnover rates[21]. Due to the confounding impacts of variable energetic content, enzymatic cost, enzymatic rates, and many other factors[22], compounds range from highly labile, with turnover rates of minutes, to highly recalcitrant, with turnover rates of decades to millenia[21,23].

Here, we develop a global ocean ecosystem model that uses these two established organizing principles to resolve picoheterotrophic functional diversity (Fig. 1). We compare the global picoheterotrophic biogeographical distributions predicted by the model to the coarse-grained patterns revealed from a global-scale sequencing dataset. Our integrated model-data synthesis identifies and explains structural and functional patterns of the picoheterotrophic metacommunity. In addition to the transition from oligotrophs to copiotrophs with latitude in the surface, model and data also demonstrate a depth transition: oligotrophy at low-latitude surface waters gives way to deep populations that grow more slowly than surface populations, but are demonstrably more copiotrophic. This vertical shift accompanies a concurrent functional shift in the model from the remineralization of labile substrates in the surface to deep remineralization of recalcitrant substrates by 'slow copiotrophs'. This functional shift correlates with the observed increase in microbial diversity with depth, which is also captured by the model.

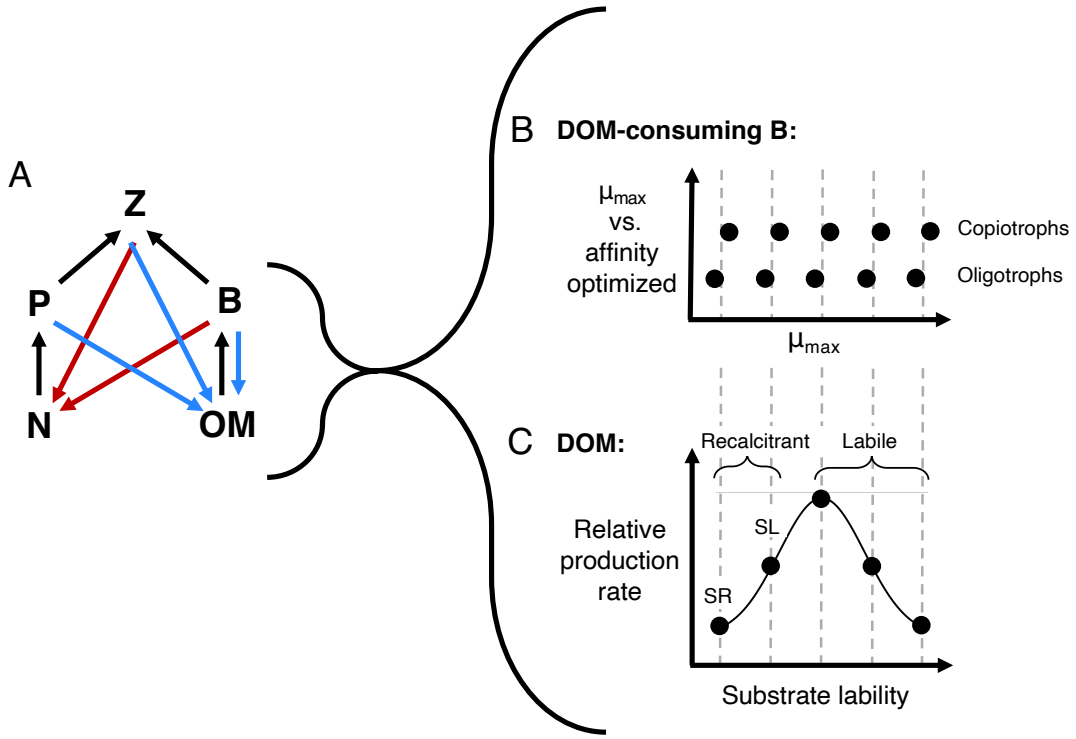

Figure 1: Schematic of the marine ecosystem model with picoheterotrophs. (A) The model resolves populations of picoheterotrophs (B, for bacteria), organic matter (OM), phytoplankton (P), inorganic nutrients (N), and zooplankton grazers (Z). Populations and substrates are connected via consumption (black arrows), respiration (red arrows), and mortality (blue arrows). (B) DOM-consuming picoheterotrophic diversity is resolved with two axes: 1) Lability, which sets the baseline maximum growth rate (μ_{max}) of consumers, and 2) a tradeoff between maximum growth rate (copiotrophy) and substrate affinity (oligotrophy). (C) Total DOM production is partitioned into five classes following a lognormal distribution[22, 24]. These five classes align with empirical and dynamical classifications[21, 22]: labile (an aggregate of the three functionally labile classes) and recalcitrant (semi-labile (SL) and semi-refractory (SR) classes).

Global-scale picoheterotroph biogeography

We generate a comprehensive, coarse-grained biogeography of the picoheterotroph metacommunity by leveraging an *in situ* metabarcoding dataset from seawater along an Arctic-to-Antarctic transect through the Pacific Ocean (Fig. 2A). Diversity and richness of the amplicon sequence variants (ASVs; unique genetic barcodes) both increase with depth (Fig. S1). These patterns remain whether considering the entire biological community (the complete SSU 16S and 18S rRNA amplicon dataset) or the picoheterotrophic subset (Methods; Fig. S1). To reduce the complexity of the dataset, we aggregated the ASVs of all non-phototrophic prokaryotes into ecologically meaningful taxonomic guilds: 21 picoheterotrophic guilds and 2 chemoautotrophic (ammonia- and nitrite-oxidizing) guilds (Methods; Figs. S2–S7; Table S1, see list of picoheterotrophic guild names in Fig. 3A). We aggregated at a level that broadly represents function[25,26], predominantly at the order-level or higher, acknowledging that much functional diversity remains unresolved within each guild. The 23 guilds cumulatively represent 92% of the non-photoautotrophic prokaryotic abundance in our dataset (Fig. S8).

The biogeographies of the 21 picoheterotrophic guilds reveal several reoccurring patterns, which we identify by broad locations of peak relative abundances (RA) (Figs. 2, 3, and S2–S6). We find that six *surface, low latitude/ubiquitous* guilds have peak RA in the nutrient-depleted ocean deserts. SAR11 is the most abundant representative of this group (Fig. 2C), and is known for its high abundance and oligotrophic traits: slow growth, small cell size, and a high affinity for DOC[27,28]. Three *surface, high latitude* guilds are identified by peak RA in high-latitude, nutrient-rich waters (e.g., *Flavobacteriales*, Fig. 2D). Seven *deep, ubiquitous* guilds have negligible RA in surface waters but are present throughout the subsurface (e.g., SAR324 and SAR202, Fig. 2E and F). Four *particle-like* guilds (e.g., Marinimicrobia) are broadly ubiquitous, generally peaking at depth but also present in surface waters, and named as such here because this biogeography is consistent with the emergent organic-particle-associated populations in the model (Fig. S5). One *deep, equatorial* guild is identified (*Enterobacterales*), which we speculate may also be associated with sinking

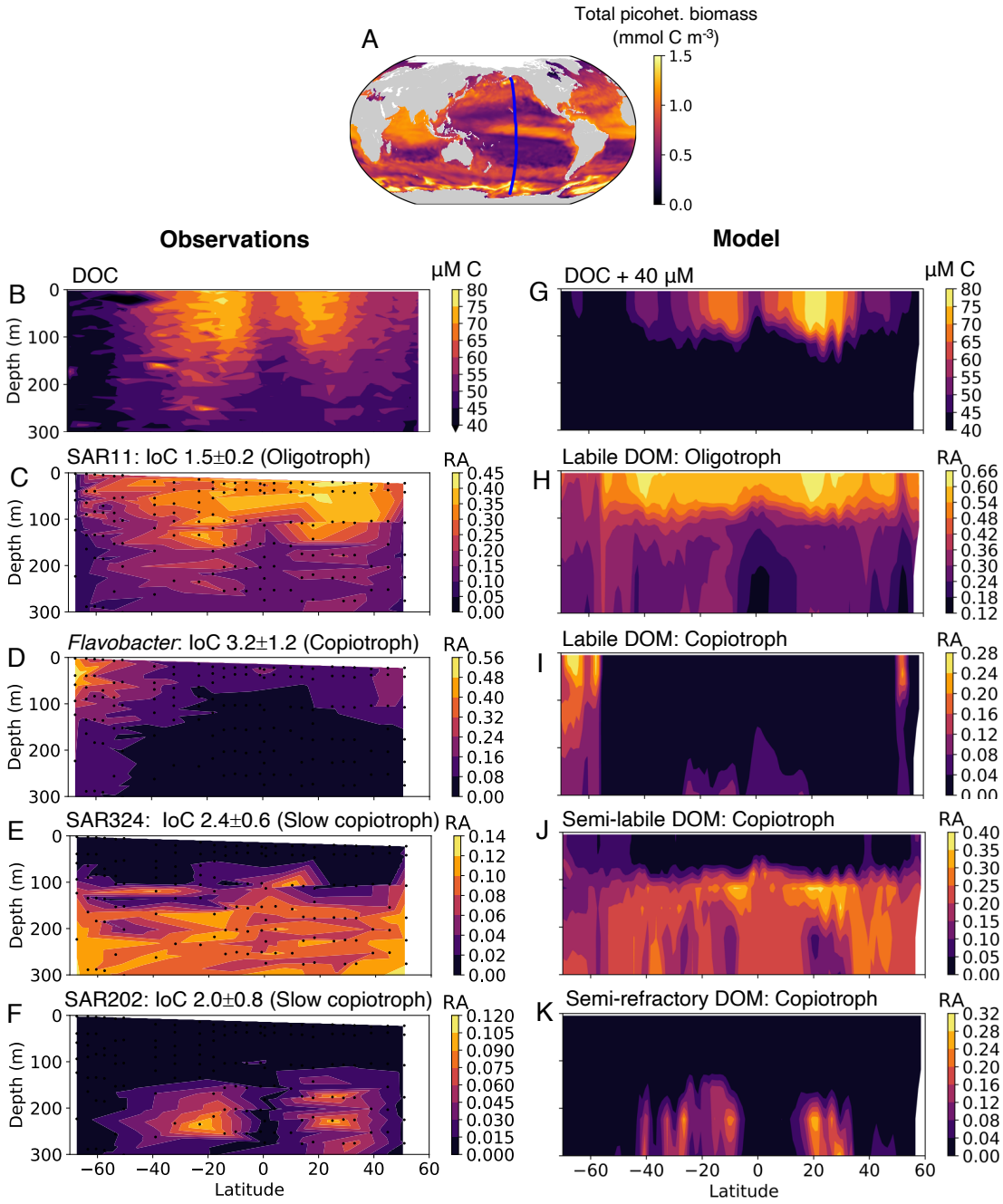

Figure 2: Observed and modeled distributions across the Pacific Ocean. (A) The Pacific transect (P16N and P16S: blue line) overlaid on the modeled surface concentration of total picoheterotrophic biomass. (B, G) Total dissolved organic carbon (DOC). (C–F) Relative abundances (RA) of four of the 21 taxonomic guilds, index of copiotrophy (IoC) value and associated classification. (H–K) RA of four modeled picoheterotrophic functional types. RA is calculated with respect to non-photosynthetic prokaryotes of ASVs (data) and functional types (model).

particles (Fig. S6). Because the guilds span the whole picoheterotrophic community, many likely include both particle-associated and free-living populations. Indeed, size-fractionated analysis reveals similarities in taxonomic composition in free-living and particle-associated fractions[29]. Though the categorization of some guilds is ambiguous (e.g., SUP05 RA is both high-latitude- and depth-intensified), the repetitive patterns hint at the potential for a coarse-grained mechanistic explanation of community structure.

Modeling picoheterotrophic functional diversity

We next compare the observed picoheterotrophic biogeographical patterns with simulations from a dynamic marine ecosystem model with explicit picoheterotrophs. The global model synthesizes the physics, biogeochemistry, and ecosystem of the ocean (Darwin-MITgcm; Methods). The ecosystem component allows for the self-assembly of microbial communities[30,31]. We extend a previous configuration in which all remineralization is due to the growth and respiration of dynamic microbial functional type populations, resolving multiple phytoplankton, zooplankton, picoheterotrophs that consume particulate and dissolved organic matter (POM and DOM), and other chemoautotrophic and anaerobic metabolic functional types (Methods; Fig. S9). Here we expand the model to include five lability classes of DOM and associated free-living picoheterotrophic consumers (Fig. 1). Lability sets a baseline maximum growth rate of the consumers[22] (Fig. 1B,C). For each lability class of DOM, we resolve the metabolic tradeoff axis with two competing consumers: small oligotrophs with maximum growth rates lower than the baseline and higher substrate affinity, and larger copiotrophs with maximum growth rates higher than the baseline and lower affinity[13,14,16,17] (Fig. 1B, Table S2). DOM is produced from the mortalities and excretions of all biomasses following a lognormal distribution (Fig. 1C), representing the average outcome from multiplicative stochasticity of complex underlying processes[22,24]. Consequently, cross-feeding and transformations among DOM classes are resolved implicitly. Modeled predation is structured by organism size, with a preferred prey size for each zooplankton grazer[32,33]. We assume that all oligotrophs, regardless of lability preferences,

are consumed by the same grazer (the smallest zooplankton). The slightly larger copiotrophs have different shared grazer (the second smallest zooplankton). We integrate the model forward in time to reach a quasi-steady state and examine the emergent microbial community structure (Methods). Model solutions are consistent with observed estimates of global net primary production, the export flux of organic carbon, DOM production, surface picoheterotrophic biomass, and surface picoheterotrophic growth rates (Figs. S10–S12).

We find that many of the emergent picoheterotrophic biogeographies in the model are remarkably consistent with the broad biogeographic patterns exhibited by the taxonomic guilds (Figs. 2, S13, S14). This suggests that taxonomy at this coarse level may indeed broadly correspond to function. We compare models and data qualitatively because we lack an accurate, common currency between model (i.e., biomass concentrations) and molecular data (i.e., sequence counts). In modeled surface waters, the most abundant picoheterotrophic types are the consumers of labile DOM (Fig. 2H and I). Modeled labile DOM, like measured labile constituents, is depleted to low (pM–nM) concentrations, reflecting the subsistence concentrations of the consumers[9,22] (Fig. S15). The model captures the latitudinal transition from oligotrophs to copiotrophs in the surface, which is consistent with previous theory and modeling of both phytoplankton and picoheterotrophs[20,33–35] (Fig. 2H and I). In contrast, all consumers of DOM that is functionally recalcitrant in the surface[22] (hereafter referred to as RDOM, aligning with previously designated semi-labile and semi-refractory classes[21] (Fig. 1C, Fig. S10), are largely excluded from the surface, broadly matching the biogeography of the *deep, ubiquitous* guilds (Fig. 2, Fig. S14).

Mechanisms shaping picoheterotrophic community structure

In the model, we diagnose that the consumers of RDOM are excluded from the surface because of predation. These slow-growing populations cannot sustain themselves under the high rates of zooplankton grazing in surface waters, which emerge dynamically in the model, supported by the high biomass of other microbial populations. This "apparent competition"[36], or "shared predation"[33],

is a robust feature of the model. Results hold over a wide range in parameter space, within some plausible constraints (Supplementary Text, Figs. S16–S18).

Because its consumers are excluded, the RDOM accumulates in modeled surface waters (Fig. 2G, Fig. S14). At depth, below the mixed layer (200-300m), predation rates wane, along with total biomass of picoheterotrophs and phytoplankton (Fig. S11), and the slower-growing populations finally become viable. Notably, the slow-growing copiotrophs, rather than the oligotrophs, dominate the deep consumption of RDOM (Fig. S14). This is because concentrations of RDOM remain relatively high below the mixed layer. The slightly higher maximum growth rates of the slow-growing copiotrophs allow them to outcompete their oligotrophic counterparts (Fig. S15). The RDOM is eventually depleted further at depth, at which point oligotrophic traits are selected for. However, at that point, DOM supply is so low that only very low abundances of any picoheterotrophic consumers are sustained (Fig. S13), which perhaps suggests that these 'slow oligotrophs' might be a part of the "rare biosphere"[37].

Thus the model paints an insightful picture of how picoheterotrophic community structure and function varies with latitude and depth: 1) a transition in surface waters from the dominance of oligotrophs at low latitudes to the prevalence of copiotrophs at high latitudes, and 2) a vertical transition from the dominance of relatively faster-growing, yet oligotrophic populations consuming labile substrates in surface waters to slower-growing, yet copiotrophic populations consuming recalcitrant substrates at depth. The first pattern is driven by the general pattern of increased primary production with latitude, and selection for rate-optimized copiotrophs in dynamic, nutrient-rich waters. The second pattern applies across the lability gradient, and is driven by the decrease in predation rates with depth. The growth-rate-optimized copiotrophs consuming lower quality (recalcitrant) substrates become sustainable at shallower depths relative to their oligotrophic counterparts.

This vertical pattern and its mechanisms are consistent with previous observations. DOC decreases by about 20-40 μM between the surface and the mesopelagic ocean[21], which is captured by the model (Fig. 2B and G). Experiments have shown that only microbial communities originating from

deep waters can consume the surface-accumulated DOC, even when controlling for the additional nutrients available at depth[38]. Indeed, time-series data demonstrate that the seasonal downward flux of surface-accumulated DOC is followed by the growth of mesopelagic microbial populations[39]. It has remained an open question why such mesopelagic populations are not active in surface waters, where their preferred substrates accumulate. The model results here propose high surface grazing rates as one explanation for this phenomenon.

"Slow copiotrophs" in the deep ocean

The modeled biogeographies of RDOM consumers qualitatively match the biogeographies and putative functions of the SAR324 and SAR202 guilds, respectively (Fig. 2). Critically, the model hypothesizes that SAR202 and SAR324 (as well as other depth-intensified guilds) may be better classified as copiotrophs than oligotrophs. They may be optimized for maximizing growth, even though absolute growth rates may be lower than other guilds because they consume recalcitrant substrates. Sequencing suggests that members of the SAR202 guild do metabolize recalcitrant substrates[40–42], and both guilds contain high metabolic versatility and diversity, which may indicate distinct function from surface clades[41,43,44]. We show that the greater ubiquity of the SAR324 model analog relative to the SAR202 model analog is because of its higher potential growth rate (set by substrate lability), rather than its higher substrate supply (Figs. S19 and S20). This allows us to speculate that SAR324 grow more quickly than SAR202.

To test the model-hypothesized shift from the dominance of oligotrophy to copiotrophy with depth, we developed a genome-based index of copiotrophy (IoC). We extracted representative genomes with rRNA sequences that matched ASVs at approximately the genus level from the novel, comprehensive Genome Taxonomy Database[45], and assembled the IoC as a function of three copiotrophic characteristics[15]: 1) the number of genes, 2) the number of carbohydrate active enzymes (CAZy database)[46], and 3) codon usage bias (CUB) (Methods; Figs. S21–S23). CUB is the biased use of alternative codons coding for a particular amino acid that, in genes coding for

ribosomal proteins, may optimize growth rate[15]. We also used CUB to estimate maximum growth rates for each genome[15], with and without the confounding impact of temperature, and considering the large uncertainties in the estimates (Methods; Figs. S24–S27).

A Gaussian mixture model of the resulting IoC values shows that the ASVs fall into two statistically significant lognormal distributions (Fig. 3, Fig. S22). We label these as the "oligotroph" and "copiotroph" clusters, in an update to the classification in previous work[15]. The oligotroph cluster contains predominantly the *surface, low latitude/ubiquitous* guilds, such as SAR11. The copiotroph cluster contains most of the *surface, high latitude* guilds, two of the *particle-like* guilds, and *Enterobacterales*. However, most of the *deep, ubiquitous* guilds (6 of 7) fall between these two extremes, and do not clearly belong to either cluster. These deep guilds appear to have the lowest maximum growth rates among the guilds (Fig. 3B), even when controlling for temperature (Fig. S28). This suggests that SAR11 and other surface guilds are more oligotrophic but may actually grow faster than deep guilds, decoupling copiotrophy from growth rate (Fig. 3B)[47]. We label these in-between guilds as 'slow copiotrophs' (or perhaps for short, 'slowiotrophs').

At the community level, both model and data show that the degree of copiotrophy increases with depth as the community-averaged maximum growth rate decreases (Fig. 4, Fig. S23, Fig. S27). Much of the picoheterotrophic community in the deep ocean appears to be evolutionarily driven to optimize growth, despite lower absolute growth rates. Our framework suggests that this pattern reflects the consumption of accumulated, low quality (recalcitrant) substrates that do not allow for fast growth in absolute terms, perhaps due to the enzymatic expense of metabolizing these substrates or their energetic content. This dynamic may apply across natural environments, such as in soils and sediments, where the quality of substrate is low, but the quantity is high (Fig. 3C).

The model reproduces the observed increase in microbial diversity with depth (Fig. 4, Fig. S1). In the model, slow-growing functional types are excluded from the surface, but few functional types are excluded at depth because continued microbial cycling supplies fresh labile DOM at low rates there, supporting low abundances of many functional types. Our results thus suggest high surface

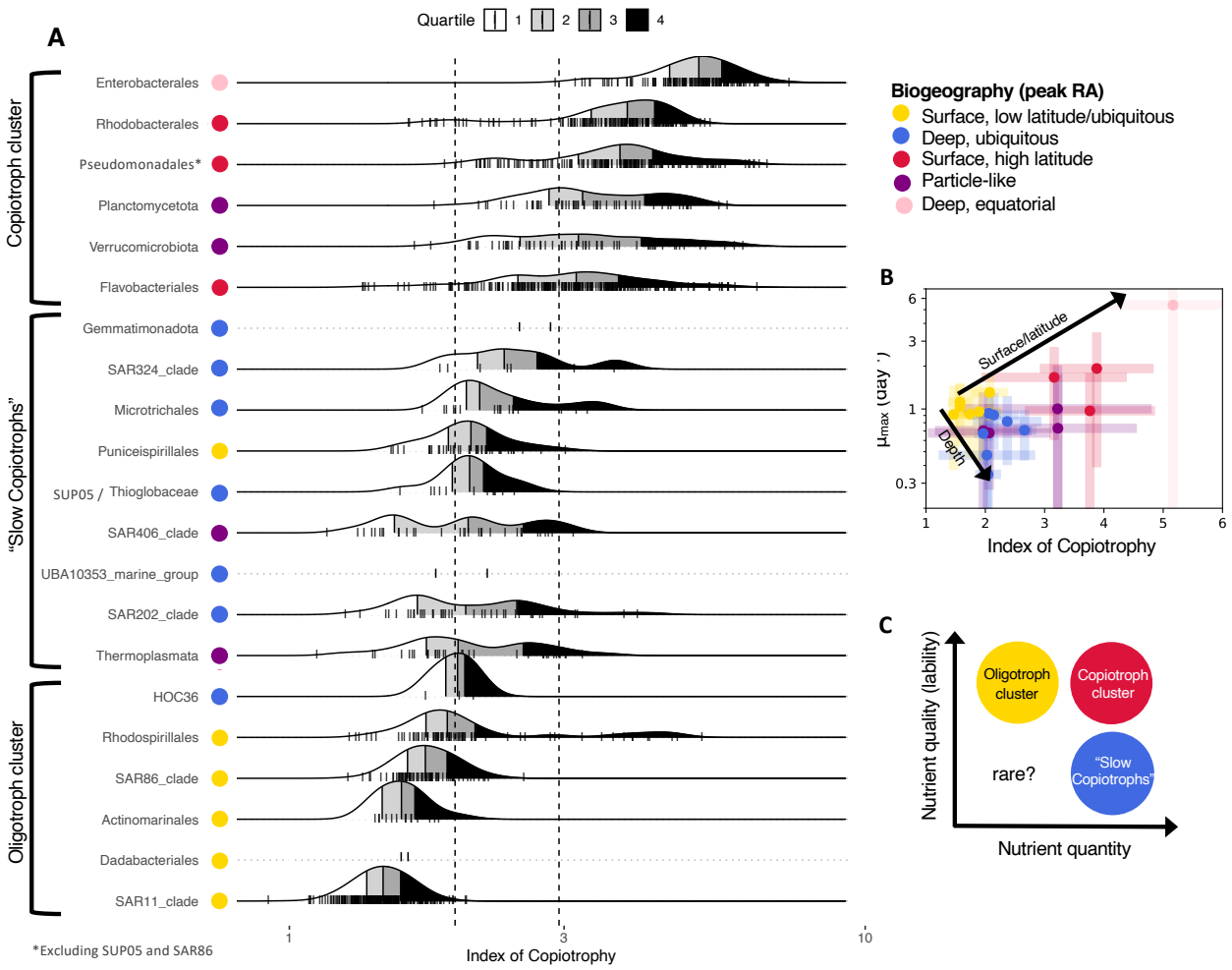

Figure 3: Biogeographical and functional patterns of marine picoheterotrophic taxonomic guilds. (A) Guilds ranked by the median value of the genome-based index of copiotrophy. Dashed lines indicate uncertainty in the clustering analysis, delineating the "slow copiotrophs." (B) Genome-based maximum growth rate (μ_{max}) estimates[15] against the index of copiotrophy for each guild (median and interquartile range), colored by biogeography. (C) Schematic of proposed broad classifications of marine picoheterotrophic guilds.

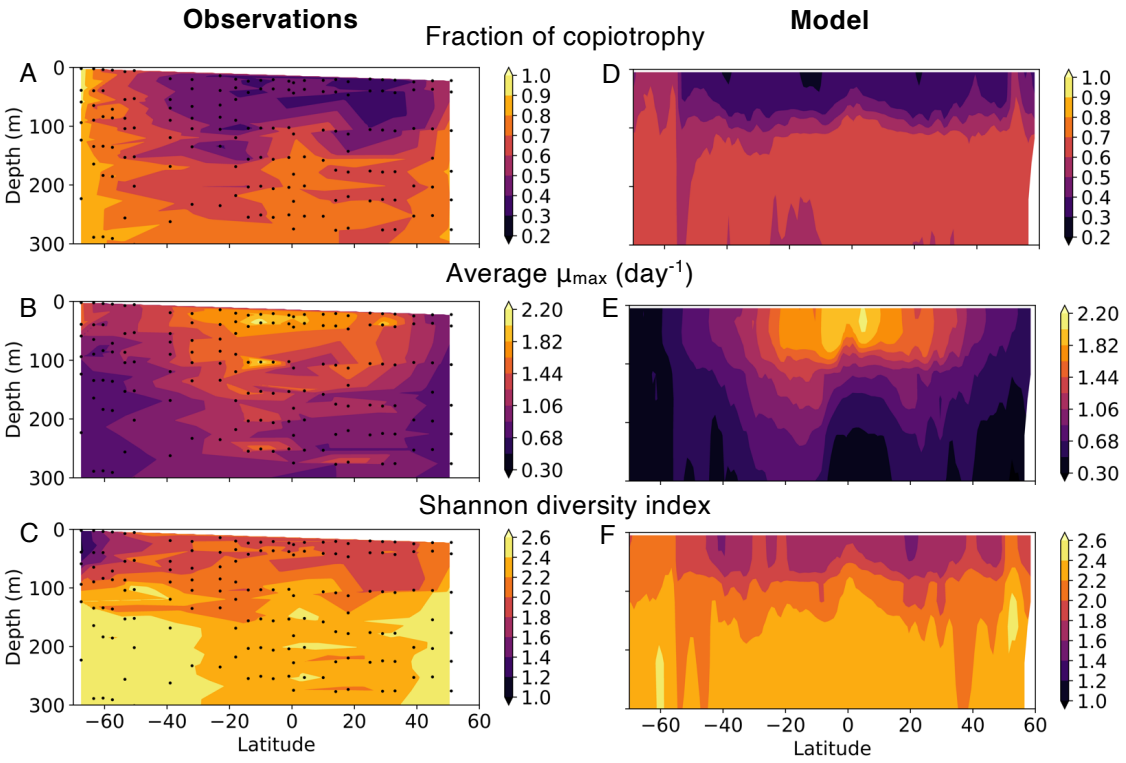

Figure 4: Observed and modeled picoheterotrophic community patterns. (A and D) Fraction of the picoheterotrophic community classified as copiotrophic. (B and E) Community-averaged maximum growth rate (μmax) estimate[15]. (C and F) Shannon diversity index, calculated with respect to the taxonomic guilds (C) and the modeled functional types (F).

predation rates as a mechanism for the higher microbial diversity in the deep ocean (apparent competition)[36]. Other factors, such as ecological succession on or around sinking particles[48], may also contribute.

Outlook

Our results demonstrate an "emergent simplicity"[26] and predictability of the complex marine microbiome at the global scale. Resolving picoheterotrophic diversity with two axes – a range of organic substrate lability and a rate-affinity tradeoff – is sufficient to broadly capture observed global patterns and simple enough to understand the mechanisms producing them. Our model suggests that gradients in predation pressure organize the patterns of DOM remineralization in terms of its lability and in the process the biogeography of diverse guilds of picoheterotrophs. Results explicitly link diversity to function and propose a new picoheterotrophic functional type: a "slow copiotroph." Incorporating other traits and tradeoffs to the framework could resolve additional functional diversity[49–52].

Competition for inorganic nutrients may also contribute the exclusion of slow-growing types from the ocean surface. Though experiments suggest that this is not the dominant mechanism[38], slow-growing consumers of RDOM would also likely lose a competition for inorganic nitrogen, phosphorus, iron, and/or other micronutrients in environments. Small oligotrophs like SAR11 are likely to also have high affinity for these nutrients, and consumers of RDOM may have a more limited proteome to allocate towards efficient nutrient uptake.

Finally, the model mechanistically resolves diverse rates and locations of CO_2 production from picoheterotroph respiration (Fig. S29). The heterogeneity suggests that different picoheterotrophic types may respond differently to environmental perturbations. Modeled labile DOM is remineralized predominantly in the surface ocean, while RDOM is remineralized predominantly at depth. This vertical structure translates into qualitative differences in the controls on remineralization. Substrate supply rate limits the remineralization rate of labile DOM, because it is depleted to low concentrations[22,53],

set by the subsistence concentrations of the consumers (Fig. S15). In contrast, remineralization of RDOM, mainly at depth, is limited by the maximum growth rates of the microbial consumers. This shift from substrate-limited (externally controlled) to growth-rate-limited (internally controlled) remineralization with depth suggests that surface vs. deep microbial communities may provide distinct feedbacks to ocean warming. For example, we speculate that deeper growth-rate-limited remineralization rates may increase as metabolic rates increase with warming. This could reduce standing stocks of DOC (analogous to soil carbon stocks[54]) as a positive feedback to warming.

We have identified the mechanisms shaping the coarse-grained structure and function of marine picoheterotrophic communities in the ocean. In doing so, we have made progress towards meeting two grand challenges in marine biogeochemistry: organizing the functional diversity of complex microbial communities along quantitative axes, and connecting sequencing datasets with ecosystem models. Additionally, the model represents a strategy for incorporating dynamic picoheterotrophic feedbacks into earth system models. This strategy can be used to investigate heterogeneity in the response of microbial communities impacting biological carbon storage in a warmer climate.

Methods

Please see the attached Supporting Information for Methods, supplemental figures, supplemental tables, and free, public repositories for data and model code.

Acknowledgments

EJZ and JLW thank the Simons Foundation Postdoctoral Fellowship in Marine Microbial Ecology for funding. EJZ and LX acknowledge the Carnegie Institution for Science for funding and support. EJZ and JAF thank NSF (Grant #2125142) for funding. JAF thanks the Simons Foundation (Grant #549943FY22) for funding. SD acknowledges funding from the Simons Collaboration on Computational Biogeochemical Modeling of Marine Ecosystem/CBIOMES (Grant #549931). CAC thanks Simons Foundation International's BIOS-SCOPE program for support. NML acknowledges funding from the Simons Foundation (Grant #542389).

References

1. Falkowski, P. G., Fenchel, T. & Delong, E. F. The microbial engines that drive Earth's biogeochemical cycles. *Science* **320,** 1034–9 (2008).
2. Volk, T. & Hoffert, M. I. in *The carbon cycle and atmospheric CO2: Natural variations Archean to present. Chapman conference papers, 1984* (eds Sundquist, E. T & Broecker, W. S.) 99–110 (American Geophysical Union, 1985).
3. Séférian, R. *et al.* Tracking Improvement in Simulated Marine Biogeochemistry Between CMIP5 and CMIP6. *Current Climate Change Reports* **6,** 95–119 (2020).
4. Wieder, W. R., Bonan, G. B. & Allison, S. D. Global soil carbon projections are improved by modelling microbial processes. *Nature Climate Change* **3,** 909–912 (2013).
5. Henson, S. A. *et al.* Uncertain response of ocean biological carbon export in a changing world. *Nature Geoscience* **15,** 248–254 (2022).
6. Hedges, J. I. *et al.* The molecularly-uncharacterized component of nonliving organic matter in natural environments. *Organic Geochemistry* **31,** 945–958 (2000).
7. Carlson, C. A. & Hansell, D. A. in *Biogeochemistry of Marine Dissolved Organic Matter* 65–126 (Elsevier Inc., 2015).

8. Steen, A. D. *et al.* High proportions of bacteria and archaea across most biomes remain uncultured. *ISME Journal* **13,** 3126–3130 (2019).

9. Moran, M. A. *et al.* Microbial metabolites in the marine carbon cycle. *Nature Microbiology* **7,** 508–523 (2022).

10. Sunagawa, S. *et al.* Structure and function of the global ocean microbiome. *Science* **348,** 1261359 (2015).

11. Coles, V. J. *et al.* Ocean biogeochemistry modeled with emergent trait-based genomics. *Science* **358,** 1149–1154 (2017).

12. Milke, F., Meyerjürgens, J. & Simon, M. Ecological mechanisms and current systems shape the modular structure of the global oceans' prokaryotic seascape. *Nature Communications* **14,** 6141 (2023).

13. Koch, A. L. Oligotrophs versus copiotrophs. *BioEssays* **23,** 657–661 (2001).

14. Lauro, F. M. *et al.* The genomic basis of trophic strategy in marine bacteria. *Proceedings of the National Academy of Sciences* **106,** 15527–15533 (2009).

15. Weissman, J. L., Hou, S. & Fuhrman, J. A. Estimating maximal microbial growth rates from cultures, metagenomes, and single cells via codon usage patterns. *Proceedings of the National Academy of Sciences* **118,** e2016810118 (2021).

16. Soler-Bistué, A., Couso, L. L. & Sanchez, I. E. The evolving copiotrophic / oligotrophic dichotomy: From Winogradsky to physiology and genomics. *Environmental Microbiology* **25,** 1232–1237 (2023).

17. Norris, N., Levine, N. M., Fernandez, V. I. & Stocker, R. Mechanistic model of nutrient uptake explains dichotomy between marine oligotrophic and copiotrophic bacteria. *PLoS Computational Biology* **17,** e1009023 (2021).

18. Wietz, M, Gram, L, Jørgensen, B & Schramm, A. Latitudinal patterns in the abundance of major marine bacterioplankton groups. *Aquatic Microbial Ecology* **61,** 179–189 (2010).

19. Yooseph, S. *et al.* Genomic and functional adaptation in surface ocean planktonic prokaryotes. *Nature* **468,** 60–66 (2010).

20. Brown, M. V., Ostrowski, M., Grzymski, J. J. & Lauro, F. M. A trait based perspective on the biogeography of common and abundant marine bacterioplankton clades. *Marine Genomics* **15,** 17–28 (2014).

21. Hansell, D. A. Recalcitrant Dissolved Organic Carbon Fractions. *Annual Review of Marine Science* **5,** 421–445 (2013).

22. Zakem, E. J., Cael, B. B. & Levine, N. M. A unified theory for organic matter accumulation. *Proceedings of the National Academy of Sciences* **118,** e2016896118 (2021).

23. Carlson, C. & Ducklow, H. Dissolved organic carbon in the upper ocean of the central equatorial Pacific Ocean, 1992: Daily and finescale vertical variations. *Deep Sea Research Part II: Topical Studies in Oceanography* **42,** 639–656 (1995).

24. Forney, D. C. & Rothman, D. H. Common structure in the heterogeneity of plant-matter decay. *Journal of The Royal Society Interface* **9,** 2255–2267 (2012).

25. Louca, S., Parfrey, L. W. & Doebeli, M. Decoupling function and taxonomy in the global ocean microbiome. *Science* **353,** 1272–1277 (2016).

26. Goldford, J. E. *et al.* Emergent simplicity in microbial community assembly. *Science* **361,** 469–474 (2018).

27. Giovannoni, S. J. SAR11 Bacteria: The Most Abundant Plankton in the Oceans. *Annual Review of Marine Science* **9,** 231–255 (2017).

28. Noell, S. E. & Giovannoni, S. J. SAR11 Bacteria Have a High Affinity and Multifunctional Glycine Betaine Transporter. *Environmental Microbiology* **21,** 2559–2575 (2019).

29. Yeh, Y.-c. & Fuhrman, J. A. Effects of phytoplankton, viral communities, and warming on free-living and particle- associated marine prokaryotic community structure. *Nature Communications* **13,** 7905 (2022).

30. Follows, M. J., Dutkiewicz, S., Grant, S. & Chisholm, S. W. Emergent biogeography of microbial communities in a model ocean. *Science* **315,** 1843–6 (2007).

31. Dutkiewicz, S. *et al.* Impact of ocean acidification on the structure of future phytoplankton communities. *Nature Climate Change* **5,** 1002–1006 (2015).

32. Ward, B. A., Dutkiewicz, S., Jahn, O. & Follows, M. J. A size-structured food-web model for the global ocean. *Limnology and Oceanography* **57,** 1877–1891 (2012).

33. Follett, C. L. *et al.* Trophic interactions with heterotrophic bacteria limit the range of Prochlorococcus. *Proceedings of the National Academy of Sciences* **119,** e2110993118 (2022).

34. Dutkiewicz, S., Follows, M. J. & Bragg, J. G. Modeling the coupling of ocean ecology and biogeochemistry. *Global Biogeochemical Cycles* **23,** GB4017 (2009).

35. Abreu, C. I., Bello, M. D., Bunse, C., Pinhassi, J. & Gore, J. Warmer temperatures favor slower-growing bacteria in natural marine communities. *Science Advances* **9,** eade8352 (2023).

36. Holt, R. D. Predation, apparent competition, and the structure of prey communities. *Theoretical Population Biology* **12,** 197–229 (1977).

37. Sogin, M. L. *et al.* Microbial diversity in the deep sea and the underexplored "rare biosphere ". *Proceedings of the National Academy of Sciences* **103,** 12115–12120 MICROBIOLOGY (2006).

38. Carlson, C. A. *et al.* Interactions among dissolved organic carbon, microbial processes, and community structure in the mesopelagic zone of the northwestern Sargasso Sea. *Limnology and Oceanography* **49,** 1073–1083 (2004).

39. Liu, S. *et al.* Linkages Among Dissolved Organic Matter Export , Dissolved Metabolites , and Associated Microbial Community Structure Response in the Northwestern Sargasso Sea on a Seasonal Scale. *Frontiers in Microbiology* **13,** 833252 (2022).

40. Landry, Z., Swa, B. K., Herndl, G. J., Stepanauskas, R. & Giovannoni, S. J. SAR202 genomes from the dark ocean predict pathways for the oxidation of recalcitrant dissolved organic matter. *mBio* **8,** 00413–17 (2017).

41. Saw, J. H. *et al.* Pangenomics Analysis Reveals Diversification of Enzyme Families and Niche Specialization in Globally Abundant. *mBio* **11,** 02975–19 (2020).

42. Liu, S. *et al.* Different carboxyl-rich alicyclic molecules proxy compounds select distinct bacterioplankton for oxidation of dissolved organic matter in the mesopelagic Sargasso Sea. *Limnology and Oceanography* **65,** 1532–1553 (2020).

43. Sheik, C. S., Jain, S. & Dick, G. J. Metabolic flexibility of enigmatic SAR324 revealed through metagenomics and metatranscriptomics. *Environmental Microbiology* **16,** 304–317 (2014).

44. Boeuf, D. *et al.* Metapangenomics reveals depth-dependent shifts in metabolic potential for the ubiquitous marine bacterial SAR324 lineage. *Microbiome* **9,** 172 (2021).

45. Parks, D. H. *et al.* GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Research* **50,** D785–D794 (2022).

46. Drula, E. *et al.* The carbohydrate-active enzyme database: functions and literature. *Nucleic Acids Research* **50,** D571–D577 (2022).

47. Couso, L. L., Soler-Bistué, A., Aptekmann, A. A. & Sánchez, I. E. Ecology theory disentangles microbial dichotomies. *Environmental Microbiology* **25,** 3052–3063 (2023).

48. Datta, M. S., Sliwerska, E., Gore, J., Polz, M. F. & Cordero, O. X. Microbial interactions lead to rapid micro-scale successions on model marine particles. *Nature Communications* **7,** 11965 (2016).

49. Green, J. L., Bohannan, B. J. M. & Whitaker, R. J. Microbial biogeography: from taxonomy to traits. *Science* **320,** 1039–1043 (2008).

50. Allison, S. D. A trait-based approach for modelling microbial litter decomposition. *Ecology Letters* **15,** 1058–1070 (2012).

51. Martiny, J. B., Jones, S. E., Lennon, J. T. & Martiny, A. C. Microbiomes in light of traits: A phylogenetic perspective. *Science* **350** (2015).

52. Litchman, E., Edwards, K. F. & Klausmeier, C. A. Microbial resource utilization traits and trade-offs: Implications for community structure, functioning, and biogeochemical impacts at present and in the future. *Frontiers in Microbiology* **6,** 254 (2015).

53. Arrieta, J. M., Mayol, E., Hansman, R. L. & Herndl, G. J. Dilution limits dissolved organic carbon utilization in the deep ocean. *Science* **348,** 331–333 (2015).

54. Crowther, T. W. *et al.* Quantifying global soil carbon losses in response to warming. *Nature* **540,** 104–108 (2016).