Supporting Information

Flombaum et al. 10.1073/pnas.1307701110

SI Materials and Methods

1.1. Mixed Layer Depth. For reasons of parsimony and based on numerical analyses, we used the absolute photosynthetic active radiation (PAR) rather than the average light intensity throughout the mixed layer. The average light intensity throughout the mixed layer has earlier been suggested as the more appropriate representation of the light perceived by surface ocean cells (1). Determining the mixed layer depth (MLD) included five decision levels, which increased the uncertainty of PAR estimation and compressed the absolute PAR range (Fig. S3*D*). The top layer of the ocean is subject to turbulent mixing, and the MLD can be used to average PAR for cells within this layer. Cells within deep turbulent layers compared with cells within shallow layers are exposed on average to less light, which affects the intracellular chlorophyll concentration (1). Thus, we evaluated the use of the MLD to estimate PAR for samples within the first layer but found no improvement in the predictions. A more detailed account is given below.

We estimated the MLD for 39,378 samples based on field records of conductivity, temperature, and depth (CTD). We defined the MLD as the change in temperature or density larger than Δ compared with a reference value. We used two alternative reference values, the average between 0 and 10 m, and the average of 10 m above a given depth in the CTD profile. We used five values for Δ (0.005, 0.01, 0.05, 0.1, and 0.5 °C or kg m^{-3}). After an MLD was detected, we tested temperature or density for 5 m below it. This neighboring test avoided the small drops in the temperature or density profile that would be confused with the MLD; if values below the MLD did not drop consistently, we continued to scan the CTD profile. Then, we looked for the closest CTD field to a Cyanobacteria sample based on date, latitude, and longitude. We set one-half of 1° as a maximal distance to associate a CTD with a field sample. For field samples with no CTD, we used a global dataset with monthly averages (www.ifremer.fr/cerweb/deboyer/mld/home.php) (2). For samples above the MLD, we estimated the mean PAR as the integral of PAR between the surface PAR and the MLD divided by the MLD. PAR at the MLD was estimated with the downward attenuation equation (3). For field samples below the MLD, we used the downward attenuation equation by Morel et al. (3) at the depth of the sample. As a whole, to use the average of PAR within the mixed layer, we introduced five decision levels (variable, reference value, Δ, neighboring test, and distance to the sample) all subject to errors. In addition, the correlation between our in situ MLD and climatological MLD was poor (2). By contrast, PAR estimated as a function of depth with the downward attenuation equation only involved using PAR, K_{490}, and the depth of the field sample in the water column (3). Because of the number of decisions involved to estimate average PAR within the mixing layer and the lack of reference with which to compare, we decided to estimate PAR solely as a function of depth (3).

1.2. Neural Network Analyses and Local Regression Model. We used back-propagation artificial neural networks (ANN) to evaluate possible predictors of *Prochlorococcus* and *Synechococcus* cell abundances. Temperature, PAR, nitrate, and phosphate concentration and their combination were analyzed (Table S2). Two hidden layers were deemed sufficient to predict the observed *Prochlorococcus* and *Synechococcus* concentrations, because additional layers did not improve the overall network performance. Before training of each ANN model, the data were randomly divided into calibration (20%) and evaluation (80%) datasets. The weights and biases of both hidden layers of each ANN model were derived by training against the calibration dataset using a total of 500 epochs. The predictive abilities of each trained ANN model were subsequently evaluated against the validation dataset. The goodness of fit was quantified using the root mean square error (RMSE) and coefficient of determination (R^2) (Table S2). Both summary statistics were computed as averages of 25 ANN calibration trials.

Preliminary runs showed a difference in the outcome of the ANN analysis when including or excluding observations with no *Prochlorococcus*. We, therefore, reported the results of both calibration strategies. For *Synechococcus*, we did not make this distinction. The predictive ability of the ANN *Prochlorococcus* models significantly increased with the zero cell abundances observations included (Table S2). This increase in R^2 was not surprising, because the frequent presence of zero abundance reduced the variance of the overall *Prochlorococcus* dataset. When observations with zero *Prochlorococcus* were included, temperature was the variable that explained the largest fraction of the observed variation. When the zero cell abundance observations were removed, the predictive abilities of temperature became similar to the predictive abilities of PAR (Table S2). This change in R^2 indicated the presence of a critical temperature threshold for the presence of *Prochlorococcus*. The RMSE and R^2 summary statistics suggested that PAR was a poor predictor for *Prochlorococcus* if samples with zero abundance were included, whereas the performance of PAR improved if these samples were excluded (Table S2). Nitrate and phosphate were weak predictors of cell abundance, irrespective of whether zero cell abundance observations were included. Joint use of two environmental variables resulted in a significant increase in the ANN goodness of fit, whereas the additional benefit of third and fourth predictors was negligible (Table S2). With zero cell abundance data excluded, temperature and PAR exhibited the highest R^2 (0.413). This value was significantly higher than the individual predictors. When phosphate was added as a third explanatory variable, the network performance only improved by 7% (Table S2). The same was true for nitrate. The ANN results for *Synechococcus* were very similar (Table S2). Temperature and PAR were the strongest predictors. The predictions improved marginally in 2% if both nutrients were included (Table S2). Therefore, we retained temperature and PAR as the main explanatory variables for predicting both *Prochlorococcus* and *Synechococcus*.

ANN is a powerful method to select the most informative variables in nonlinear regression. However, ANNs are black box models and less suitable for identifying analytical relationships between input and output variables. To develop such analytical relationships, we introduced a local nonparametric regression model to find the shape and thresholds of cell abundance as a function of temperature and PAR. To find the shapes and thresholds, we defined a computational grid in temperature and PAR space, and for each grid point, we searched for the closest 200 cell abundance values and estimated the local expected mean. Because the local regression model was sensitive to data coverage, we then formulated a parametric regression model to smooth out any local irregularities with little biological support (Fig. S3*A*). The performances of the local and parametric regression models were similar, indicating that the parametric models could be used to describe the abundance of both lineages. For *Prochlorococcus*, the parametric model accounted for 98.12% of the variability explained by the local regression model and 96.20% of the residual

sum of squares (Table S3). For *Synechococcus*, the corresponding percentages were 83.19% and 89.52% (Table S3).

1.3. Parametric Regression Model. The parametric regression model included the probability P of nonzero abundance and the concentration $C+$ for the case of positive cell abundance (Eq. **1**). Zero *Prochlorococcus* abundances were recorded mostly below 13 °C (Fig. 1*B* and Fig. S3*B*). To estimate P and $C+$, we used logarithmically transformed observations of PAR and untransformed measurements of temperature. For P, we used presence and absence data (Fig. S6 *A* and *B*) to fit a logistic equation with two thresholds for temperature (8 °C and 13 °C) (Eq. **2** and Fig. S6 *A* and *B*). For $C+$, we included linear combinations of PAR and temperature with thresholds at 0.1 E m^{-2} d^{-1} to produce a concave curve at high PAR and at 20 °C, produce a change in the slope of the function at higher temperatures (Fig. S6*B*). The regression model proposed for *Synechococcus* included a region between 0.01 and 1 E m^{-2} d^{-1} with a log-linear dependence between cell concentrations. Outside this region, no apparent influence of PAR was found (Fig. S6*B*). We also identified a region of log-linear increase in cell concentration with temperature up to 10 °C and saturation thereafter (Fig. S6*B*). Finally, the model included a nonlinear interaction between PAR and temperature (Eq. **4**).

1.4. Prediction of Global Abundance and Distribution. To predict the global abundance of *Prochlorococcus* and *Synechococcus* with our parametric regression models, we used as input monthly average temperatures from the World Ocean Atlas 2005 (1° × 1° resolution) (4) and 8-d average PAR and KD_{490} values derived from satellite data (SeaWiFS 0.083° × 0.083°) (5). The environmental data were structured in a 3D matrix with layers from 0 to 200 m depth. The layers have the same depth intervals in the water column used by the World Ocean Atlas (0–10, 10–20, 20–30, 30–50, 50–75, 75–100, 100–125, 125–150, and 150–200 m). The downward light attenuation (K_{PAR}) used to estimate PAR at depth was based on the light attenuation coefficient (K_{490}) and chlorophyll *a* concentration (3). For each depth layer, the value of PAR was derived by averaging values from the top and the bottom of the layer at each geographical grid point. The output of the regression models was the mean cell abundance per milliliter in logarithm scale. We converted logarithm to linear scale using the bias correction factor ($\gamma_{PRO} = 1.8095$, $\gamma_{SYN} = 3.6415$). We multiplied the volume of each layer by its predicted cell abundance and summed over all layers to obtain total estimates at each grid point.

To estimate the surface distribution, we used sea surface temperatures and PAR satellite data (SeaWiFS 0.083° × 0.083°) (5) as input variables in our regression model. We obtained monthly maps of cell abundance on a 0.083° × 0.083° grid that were subsequently averaged to a 1° × 1° resolution. These monthly assessments (Fig. S5) were averaged to obtain an annual distribution map of *Prochlorococcus* and *Synechococcus* (Fig. 2).

1. Bouman HA, et al. (2006) Oceanographic basis of the global surface distribution of *Prochlorococcus* ecotypes. *Science* 312(5775):918–921.
2. Montegut CD, Madec G, Fischer AS, Lazar A, Iudicone D (2004) Mixed layer depth over the global ocean: An examination of profile data and a profile-based climatology. *J Geophys Res: Oceans* 109(C12):C12003.
3. 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.
4. Locarnini RA, Mishonov AV, Antonov JI, Boyer TP, Garcia HE (2006) *World Ocean Atlas 2005, Volume 1: Temperature* (US Government Printing Office, Washington, DC).
5. 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.

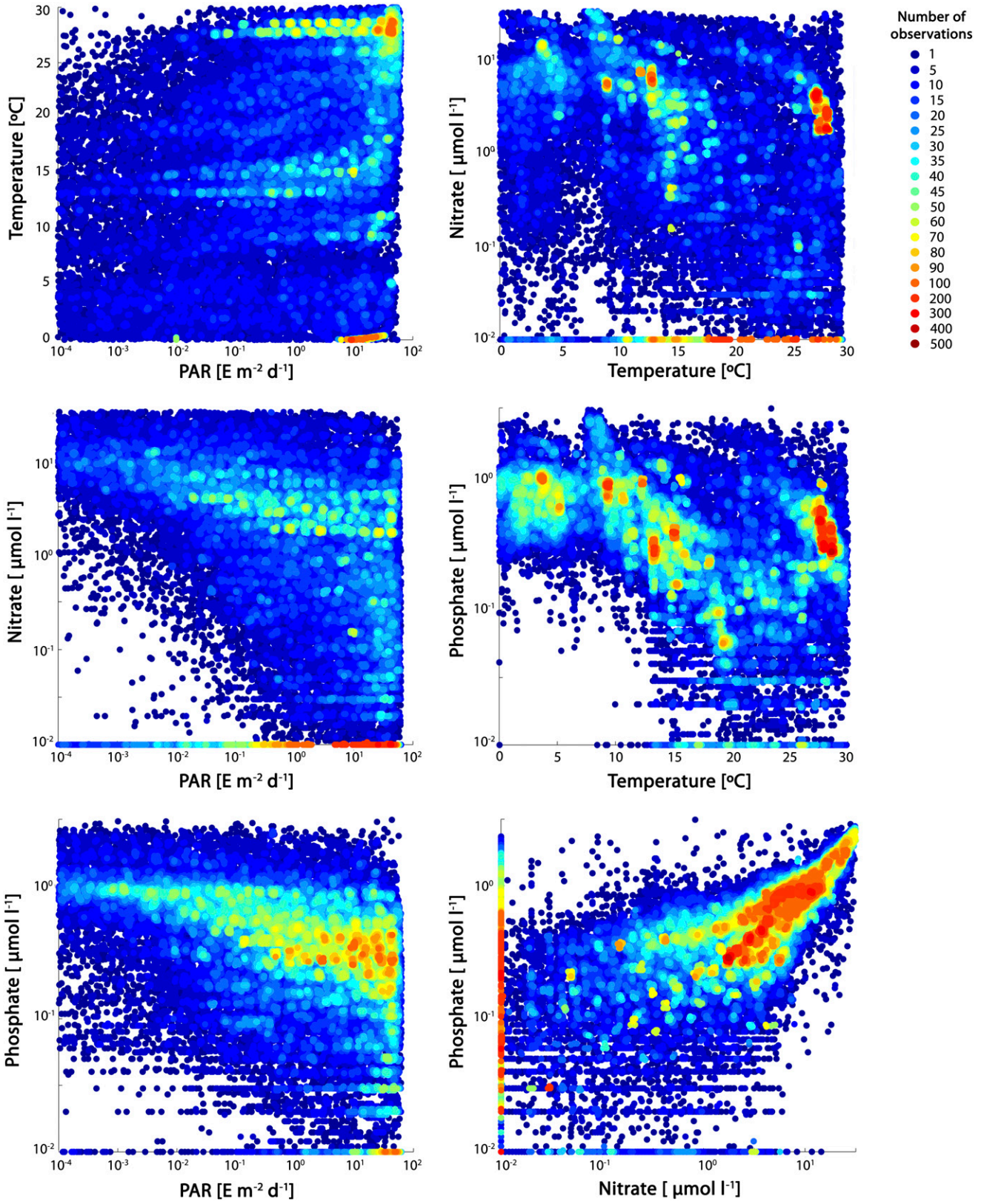

Fig. S1. Environmental variables and boundaries ordered by pairs. Minimum and maximum values were PAR = 10^{-4} and $10^{1.8}$ E m^{-2} d^{-1}, temperature = −2 °C and 30 °C, and nitrate and phosphate = 10^{-2} µmol $L^{-1}$and no upper limit. The numbers of discarded observations were 4,296 for values with PAR < PAR_{min}, 28 for values with PAR > PAR_{max}, and 305 for values with temperature > 30 °C. Symbol color represents the number of overlapping observations.

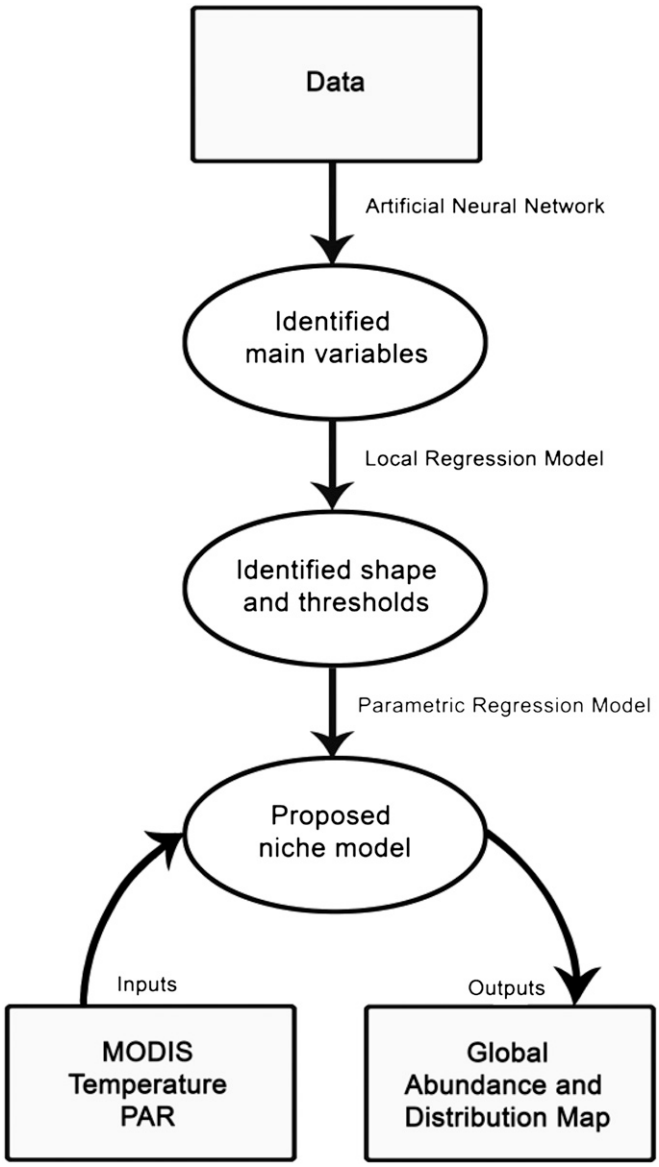

Fig. S2. Overview of data analysis. First, we identified the most informative explanatory variables using the ANN analysis. Second, we identified the shape and thresholds of the observed cell abundances as a function of environmental variables using a local regression analysis. Third, we proposed a parametric regression model similar in shape to the local regression analysis. The relatively simple regression model accurately predicts the major trends of the observed *Prochlorococcus* and *Synechococcus* data with a single closed-form equation.

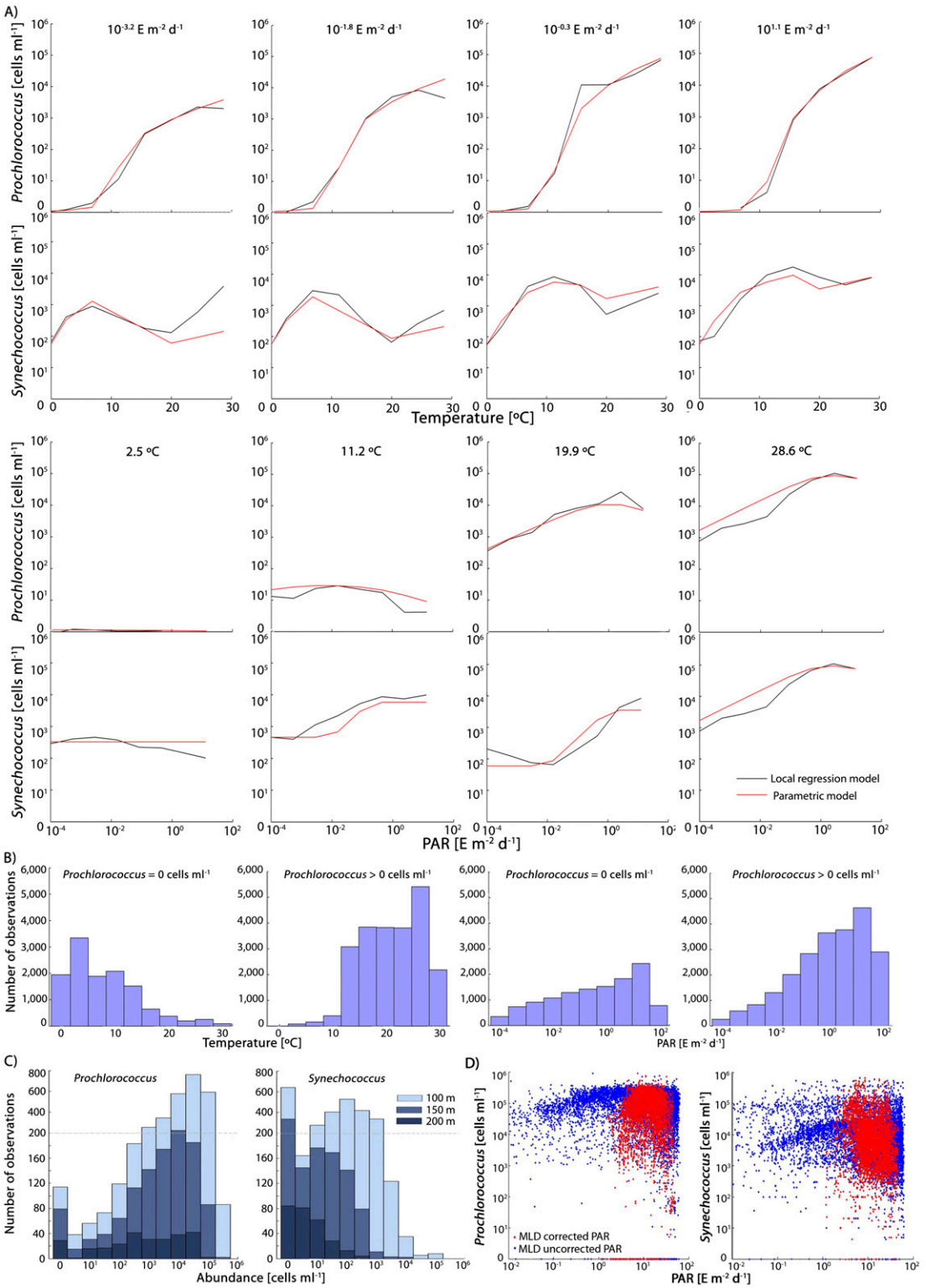

Fig. S3. Analyses of *Prochlorococcus* and *Synechococcus* abundance as a function of temperature, PAR, and depth. (*A*) Abundance as a function of temperature and PAR for local regression and parametric models. The local regression model used the input of the 200 nearest observations in a temperature and PAR computational grid. The parametric regression used the whole dataset. (*B*) Number of observations of *Prochlorococcus* zero and larger than zero cell abundance as a function of temperature and PAR. (*C*) Number of observations as a function of the cell abundance and water column depth. Samples were constrained to the latitudinal band between 30°N and 30°S. (*D*) Abundance as a function of corrected and uncorrected PAR. The corrected PAR is the average of PAR intensity throughout the MLD. The uncorrected PAR estimated PAR intensity using a downward attenuation model (1).

1. 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.

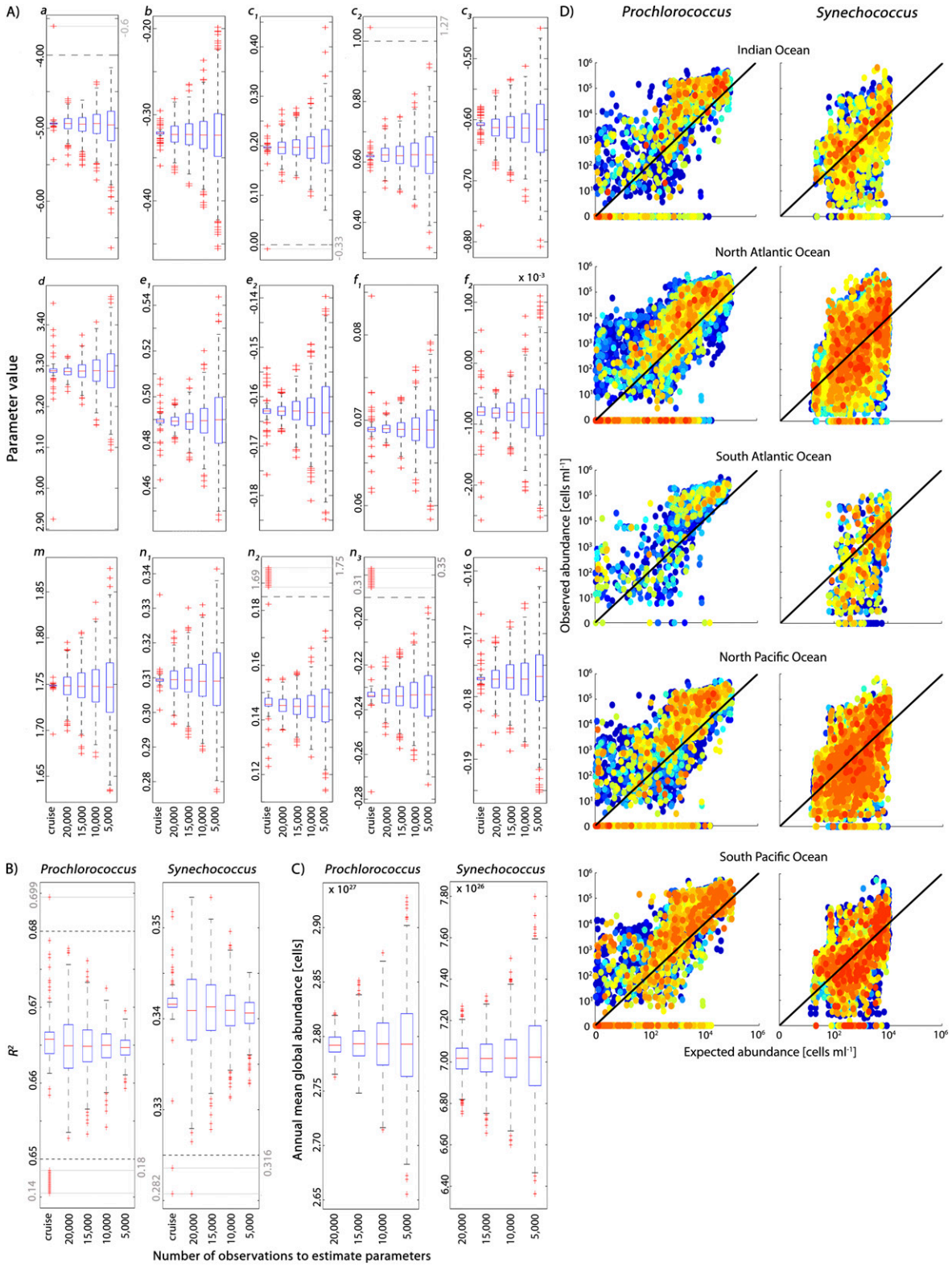

Fig. S4. Sensitivity analysis of the parametric model. We evaluated the effects of single cruises and time series on the model parameters by removing one cruise at a time (labeled cruise; split sampling). The model parameter values were not affected by removing different cruises, showing that the parametric model was well-defined. We evaluated the effects of sample size on the final optimized parameter estimates using an iterative jackknife procedure with 20,000, 15,000, 10,000, and 5,000 observations. Box plot for (*A*) the parameters (Eqs. **2–4**), (*B*) R^2, and (*C*) mean annual global cell abundance of the parametric model. The optimized model parameters seemed insensitive to down sampling of the original global dataset, but their posterior uncertainty increased with decreasing sample size. (*D*) Observed vs. estimated cell abundance for major oceans. We estimated cell abundance using temperature and PAR records associated with each *Prochlorococcus* and *Synechococcus* field sample. Solid black line is the 1:1 relationship.

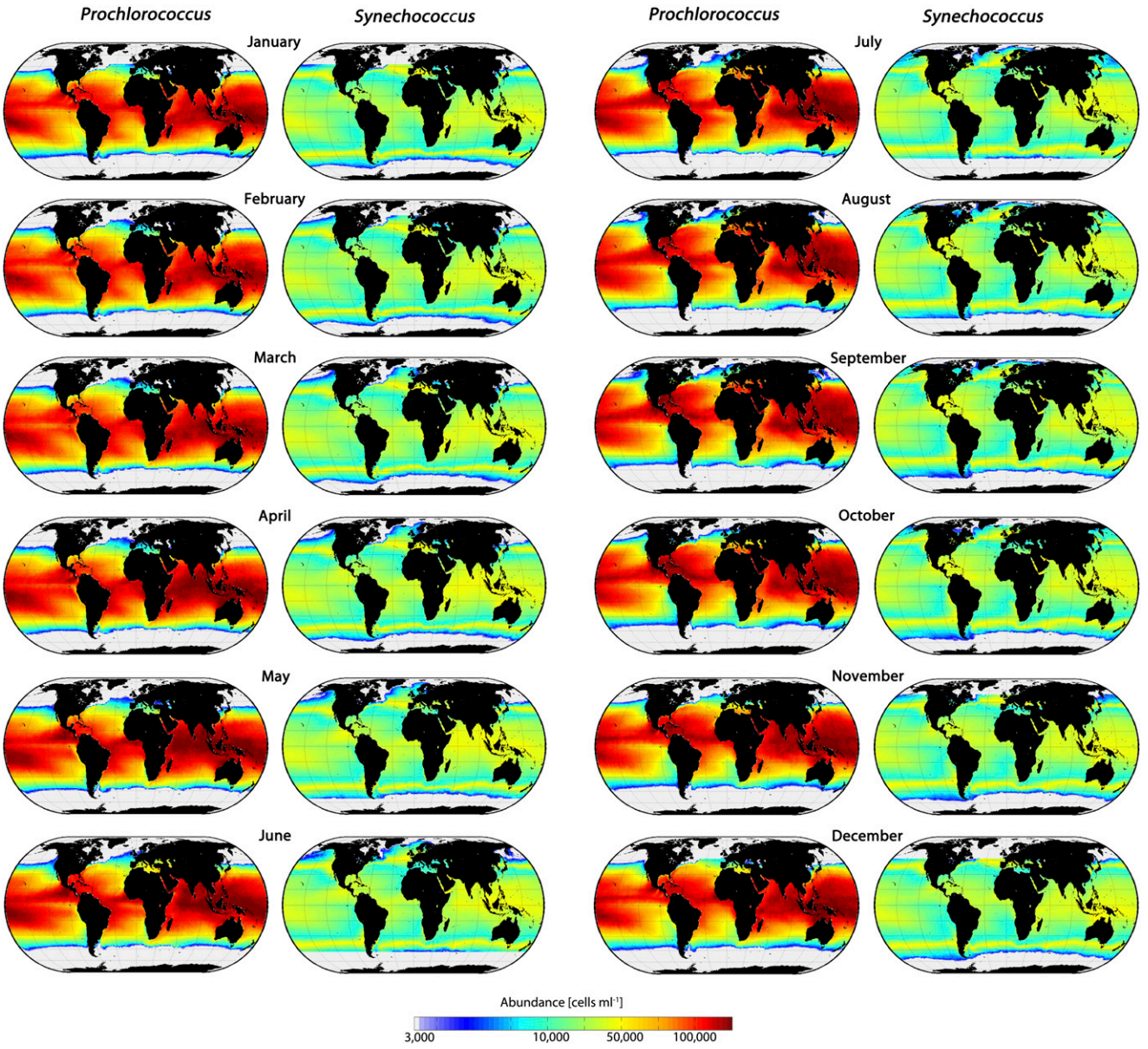

Fig. S5. Monthly sea surface distribution of *Prochlorococcus* and *Synechococcus*. Mean monthly abundances were estimated using SeaWiFS PAR and sea surface temperature as inputs to the niche models.

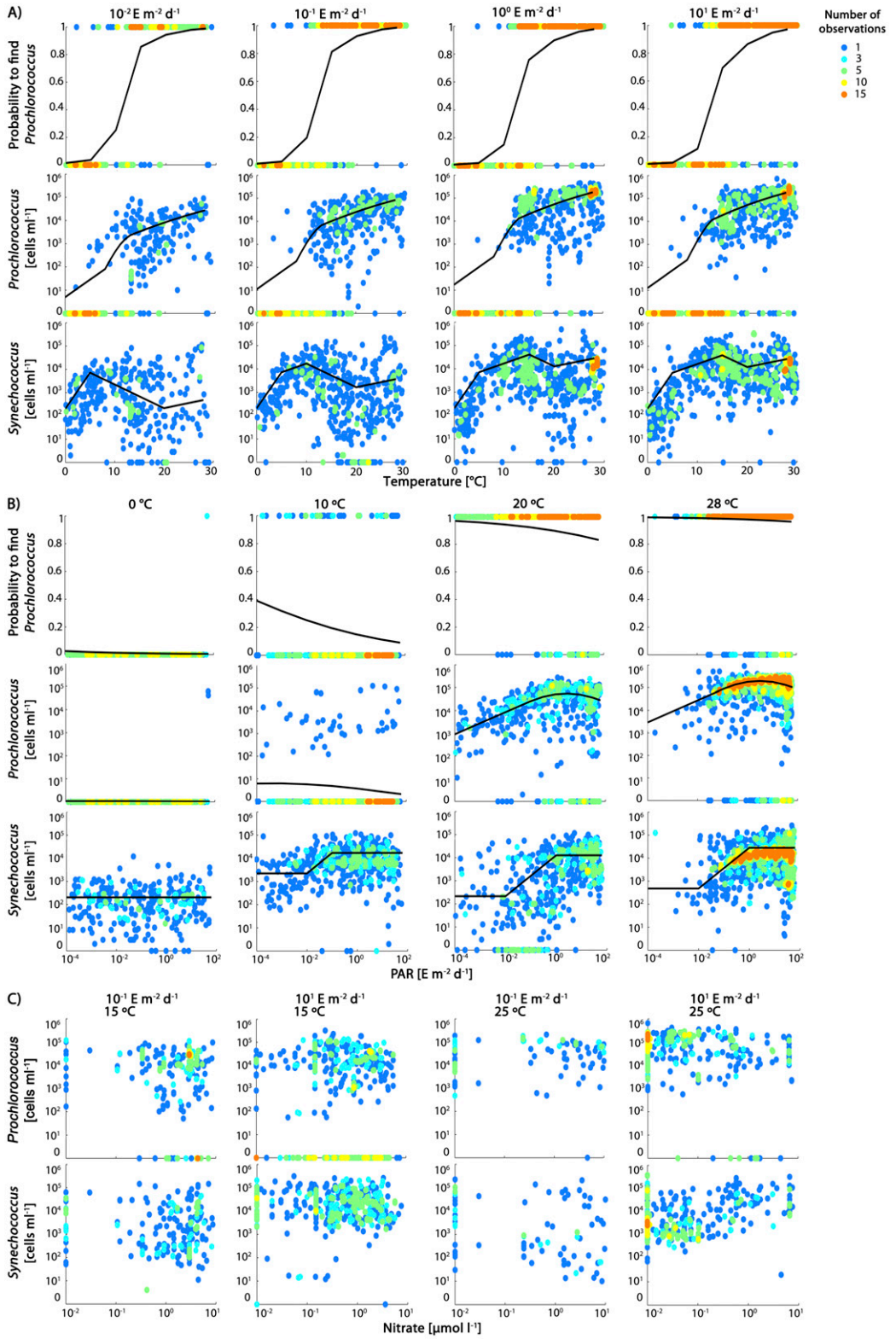

Fig. S6. Observed and estimated *Prochlorococcus* and *Synechococcus* abundances and probability of presence as a function of temperature, PAR, and nitrate concentration. (*A*) Temperature gradients at constant PAR intervals ($10^{x \pm 0.05}$ E m^{-2} d^{-1}). (*B*) PAR gradients at constant temperature intervals ($x \pm 0.05$ °C). (*C*) Nitrate gradients at constant temperature and PAR intervals ($x \pm 0.25$ °C and $10^{x \pm 0.05}$ E m^{-2} d^{-1}, respectively). Symbols represent observed abundance or probability, and symbol color the number of overlapped observations. Lines represent expected abundance or probability. In *A* and *B*, the observed probability to find *Prochlorococcus* equals one for samples with cells abundance $mL^{-1} > 0$ and zero for cells abundance $mL^{-1} = 0$.

Table S1. List of cruises and time series used in this study

ID	Year range		Latitudinal range (°)		Longitudinal range (°)		Number of observations		Ref.
	First	Last	Low	High	Low	High	PRO	SYN	
1	1996	1999	16.1	16.5	119.9	120.4	0	139	1
2; 5	1998	2003	−33.6	49.8	0.6	357.4	518	575	2
3	2003	2003	−47.8	48.0	308.6	348.5	283	283	3
4	2004	2004	−47.0	49.3	309.7	353.4	945	932	4
6	1999	1999	−46.1	−42.5	61.9	64.8	1,035	1,035	5
7	1994	1994	7.9	26.0	56.6	67.1	311	298	6
8	2002	2002	70.5	79.5	4.0	20.0	0	41	7
9	1990	1993	46.1	81.2	144.9	353.3	0	233	8
11	2005	2005	−12.7	17.5	189.6	200.3	119	119	9
12	1989	2006	23.9	61.0	292.6	349.9	13,179	12,791	10
13	2002	2003	43.6	43.8	354.4	354.4	252	252	11
14	2007	2007	−16.0	16.9	190.0	201.0	46	46	12
15	2003	2003	23.6	54.4	202.7	208.0	221	221	13
17–25	2000	2003	−23.1	−20.0	166.1	167.6	443	443	14
26	1990	2005	43.3	43.4	7.5	7.9	407	494	15
27	1995	1995	43.3	43.3	7.5	7.5	286	286	16
28–30	1991	2003	−23.1	21.5	166.1	341.7	1,015	1,033	17
31	1996	1996	−7.6	8.0	179.5	180.5	1,537	1,538	18
32; 43	1992	2003	−23.1	21.5	164.6	341.4	1,970	1,889	19
33	1997	1998	35.2	36.5	358.0	359.9	732	677	20
34	1999	1999	42.1	42.2	293.4	293.4	0	137	21
35	1996	1997	40.6	40.6	22.9	22.9	0	24	22
37	1988	2006	22.8	22.8	202.0	202.0	1,338	1,347	23
38	2001	2004	33.0	33.0	132.5	132.5	422	422	24
39	1995	1995	10.0	24.3	56.5	68.8	2,001	1,443	25
40	1999	2001	29.6	30.6	272.2	275.1	171	162	26
42	2000	2000	5.2	5.2	60.3	60.3	0	30	27
44	2001	2001	39.0	44.5	339.3	343.3	102	102	28
45	1996	1996	36.0	59.3	338.8	341.0	374	374	29
46	1999	1999	31.0	43.4	3.8	359.1	563	610	30
47	1992	1992	−19.6	8.0	164.2	166.2	357	322	31
48	2000	2002	32.9	32.9	242.8	242.8	67	67	32
49	1994	1996	29.8	35.1	235.7	242.7	0	491	33
50	1999	2002	0.0	0.1	145.0	200.0	592	592	34
51	1999	2002	−39.0	50.1	40.9	359.5	1,411	1,578	35
52	1992	1992	−12.2	12.0	219.5	225.5	144	100	36
53	1997	1998	−71.3	−53.0	185.3	194.1	0	259	37
54	1989	2007	20.7	36.9	294.0	303.8	1,384	1,383	38
55	2001	2001	−2.4	26.0	56.6	67.0	375	375	39
56	2004	2004	−34.6	−8.3	218.7	287.6	1,436	1,582	40
57	2001	2002	32.4	33.6	32.3	34.0	112	69	41
58	2005	2005	27.8	66.6	280.5	344.5	183	183	42
59	2000	2003	33.5	33.5	241.6	241.6	132	128	43
60	2002	2004	68.5	73.7	192.6	208.2	0	113	44
61	1999	1999	−64.9	70.4	173.7	200.1	34	34	45
62	1987	1996	−30.0	43.4	2.0	355.9	1,262	1,270	46
63	1989	1995	−43.0	62.8	8.3	340.4	442	521	47
64	2000	2000	−24.0	14.2	0.9	358.4	184	184	48
65	2005	2005	71.3	71.8	202.7	206.4	0	69	49
66; 74; 100	2000	2004	−23.1	44.7	166.1	235.8	729	789	50
68	2000	2000	7.4	8.7	269.2	269.5	10	10	51
69	1999	1999	27.2	29.5	34.1	35.0	63	63	52
70	1999	2000	34.9	34.9	29.5	29.5	90	90	53
71	2005	2005	−7.4	15.1	260.7	279.0	20	20	54
72	1998	1998	22.8	50.0	147.1	202.0	89	89	55
73	2001	2001	24.9	31.1	241.2	247.2	62	62	56
75	1998	2000	44.0	44.0	155.0	155.0	104	104	57
78	1991	1991	−0.0	0.2	165.0	264.6	0	45	58
79	1992	1994	47.1	49.7	294.0	298.0	0	0	59
80; 96	2000	2003	−23.1	44.7	166.1	235.8	1,013	2,148	60
81	2001	2002	10.0	40.3	289.6	314.7	157	145	61
82	2001	2001	34.4	40.0	286.8	288.2	49	41	62

Table S1. Cont.

ID	Year range		Latitudinal range (°)		Longitudinal range (°)		Number of observations		Ref.
	First	Last	Low	High	Low	High	PRO	SYN	
83; 87	2000	2008	−23.1	71.4	166.1	288.6	725	434	63
84	2007	2007	23.2	24.5	200.8	204.0	184	152	64
85	2004	2007	29.8	35.1	235.7	242.7	1,728	1,955	65
86	2003	2004	−34.6	−20.0	1.7	359.3	198	198	66
88	2003	2003	−35.8	−12.2	23.7	113.5	65	57	67
89	1998	1998	26.0	40.2	289.2	295.9	42	51	68
90	2005	2005	46.0	51.0	160.0	215.0	0	48	69
91	1996	1998	−78.0	−52.9	168.9	194.1	0	155	70
92	2001	2001	31.8	33.2	295.1	295.8	11	12	71
93	1999	2000	−32.2	29.9	309.7	342.7	312	312	72
94	1995	1995	36.0	37.1	352.7	355.9	196	197	73
95	2007	2008	31.7	31.7	243.3	243.3	16	19	74
97	2003	2003	−26.0	48.5	140.7	274.2	77	222	75
98	2005	2005	−33.7	−11.0	168.9	179.4	65	65	76
99	2007	2007	42.0	42.0	145.2	145.3	204	204	77
101	1997	1997	6.0	32.0	108.8	129.0	352	345	78
102	1998	1998	24.5	32.0	121.0	129.0	222	222	79
103	2003	2005	−37.7	47.1	2.6	359.7	457	454	80
104	2001	2002	20.1	22.3	113.2	116.2	159	161	81
105	2002	2002	28.6	35.0	121.0	126.8	117	117	82
106	2003	2003	28.4	31.9	122.6	128.9	42	42	83
107	2002	2003	29.0	32.0	122.0	123.5	196	196	84
108	1998	1998	6.3	23.9	110.0	120.0	121	126	85
109	2005	2005	21.7	23.2	116.8	118.9	46	46	86
110	2002	2005	18.5	60.9	111.0	182.3	313	313	80

Longitude and latitude are expressed in decimal scale. ID refers to the identification label in our dataset. PRO and SYN refer to the number of observations with *Prochlorococcus* and *Synechococcus* reported in each dataset.

1. Agawin NSR, Duarte CM, Agusti S, Mc Manus L (2003) Abundance, biomass and growth rates of *Synechococcus sp*. in a tropical coastal ecosystem (Philippines, South China Sea). *Estuar Coast Shelf Sci* 56(3–4):493–502.
2. Heywood JL, Zubkov MV, Tarran GA, Fuchs BM, Holligan PM (2006) Prokaryoplankton standing stocks in oligotrophic gyre and equatorial provinces of the Atlantic Ocean: Evaluation of inter-annual variability. *Deep Sea Res Part 2 Top Stud Oceanogr* 53(14–16):1530–1547.
3. Johnson ZI, et al. (2006) Niche partitioning among *Prochlorococcus* ecotypes along ocean-scale environmental gradients. *Science* 311(5768):1737–1740.
4. Zubkov MV, et al. (2007) Microbial control of phosphate in the nutrient-depleted North Atlantic subtropical gyre. *Environ Microbiol* 9(8):2079–2089.
5. Dubreuil C, Denis M, Conan P, Roy S (2003) Spatial-temporal variability of ultraplankton vertical distribution in the Antarctic frontal zones within 60-66°E, 43-46°S. *Polar Biol* 26(11): 734–745.
6. Tarran GA, Burkill PH, Edwards ES, Woodward EMS (1999) Phytoplankton community structure in the Arabian Sea during and after the SW monsoon, 1994. *Deep Sea Res Part 2 Top Stud Oceanogr* 46(3–4):655–676.
7. Not F, et al. (2005) Late summer community composition and abundance of photosynthetic picoeukaryotes in Norwegian and Barents Seas. *Limnol Oceanogr* 50(5):1677–1686.
8. 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.
9. Karl DM (2005) *BEACH BASH*. Available at http://cmore.soest.hawaii.edu/datasearch/data.php. Accessed February 12, 2008.
10. Li WKW (1998) Annual average abundance of heterotrophic bacteria and *Synechococcus* in surface ocean waters. *Limnol Oceanogr* 43(7):1746–1753.
11. Calvo-Díaz A, Moran XAG (2006) Seasonal dynamics of picoplankton in shelf waters of the southern Bay of Biscay. *Aquat Microb Ecol* 42(2):159–174.
12. Karl DM (2007) *C-MORE Bula*. Available at http://cmore.soest.hawaii.edu/datasearch/data.php. Accessed November 26, 2008.
13. Karl DM (2003) *COOK BOOK1*. Available at http://cmore.soest.hawaii.edu/datasearch/data.php. Accessed March 20, 2008.
14. Neveux J, Tenorio MMB, Dupouy C, Villareal TA (2006) Spectral diversity of phycoerythrins and diazotroph abundance in tropical waters. *Limnol Oceanogr* 51(4):1689–1698.
15. Marty JC, Chiaverini J, Pizay MD, Avril B (2002) Seasonal and interannual dynamics of nutrients and phytoplankton pigments in the western Mediterranean Sea at the DYFAMED time-series station (1991–1999). *Deep Sea Res Part 2 Top Stud Oceanogr* 49(11):1965–1985.
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. Partensky F, Blanchot J, Lantoine F, Neveux J, Marie D (1996) Vertical structure of picophytoplankton at different trophic sites of the tropical northeastern Atlantic Ocean. *Deep Sea Res Part 1 Oceanogr Res Pap* 43(8):1191–1213.
18. Le Borgne R, Landry MR (2003) EBENE: A JGOFS investigation of plankton variability and trophic interactions in the equatorial Pacific (180°). *J Geophys Res: Oceans* 108(C12):8136.
19. Blanchot J, Andre JM, Navarette C, Neveux J (1997) Picophytoplankton dynamics in the equatorial Pacific: Diel cycling from flow-cytometer observations. *C R Acad Sci III* 320(11): 925–931.
20. Jacquet S, Prieur L, Avois-Jacquet C, Lennon JF, Vaulot D (2002) Short-timescale variability of picophytoplankton abundance and cellular parameters in surface waters of the Alboran Sea (western Mediterranean). *J Plankton Res* 24(7):635–651.
21. Sieracki ME, Haugen EM, Cucci TL (1995) Overestimation of heterotrophic bacteria in the Sargasso Sea: Direct evidence by flow and imaging cytometry. *Deep Sea Res Part 1 Oceanogr Res Pap* 42(8):1399–1409.
22. Mihalatou HM, Moustaka-Gouni M (2002) Pico-, nano-, microplankton abundance and primary productivity in a eutrophic coastal area of the Aegean Sea, Mediterranean. *Int Rev Hydrobiol* 87(4):439–456.
23. Campbell L, Nolla HA, Vaulot D (1994) The importance of *Prochlorococcus* to community structure in the central North Pacific Ocean. *Limnol Oceanogr* 39(4):954–961.
24. Katano T, Kaneda A, Takeoka H, Nakano SI (2005) Seasonal changes in the abundance and composition of picophytoplankton in relation to the occurrence of 'Kyucho' and bottom intrusion in Uchiumi Bay, Japan. *Mar Ecol Prog Ser* 298:59–67.
25. Partensky F, Hess WR, Vaulot D (1999) *Prochlorococcus*, a marine photosynthetic prokaryote of global significance. *Microbiol Mol Biol Rev* 63(1):106–127.
26. Murrell MC, Lores EM (2004) Phytoplankton and zooplankton seasonal dynamics in a subtropical estuary: Importance of Cyanobacteria. *J Plankton Res* 26(3):371–382.
27. Larsen A, et al. (2004) Spring phytoplankton bloom dynamics in Norwegian coastal waters: Microbial community succession and diversity. *Limnol Oceanogr* 49(1):180–190.
28. Maixandeau A, et al. (2005) Microbial community production, respiration, and structure of the microbial food web of an ecosystem in the northeastern Atlantic Ocean. *J Geophys Res: Oceans* 110(C7):C07S17.

29. Tarran GA, Zubkov MV, Sleigh MA, Burkill PH, Yallop M (2001) Microbial community structure and standing stocks in the NE Atlantic in June and July of 1996. *Deep Sea Res Part 2 Top Stud Oceanogr* 48(4–5):963–985.
30. Garczarek L, et al. (2007) High vertical and low horizontal diversity of *Prochlorococcus* ecotypes in the Mediterranean Sea in summer. *FEMS Microbiol Ecol* 60(2):189–206.
31. Blanchot J, Rodier M (1996) Picophytoplankton abundance and biomass in the western tropical Pacific Ocean during the 1992 El Niño year: Results from flow cytometry. *Deep Sea Res Part 1 Oceanogr Res Pap* 43(6):877–895.
32. Worden AZ, Nolan JK, Palenik B (2004) Assessing the dynamics and ecology of marine picophytoplankton: The importance of the eukaryotic component. *Limnol Oceanogr* 49(1): 168–179.
33. Collier JL, Palenik B (2003) Phycoerythrin-containing picoplankton in the Southern California Bight. *Deep Sea Res Part 2 Top Stud Oceanogr* 50(14–16):2405–2422.
34. Matsumoto K, Furuya K, Kawano T (2004) Association of picophytoplankton distribution with ENSO events in the equatorial Pacific between 145°E and 160°W. *Deep Sea Res Part 1 Oceanogr Res Pap* 51(12):1851–1871.
35. Dandonneau Y, et al. (2004) Seasonal and interannual variability of ocean color and composition of phytoplankton communities in the North Atlantic, equatorial Pacific and South Pacific. *Deep Sea Res Part 2 Top Stud Oceanogr* 51(1–3):303–318.
36. Landry MR, Kirshtein J, Constantinou J (1996) Abundances and distributions of picoplankton populations in the central equatorial pacific from 12°N to 12°S, 140°W. *Deep Sea Res Part 2 Top Stud Oceanogr* 43(4–6):871–890.
37. Landry MR (2002) *Abundances of Picoplankton, Phytoplankton and Bacteria by Flow Cytometry. (United States Joint Global Ocean Flux Study)*. Available at http://usjgofs.whoi.edu/jg/dir/jgofs/southern/rr-kiwi;6/. Accessed February 3, 2007.
38. DuRand MD, Olson RJ, Chisholm SW (2001) Phytoplankton population dynamics at the Bermuda Atlantic Time-series station in the Sargasso Sea. *Deep Sea Res Part 2 Top Stud Oceanogr* 48(8–9):1983–2003.
39. Zubkov MV, Fuchs BM, Tarran GA, Burkill PH, Amann R (2003) High rate of uptake of organic nitrogen compounds by *Prochlorococcus* Cyanobacteria as a key to their dominance in oligotrophic oceanic waters. *Appl Environ Microbiol* 69(2):1299–1304.
40. Grob C, et al. (2007) Picoplankton abundance and biomass across the eastern South Pacific Ocean along latitude 32.5°S. *Mar Ecol Prog Ser* 332:53–62.
41. Psarra S, et al. (2005) Phytoplankton response to a Lagrangian phosphate addition in the Levantine Sea (Eastern Mediterranean). *Deep Sea Res Part 2 Top Stud Oceanogr* 52(22–23): 2944–2960.
42. Michelou VK, Cottrell MT, Kirchman DL (2007) Light-stimulated bacterial production and amino acid assimilation by Cyanobacteria and other microbes in the North Atlantic Ocean. *Appl Environ Microbiol* 73(17):5539–5546.
43. Fuhrman JA, et al. (2006) Annually reoccurring bacterial communities are predictable from ocean conditions. *Proc Natl Acad Sci USA* 103(35):13104–13109.
44. Campbell RG, et al. (2009) Mesozooplankton prey preference and grazing impact in the western Arctic Ocean. *Deep Sea Res Part 2 Top Stud Oceanogr* 56(17):1274–1289.
45. Meador JA, et al. (2009) Sunlight-induced DNA damage in marine micro-organisms collected along a latitudinal gradient from 70 ° N to 68 ° S. *Photochem Photobiol* 85(1):412–420.
46. Shimada A, Maruyama T, Miyachi S (1996) Vertical distributions and photosynthetic action spectra of two oceanic picophytoplankters, *Prochlorococcus marinus* and *Synechococcus* sp. *Mar Biol* 127(1):15–23.
47. McManus GB, Dawson R (1994) Phytoplankton pigments in the deep chorophyll maximum of the Caribbean Sea and the western tropical Atlantic Ocean. *Mar Ecol Prog Ser* 113: 199–206.
48. Veldhuis MJW, Timmermans KR, Croot P, Van Der Wagt B (2005) Picophytoplankton; A comparative study of their biochemical composition and photosynthetic properties. *J Sea Res* 53(1–2):7–24.
49. Sherr E (2006) *Environmental Variability, Bowhead Whale Distributions, and Inupiat Subsistence Whaling—Whaling Linkages and Resilience of an Alaskan Coastal System*. Available at http://data.eol.ucar.edu/datafile/nph-get/106.219/SNACS06_FCM_Nutrients-readme.html. Accessed August 7, 2007.
50. Longnecker K, Sherr BF, Sherr EB (2005) Activity and phylogenetic diversity of bacterial cells with high and low nucleic acid content and electron transport system activity in an upwelling ecosystem. *Appl Environ Microbiol* 71(12):7737–7749.
51. Saito MA, Rocap G, Moffett JW (2005) Production of cobalt binding ligands in a *Synechococcus* feature at the Costa Rica upwelling dome. *Limnol Oceanogr* 50(1):279–290.
52. Stambler N (2006) Light and picophytoplankton in the Gulf of Eilat (Aqaba). *J Geophys Res: Oceans* 111(C11):C11009.
53. Chung CC, et al. (2011) Effects of Asian dust storms on *Synechococcus* populations in the subtropical Kuroshio Current. *Mar Biotechnol (NY)* 13(4):751–763.
54. Thompson A (2009) Iron and *Prochlorococcus*. PhD thesis (Massachusetts Institute of Technology, Cambridge, MA).
55. Selph KE, et al. (2001) Microbial community composition and growth dynamics in the Antarctic Polar Front and seasonal ice zone during late spring 1997. *Deep Sea Res Part 2 Top Stud Oceanogr* 48(19–20):4059–4080.
56. Millan-Nuñez E, et al. (2004) Specific absorption coefficient and phytoplankton biomass in the southern region of the California Current. *Deep Sea Res Part 2 Top Stud Oceanogr* 51 (6–9):817–826.
57. Liu H, Suzuki K, Minami C, Saino T, Watanabe M (2002) Picoplankton community structure in the subarctic Pacific Ocean and the Bering Sea during summer 1999. *Mar Ecol Prog Ser* 237:1–14.
58. Bruyant F, et al. (2001) An axenic cyclostat of *Prochlorococcus* PCC 9511 with a simulator of natural light regimes. *J Appl Phycol* 13(2):135–142.
59. Doyon P, et al. (2000) Influence of wind mixing and upper-layer stratification on phytoplankton biomass in the Gulf of St. Lawrence. *Deep Sea Res Part 2 Top Stud Oceanogr* 47(3–4): 415–433.
60. Sherr EB, Sherr BF, Wheeler PA (2005) Distribution of coccoid Cyanobacteria and small eukaryotic phytoplankton in the upwelling ecosystem off the Oregon coast during 2001 and 2002. *Deep Sea Res Part 2 Top Stud Oceanogr* 52(1–2):317–330.
61. Zinser ER, et al. (2006) *Prochlorococcus* ecotype abundances in the North Atlantic Ocean as revealed by an improved quantitative PCR method. *Appl Environ Microbiol* 72(1):723–732.
62. Binder BJ, DuRand MD (2002) Diel cycles in surface waters of the equatorial Pacific. *Deep Sea Res Part 2 Top Stud Oceanogr* 49(13–14):2601–2617.
63. Cottrell MT, Kirchman DL (2009) Photoheterotrophic microbes in the Arctic Ocean in summer and winter. *Appl Environ Microbiol* 75(15):4958–4966.
64. Karl DM (2007) *C-MORE Bloomer*. Avalable at http://cmore.soest.hawaii.edu/datasearch/data.php. Accessed March 11, 2008.
65. Landry MR, et al. (2008) Depth-stratified phytoplankton dynamics in Cyclone Opal, a subtropical mesoscale eddy. *Deep Sea Res Part 2 Top Stud Oceanogr* 55(10–13):1348–1359.
66. Bouman HA, et al. (2006) Oceanographic basis of the global surface distribution of *Prochlorococcus* ecotypes. *Science* 312(5775):918–921.
67. Zwirglmaier K, et al. (2008) Global phylogeography of marine *Synechococcus* and *Prochlorococcus* reveals a distinct partitioning of lineages among oceanic biomes. *Environ Microbiol* 10(1):147–161.
68. Cavender-Bares KK, Karl DM, Chisholm SW (2001) Nutrient gradients in the western North Atlantic Ocean: Relationship to microbial community structure and comparison to patterns in the Pacific Ocean. *Deep Sea Res Part 1 Oceanogr Res Pap* 48(11):2373–2395.
69. Fujiki T, Matsumoto K, Honda MC, Kawakami H, Watanabe S (2009) Phytoplankton composition in the subarctic North Pacific during autumn 2005. *J Plankton Res* 31(2):179–191.
70. Landry MR, Kirchman DL (2002) Microbial community structure and variability in the tropical Pacific. *Deep Sea Res Part 2 Top Stud Oceanogr* 49(13–14):2669–2693.
71. Zinser ER, et al. (2007) Influence of light and temperature on *Prochlorococcus* ecotype distributions in the Atlantic Ocean. *Limnol Oceanogr* 52(5):2205–2220.
72. Agawin NSR, Agusti S, Duarte CM (2002) Abundance of Antarctic picophytoplankton and their response to light and nutrient manipulation. *Aquat Microb Ecol* 29(2):161–172.
73. Echevarria F, et al. (2009) Spatial distribution of autotrophic picoplankton in relation to physical forcings: The Gulf of Cadiz, Strait of Gibraltar and Alboran Sea case study. *J Plankton Res* 31(11):1339–1351.
74. Linacre LP, Landry MR, Lara-Lara JR, Hernandez-Ayon JM, Bazan-Guzman C (2010) Picoplankton dynamics during contrasting seasonal oceanographic conditions at a coastal upwelling station off Northern Baja California, Mexico. *J Plankton Res* 32(4):539–557.
75. Sato M, Takeda S, Furuya K (2006) Effects of long-term sample preservation on flow cytometric analysis of natural populations of pico-and nanophytoplankton. *J Oceanogr* 62(6): 903–908.
76. Hashihama F, Sato M, Takeda S, Kanda J, Furuya K (2010) Mesoscale decrease of surface phosphate and associated phytoplankton dynamics in the vicinity of the subtropical South Pacific islands. *Deep Sea Res Part 1 Oceanogr Res Pap* 57(3):338–350.
77. Sato M, Furuya K (2010) Pico- and nanophytoplankton dynamics during the decline phase of the spring bloom in the Oyashio region. *Deep Sea Res Part 2 Top Stud Oceanogr* 57(17–18):1643–1652.
78. Jiao N, et al. (2001) Microscopic overestimation of heterotrophic bacteria in open waters of China Seas. *J Microbiol Biotechnol* 11(5):899–901.
79. Jiao N, Yang Y (2002) Ecological studies on *Prochlorococcus* in China Seas. *Chin Sci Bull* 47(15):1243–1250.
80. Jiao N, et al. (2007) Distinct distribution pattern of abundance and diversity of aerobic anoxygenic phototrophic bacteria in the global ocean. *Environ Microbiol* 9(12):3091–3099.
81. Wu J (2009) A preliminary study of the variation of phytoplankton absorption coefficients in the northern South China Sea. *Acta Oceanolog Sinica* 28(5):17–29.
82. Jiao N, et al. (2007) Ecological anomalies in the East China Sea: Impacts of the Three Gorges Dam? *Water Res* 41(6):1287–1293.
83. Pan LA, Zhang LH, Zhang J, Gasol JM, Chao M (2005) On-board flow cytometric observation of picoplankton community structure in the East China Sea during the fall of different years. *FEMS Microbiol Ecol* 52(2):243–253.
84. Zhang Y, Jiao N (2007) Dynamics of aerobic anoxygenic phototrophic bacteria in the East China Sea. *FEMS Microbiol Ecol* 61(3):459–469.

85. Ning X, et al. (2004) Physical-biological oceanographic coupling influencing phytoplankton and primary production in the South China Sea. *J Geophys Res: Oceans* 109(C10):C10005.
86. Hong HS, Wang YJ, Wang DZ (2011) Response of phytoplankton to nitrogen addition in the Taiwan Strait upwelling region: Nitrate reductase and glutamine synthetase activities. *Cont Shelf Res* 31(6):S57–S66.

Table S2. Selection of environmental factors

	Prochlorococcus					
	Cells per milliliter ≥ 0		Cells per milliliter > 0		*Synechococcus* cells per milliliter ≥ 0	
Predictors	RMSE	R^2	RMSE	R^2	RMSE	R^2
T	1.298	0.662	0.821	0.261	1.072	0.146
N	2.040	0.141	0.888	0.130	1.129	0.055
P	1.943	0.231	0.935	0.030	1.153	0.010
PAR	2.102	0.091	0.804	0.282	1.071	0.145
T N	1.267	0.668	0.796	0.328	1.103	0.085
T P	1.294	0.655	0.805	0.306	1.053	0.164
T PAR	**1.269**	**0.662**	**0.723**	**0.413**	**0.937**	**0.349**
T N P	1.276	0.659	0.806	0.287	1.069	0.147
T N PAR	1.273	0.668	0.729	0.414	0.959	0.319
T P PAR	1.257	0.679	0.732	0.442	0.950	0.335
T N P PAR	1.262	0.674	0.743	0.392	0.938	0.355

ANN analysis for *Prochlorococcus* and *Synechococcus* as a function of environmental variables and all possible combinations (the selected option is in bold). N, nitrate; P, phosphate; R^2, coefficient of determination; T, temperature.

Table S3. Comparison between parametric and local regression models

	Prochlorococcus cells per milliliter ≥ 0		*Synechococcus* cells per milliliter ≥ 0	
Models	RSS	R^2	RSS	R^2
Parametric model	59,630	0.665	32,881	0.341
Local regression model	54,766	0.678	29,436	0.415
Total	169,955		49,920	
Ratio	0.962	0.981	0.895	0.832

Explained variance (R^2) and residual sum of squares (RSS) for the global abundance model. The parametric regression models explained a significant fraction of total variance with two environmental variables and 5 or 10 parameters (Eqs. **1–4**) compared with the local regression model with the same environmental variables but an unconstrained number of parameters. A ratio of R^2 and RSS values close to one indicated a similar performance of the regression and the local regression model. Ratio R^2: parametric divided by local; ratio RSS: local divided by parametric.

Table S4. Fitted parameters for the parametric regression model

	Parameter	Value
Prochlorococcus	a	−4.9348
	b	−0.3224
	c	0.1956
	c_1	0.6243
	c_2	−0.6127
	d	3.2867
	e	0.4893
	e_1	−0.1629
	f	0.0690
	f_1	-8.11×10^{-4}
Synechococcus	m	1.749
	n	0.309
	n_1	0.145
	n_2	−0.234
	o	−0.177

We assumed that C_{PRO}, $C+$, and C_{SYN} are on $\log_{10}$ scale (Eqs. **1–4**). The correction factor gamma (γ) was applied to avoid bias from the logarithm to the decimal scale (γ_{PRO} = 1.8095 and γ_{SYN} = 3.6415). Parameter letters correspond to the parametric regression model (Eqs. **1–4**).