Emergent Biogeography of Microbial Communities in a Model Ocean
Michael J. Follows, et al.
Science 315, 1843 (2007);
DOI: 10.1126/science.1138544

The following resources related to this article are available online at www.sciencemag.org (this information is current as of April 25, 2007):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:
http://www.sciencemag.org/cgi/content/full/315/5820/1843

Supporting Online Material can be found at:
http://www.sciencemag.org/cgi/content/full/315/5820/1843/DC1

This article cites 8 articles, 5 of which can be accessed for free:
http://www.sciencemag.org/cgi/content/full/315/5820/1843#otherarticles

This article appears in the following subject collections:
Ecology
http://www.sciencemag.org/cgi/collection/ecology

Information about obtaining reprints of this article or about obtaining permission to reproduce this article in whole or in part can be found at:
http://www.sciencemag.org/about/permissions.dtl
Emergent Biogeography of Microbial Communities in a Model Ocean

Michael J. Follows,1,* Stephanie Dutkiewicz,1 Scott Grant,1,2 Sallie W. Chisholm3

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.

References and Notes
21. The authors wish to thank members of the Desplan and Small laboratories for support and advice. This project was supported by NIH grants GM64864, awarded to C.D., and GM51946, awarded to S.S. A.E.B is a Damon Runyon Fellow, supported by the Damon Runyon Cancer Research Foundation (DRG-1870-05).

Supporting Online Material
www.sciencemag.org/cgi/content/full/315/5820/1841/DC1
Materials and Methods
SOM Text
Fig. S1
References
13 November 2006; accepted 7 March 2007
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 utilizing 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
Distributions of the four most abundant Prochlorococcus lines indicate isotherms. (observations (Fig. 2). Observed and modeled properties along the AMT13 cruise track. Left column shows 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 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 10¹⁶ 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 temperature or light intensity at which growth rates are significant abundances along the AMT transect (solid circles, Fig. 3), but those that maintained significant optima of optimal temperature and photon fluxes (all logs 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 $T_{opt} > 15^\circ$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.

### Cascading Effects of the Loss of Apex Predatory Sharks from a Coastal Ocean

Ransom A. Myers,¹ Julia K. Baum,¹ Travis D. Shepherd,¹ Sean P. Powers,² Charles H. Peterson³*

Impacts of chronic overfishing are evident in population depletions worldwide, yet indirect ecosystem effects induced by predator removal from oceanic food webs remain unpredictable. As abundances of all 11 great sharks that consume other elasmobranchs (rays, skates, and small sharks) fell over the past 35 years, 12 of 14 of these prey species increased in coastal northwest Atlantic ecosystems. Effects of this community restructuring have cascaded downward from the cownose ray, whose enhanced predation on its bay scallop prey was sufficient to terminate a century-long scallop fishery. Analogous top-down effects may be a predictable consequence of eliminating entire functional groups of predators.

Ecological impacts of eliminating top predators can be far-reaching (1) and include release of mesopredator prey populations from predatory control (2) and induction of subsequent cascades of indirect trophic interactions (3–5). In the oceans, fishing has dispropor-
Supporting Online Material for

Emergent Biogeography of Microbial Communities in a Model Ocean

Michael J. Follows,* Stephanie Dutkiewicz, Scott Grant, Sallie W. Chisholm

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

DOI: 10.1126/science.1138544

This PDF file includes

Materials and Methods
SOM Text
Figs. S1 to S4
Table S1
References
Supporting Online Material

METHODS:

S1. Ecosystem Model Algorithms.

We formulate the ecosystem model in a generalized framework which represents an arbitrary
number of nutrients, $N_i$, phytoplankton types, $P_j$, and grazers, $Z_{ki}$. Each nutrient element also has
an associated particulate organic and dissolved organic matter pool, $POM_i$ and $DOM_i$,
respectively. The rates of change of these prognostic variables are described by the following set
of equations:

\begin{align}
\frac{\partial N_i}{\partial t} + \nabla \cdot (u N_i) &= \nabla \cdot \left( \kappa \nabla N_i \right) - \sum_j \left[ \mu_j \gamma_j^N \gamma_j^N P_j R_{ij} \right] + S_i \\
\frac{\partial P_j}{\partial t} + \nabla \cdot (u P_j) &= \nabla \cdot \left( \kappa \nabla P_j \right) + \mu_j \gamma_j^P \gamma_j^P P_j - \frac{\partial (w_{Pj}^P P_j)}{\partial z} - m_j^P P_j - \sum_k g_{jk} \frac{P_j}{P_j + k_j^P} Z_{ki} \\
\frac{\partial Z_{ki}}{\partial t} + \nabla \cdot (u Z_{ki}) &= \nabla \cdot \left( \kappa \nabla Z_{ki} \right) + Z_{ki} \sum_j g_{jk} R_{ij} \frac{P_j}{P_j + k_j^P} - m_i^Z Z_{ki} \\
\frac{\partial POM_i}{\partial t} + \nabla \cdot (u POM_i) &= \nabla \cdot \left( \kappa \nabla POM_i \right) - r_{POM} POM_i - \frac{\partial (w_{POM} POM_i)}{\partial z} + S_{POM} \\
\frac{\partial DOM_i}{\partial t} + \nabla \cdot (u DOM_i) &= \nabla \cdot \left( \kappa \nabla DOM_i \right) - r_{DOM} DOM_i + S_{DOM}
\end{align}

Symbols are defined in the text below and parameter values or ranges are provided in Table S1.
Units are $\mu M$ P for Eq. S2, and $\mu M$ P, $\mu M$ N, $\mu M$ Si, or $\mu M$ Fe (element represented by subscript
$i$) for Eqs. S1, S3, S4 and S5. Here $R_{ij}$ denotes the ratio of element, $i$, relative to phosphorus, for
each phytoplankton type, $j$. Separate zooplankton pools are carried for each element, $Z_{ki}$, where $k$
is the zooplankton type and $i$ the nutrient element, accounting for the ingestion of prey with
different elemental ratios. Subscript $i=1$ refers to phosphorus.

Tracers are transported by the currents, $u$, and mixing coefficients, $\kappa$, from the ECCO
(“Estimating the Circulation and Climate of the Ocean”) state estimate of ocean circulation ($SI$)
based on a moderate resolution ($1^\circ \times 1^\circ$, 23 vertical levels), global configuration of the MIT ocean
circulation model (S2) constrained to be consistent with observations of large-scale hydrography and altimetry. Nutrient distributions are initialized from observed climatologies (S3) or previous simulations (S4).

**S1.1 Parameterizations of Phytoplankton Physiology**

While the approach to the organization and complexity of the ecosystem model are novel, the idealized descriptions of phytoplankton physiological processes are similar to those applied in previous studies (S4-S7). Phytoplankton growth is determined by a maximum intrinsic growth rate, \( \mu_j \), modulated by non-dimensional factors which reflect sensitivities to ambient temperature, photon flux and essential nutrients (Fig. S1). Nutrient limitation of growth is determined by the most limiting resource,

\[
\gamma^N_j = \varphi \min (N_1^{lim}, N_2^{lim}, ...)
\]

(S6)

where the nutrients considered are phosphate, iron, silicic acid and nitrate, nitrite and ammonia. The effect on growth rate of ambient phosphate, iron or silicic acid concentrations is represented by a Michaelis-Menton function

\[
N_i^{lim} = \frac{N_i}{N_i + k_{ij}}
\]

(S7)

where the \( K_{ij} \) are half-saturation constants for phytoplankton type \( j \) with respect to the ambient concentration of nutrient \( i \) (Fig. S1C). We resolve three potential sources of inorganic nitrogen (ammonia, nitrite and nitrate) though modeled phytoplankton may be able to assimilate ammonia only, ammonia and nitrite, or all three (S8). Since it is energetically more expensive to utilize nitrate relative to the other sources we represent nitrogen limitation by the following function:

\[
N_N^{lim} = \frac{NO_3}{NO_3 + k_{NO3j}} e^{-\psi NH4 - \psi NO2} + \frac{NO_2}{NO_2 + k_{NO2j}} e^{-\psi NH4} + \frac{NH_4}{NH_4 + k_{NH4j}}
\]

(S8)

where \( \psi \) reflects the inhibition of nitrate or nitrite uptake (S9). Growth rate is enhanced when utilizing only ammonia, or ammonia and nitrite:

\[
\varphi = (v + (1-v)(NO_2^{lim} + NH_4^{lim})/N_N^{lim})
\]

(S9)

where \( NO_2^{lim} \) and \( NH_4^{lim} \) represent the second and third terms on the right of Eq. S8. A phytoplankton type utilizing only nitrate thus has growth rate reduced by a factor \( v \) relative to one using no nitrate (S10).
Temperature modulation of growth is represented by a non-dimensional factor

\[ \gamma_j^T = \frac{1}{\tau_1} \left( A^T e^{-B(T-T_0)^C} - \tau_2 \right) \]  
(S10)

which sets a temperature range over which each phytoplankton type can grow efficiently (Fig. S1A), and there is a general decrease in growth efficiency with temperature (SI1). Coefficients \( \tau_1 \) and \( \tau_2 \) normalize the maximum value, while \( A, B, T_0, \) and \( C \) regulate the form of the sensitivity envelope.

We incorporate a very simple radiative transfer model (S4) which captures self-shading but does not resolve spectral bands. The light sensitivity of growth rate is parameterized using the function (SI2):

\[ \gamma_j^l = \frac{1}{F_{\text{max}}} \left(1 - e^{-k_{\text{PAR}j}} \right) e^{-k_{\text{ss}ad}l} \]  
(S11)

where \( I(z) \) is the local, vertical flux of photosynthetically active radiation, PAR, and

\[ F_{\text{max}} = \frac{k_{\text{PAR}} + k_{\text{inhib}}}{k_{\text{PAR}}} \exp \left( - \frac{k_{\text{inhib}}}{k_{\text{PAR}}} \ln \left( \frac{k_{\text{inhib}}}{k_{\text{PAR}} + k_{\text{inhib}}} \right) \right) \]

is chosen to normalize the maximum value of \( \gamma_j^l \) to 1 (Fig. S1B). The parameter \( k_{\text{PAR}} \) defines the increase of growth rate with light at low levels of irradiation while \( k_{\text{inhib}} \) regulates the rapidity of the decline of growth efficiency at high PAR, or photo-inhibition (SI2). This highly idealized parameterization of light sensitivity captures variations in optimal light intensity, and their ecological implications, but does not explicitly account for photo-acclimation, differences in accessory pigments and other factors which might lead to variability in the maximum light dependent growth factor. We note that, while the function \( \gamma_j^l \) is normalized to a maximum value of 1 for all phytoplankton types, large size-class phytoplankton are given a higher maximum intrinsic growth rate, \( \mu_j \).

We impose fixed elemental ratios for each phytoplankton type, \( R_{ij} \), though these may vary between types (e.g. some require silica while others do not). To restrict the niche dimension and computational expense of this initial study, we have imposed an average, Redfieldian N:P stoichiometry of 16:1 for all phytoplankton types. We note that in nature elemental ratios are flexible and Prochlorococcus, for example, can significantly exceed this value (SI3). Formulating the model with dynamic nutrient quotas would capture flexible stoichiometry and is more physiologically appropriate (SI4,S15) but also would significantly increase the number of three-
dimensional arrays required to describe each phytoplankton type, dramatically increasing the computational expense. Hence we have not used this approach in this initial illustration.

**S1.2 Assignment of Physiological Functionality and Growth Rate Sensitivities.**

At the heart of this modeling strategy is the self-organization of a stochastically generated phytoplankton community. The physiological functionality and sensitivity of growth to temperature, light and ambient nutrient abundance for each modeled phytoplankton type is governed by several true/false parameters, the values of which are based on a virtual “coin-toss” at the initialization of each phytoplankton type. These determine the size class of each phytoplankton type (“large” or “small”), whether the organism can assimilate nitrate, whether the organism can assimilate nitrite, and whether the organism requires silicic acid. Parameter values which regulate the effect of temperature, light and nutrient availability on growth, are then assigned stochastically. $T_o$, which controls the optimum temperature for growth, and $K_{PO4}$, the phosphate half-saturation coefficient (to which other half-saturations are indexed by the fixed elemental ratios), are drawn from prescribed ranges using a random number generator. Values for $k_{par}$ and $k_{inhib}$ are also randomly chosen, drawn from prescribed normal distributions. Some simple allometric trade-offs are imposed (Fig. S1): Phytoplankton in the large size class are distinguished by higher intrinsic maximum growth rates and faster sinking speeds (S16). They also draw parameter values from distributions with higher nutrient half-saturations (assuming they are less efficient at acquiring nutrients, S17) and are assumed to be high-light adapted due to packaging effects (S18, S19). These trade-offs are implemented by randomly selecting parameter values from different (though overlapping) distributions for large and small phytoplankton.

We note that, since the values of the governing coefficients are initialized stochastically from given distributions rather than prescribed specifically for each phytoplankton functional type, the total number of externally prescribed parameters in this approach (Table S1) is the same whether 10 or 10,000 phytoplankton types are initialized. The diversity of the “successful” population, and the parameter values that govern those organisms, are self-selected during the initial adjustment of the ecosystem model.

**S1.3 Grazing, Mortality, Remineralization and Biogeochemical Cycles.**

Parameterizations of grazing and other forms of heterotrophy are simplified in this study, which focuses on complexity and selection in the photo-autotrophs. None of the parameters regulating grazing and remineralization processes are stochastic in the simulations presented here. We prescribe a simple grazer community with two size classes. Large zooplankton preferentially graze ($g_{fast}$) on large phytoplankton, but can graze on small phytoplankton ($g_{slow}$) and visa versa for small zooplankton. A half-saturation coefficient ($K^P$) regulates grazing efficiency at high prey concentrations. Excretion and non-grazing mortality are represented as linear loss terms for both phytoplankton and grazers, with coefficients $m^P$ and $m^z$ respectively. This simplified, low diversity grazer community is chosen to facilitate a computationally and intellectually tractable study in this initial illustration. Future studies should examine, for example, a greater diversity of grazers with a variety of stochastically appointed feeding strategies broadening the general strategy to include the next trophic level.

The term $S_i$ (Eq. S1) represents the source of inorganic nutrient due to the remineralization of organic forms as well as external sources and non-biological transformations (S4,S17). Heterotrophic microbes are not explicitly represented and the remineralization of dissolved and particulate organic detritus pools is treated as a simple linear decay with respective prescribed timescales $1/r_{POM}$ and $1/r_{DOM}$ (S4). $S_{POM}$ (Eq. S4) and $S_{DOM}$ (Eq. S5) are the sources of particulate and dissolved organic detritus arising from mortality and excretion of all phytoplankton types and
grazers (in Eq. S2 and S3), closing the nutrient budgets. Here we simply define a fixed fraction ($f_{DOM}$) of mortality and excretion to pass into each organic detritus pool, assuming that large phytoplankton and zooplankton contribute a larger fraction of their detritus to the $POM_i$ pool than do the small phytoplankton. All silica is assumed to go to a POM pool, there is no dissolved organic silica.

The remineralization of organic phosphorus and iron produce phosphate and dissolved iron respectively, while the remineralization of organic nitrogen is assumed to produce ammonia which may then be nitrified to nitrite and, subsequently, nitrate. The microbial process of nitrification is also treated simply as first order reactions with fixed rate coefficients ($\zeta_{NO2}$, $\zeta_{NO3}$) resulting in qualitatively reasonable distributions of the nitrogen species. Due to the relatively short timescale of the integrations and to restrict the complexity of this initial study we do not represent diazotrophy. Simplified one dimensional studies indicate that enabling diazotrophy as a possible functionality for the modeled phytoplankton types enhances the availability of more reduced forms of nitrogen in the subtropical regions resulting in an increase the abundance of Prochlorococcus analogs.

Iron chemistry in seawater is parameterized (S20) with a complexation to an organic ligand (binding strength, $\beta_{Fe}$) and scavenging to falling particles (rate, $c_{Fe}$). Dust (S21) deposited in the surface (solubility, $\alpha_{Fe}$) is a source of iron.

**SUPPORTING TEXT**

**S2. Supplementary Model Results.**

An ensemble of model integrations was performed, each with a different randomization of physiological characteristics but identical initialization and physical environment. 78 phytoplankton types were initialized in each integration: Experimentation suggested that the modeled community structure would be less robust with fewer than 30, and practical computational considerations placed an upper limit at 78. Computational cost also limited the ensemble to only 10 members. Fig. S2 shows the annual mean concentration, at year 10, of phosphorus in biomass of the 78 phytoplankton from a single ensemble member. All ensemble members exhibit a similar set of occupied habitats which are collectively reminiscent of the previously proposed biogeographical provinces (S22). All ensemble members produce very similar total primary production and nutrient fields (shown for one member in Fig. S3), and these compare favorably to observations. The similarity in the total primary production reflects the significant regulation of physical nutrient supply and light on gyre and basin scales.

The general biogeography of the model (depicted for a single ensemble member in Fig. 1B and Fig. S2) is robust between ensemble members. While various categorizations of “types” into functional groups might be considered, the classification here (Fig. 1B) reflects groupings of general interest and is tailored to reflect our particular interest in Prochlorococcus. In general, the habitats of the emergent Prochlorococcus-analogs bear some qualitative resemblance to those observed but are much more sharply defined (Fig. 2, Fig. S4). Indeed, very low background abundances and sharply defined habitats of all the abundant, modeled phytoplankton types suggest that the model ecosystem is closer to complete competitive exclusion than is the real world (S23). This may reflect the relatively small number of
physiological specializations (niche dimensions) in the model, the comparatively smooth, coarse
resolution, physical environment (S24) or the low diversity of predatorial strategies (S23).

Though each of the ten members of the ensemble of solutions are initialized with different
randomization of the characteristics of the phytoplankton population, the emergent community
structures and biogeography are relatively robust. For example, in each solution the four most
abundant, emergent Prochlorococcus-analogs are relatively consistent (Fig. S3): the most
abundant is typically of m-e1 classification and the second most abundant typically m-e2, with m-
e3 type analogs at lower abundances. Although our model does not exhibit a significant deep (low
light) biomass of Prochlorococcus-analogs (Fig. S4), there is a deep biomass maximum at the
nutricline in the equatorial regions, comprised of “other small phytoplankton” types. Some of the
phytoplankton types which make up this deep maximum might represent nitrate consuming
Prochlorococcus strains which have been suggested from field observations (S25) but not yet
cultured. Such organisms, though present in the model, are not classified as Prochlorococcus in
our rather crude definition of functional groups.
**Fig. S1** Functional forms of the sensitivity of phytoplankton growth to (A) temperature, (B) flux of photosynthetically active radiation, and (C) ambient phosphate concentration expressed as normalized, non-dimensional growth factors, $\gamma_j$, which modulate the maximum intrinsic growth rate. The collection of curves in each panel is chosen to illustrate the ranges from initialized sensitivities are selected. Simple allometric trade-offs are indicated by the different ranges for the small phytoplankton class (blue curves) and large phytoplankton class (red curves). The highly idealized parameterization of light sensitivity captures variations in optimal light intensity but does not explicitly represent variability in the maximum light dependent growth factor. However, larger phytoplankton are given a higher intrinsic growth rate, $\mu_j$. Optimal temperature and light intensity for growth, $T_{opt}$ and $I_{opt}$, are illustrated for a single phytoplankton type (dashed black curves).
Fig. S2. Phytoplankton abundance (μM P; average 0-50m, logarithmic color-scale) for each of 78 initialized types in a single ensemble member. Annual mean of tenth year of integration.
Fig S3: Comparison of one ensemble member annual (0-50m) fields (right column) to observations (left column). (A,B) Primary Production (gC/m²/y); (C,D) Phosphate (µM P); (E,F) Nitrate (µM N); (G,H) Silicic Acid (µM Si). Observational euphotic layer primary production was calculated for 2005 using the Vertically Generalized Productivity Model (S26) and SeaWiFS-derived Chl. Data for this panel was downloaded from http://science.oregonstate.edu/ocean.productivity. Observational nutrients are from climatology of in situ data (S3) and are averaged over 0-50m.
Fig. S4. The four most abundant *Prochlorococcus*-analogs (log(cells ml\(^{-1}\))) for the month of September along the AMT13 track from four of the ten member ensemble of integrations. “Type” number indicates the numerical designation of each of the 78 stochastically initialized phytoplankton types in each ensemble member. Analogs are classified into model-ecotypes as described in the main text. Model biomass is converted to cell density assuming a nominal phosphorus quota of 1 fg cell\(^{-1}\) for *Prochlorococcus* (13). Black contours are isotherms.
### Table S1: Parameters of the ecosystem model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Symbol</th>
<th>Fixed Value</th>
<th>Range</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum phytoplankton growth rate</td>
<td>$\mu$</td>
<td>Small: 1.4</td>
<td>Large: 2.2</td>
<td>d$^{-1}$</td>
</tr>
<tr>
<td>Phytoplankton mortality rate</td>
<td>$m^p$</td>
<td>Small: 0.1</td>
<td>Large: 0.1</td>
<td>d$^{-1}$</td>
</tr>
<tr>
<td>PAR saturation coefficient</td>
<td>$k_{sat}$</td>
<td>Small: mean 0.012, std 0.01 Large: mean 0.012, std 0.003</td>
<td>($\mu$Ein m$^{-1}$ s$^{-1}$)$^{-1}$</td>
<td></td>
</tr>
<tr>
<td>PAR inhibition coefficient</td>
<td>$k_{inhib}$</td>
<td>Small: mean 6<em>10$^{-3}$, std 1</em>10$^{-4}$ Large: mean 1<em>10$^{-3}$, std 5</em>10$^{-5}$</td>
<td>($\mu$Ein m$^{-1}$ s$^{-1}$)$^{-1}$</td>
<td></td>
</tr>
<tr>
<td>Temperature curve coefficient</td>
<td>$A$</td>
<td>1.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Temperature optimum coefficient</td>
<td>$T_o$</td>
<td>-2 to 30</td>
<td></td>
<td>°C</td>
</tr>
<tr>
<td>Temperature range coefficient</td>
<td>$B$</td>
<td>Small: 1*10$^{-3}$</td>
<td>Large: 3*10$^{-5}$</td>
<td>°C$^{-1}$</td>
</tr>
<tr>
<td>Temperature decay coefficient</td>
<td>$C$</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Temperature normalization coefficients</td>
<td>$\tau_1, \tau_2$</td>
<td>0.33, 0.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phosphate half saturation</td>
<td>$K_{PO4}$</td>
<td>Small: 1.35<em>10$^{-2}$ to 3.5</em>10$^{-2}$ Large: 3.5<em>10$^{-2}$ to 5.5</em>10$^{-2}$</td>
<td>μM P</td>
<td></td>
</tr>
<tr>
<td>Nitrate half saturation</td>
<td>$K_{NO3}$</td>
<td>Small: 0.24 to 0.56</td>
<td>Large: 0.56 to 0.88</td>
<td>μM N</td>
</tr>
<tr>
<td>Nitrite half saturation</td>
<td>$K_{NO2}$</td>
<td>Small: 0.16 to 0.42</td>
<td>Large: 0.42 to 0.66</td>
<td>μM N</td>
</tr>
<tr>
<td>Ammonium half saturation</td>
<td>$K_{NH4}$</td>
<td>Small: 4.3*10$^{-2}$ to 0.112 Large: 0.112 to 0.132</td>
<td>μM N</td>
<td></td>
</tr>
<tr>
<td>Silicic acid half saturation</td>
<td>$K_{Si}$</td>
<td>Non-diatom: 0</td>
<td>Diatom: 2</td>
<td>μM Si</td>
</tr>
<tr>
<td>Iron half saturation</td>
<td>$K_{Fe}$</td>
<td>Small: 1.7<em>10$^{-5}$ to 4.4</em>10$^{-5}$ Large: 4.4<em>10$^{-5}$ to 6.9</em>10$^{-5}$</td>
<td>μM Fe</td>
<td></td>
</tr>
<tr>
<td>Phytoplankton elemental ratios</td>
<td>$R_{Si:P}$</td>
<td>16</td>
<td>1.25*10$^{-3}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$R_{N:P}$</td>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$R_{Fe:P}$</td>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ammonia/nitrite inhibition</td>
<td>$\psi$</td>
<td>4.6</td>
<td></td>
<td>($\mu$M N)$^{-1}$</td>
</tr>
<tr>
<td>Nitrate consumption cost</td>
<td>$\nu$</td>
<td>0.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phytoplankton sinking rate</td>
<td>$w^p$</td>
<td>Small: 0</td>
<td>Large: 0.5</td>
<td>m d$^{-1}$</td>
</tr>
<tr>
<td>Phytoplankton partitioning DOM/POM</td>
<td>$f_{DOM}$</td>
<td>Small: 0.2</td>
<td>Large: 0.5</td>
<td></td>
</tr>
<tr>
<td>Parameter</td>
<td>Value</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------------------------------------------</td>
<td>------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zooplankton fast grazing rate $g_{fast}$</td>
<td>0.2 d$^{-1}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zooplankton slow grazing rate $g_{slow}$</td>
<td>0.033 d$^{-1}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zooplankton mortality rate $m^z$</td>
<td>0.033 d$^{-1}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phytoplankton half saturation $K^p$</td>
<td>0.1 $\mu$M P</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DOM remineralization rate $r_{DOP}$, $r_{DON}$, $r_{DOFe}$</td>
<td>2.8*10$^{-3}$ d$^{-1}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>POM remineralization rate $r_{POP}$, $r_{PON}$, $r_{POFe}$, $r_{POSi}$</td>
<td>0.033, 0.033, 3.3*10$^{-3}$ d$^{-1}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>POM sinking rate $w_{POM}$</td>
<td>10 m d$^{-1}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NH$_4$ to NO$<em>2$ oxidation rate $\zeta</em>{NO2}$</td>
<td>0.1 d$^{-1}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NO$_2$ to NO$<em>3$ oxidation rate $\zeta</em>{NO3}$</td>
<td>0.033 d$^{-1}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Iron solubility constant $\alpha_{Fe}$</td>
<td>0.04</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Iron scavenging rate $c_{Fe}$</td>
<td>1.1*10$^{-3}$ d$^{-1}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ligand binding strength $\beta_{Fe}$</td>
<td>2*10$^{5}$ (µM Fe)$^{-1}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PAR attenuation coefficient $k_o$</td>
<td>0.04 m$^{-1}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PAR attenuation coefficient from phytoplankton $k_{phyto}$</td>
<td>0.64 (µM P)$^{-1}$ m$^{-1}$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
SUPPORTING REFERENCES:


Supporting Online Material

www.sciencemag.org

Methods

Supporting Text

Figs. S1, S2, S3, S4

Table S1