Emergent Biogeography of Microbial Communities in a Model Ocean

Michael J. Follows,[1]* Stephanie Dutkiewicz,[1] Scott Grant,[1,2] Sallie W. Chisholm[3]

A marine ecosystem model seeded with many phytoplankton types, whose physiological traits were randomly assigned from ranges defined by field and laboratory data, generated an emergent community structure and biogeography consistent with observed global phytoplankton distributions. The modeled organisms included types analogous to the marine cyanobacterium *Prochlorococcus*. Their emergent global distributions and physiological properties simultaneously correspond to observations. This flexible representation of community structure can be used to explore relations between ecosystems, biogeochemical cycles, and climate change.

A significant challenge in understanding the changing earth system is to quantify and model the role of ocean ecosystems in the global carbon cycle. The structure of microbial communities in the surface ocean is known to regulate important biogeochemical pathways, including the efficiency of export of organic carbon to the deep ocean. Although there is extraordinary diversity in the oceans, the biomass of local microbial communities at

any location is typically dominated by a smaller subset of strains. Their relative fitness and ecosystem community structure are regulated by a variety of factors, including physical conditions, dispersal, predation, competition for resources, and the variability of the environment (1–3). Models reflecting this conceptual view have been examined in idealized ecological settings (4) and have been applied to studies of terrestrial ecosystems (5). We have used this approach in a marine ecosystem model that embraces the diversity of microbes and their genomic underpinnings, a model in which microbial community structure "emerges" from a wider set of possibilities and, thus, mimics aspects of the process of natural selection. The system is flexible enough to respond to changing ocean environments and can be used to interpret the structure and development of marine microbial communities and to reveal critical links between marine ecosystem structure, global biogeochemical cycles, and climate change.

Recent ocean models have begun to resolve community structure by the explicit representation of three or four classes, or functional groups, of phytoplankton (6–9), but significant challenges remain (10, 11). First, the specification of functional groups and diversity of the model ecosystem is subjective and somewhat arbitrary. Second, it is difficult to evaluate the parameters controlling such models because quantitative, physiological information from laboratory cultures is extremely limited. Third, observations of microbial community structure with which to evaluate global-scale models are still relatively sparse. Finally, model ecosystem structures optimized to reflect today's ocean may not be sufficiently dynamic to adapt appropriately to a changing climate where radical shifts in community structure might be possible.

To circumvent some of these difficulties, we formulated a marine ecosystem model that represents a large number of potentially viable phytoplankton types whose physiological characteristics were determined stochastically. The initialized organism types interacted with one another and their environment, evolving into a sustainable ecosystem where community structure and diversity were not imposed, but were emergent properties.

The ecosystem model consisted of a set of coupled prognostic equations (eqs. S1 to S5), with idealized representations of the transformations of inorganic and organic forms of phosphorus, nitrogen, iron, and silica. Many tens of phytoplankton types (here, 78) were initialized in each simulation, each type distinguished by

its physiological capabilities and the values of coefficients that control the rates and sensitivities of metabolic processes. These were provided by random drawing from broad ranges guided by laboratory and field studies (table S1). We focused these choices on light, temperature, and nutrient requirements (fig. S1), the niche dimensions for phytoplankton thought to be most important in regulating growth. To facilitate a test of the approach, we also specifically addressed functions that differentiate *Prochlorococcus* spp. from other phytoplankton, including their small size and inability to assimilate nitrate. Other functions could be emphasized depending on the aim of the study. Ecological trade-offs were imposed through highly simplified allometric constraints [see supporting online material (SOM)]. To reflect the extra energetic expense of using nitrate, relative to other inorganic nitrogen sources, we allowed the maximum growth rate to increase slightly when nitrate was not the major nitrogen source (12). Organisms incapable of utilizing nitrate were given a slightly lower nutrient half-saturation. We explicitly represented predation by two classes of grazer and, for the action of heterotrophic microbes, we used a simple remineralization rate (SOM).

A global ocean circulation model constrained by observations (13) provided flow fields and mixing coefficients that transport all biological

and chemical tracers. All phytoplankton types were initialized with identical distributions of biomass, and the model was integrated forward for 10 years, over which time a repeating annual cycle in ecosystem structure emerged. We repeated the integration 10 times, each time with a different random selection of phytoplankton physiologies, forming an ensemble of 10 members. Although each ensemble member produced a unique emergent ecosystem, the broad-scale patterns of productivity, community structure, and biogeography were robust across all 10. Global patterns of open-ocean biomass (Fig. 1A), primary production, and nutrients (fig. S3) were qualitatively consistent with in situ and remote observations. The ensemble mean globally integrated, annual primary production was 44 gigatons C per year, with a standard deviation of less than 5%. This small standard deviation suggested that sufficient phytoplankton "types" were initialized for consistent emergent solutions and also reflects the large-scale regulation by the physical transport of nutrients.

After an initial adjustment, the biomass of some phytoplankton types fell below the threshold of numerical noise, and these types were assumed to have become "extinct." In all ensemble members, about 20 phytoplankton types accounted for almost all of the total global biomass (fig. S2). We classified the phytoplankton types into four broad functional groups, each a

Fig. 1. Annual mean biomass and biogeography from single integration. (**A**) Total phytoplankton biomass (μM P, 0 to 50 m average). (**B**) Emergent biogeography: Modeled photo-autotrophs were categorized into four functional groups; color coding is according to group locally dominating annual mean biomass. Green, analogs of *Prochlorococcus*; orange, other small photo-autotrophs; red, diatoms; and yellow, other large phytoplankton. (**C**) Total biomass of *Prochlorococcus* analogs (μM P, 0 to 50 m average). Black line indicates the track of AMT13.
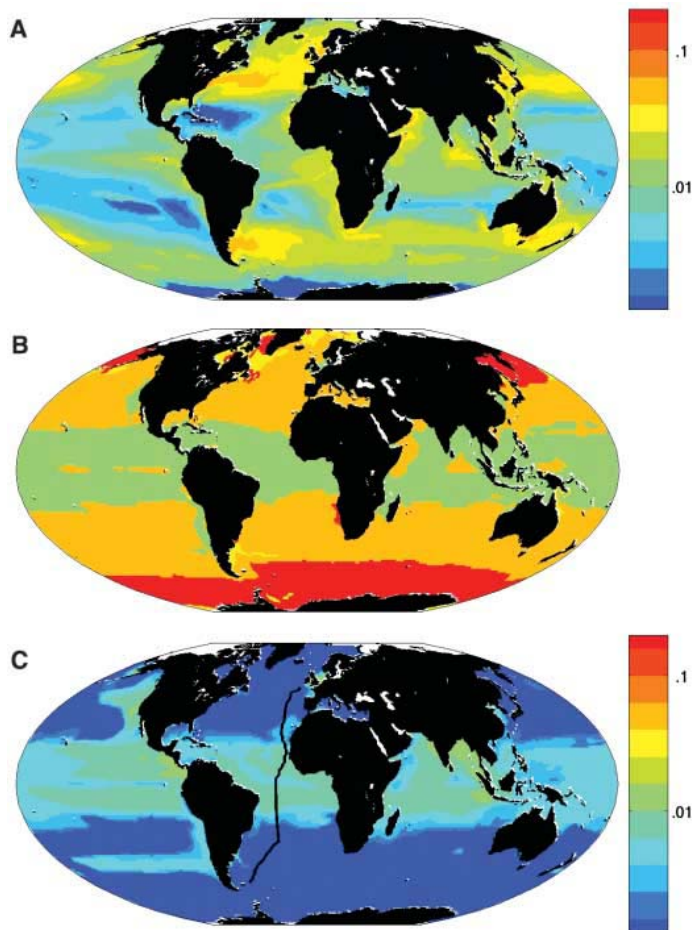
[1]Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, 54-1514 MIT, Cambridge, MA 02139, USA. [2]Department of Oceanography, University of Hawaii, 1000 Pope Road, Honolulu, HI 96822, USA. [3]Departments of Civil and Environmental Engineering and Biology, Massachusetts Institute of Technology, 48-419 MIT, Cambridge, MA 02139, USA

*To whom correspondence should be addressed. E-mail: mick@mit.edu

composite of several types, according to aspects of their physiology: (i) diatom analogs—large phytoplankton that require silica, (ii) other large eukaryotes, (iii) *Prochlorococcus* analogs—small phytoplankton that cannot assimilate nitrate, and (iv) other small photo-autotrophs. The large-scale biogeography of the emergent phytoplankton community was plausible with respect to observations (Fig. 1B) and consistent among the 10 ensemble members. The model successfully captured the domination of annual biomass by large phytoplankton in subpolar upwelling regions, where both light and macronutrients are seasonally plentiful. The subtropical oceans were dominated by small phytoplankton functional types (*14*). Large areas of the tropics and subtropics were dominated by several *Prochlorococcus* analogs (Fig. 1C), also in accord with observations (*15*, *16*). Along the cruise track of Atlantic Meridional Transect 13 (AMT13), total *Prochlorococcus* abundance (the sum of all *Prochlorococcus* analogs) qualitatively and quantitatively reflected the major features of the observed distribution with highest abundances in the most oligotrophic (nutrient-depleted) waters (*15*, *17*) (Fig. 2, A to D).

Real-world *Prochlorococcus* exhibit genetic diversity, which leads to differences in light and temperature sensitivities (*17–20*), as well as nitrogen assimilation abilities (*21*). The strains, or ecotypes, of *Prochlorococcus* exhibit distinct patterns of abundance along ocean gradients (*15*, *17*), and observations on AMT13 (*17*) (Fig. 2, E, G, I, and K) provide an ideal test for the stochastic modeling strategy: Do the emergent model analogs of *Prochlorococcus* reflect the geographic distributions, relative abundances, and physiological properties of their real-world counterparts?

Of the *Prochlorococcus* analogs initialized in each model solution, between three and six variants persisted with significant abundances (fig. S4). We grouped the analogs by defining three "model ecotypes" based only on distinct geographic habitats, without regard to physiology, which had a qualitative resemblance to the observed distributions of ecotypes along AMT13. In any ensemble member, more than one emergent *Prochlorococcus* analog may fall into a particular model-ecotype classification, and some were ambiguous. Model ecotype *m-e1* (Fig. 2F) was defined to include emergent analogs with significant biomass in the upper 25 m along the transect between 15°N and 15°S, qualitatively corresponding to the habitat of real-world ecotype eMIT9312 (Fig. 2E). Model ecotype *m-e2* (Fig. 2H) included analogs that had significant biomass in surface waters polewards of 15° but low biomass within 15° of the equator, broadly reflecting eMED4 (Fig. 2G). Finally, model ecotype *m-e3* (Fig. 2J) was defined to include analogs that had a subsurface maximum biomass, in common with eMIT9313 and eNATL2A (Fig. 2, I and K). The observed widespread distribution of deep maxima with low abundance associated with eMIT9313 and eNATL2A was not clearly reflected in the model analogs. This might be explained by the tendency toward unrealistically complete competitive exclusion typical in ecosystem models (*22*, *23*), precluding persistent populations at low abundance. There is a deep, high biomass layer in the model made up of other, nitrate-consuming, small phytoplankton. This may partially reflect a contribution from nitrate-utilizing *Prochlorococcus*, which have recently been inferred from ocean observations (*24*), but which have not yet been seen in culture.
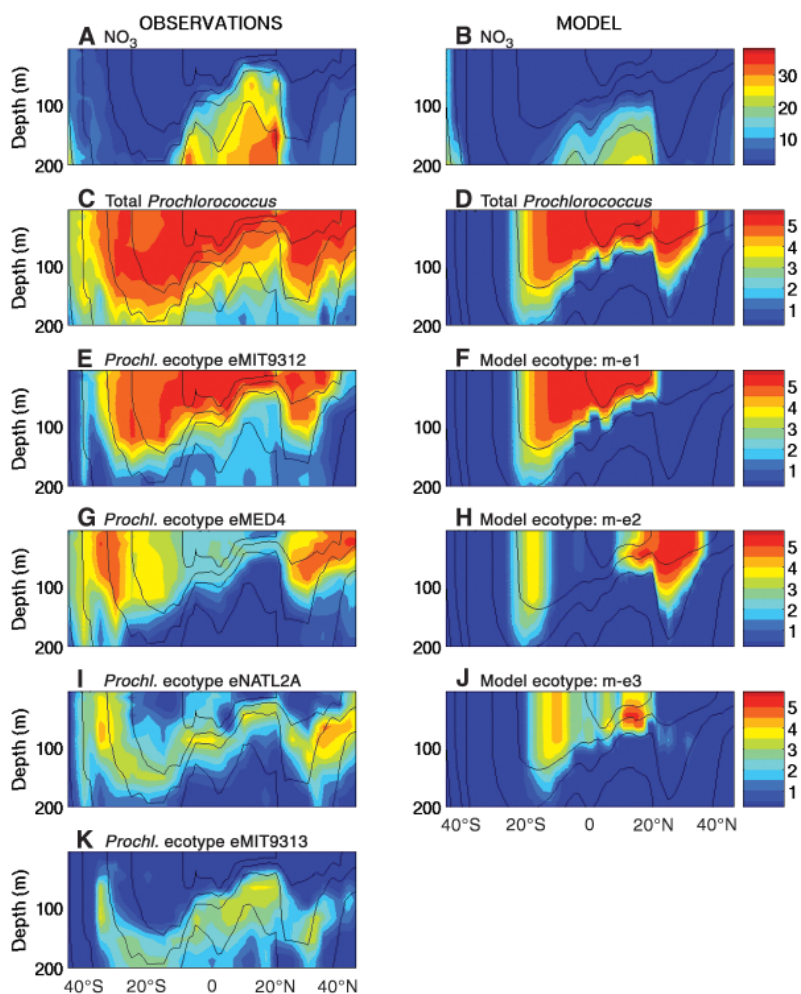
Fig. 2. Observed and modeled properties along the AMT13 cruise track. Left column shows observations (*17*), right column shows results from a single model integration. (**A** and **B**) Nitrate (μmol kg^{-1}); (**C** and **D**) total *Prochlorococcus* abundance [log (cells ml^{-1})]. (**E**, **G**, **I**, and **K**) Distributions of the four most abundant *Prochlorococcus* ecotypes [log (cells ml^{-1})] ranked vertically. (**F**, **H**, and **J**) The three emergent model ecotypes ranked vertically by abundance. Model *Prochlorococcus* biomass was converted to cell density assuming a quota of 1 fg P cell^{-1} (*27*). Black lines indicate isotherms.
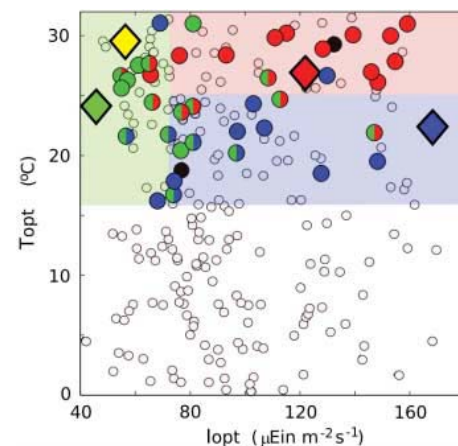
Fig. 3. Optimum temperature and light intensity for growth, T_{opt} and I_{opt}, of all initialized *Prochlorococcus* analogs (all circles) from the ensemble of 10 model integrations. Large circles indicate the analogs that exceeded a total biomass of 10^6 mol P along AMT13 in the 10th year. Colors indicate classification into model ecotypes (see main text): Red circles, *m-e1*; blue circles, *m-e2*; green circles, *m-e3*. Mixed-color and solid black circles denote ambiguity in model-ecotype classification. Bold diamonds indicate real-world *Prochlorococcus* ecotypes (red, eMIT9312; blue, eMED4; green, eNATL2A; and yellow, eMIT9313).

Within each ensemble member, emergent model ecotypes typically followed the abundance ranking of their geographically identified real-world counterparts (Fig. 2 and fig. S4): Model ecotypes *m-e1* and *m-e2* ranked first and second (compare these with eMIT9312 and eMED4, respectively), with *m-e3* consistently at lower abundances (compare this with ecotypes eNATL2A and eMIT9313).

There is a simultaneous correspondence between the physiological characteristics of emergent, modeled ecotypes and cultured representatives of the wild population. Each cultured strain of *Prochlorococcus* and the emergent model ecotypes from all 10 ensemble members were characterized by an optimal temperature (T_{opt}) and photon flux (I_{opt}) for growth, the temperature or light intensity at which growth rates are greatest if all other limitations are set aside (fig. S1). Potentially viable *Prochlorococcus* analogs were seeded in the model over wide ranges of optimal temperature and photon fluxes (all circles, Fig. 3), but those that maintained significant abundances along the AMT transect (solid large circles, Fig. 3) were all characterized by $T_{opt} > 15°C$. This is consistent with the observations of *Prochlorococcus* in warmer waters and with the warm T_{opt} of cultured strains (*17*). Our model indicates that the oligotrophic conditions confined *Prochlorococcus* analogs to warmer waters and selected for warm T_{opt}, an emergent "adaptation" driven by other environmental factors. In the cooler waters of the model, nutrients are typically abundant, and so larger phytoplankton, with higher intrinsic maximum growth rates, have an advantage. In the highly oligotrophic (typically warmer) regions, the *Prochlorococcus* analogs' lower half-saturation (consistent with their very small size) is advantageous.

Across the ensemble of 10 integrations, the geographically defined model ecotypes were clustered in optimal temperature and light parameter space (Fig. 3): Model ecotype *m-e1* (red circles) generally occupied the warmest area of parameter space over a broad, upper range of optimal photon fluxes; *m-e2* (blue circles) generally had a lower T_{opt} but a similar range of I_{opt}. This is consistent with their surface-oriented habitats and latitudinal (or temperature) separation. In contrast, *m-e3* (green circles) occupied a wider range of T_{opt} but only in the region of lowest I_{opt}, consistent with its expression of subsurface maxima. Although there were exceptions, the clustering of geographically defined model ecotypes in physiological parameter space indicated that robust ecological controls were operating across the 10 integrations. The physiological characteristics (T_{opt}, I_{opt}) of real-world ecotypes (colored diamonds, Fig. 3) are notably consistent with the grouping of their model counterparts. This correspondence was not imposed, but emerged as a feature of the model solution.

Significantly, there was simultaneous consistency between the geographical habitat, rank abundance, and physiological specialization of the emergent *Prochlorococcus* model ecotypes and their real-world counterparts. These parallels indicate that the stochastic, self-organizing representation of marine ecosystems reflects real-world processes and is suitable for application in ecological and biogeochemical studies. This approach circumvents some of the obstacles facing most current ocean ecosystem models, such as the a priori imposition of low diversity, the prescription of dominant functional types, and the difficulty of specifying the physiological rate coefficients that define them. This function-based approach can naturally evolve to exploit the growing body of genomic and metagenomic data mapping the oceans in terms of genes and their encoded physiological functionality (*25*, *26*). Finally, because the ecosystem structure and function are, by design, emergent and not tightly prescribed, this modeling approach is ideally suited for studies of the relations between marine ecosystems, evolution, biogeochemical cycles, and past and future climate change.

References and Notes
1. D. Tilman, *Ecology* **58**, 338 (1977).
2. R. Margalef, *Perspectives in Ecological Theory* (Univ. of Chicago Press, Chicago, 1968).
3. C. Pedrós-Alió, *Trends Microbiol.* **14**, 257 (2006).
4. S. L. Pimm, J. H. Lawton, *Nature* **268**, 329 (1977).
5. A. Kleidon, H. A. Mooney, *Glob. Change Biol.* **6**, 507 (2000).
6. J. K. Moore, S. Doney, J. Kleyplas, D. Glover, I. Fung, *Deep-Sea Res. II* **49**, 403 (2001).
7. W. W. Gregg, P. Ginoux, P. S. Schopf, N. W. Casey, *Deep-Sea Res. II* **50**, 3143 (2003).
8. E. Litchman, C. A. Klausmeier, J. R. Miller, O. M. Schofield, P. G. Falkowski, *Biogeosciences* **3**, 585 (2006).
9. C. LeQuere *et al.*, *Glob. Change Biol.* **11**, 2016 (2006).
10. T. R. Anderson, *J. Plankton Res.* **27**, 1073 (2005).
11. R. R. Hood *et al.*, *Deep-Sea Res. II* **53**, 459 (2006).
12. P. A. Thompson *et al.*, *Limnol. Oceanogr.* **34**, 1014 (1989).
13. C. Wunsch, P. Heimbach, *Physica D* 10.1016/j.physd.2006.09.040 (2006).
14. Y. Dandonneau, Y. Montel, J. Blanchot, J. Giraudeau, J. Neveux, *Deep-Sea Res. I* **53**, 689 (2006).
15. M. V. Zubkov, M. A. Sleigh, P. H. Burkill, R. J. G. Leakey, *Prog. Oceanogr.* **45**, 369 (2000).
16. H. A. Bouman *et al.*, *Science* **312**, 918 (2006).
17. Z. I. Johnson *et al.*, *Science* **311**, 1737 (2006).
18. L. R. Moore, G. Rocap, S. W. Chisholm, *Nature* **393**, 464 (1998).
19. G. Rocap *et al.*, *Nature* **424**, 1042 (2003).
20. L. R. Moore, S. W. Chisholm, *Limnol. Oceanogr.* **44**, 628 (1999).
21. L. R. Moore, A. F. Post, G. Rocap, S. W. Chisholm, *Limnol. Oceanogr.* **47**, 989 (2002).
22. G. E. Hutchinson, *Am. Nat.* **95**, 137 (1961).
23. R. A. Armstrong, R. McGehee, *Am. Nat.* **115**, 151 (1980).
24. M. W. Lomas, F. Lipschultz, *Limnol. Oceanogr.* **51**, 2453 (2006).
25. J. C. Venter *et al.*, *Science* **304**, 66 (2004).
26. E. F. DeLong *et al.*, *Science* **311**, 496 (2006).
27. S. Bertilsson, O. Berglund, D. M. Karl, S. W. Chisholm, *Limnol. Oceanogr.* **48**, 1721 (2003).
28. Thanks to J. Marshall, R. Williams, P. Falkowski, J. Cullen, and J. Bragg for inspiration and encouragement. Thanks also to M. Coleman, R. Hood, and three anonymous reviewers for stimulating comments on the manuscript; to C. Hill for computing guidance; and to P. Heimbach, C. Wunsch, and the ECCO group for ocean circulation state estimates. We are grateful for funding from the PARADIGM consortium of the National Ocean Partnership Program, NSF (M.J.F., S.D.), NSF, DOE (S.W.C.), and the Gordon and Betty Moore Foundation (S.W.C., M.J.F.). M.J.F. is also grateful for the MIT Global Habitat Longevity Award. We acknowledge the Atlantic Meridional Transect consortium (NER/O/S/2001/00680), which enabled the biogeographical observations first published in (*17*) (AMT contribution no. 107).