Present and future global distributions of the marine Cyanobacteria *Prochlorococcus* and *Synechococcus*

Pedro Flombaum[a,b], José L. Gallegos[a], Rodolfo A. Gordillo[a], José Rincón[a], Lina L. Zabala[b], Nianzhi Jiao[c], David M. Karl[d,1], William K. W. Li[e], Michael W. Lomas[f], Daniele Veneziano[g], Carolina S. Vera[b], Jasper A. Vrugt[a,h], and Adam C. Martiny[a,i,1]

Departments of [a]Earth System Science, [h]Civil Engineering, and [i]Ecology and Evolutionary Biology, University of California, Irvine, CA 92697; [b]Centro de Investigaciones del Mar y la Atmósfera, Departamento de Ciencias de la Atmósfera y los Océanos, and Instituto Franco Argentino sobre Estudios del Clima y sus Impactos, Consejo Nacional de Investigaciones Científica y Tecnológicas and Universidad de Buenos Aires, 1428 Buenos Aires, Argentina; [c]Institute of Microbes and Ecosphere, State Key Lab for Marine Environmental Sciences, Xiamen University, Xiamen 361005, People's Republic of China; [d]Center for Microbial Oceanography: Research and Education (C-MORE), University of Hawaii, Honolulu, HI 96822; [e]Fisheries and Oceans Canada, Bedford Institute of Oceanography, Dartmouth, NS, Canada B2Y 4A2; [f]Bigelow Laboratory for Ocean Sciences, East Boothbay, ME 04544; and [g]Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139

Contributed by David M. Karl, April 25, 2013 (sent for review January 22, 2013)

The Cyanobacteria *Prochlorococcus* and *Synechococcus* account for a substantial fraction of marine primary production. Here, we present quantitative niche models for these lineages that assess present and future global abundances and distributions. These niche models are the result of neural network, nonparametric, and parametric analyses, and they rely on >35,000 discrete observations from all major ocean regions. The models assess cell abundance based on temperature and photosynthetically active radiation, but the individual responses to these environmental variables differ for each lineage. The models estimate global biogeographic patterns and seasonal variability of cell abundance, with maxima in the warm oligotrophic gyres of the Indian and the western Pacific Oceans and minima at higher latitudes. The annual mean global abundances of *Prochlorococcus* and *Synechococcus* are $2.9 \pm 0.1 \times 10^{27}$ and $7.0 \pm 0.3 \times 10^{26}$ cells, respectively. Using projections of sea surface temperature as a result of increased concentration of greenhouse gases at the end of the 21st century, our niche models projected increases in cell numbers of 29% and 14% for *Prochlorococcus* and *Synechococcus*, respectively. The changes are geographically uneven but include an increase in area. Thus, our global niche models suggest that oceanic microbial communities will experience complex changes as a result of projected future climate conditions. Because of the high abundances and contributions to primary production of *Prochlorococcus* and *Synechococcus*, these changes may have large impacts on ocean ecosystems and biogeochemical cycles.

climate change | marine biogeochemistry | microbial biogeography

Marine phytoplankton contribute roughly one-half of the global net primary production and play a key role in regulating global biogeochemical cycles (1). Marine phytoplankton are very diverse (2), including phylogenetic, biochemical, metabolic, and ecological variability (3–6). Thus, understanding the contribution of different phytoplankton groups to ecosystem functioning is central to predicting the biogeochemical impact of future environmental changes (7). However, our limited quantitative understanding of the global distribution and abundance of most phytoplankton groups constrains our ability to incorporate phytoplankton diversity into ocean biogeochemical models.

The marine Cyanobacteria *Prochlorococcus* and *Synechococcus* are abundant in many ocean regions. The known geographical distributions of the two lineages are based primarily on individual cruises or time series observations and secondarily on macroecological descriptions, indicating central tendencies and boundary constraints related to light, temperature, nutrients, and chlorophyll *a* concentration (8–10). *Prochlorococcus* is present from the surface to a depth of ~150 m in the open ocean between 40°N and 40°S. The population size declines beyond these latitudinal limits, and *Prochlorococcus* is thought to be absent at temperatures below 15 °C (11). Furthermore, the lineage is believed to be outcompeted by other phytoplankton in high-nutrient waters (12, 13). *Synechococcus* does not extend as deep in the water column as *Prochlorococcus*, but it has a wider geographical distribution that covers both polar and high-nutrient waters (14). Mortality induced by grazers and viruses balance growth and is, in part, responsible for a marked diel cycle (15, 16); however, the mortality effect on the geographical distribution is unclear. Whereas much knowledge is available about the ecology of *Prochlorococcus* and *Synechococcus*, quantitative niche models have not been developed. Such models are essential for projecting the distribution of *Prochlorococcus* and *Synechococcus* in response to future global environmental changes (17). To identify the present and future ocean distributions, we compiled a global dataset describing the abundance of *Prochlorococcus* and *Synechococcus* and associated environmental variables. This dataset was then used to build a realized niche model for each lineage and address the following questions: (*i*) what are the global quantitative distributions of *Prochlorococcus* and *Synechococcus*, (*ii*) how do environmental factors influence their abundances, and (*iii*) what are the future global distributions as a result of climate change?

Results

Global Distributions of Cyanobacteria. To predict the global distributions of *Prochlorococcus* and *Synechococcus*, we first compiled a global dataset. The dataset consisted of 35,129 and 37,699 observations of abundance, respectively, and contained data from a total of 103 cruises and time series with samples from all major ocean regions (Fig. 1*A* and Table S1). We also compiled additional data on habitat temperature, light [described as photosynthetically active radiation (PAR)], and nutrients (i.e., nitrate and phosphate concentration). This extensive dataset covered nearly the full environmental range of these variables (Fig. 1*B*) as well as their combinations (Fig. S1). We then applied a three-step approach to formulate quantitative niche models (Fig. S2). First, we identified the environmental factors that explained most of the variance using artificial neural network (ANN) models. Temperature and PAR were found to be important for predicting both *Prochlorococcus* and *Synechococcus* (ANN, $R^2_{Pro} = 0.66$ and $R^2_{Syn} = 0.35$), whereas nutrients explained little additional

Author contributions: P.F., J.A.V., and A.C.M. designed research; P.F., D.V., J.A.V., and A.C.M. performed research; N.J., D.M.K., W.K.W.L., M.W.L., D.V., and C.S.V. contributed new reagents/analytic tools; P.F., J.L.G., R.A.G., J.R., L.L.Z., D.V., and A.C.M. analyzed data; and P.F., N.J., D.M.K., W.K.W.L., M.W.L., D.V., J.A.V., and A.C.M. wrote the paper.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

[1]To whom correspondence may be addressed. E-mail: dkarl@hawaii.edu or amartiny@uci.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1307701110/-/DCSupplemental.

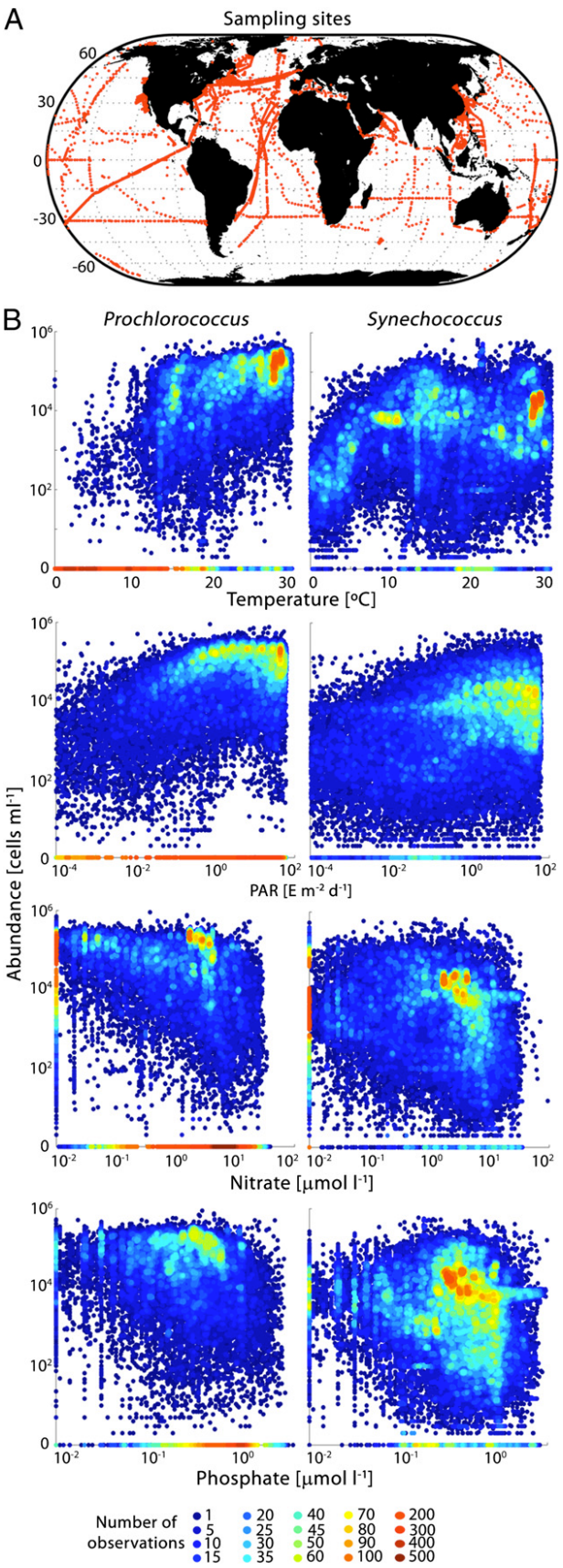


Fig. 1. Observations of *Prochlorococcus* and *Synechococcus* cell abundance. (*A*) Geographic distribution of samples used in this study. (*B*) *Prochlorococcus* and *Synechococcus* as a function of temperature, PAR, nitrate, and phosphate concentration. Symbol color represents the number of overlapping observations.

variance (Table S2). Second, we predicted cell abundance for each combination of temperature and PAR using nonparametric local regression models and found that the performances were similar to the ANN models ($R^2_{Pro} = 0.68$ and $R^2_{Syn} = 0.42$). Third, we closely reproduced the nonparametric predictions using more robust parametric regression models (Fig. S3*A*). The parametric models for *Prochlorococcus* and *Synechococcus* explained a high fraction of the total variances based on temperature and PAR ($R^2_{Pro} = 0.67$ and $R^2_{Syn} = 0.34$) (Table S3). Thus, the variances of the parametric models were comparable with both the ANN and nonparametric models. A subsequent sensitivity analysis using split sampling showed that the model parameters, R^2, and predicted abundance were robust relative to the selection of input data and cruises (Fig. S4 *A*–*C*). No regional bias was detected for either lineage (Fig. S4*D*).

Using the parametric models, we then estimated the global quantitative distributions of *Prochlorococcus* and *Synechococcus*. On average, *Prochlorococcus* was most abundant in the warm oligotrophic waters, especially the Indian and western Pacific Ocean subtropical gyres (maximum = 2.5×10^5 and 2.1×10^5 cells mL^{-1}, respectively) (Fig. 2*A*), and the highest monthly average value estimated was 2.8×10^5 cells mL^{-1}. The mean abundance declined from 4.4×10^4 to 2.8×10^3 cells mL^{-1} from lower latitude waters (30°N to 30°S) to the latitudinal band 30° to 40° (either N or S) and was low at latitudes above 40°. The analysis also suggested lower abundances in areas dominated by cold surface currents and strong upwelling. The median abundance of *Prochlorococcus* in these areas was 1.0×10^4 cells mL^{-1}, about one order of magnitude lower than in the Indian and western Pacific oligotrophic waters. The niche model predicted that *Synechococcus* was also most abundant in the Indian and western Pacific Oceans (maximum = 3.4×10^4 and 4.0×10^4 cells mL^{-1}, respectively), had a peak at midlatitudes, and had low abundance in the Arctic and Southern Oceans (Fig. 2*B*). For this lineage, we observed a much smaller

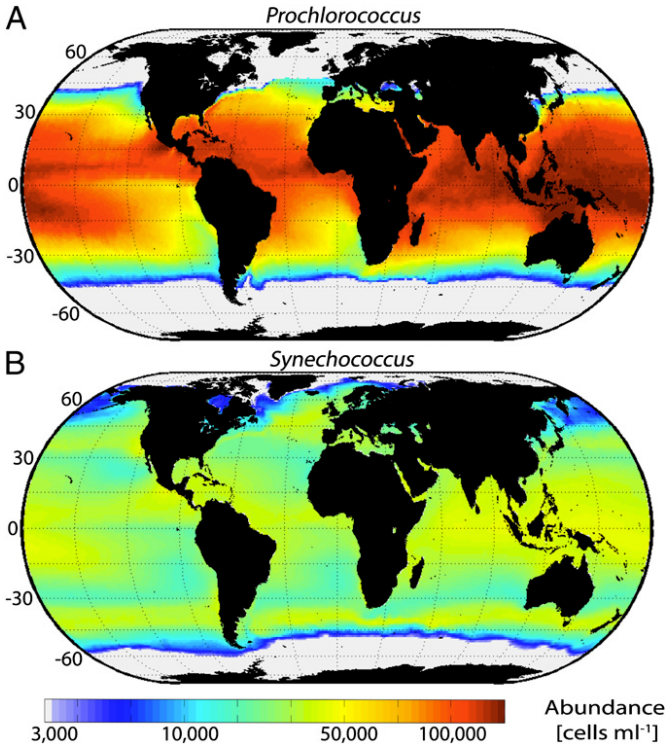


Fig. 2. Present global distribution of *Prochlorococcus* and *Synechococcus* abundance. (*A*) *Prochlorococcus* and (*B*) *Synechococcus* mean annual abundances at the sea surface.

ECOLOGY

decline in abundance in cold current or upwelling regions. We also found strong seasonal variations in both the distributions and abundances of *Prochlorococcus* and *Synechococcus*. The latitudinal range of cell abundance above 10^4 cells mL^{-1} for *Prochlorococcus* shifted 18.0°N and 8.1°S and for *Synechococcus* shifted 23.3°N and 6.0°S (Fig. S5). The global mean annual abundance for *Prochlorococcus* and *Synechococcus* was $2.9 \pm 0.1 \times 10^{27}$ and $7.0 \pm 0.3 \times 10^{26}$ cells, respectively. The highest/lowest global mean monthly abundance was April/August ($3.0/2.7 \times 10^{27}$ cells) for *Prochlorococcus* and March/July ($7.5/6.4 \times 10^{26}$ cells) for *Synechococcus*.

Next, we estimated sea surface abundances and distributions for the end of the 20th and 21st centuries using climate model projections of sea surface temperatures (18). The projection for the 21st century used the Representative Concentration Pathway (RCP) 4.5 scenario, which was based on a stabilization of the radiative forcing of greenhouse gases at what is equivalent to 650 ppm CO_2. Because of the rise in temperature, the multimodel ensemble mean projected a 28.7% and 13.9% increase in *Prochlorococcus* and *Synechococcus* abundance, respectively. However, projections slightly differed in the total increase with a range of 18.4–35.4% and 9.8–16.4% for *Prochlorococcus* and *Synechococcus*, respectively. We also found that these projected changes were geographically uneven (Fig. 3 *A* and *B*). The multimodel ensemble mean for *Prochlorococcus* abundance increased $25.8 \pm 7.0\%$ in a band between 12°N and 26°S and $37.9 \pm 10.4\%$ up to 40° (Fig. 3*A*). For *Synechococcus*, the increase was $20.0 \pm 4.5\%$ in the tropical band (20°N to 20°S) as well as a slight expansion at high latitudes (Fig. 3*B*). However, for *Synechococcus*, the biggest changes were observed from ∼30°N to 60°N or ∼30°S to 60°S. In a zone centered on 35°N or 35°S, we found a decline in abundance. Although most significant in the Pacific Ocean, we also detected a zone centered on 45°N with a large increase in abundance. Projections of the four climate models largely covaried in the increase in cell number within the tropics, and in the poleward shifts of the abundance peak at mid-latitudes (Fig. 3 *C* and *D*). In a north–south transect along the Atlantic Ocean, the models depicted two consistent crests around the equator for both lineages (Fig. 3 *C* and *D*). For *Synechococcus*, the models also showed a peak around 45° of latitude (Fig. 3*D*). In this transect, the uncertainty caused by climate model differences for *Prochlorococcus* abundance was larger in the Northern Hemisphere and the Equator, and it decreased toward the southern distribution limit (Fig. 3*C*). For *Synechococcus*, the uncertainty was associated with differences in the projected abundance and the specific latitude for the secondary peak (Fig. 3*D*). All four models projected a slight expansion of distribution areas with high cell abundance (here defined as 10^4 cells mL^{-1}) of 3.3 ± 0.4°N and 2.5 ± 0.2°S for *Prochlorococcus* and 5.6 ± 2.1°N and 2.5 ± 0.3°S for *Synechococcus* (Fig. 3 *C* and *D*). Thus, we project an overall increase of both *Prochlorococcus* and *Synechococcus* but with substantial regional variations in cell abundances.

***Prochlorococcus* and *Synechococcus* Niches.** We next identified the niches for *Prochlorococcus* and *Synechococcus* along temperature, PAR, and nutrient gradients (Figs. 1*B* and 4). The response to temperature for *Prochlorococcus* could be divided into three zones. Below 8 °C, *Prochlorococcus* was present in 230 of 7,213 (3.2%) field samples (Fig. S3*B*), and this low likelihood of observing *Prochlorococcus* was reflected in the niche model (Fig. S6*A*). Between 8 °C and 13 °C, the probability of finding *Prochlorococcus* increased from 20% to 80%, and the estimated relationship between cell abundance and temperature was the strongest irrespective of PAR (Fig. S6*A*). Above 13 °C, the probability of finding *Prochlorococcus* increased from 80% to ∼100%, and both the observed and estimated abundances changed from 10^4 to 10^5 cells mL^{-1} (Fig. 4*A* and Fig. S6*A*). As expected, the abundance was also strongly related to PAR intensity (Fig. 4*B* and

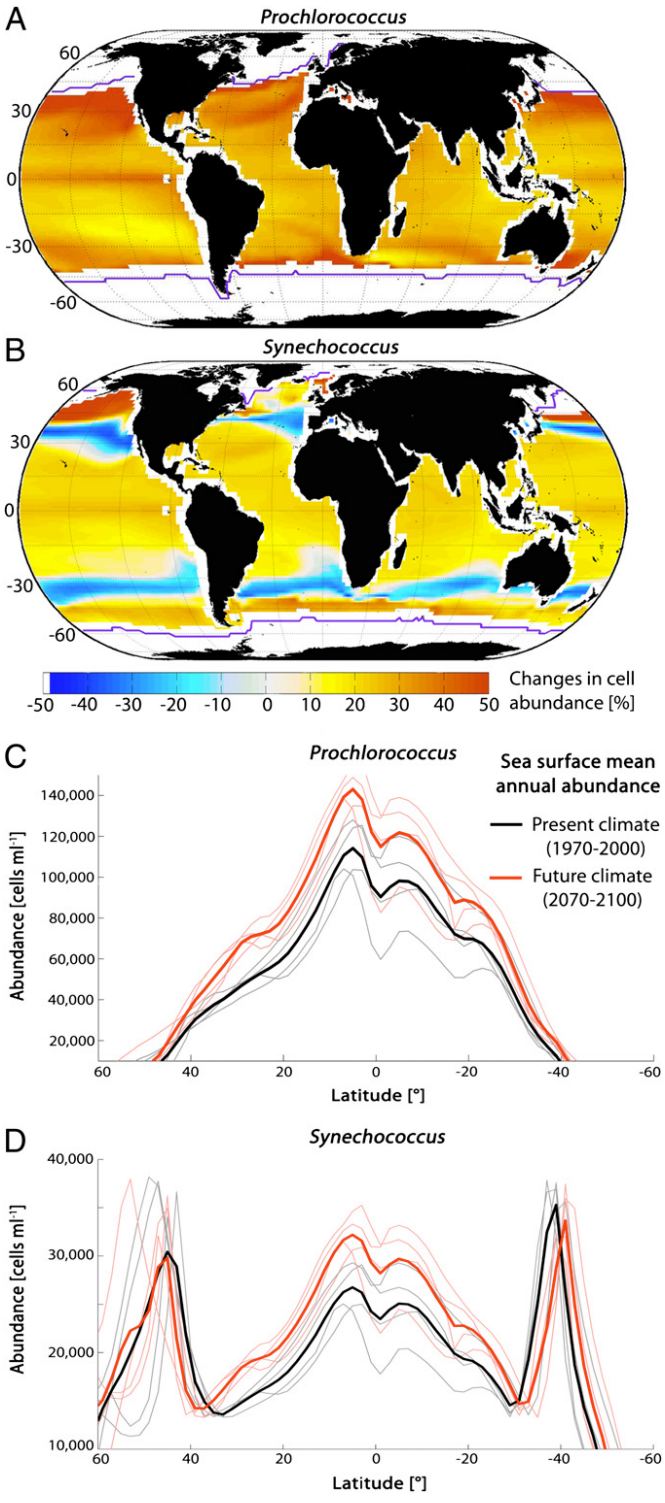


Fig. 3. Projected change in global abundance and distribution of *Prochlorococcus* and *Synechococcus* for 2100. Percent change in mean annual abundance between present and future climate (end of 20th and 21st century) at the sea surface for (*A*) *Prochlorococcus* and (*B*) *Synechococcus*. Colored areas represent the change in abundance in regions with $>10^4$ cells mL^{-1} at present climate. Purple lines represent the distribution limit of 10^4 cells mL^{-1} under future climate. Mean annual abundance estimated for present and future climate for a north–south transect at the Atlantic Ocean (330° meridian) for (*C*) *Prochlorococcus* and (*D*) *Synechococcus*. Lines represent the annual mean for the multimodel ensemble (thick) and each of the four models (thin).

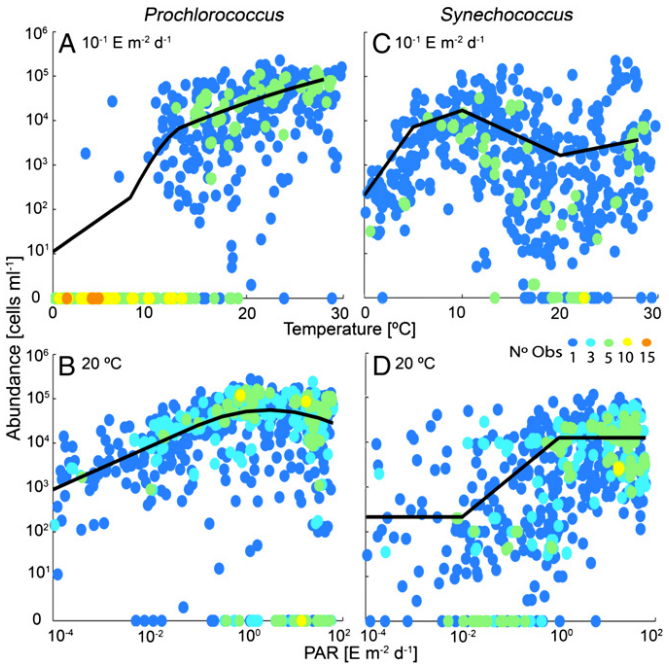


Fig. 4. *Prochlorococcus* and *Synechococcus* observations and quantitative niche model for temperature and PAR. Cell abundance as a function of (*A* and *C*) temperature at constant PAR ($10^{-1 \pm 0.05}$ E m^{-2} d^{-1}) and (*B* and *D*) PAR at constant temperature (20 ± 0.05 °C). Symbol color represents the number of overlapping observations, and the lines show the model output.

Fig. S6*B*). At the high PAR extreme, cell abundances were reduced by 31% relative to the peak (Fig. 4*B* and Fig. S6*B*). At the low PAR extreme, there was a substantial cell concentration at the base of the euphotic zone (Fig. S3*C*). For example, at depths below 150 m and a median PAR of 0.022 E m^{-2} d^{-1}, the median *Prochlorococcus* abundance was 6,227 cells mL^{-1}. In the tropics, cell abundance remains detectable (median = 594 cells mL^{-1}) in deep-water environments characterized by PAR well below the threshold for growth of *Prochlorococcus* cultures (19). Finally, we examined the relationship between nutrient availability and *Prochlorococcus* abundance, while controlling for PAR and temperature. As identified by the ANN analysis, we did not see a significant link between nitrate concentration and abundance (Fig. S6*C*). The abundance of *Prochlorococcus* was almost constant across three orders of magnitude in nitrate concentration, and our global dataset contains 6,392 observations of *Prochlorococcus* with abundance above 10^4 cells mL^{-1} in environments with more than 1 μM nitrate. Thus, variations in nutrient concentration at constant PAR and temperature seemed to have little control on the abundance of *Prochlorococcus* (20).

Synechococcus abundance was also strongly linked to temperature, but the shape of the relationship was different compared with *Prochlorococcus* (Fig. 4 *A* and *C*). *Synechococcus* was largely absent in subzero waters and found to have a maximum in abundance at 10 °C. Above 10 °C, the abundance declined and displayed a local minimum at ∼20 °C. Above this temperature, there was a small increase in cell concentration. We observed a nonlinear positive relationship of *Synechococcus* abundance with PAR (i.e., flat/positive/flat). This relationship was dependent on temperature, because the parametric regression model failed to identify a link between abundance and PAR at low temperatures (Fig. S6*B*). We also observed detectable cell abundance in deep, low PAR samples (depth > 100 m, mean PAR = 0.15 E m^{-2} d^{-1}, mean abundance = 172 cells mL^{-1}) (Fig. S3*C*). Similar to *Prochlorococcus*, we did not find a clear relationship between nutrient concentration and *Synechococcus* abundance after controlling for temperature and PAR (Fig. S6*C*).

Discussion

Our analyses build on the summed effort of 35 y of research cruises and time series and provide a basis for quantifying the global abundance of *Prochlorococcus* and *Synechococcus* within their realized niche. Our dataset differs from previous efforts (21) in having a wider geographic and environmental representation, and it includes multiple sample sets from the eastern and western Pacific, southern Atlantic, and southern Indian Oceans. We also recognize that slightly different techniques have been used to count cells, but the large size of our dataset and our sensitivity analyses suggest that the estimates are robust against such possible biases.

We find mean annual averages of 2.9×10^{27} and 7.0×10^{26} cells for *Prochlorococcus* and *Synechococcus*, respectively, which are in good agreement with the near surface estimate of 2.9×10^{27} cells of total prokaryotic autotrophs (22). Our estimates suggest that marine Cyanobacteria constitute about 10% of the total ocean marine picoplankton in the upper 200 m (assuming 3.6×10^{28} cells) (22). We predict the global net primary production of 4 and 8 Gt C y^{-1} [8.5% and 16.7% of ocean net primary production (1)] for *Prochlorococcus* and *Synechococcus* respectively, using estimates of hourly carbon fixation rates (3). We acknowledge a high uncertainty of the global variation in fixation rates, but our results suggest that marine Cyanobacteria could be responsible for ∼25% of ocean net primary productivity. This value is within the range of local observations of ocean productivity for the two lineages (23).

We observe complex changes in the abundances and distributions of *Prochlorococcus* and *Synechococcus* as a result of the climate change projected by the end of 21st century using the RCP4.5 scenario. As proposed previously (24), we expect an overall increase in cell abundances of *Prochlorococcus* and *Synechococcus*. However, the changes in abundance are geographically differentiated. Mid-latitude regions can experience large shifts in community structure, including regions with a large increase in both lineages but also places with a decline in *Synechococcus* abundance (around 35°N or 35°S). At low latitudes, we expect that both lineages increase their abundance. Our analysis supports an overall increase in Cyanobacteria abundance explained primarily by the increase in areas where *Prochlorococcus* and *Synechococcus* are currently abundant and secondarily by the geographical expansion of these lineages. Large model uncertainties in the distribution and abundance of cells for future projections are mostly at high latitudes. However, the sign of change is consistent among the four climate models. It is also worth noting that other scenarios predict a much larger increase in the atmospheric CO_2 concentration and resulting ocean temperature. Thus, it is quite likely that marine ecosystems may experience even larger changes in the phytoplankton community structure than suggested here.

It can be challenging to differentiate between the individual effects of environmental factors because of a high degree of covariance in the ocean (25). Yet, the diversity and range of our large dataset provides an unprecedented opportunity to determine the importance of the different environmental variables. We find that temperature is the main control of the regional distributions of both *Prochlorococcus* and *Synechococcus*, but there are important quantitative differences. For *Prochlorococcus*, the inclusion of observations with zero cell abundance is key to delineate the sharp geographic limits and identify a steeper than previously thought response to temperature between 10 °C and 15 °C (26). *Synechococcus* displays a similar positive step response to temperature but with a threshold around 0 °C instead of 10 °C. This lower temperature boundary results in predicting significant *Synechococcus* populations in arctic waters but an absence in subzero waters around Antarctica—consistent with observations (27, 28). In addition, we observe a local minimum around 20 °C, which is

supported by specific latitudinal transects (11, 29). This intermediate temperature minimum could be the result of ecotype replacements with different temperature optima or a competitive interaction with *Prochlorococcus*, which displays a large increase in abundance at this temperature (Fig. 4 *A* and *C* and Fig. S6*A*).

PAR displays a complex control on *Prochlorococcus* and *Synechococcus*. At high PAR intensities (>10 E m^{-2} d^{-1}), such as in tropical surface waters, we detect a 31% decline in cell abundance for *Prochlorococcus*. This observation is corroborated by quantitative PCR (26) and thus, likely not an artifact of counting populations with low chlorophyll content with flow cytometry. Photoinhibition or UV radiation damage can result in lower growth rates and overall lower abundances at these light levels (30). Alternatively, these cells receive a lower flux of nutrients compared with lower depths and PAR. At low PAR levels, we found a substantial number of *Prochlorococcus* and *Synechococcus* cells. Culture studies predict that both lineages have a compensation light level at about 0.06 E m^{-2} d^{-1} (19). The compensation light suggests that cells are unable to support autotrophic growth below this PAR level. Perhaps cells have been vertically advected out of the photic zone rather than actively growing below the compensation depth. Alternatively, some uncultured ecotypes may be capable of heterotrophy or growing at extremely low light levels. We posit vertical advection, because it is consistent with a suggested downward flux of picophytoplankton (5, 31).

Supported by the ANN and regression analyses, nutrient concentrations seem to have a limited influence on the biogeography of both *Prochlorococcus* and *Synechococcus* (Fig. S6*C* and Tables S2 and S3). The limited role of nutrients is in agreement with the presence of *Prochlorococcus* and *Synechococcus* ecotypes that are adapted to different nutrient levels and thus, enable each lineage as a whole to proliferate over a wide range of nutrient concentrations (20, 32, 33). However, it is in contrast to the commonly held view that *Prochlorococcus* is at a competitive disadvantage in elevated nutrient environments (13). It is important to recognize that we may underestimate the bioavailable nutrients, because most of the nitrogen and phosphate in oligotrophic waters are linked to dissolved organic matter. In addition, we have analyzed the relationship between abundance and nutrient stocks rather than fluxes. These two aspects could influence the observed relationship between nutrient availability and abundance, and therefore, it is possible that nutrients can influence *Prochlorococcus* and *Synechococcus* distributions in other ways. Nevertheless, the widespread presence of *Prochlorococcus* in elevated nutrient environments suggests that the linkage between abundance and nutrient availability is more complex for this lineage than previously thought.

Niche-based models have been widely used to predict the response of plant species to future climate change (17). Niche models express species abundance (or just presence) as a function of physical variables of the environment. They are based on collections of observations and their associated metadata describing the environment. These models represent the realized niche of species, because they reflect the use of resources as the net outcome of bottom-up and top-down processes. In this way, niche models do not address mechanisms such as predation or biotic interactions, which also can influence distribution and abundance (17). Projections of species distribution in future climate scenarios assume that species presence and abundance will be linked to similar factors as under current climate condition. We then use these principles to predict the global abundances and distributions of *Prochlorococcus* and *Synechococcus*. Our models provide a quantitative basis for predicting and understanding shifts in *Prochlorococcus* and *Synechococcus* distributions and suggest that future changes in ocean temperature can cause large-scale changes in phytoplankton community structures. Because of the high abundance of *Prochlorococcus* and *Synechococcus*, these changes can have a large impact on ocean ecosystems and global biogeochemical cycles.

Materials and Methods

Dataset. All analyses were done using Matlab (Mathworks). We obtained *Prochlorococcus* and *Synechococcus* observations from available public repositories and primary sources (Fig. 1*A* and Table S1). We only considered cell counts by flow cytometry for *Prochlorococcus* because of their weak autofluorescence, but we also included microscope counts for *Synechococcus*. Samples covered a latitudinal range from 81°N to 69°S up to 400 m depth; 79% of the samples were collected in the Northern Hemisphere. Ancillary temperature/nitrate/phosphate records were available for all but 3,600/8,000/14,000 observations, respectively, which we complemented with 1° monthly depth-dependent averages from the World Ocean Atlas (www.nodc.noaa.gov). To avoid issues with detection limits, we imposed a minimum nutrient concentration of 0.01 μM. We calculated surface PAR (8 d averaged, 0.047° grid cell) using SeaWiFS and MODIS observations. Downward PAR was calculated using the attenuation coefficient K_{490} from MODIS and SeaWiFS (http://oceancolor.gsfc.nasa.gov) and corrected for chlorophyll *a* (34). Numerical results illustrated that the use of absolute PAR was better suited for our models than the average light intensity through the mixed layer (*SI Materials and Methods*, section 1.1), because the averaging in the mixed layer involved a number of parameters to estimate the mixed layer depth that did not result in a noticeable improvement in the predictive power of PAR. We discarded infrequently high and low values for all variables (Fig. S1). The dataset is available at http://gdps.cima.fcen.uba.ar/.

ANN and Local Regression Models. We used back-propagation feed-forward ANN analyses to identify which input variables explained the observed *Prochlorococcus* and *Synechococcus* abundances. Temperature, PAR, nitrate, and phosphate concentrations and their possible combinations were tested as predictors (Table S2). For both lineages, temperature and PAR exhibited the highest predictive powers, and they were subsequently included in the regression models. We next applied local regression models with temperature and PAR as independent variables. Local abundances were estimated by fitting a linear regression to 200 observations in the neighborhood of each temperature and PAR combination. This local analysis provided the shape of the response curve of the cell abundance to environmental variables and highlighted thresholds on both axes. Based on the shapes and thresholds, we formulated parametric regression models (*SI Materials and Methods*, section 1.2).

Parametric Regression Model. The parametric regression model for *Prochlorococcus* cell abundance (C_{PRO}) consists of two components: an indicator variable *I* for the probability of presence (*P*) and cell abundance *C*+ (Eq. **1**):

$$C_{PRO} = P \times C+. \quad [1]$$

We assumed that C_{PRO} and *C*+ are on $\log_{10}$ scale. To obtain $P = Pr[I = 1]$ and *C*+, we used $\log_{10}$-transformed PAR and temperature (*T*) as input variables. For *P*, we used a logistic fit with two different temperature thresholds (Eq. **2**):

$$P = a + b * \mathrm{PAR} + c * T + c_1^* T * \mathrm{I}_1 + c_2^* T * \mathrm{I}_2, \quad [2]$$

where $\mathrm{I}_1(\mathrm{I}_2) = 0$ if $T \le 8$ (13) and $\mathrm{I}_1(\mathrm{I}_2) = 1$ otherwise. For *C*+, we used nonzero observations and predicted the cell abundance of *Prochlorococcus* as follows:

$$C+ = d + e * \mathrm{PAR} + e_1^* \mathrm{PAR}^2 * \mathrm{I}_3 + \; f * T + f_1^* T^2 * I_4, \quad [3]$$

where $\mathrm{I}_3(\mathrm{I}_4) = 0$ if $\mathrm{PAR} \le 1 (T \le 20)$ and $\mathrm{I}_3(\mathrm{I}_4) = 1$ otherwise. The regression model proposed for *Synechococcus* cell abundance (C_{SYN}) included temperature and PAR as explanatory variables (Eq. **4**). We assumed C_{SYN} on a $\log_{10}$ scale, unless stated differently. We defined a region of linear dependence ($-2 < \mathrm{PAR} < 0$ $\log_{10}$ E d^{-1} m^{-2}) surrounded by two regions without influence of PAR. The model assumed a nonlinear interaction between temperature and PAR:

$$C_{SYN} = m + n * T + n_1^* T * \mathrm{I}_5 + n_3^* T * \mathrm{I}_6 + o * [T - 5 * (1 + \mathrm{PAR} * \mathrm{I}_7)] * \mathrm{I}_8, \quad [4]$$

where $\mathrm{I}_5(\mathrm{I}_6) = 0$ if $T \le 20$ (5) and $\mathrm{I}_5(\mathrm{I}_6) = 1$ otherwise; $\mathrm{I}_7 = 0$ if $-2 \le \mathrm{PAR} \le 0$ and $\mathrm{I}_7 = 1$ otherwise; and $\mathrm{I}_8 = 0$ if $T - 5*(1 + \mathrm{PAR}*\mathrm{I}_7) \le 0$ and $\mathrm{I}_8 = 1$ otherwise. Fitted parameter values are listed in Table S4, and additional details are in *SI Materials and Methods*, section 1.3.

Sensitivity Analysis. We evaluated the sensitivity of the parametric model to input data in three ways. First, a jackknife analysis showed that the variance in each parameter decreased as a function of observations, but the means were stable (Fig. S4*A*). The reduced number of samples influenced the spread of the determination coefficient and the variance of the mean global abundance

but not their mean values (Fig. S4 *B* and *C*). Second, we performed a split sample test to evaluate the influence of individual cruises or time series on the parametric regression models. This analysis would reveal possible biases originating from different measurement methods. The split sampling was done by stepwise removing datasets. This split-sampling approach hardly affected the coefficients of our regression models (Fig. S4 *A* and *B*) except for ID12 (35). ID12 included 11,056 samples with zero *Prochlorococcus* abundance (set as zero in the $\log_{10}$ scale) and thus, influenced the logistic model (Eq. **2** and Fig. S6*A*). Third, we tested for any possible geographical bias contrasting observed and expected cell abundances (Fig. S4*D*).

Present and Future Global Abundances and Distributions. To estimate the mean global cell abundances of *Prochlorococcus* and *Synechococcus*, we used as input to our parametric regression models monthly average temperatures from the World Ocean Atlas 2005 (1° × 1° resolution) (36) and 8-d average PAR and KD_{490} values derived from satellite data (SeaWiFS 0.083° × 0.083°) (37) (*SI Materials and Methods*, section 1.4). To estimate the sea surface distributions (Fig. 3 *A* and *B*), we used sea surface temperatures and PAR satellite data (SeaWiFS 0.083° × 0.083°) (37) as input variables in our regression models (*SI Materials and Methods*, section 1.4). To estimate abundance changes in the context of climate change, we used as inputs to our model monthly PAR from MODIS (averaged 2002–2011) and sea surface temperature simulations from an ensemble of four global circulation models (CSIRO Mk3 6 0, CanESM2, HadGEM2 ES, and GISS E2 H) included in the Coupled Model Intercomparison Project (18). From the RCP4.5 scenario (38), we obtained the monthly sea surface temperature means of the last 30 y of the 20th and 21st century to represent present and future climate conditions, respectively. The RCP4.5 scenario predicts long-term global emissions of greenhouse gases, short-lived species, and land use changes, which stabilized radiative forcing at 4.5 Wm^2 or the equivalent of 650 ppm CO_2. The percentage change in cell abundance was calculated based on the difference between the annual averages of future and present estimates. Uncertainty associated with each value represented the difference in the projection of each model (multimodel ensemble mean ± 1 SD; $n = 4$ models).

ACKNOWLEDGMENTS. We thank the many contributing researchers for the oceanographic data and Coupled Model Intercomparison Project (CMIP5) model outputs. Financial support for this work was provided by the Agencia Nacional de Promoción Científica y Tecnológica and Consejo Nacional de Investigaciones Científicas y Técnicas de Argentina (P.F.), University of California at Irvine Undergraduate Research Opportunity Program awards (to J.L.G. and R.A.G.), a National Aeronautics and Space Administration Inspiring the Next Generation of Earth Explorers grant to the Earth System Science as well as California Alliance for Minority Participation at University of California at Irvine (to J.L.G., R.A.G., and J.R.), National Major Scientific Research Program and Natural Science Foundation of China (N.J.), National Science Foundation Emerging Frontiers program (D.M.K.), the Gordon and Betty Moore Foundation (D.M.K.), the National Science Foundation Dimensions of Biodiversity and Biological Oceanography programs (M.W.L. and A.C.M.), and the University of California at Irvine Environment Institute (J.A.V. and A.C.M.).

1. Field CB, Behrenfeld MJ, Randerson JT, Falkowski P (1998) Primary production of the biosphere: Integrating terrestrial and oceanic components. *Science* 281(5374): 237–240.
2. Zeidner G, et al. (2003) Molecular diversity among marine picophytoplankton as revealed by psbA analyses. *Environ Microbiol* 5(3):212–216.
3. Li WKW (1994) Primary production of Prochlorophytes, Cyanobacteria, and eucaryotic ultraphytoplankton: Measurements from flow cytometric sorting. *Limnol Oceanogr* 39(1):169–175.
4. Arrigo KR, et al. (1999) Phytoplankton community structure and the drawdown of nutrients and CO_2 in the Southern Ocean. *Science* 283(5400):365–367.
5. Richardson TL, Jackson GA (2007) Small phytoplankton and carbon export from the surface ocean. *Science* 315(5813):838–840.
6. Thomas MK, Kremer CT, Klausmeier CA, Litchman E (2012) A global pattern of thermal adaptation in marine phytoplankton. *Science* 338(6110):1085–1088.
7. Le Quéré C, et al. (2005) Ecosystem dynamics based on plankton functional types for global ocean biogeochemistry models. *Glob Chang Biol* 11(11):2016–2040.
8. Li WKW (2009) From cytometry to macroecology: A quarter century quest in microbial oceanography. *Aquat Microb Ecol* 57(3):239–251.
9. Li WKW (2007) Macroscopic patterns in marine plankton. *Encyclopedia of Biodiversity*, ed Simon AL (Elsevier, New York), pp 1–16.
10. Li WKW (2009) Plankton populations and communities. *Marine Macroecology*, eds Witman JD, Roy K (Univ of Chicago Press, Chicago), pp 29–64.
11. Johnson ZI, et al. (2006) Niche partitioning among *Prochlorococcus* ecotypes along ocean-scale environmental gradients. *Science* 311(5768):1737–1740.
12. Jiao N, et al. (2005) Dynamics of autotrophic picoplankton and heterotrophic bacteria in the East China Sea. *Cont Shelf Res* 25(10):1265–1279.
13. Partensky F, Hess WR, Vaulot D (1999) *Prochlorococcus*, a marine photosynthetic prokaryote of global significance. *Microbiol Mol Biol Rev* 63(1):106–127.
14. Partensky F, Blanchot J, Vaulot D (1999) Differential distribution and ecology of *Prochlorococcus* and *Synechococcus* in oceanic waters: A review. *Bull Inst Oceanogr* 19:457–475.
15. Coleman ML, Chisholm SW (2007) Code and context: *Prochlorococcus* as a model for cross-scale biology. *Trends Microbiol* 15(9):398–407.
16. Denis M, Martin V, Andersen V (2000) Short-term variations of the vertical distribution of Cyanobacteria in the open Mediterranean Sea. *Sci Mar* 64(2):157–163.
17. Morin X, Thuiller W (2009) Comparing niche- and process-based models to reduce prediction uncertainty in species range shifts under climate change. *Ecology* 90(5): 1301–1313.
18. Taylor KE, Stouffer RJ, Meehl GA (2012) An overview of CMIP5 and the experiment design. *Bull Am Meteorol Soc* 93(4):485–498.
19. Moore LR, Chisholm SW (1999) Photophysiology of the marine cyanobacterium *Prochlorococcus*: Ecotypic differences among cultured isolates. *Limnol Oceanogr* 44(3): 628–638.
20. Olson RJ, Chisholm SW, Zettler ER, Altabet MA, Dusenberry JA (1990) Spatial and temporal distributions of prochlorophyte picoplankton in the North Atlantic Ocean. *Deep Sea Res Part 1 Oceanogr Res Pap* 37(6):1033–1051.
21. Buitenhuis ET, et al. (2012) Picophytoplankton biomass distribution in the global ocean. *Earth System Science Data Discussions* 5(1):221–242.
22. Whitman WB, Coleman DC, Wiebe WJ (1998) Prokaryotes: The unseen majority. *Proc Natl Acad Sci USA* 95(12):6578–6583.
23. Zhang Y, Jiao N, Hong N (2008) Comparative study of picoplankton biomass and community structure in different provinces from subarctic to subtropical oceans. *Deep Sea Res Part 2 Top Stud Oceanogr* 55(14–15):1605–1614.
24. Morán XAG, López-Urrutia Á, Calvo-Díaz A, Li WKW (2010) Increasing importance of small phytoplankton in a warmer ocean. *Glob Chang Biol* 16(3):1137–1144.
25. Kamykowski D (2008) Estimating upper ocean phosphate concentrations using ARGO float temperature profiles. *Deep Sea Res Part 1 Oceanogr Res Pap* 55(11):1580–1589.
26. Zinser ER, et al. (2007) Influence of light and temperature on *Prochlorococcus* ecotype distributions in the Atlantic Ocean. *Limnol Oceanogr* 52(5):2205–2220.
27. Robineau B, et al. (1999) Ultraphytoplankton abundances and chlorophyll *a* concentrations in ice-covered waters of northern seas. *J Plankton Res* 21(4):735–755.
28. Letelier RM, Karl DM (1989) Phycoerythrin-containing Cyanobacteria in surface waters of the Drake Passage during February 1987. *Antarct J US* 24(5):185–188.
29. Zubkov MV, Sleigh MA, Burkill PH, Leakey RJG (2000) Picoplankton community structure on the Atlantic Meridional Transect: A comparison between seasons. *Prog Oceanogr* 45(3–4):369–386.
30. Llabrés M, Agustí S (2006) Picophytoplankton cell death induced by UV radiation: Evidence for oceanic Atlantic communities. *Limnol Oceanogr* 51(1I):21–29.
31. Jackson GA, Waite AM, Boyd PW (2005) Role of algal aggregation in vertical carbon export during SOIREE and in other low biomass environments. *Geophys Res Lett* 32(13):L13607.
32. Martiny AC, Kathuria S, Berube PM (2009) Widespread metabolic potential for nitrite and nitrate assimilation among *Prochlorococcus* ecotypes. *Proc Natl Acad Sci USA* 106(26):10787–10792.
33. Scanlan DJ, et al. (2009) Ecological genomics of marine picocyanobacteria. *Microbiol Mol Biol Rev* 73(2):249–299.
34. Morel A, et al. (2007) Examining the consistency of products derived from various ocean color sensors in open ocean (Case 1) waters in the perspective of a multi-sensor approach. *Remote Sens Environ* 111(1):69–88.
35. Li WKW (1998) Annual average abundance of heterotrophic bacteria and *Synechococcus* in surface ocean waters. *Limnol Oceanogr* 43(7):1746–1753.
36. Locarnini RA, Mishonov AV, Antonov JI, Boyer TP, Garcia HE (2006) *World Ocean Atlas 2005, Volume 1: Temperature* (US Government Printing Office, Washington, DC).
37. McClain CR, Feldman GC, Hooker SB (2004) An overview of the SeaWiFS project and strategies for producing a climate research quality global ocean bio-optical time series. *Deep Sea Res 2 Top Stud Oceanogr* 51(1–3):5–42.
38. Moss RH, et al. (2010) The next generation of scenarios for climate change research and assessment. *Nature* 463(7282):747–756.