ECOLOGY

Global pattern of phytoplankton diversity driven by temperature and environmental variability

Damiano Righetti[1]*, Meike Vogt[1], Nicolas Gruber[1], Achilleas Psomas[2], Niklaus E. Zimmermann[2,3]

Despite their importance to ocean productivity, global patterns of marine phytoplankton diversity remain poorly characterized. Although temperature is considered a key driver of general marine biodiversity, its specific role in phytoplankton diversity has remained unclear. We determined monthly phytoplankton species richness by using niche modeling and >540,000 global phytoplankton observations to predict biogeographic patterns of 536 phytoplankton species. Consistent with metabolic theory, phytoplankton richness in the tropics is about three times that in higher latitudes, with temperature being the most important driver. However, below 19°C, richness is lower than expected, with ~8°– 14°C waters (~35° to 60° latitude) showing the greatest divergence from theoretical predictions. Regions of reduced richness are characterized by maximal species turnover and environmental variability, suggesting that the latter reduces species richness directly, or through enhancing competitive exclusion. The nonmonotonic relationship between phytoplankton richness and temperature suggests unanticipated complexity in responses of marine biodiversity to ocean warming.

INTRODUCTION

Marine phytoplankton dominate primary production across ~70% of Earth's surface (1), play a pivotal role in channeling energy and matter up the food chain, and control ocean carbon sequestration (2). The diversity of phytoplankton species in open waters has intrigued ecologists for at least half a century (3), but the global pattern of this diversity and its underlying drivers have been unclear (4, 5). This is a critical gap in our understanding of the oceans since the richness of phytoplankton species, a key element of their diversity, may enhance resource use efficiency (6), and thus primary production, as often seen in terrestrial systems (7). For those oceanic taxa that have been investigated, including foraminifera, fish, and invertebrates, species richness tends to peak at low to mid-latitudes and to decline sharply toward the poles (8–11). This decline is consistent with the metabolic theory of ecology (12, 13); i.e., the hypothesis that temperature exerts a key control on metabolic rates and thus promotes speciation and increased species richness in warm tropical areas through time (14). However, a recent large-scale study on marine phytoplankton richness is at odds with the prediction by metabolic theory (4), and latitudinal richness gradients identified for individual phytoplankton groups have taken various shapes (5, 15–17). Observed discrepancies may originate from other factors such as resource competition (17, 18), differences in body size (15, 19), or undersampling of richness (20). Empirical tests of the competing theories explaining global diversity patterns have so far been impeded by the paucity of in situ observations and the lack of systematic sampling schemes for open-ocean phytoplankton (21).

Here, we overcome these limitations and provide the first analysis of marine phytoplankton species richness and its ecological drivers at the global scale, using 1,056,363 presence observations of 1300 species compiled from multiple sources as the basis for our analysis. These data span all major taxa, ocean basins, latitudes, and most seasons (fig. S1). To address the strong spatial and seasonal bias in sampling effort, we analyze the subset of data of species with at least 24 presences (553 species and 699,387 observations) using species distribution models (SDMs), which have been set up specifically to account for uneven sampling and fit each species' ecological niche as a function of multiple environmental predictors. Our statistical approach thus aims at filtering out spurious patterns in the raw data while integrating all observational evidence. We successfully build probabilistic SDMs (fig. S2) for 536 species using generalized additive models (GAM) and use generalized linear models (GLM) and random forest models (RF) to assess the robustness of our findings. We project the species' niches to the global ocean at a 1° resolution and at monthly scales and diagnose richness from the overlap of species' presence-absence projections. Since the diversity of short-lived phytoplankton may change over the course of each year, we project species richness for each month and map its annual mean state, as well as monthly species turnover (see Materials and Methods).
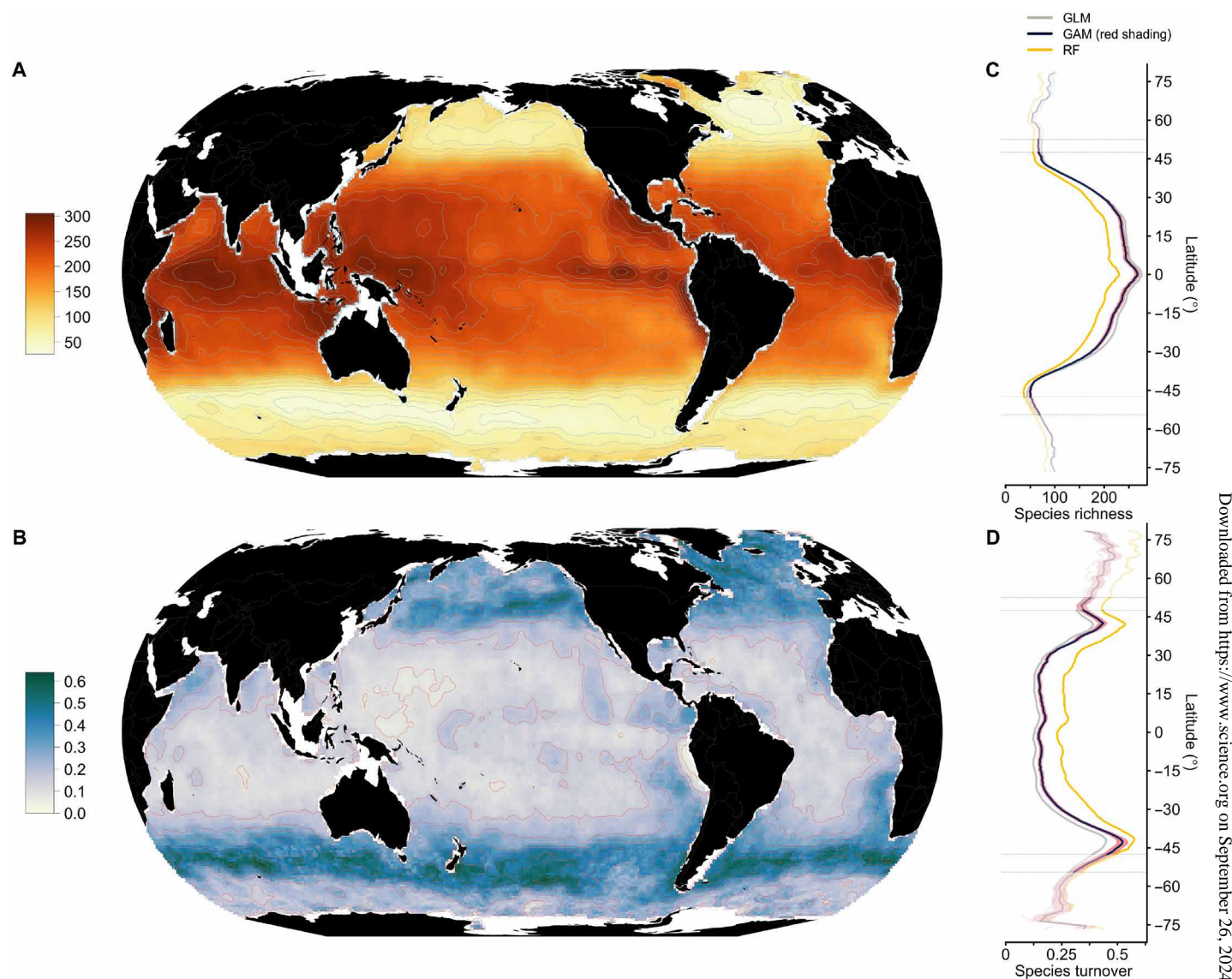
RESULTS AND DISCUSSION

Our diagnosed annual mean of monthly phytoplankton richness varies strongly with latitude, while longitudinal differences are comparably small (Fig. 1A). Phytoplankton richness is highest and least variable throughout the year in the inner tropics (<5°N and S; Fig. 3A), reaching more than 240 species on average. Richness hotspots, where roughly half of the total species analyzed occur simultaneously, are found in the central Indian, the equatorial and west Pacific, and the Indo-Australian Archipelago (Fig. 1A). Thus, hotspots of phytoplankton richness tend to be more tropical than those of foraminifera and other oceanic taxa (8, 9). Analyzed by latitude, richness declines steeply poleward of 30° (Fig. 1C), reaches its minima (~50 species) and associated inflexion points at mid-latitudes (~45° to 65°N and ~45°S), and increases slightly toward the poles. This latitudinal pattern is composed of species with notable wide thermal

Fig. 1. Global patterns of monthly phytoplankton species richness and species turnover. (**A**) Annual mean of monthly species richness and (**B**) month-to-month species turnover projected by SDMs. Latitudinal gradients of (**C**) richness and (**D**) turnover. Colored lines (regressions with local polynomial fitting) indicate the means per degree latitude from three different SDM algorithms used (red shading denotes ±1 SD from 1000 Monte Carlo runs that used varying predictors for GAM). Poleward of the thin horizontal lines shown in (C) and (D), the model results cover only <12 or <9 months, respectively.

ranges (15.8° ± 6.8°C, mean ± SD; Fig. 2B) and broad geographic distributions (Fig. 2C), with more than 60% of high-latitude species (>70°N and S) recorded close to the equator as well (table S1). The latitudinal ranges of most species ($n > 400$) tend to be aggregated between 30°N and 30°S, with relatively few specialist species populating the extratropics (Fig. 2C).

The raw observations corroborate the global latitudinal richness gradient at monthly resolution (fig. S4). The global richness pattern (Fig. 1A), which is diagnosed from these observations by our models, shows seasonal variation that is strongest from subtropical to temperate latitudes (Fig. 3A). Yet, its general form and, in particular, inflexion points (Fig. 1C) are persistent throughout the year and robust to corrections applied to balance the representation of major taxa in our model analysis (i.e., the proportion of species of a specific

taxon captured in our SDMs; fig. S3A). Moreover, the diagnosed richness pattern is decoupled from the spatial density of phytoplankton raw observations ($R^2 = 0.015$, $P < 0.001$, $n = 536$ species), robust against the model algorithm used (GAM versus GLM, $R^2 = 0.99$, $P < 0.001$; GAM versus RF, $R^2 = 0.99$, $P < 0.001$), and largely shared between major taxa (diatom versus dinoflagellate richness, $R^2 = 0.75$, $P < 0.001$; diatom versus haptophyte richness, $R^2 = 0.53$, $P < 0.001$; haptophyte versus dinoflagellate richness, $R^2 = 0.83$, $P < 0.001$). Last, the diagnosed latitudinal richness gradient does not depend on the particular predictor set used in SDMs (Fig. 1C; red shading) and is also robust to choices with respect to the sampling of environmental data for SDMs, here referred to as environmental background (fig. S3B). While we consider the overall gradient as robust, the slopes poleward of ~50° latitude are prone to larger
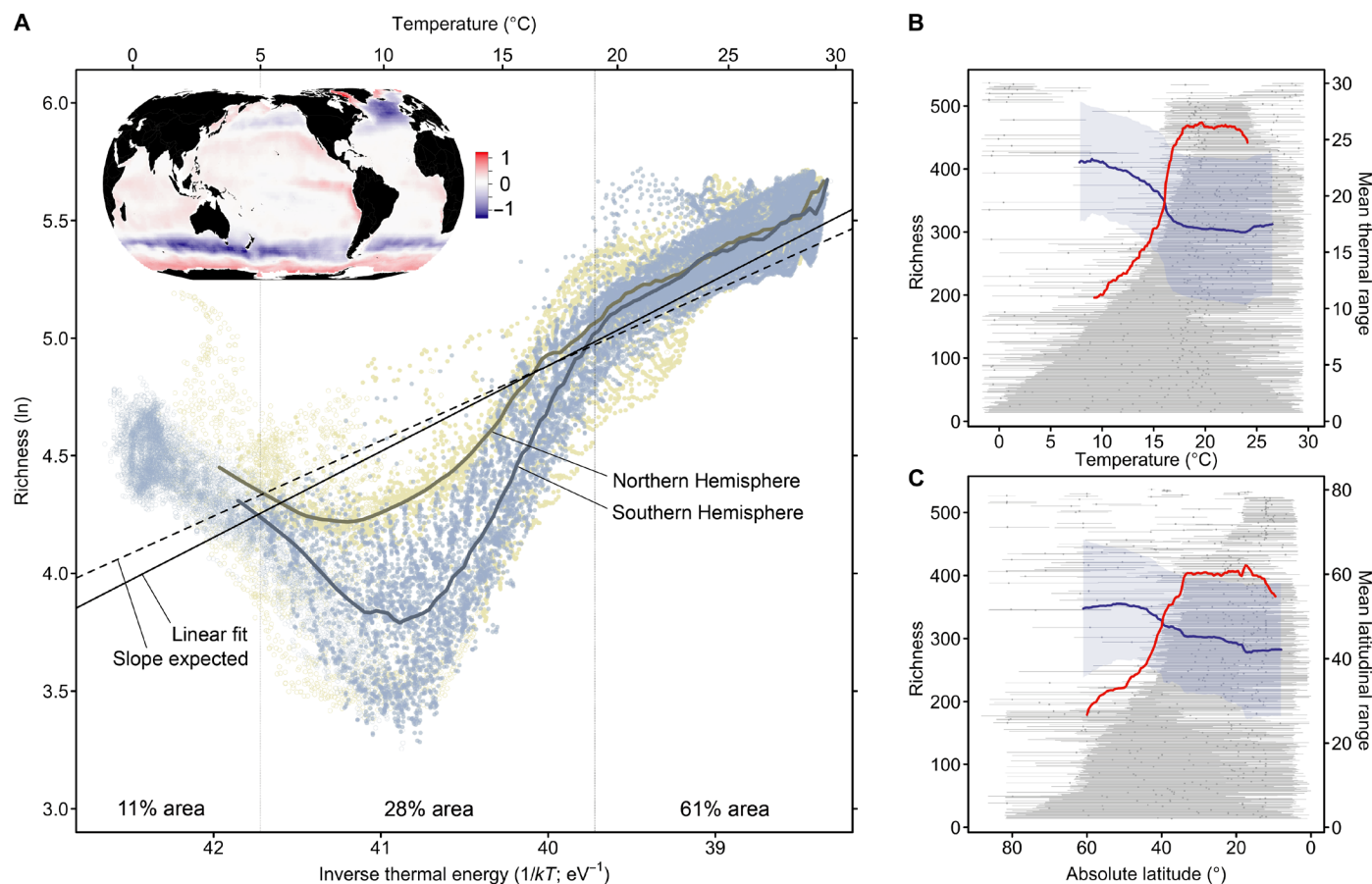
Fig. 2. Relationships between species richness and temperature or latitude. (**A**) The natural logarithm of the annual mean of monthly phytoplankton richness is shown as a function of sea temperature (*k*, Boltzmann's constant; *T*, temperature in kelvin). Filled and open circles indicate areas where the model results cover 12 or less than 12 months, respectively. Trend lines are shown separately for each hemisphere (regressions with local polynomial fitting). The solid black line represents the linear fit to richness, and the dashed black line indicates the slope expected from metabolic theory (−0.32). The map inset visualizes richness deviations from the linear fit. The relative area of three different thermal regimes (separated by thin vertical lines) is given at the bottom of the figure. Observed thermal (**B**) and latitudinal (**C**) ranges of individual species are displayed by gray horizontal bars (minimum to maximum, dots for median) and ordered from wide-ranging (bottom) to narrow-ranging (top). The *x* axis in (C) is reversed for comparison with (B). Red lines show the expected richness based on the overlapping ranges, and blue lines depict the species' average range size (±1 SD, blue shading) at any particular *x* value. Lines are shown for areas with higher confidence.

uncertainty, as winter months lack predictor data in these areas (Fig. 1C).

Sea surface temperature is the most important driver for phytoplankton richness in our data. It explains more than two-thirds of the global variation in diagnosed richness ($R^2 = 0.76$, $P < 0.001$) and is the most powerful predictor for species richness in the underlying raw observations (table S2). Temperature is also the single most important environmental predictor for habitat suitability at the species level (table S2).

The strong role of temperature is consistent with the metabolic theory of ecology (*12*), but while this theory implies a single negative and linear relationship with a slope of −0.32 between logarithmic species richness and inverse thermal energy (*22, 23*), we find three distinctly different regimes across the global ocean (Fig. 2A). Results match metabolic theory best for the ~60% of the ocean surface with annual mean temperatures above 19°C, with a linear regression slope of −0.37 ($R^2 = 0.65$, $P < 0.001$). Between ~19° and 11°C, the slope steepens to −1.14 ($R^2 = 0.69$, $P < 0.001$), and it reverses below ~11°C, with a value of 0.48 ($R^2 = 0.56$, $P < 0.001$). Thus, monthly

phytoplankton richness forms a V-shaped pattern at intermediate to low temperatures. This nonmonotonic relationship is similarly evident in the raw observations (fig. S5, A to E) and consistent with results based on independent in situ data (fig. S5F).

The globally nonmonotonic relationship between phytoplankton richness and temperature (Fig. 2A) suggests that temperature and the metabolic theory of ecology (*12*) provide incomplete explanations for the global-scale variation in phytoplankton richness. We next test whether such variation can be explained by the "physiological tolerance hypothesis" (*24*), which posits that variation in richness is a function of the number of overlapping ecological niches of species. If species had strong affinities for either warm or cold, rather than intermediate temperatures, then a nonmonotonic richness response to temperature may emerge because fewer species tolerate mid-temperatures. On the basis of direct analysis of the thermal ranges of all species in the raw observations, we reject this hypothesis, as the richness obtained by overlapping and summing up these ranges declines monotonically below ~19°C (Fig. 2B). Alternatively, changes in the magnitude of the richness-temperature slope may be in line with metabolic theory
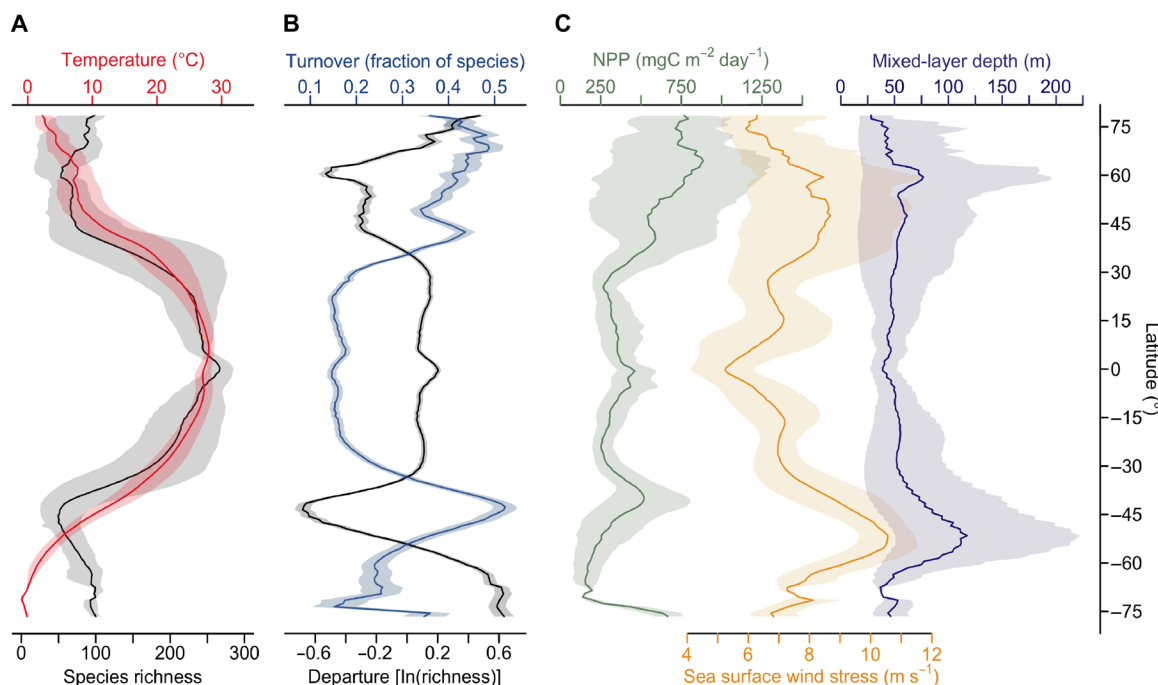
Fig. 3. Latitudinal trends in phytoplankton richness and selected environmental variables. (**A**) Annual mean of monthly species richness (black line) and sea temperature (red line). Shadings indicate the annual amplitude of monthly richness (gray) and temperature (red). (**B**) Departure of richness (black line) from the linear fit (see Fig. 2A) versus species turnover (blue line) by latitude. Shadings denote ±1 SD from Monte Carlo runs. (**C**) Net primary production (NPP; green), sea surface wind stress (orange), and mixed-layer depth (slate blue). Shadings indicate the annual amplitude (minimum to maximum of monthly patterns) for each variable.

but reflect distinct species pools with differing activation energies (*12, 13*). For example, linear slopes shallower than those predicted by metabolic theory have been reported for both marine phytoplankton (*4*) and zooplankton (*10*) richness. We reject this second alternative because the sorting of the thermal ranges within global temperatures (Fig. 2B) does not indicate a thermal separation of species pools, and the ranges of numerous tropical species reach latitudes with temperatures below 19°C (Fig. 2C), suggesting rather smooth shifts in species composition. Thus, we interpret the low richness between ~8° and 14°C (Fig. 2A) as an intermittent, non–temperature-driven suppression of richness, relative to the linear slope predicted by metabolic theory (*12*). This interpretation is favored by the fact that the global linear fit to the modeled richness (slope, −0.37, $R^2 = 0.66$, $P < 0.001$; Fig. 2A) and linear fits to richness in the underlying raw data (fig. S5, A to C) closely match the slope of −0.32 predicted by metabolic theory (*22, 23*). The reversed richness slope at lowest temperatures may signal a relaxation of intensity in the factors driving this suppression.

What could lead to suppressed phytoplankton species richness at temperate latitudes? The most notable feature associated with the global pattern of phytoplankton species richness is its strong and negative correlation with month-to-month species turnover ($R^2 = 0.62$, $P < 0.001$; Fig. 1B). This turnover quantifies the monthly change in species composition in each 1° cell, as projected by our SDMs. While turnover is lowest in the tropics with only ~15% in species being replaced on average between months, it strongly increases poleward of 20° and it peaks at ~40° latitude with monthly turnover rates approaching 50% (Fig. 1D). These areas of maximum turnover are congruent with areas characterized by low richness [Figs. 2A (map) and 3B]. In our analysis, maxima in turnover denote partic-

ularly strong month-to-month variability of those environmental factors that represent statistically determined key dimensions of species' ecological niches. This suggests that temporal variability of the environment is a critical determinant of richness.

Seasonal and short-term environmental variability may reduce richness through abiotic factors, including strong turbulence in the water column, which may select for few species (*25*). This is supported by the negative relationship between richness and sea surface wind stress ($R^2 = 0.63$, $P < 0.001$) or mixed-layer depth (MLD; $R^2 = 0.35$, $P < 0.001$), two proxies for turbulence and light limitation in the water column with strong seasonality at mid-latitudes (Fig. 3C). Furthermore, we find progressively wider tolerance ranges of species in the observational data to increasing wind stress, deeper mixing, and higher nitrate levels (fig. S6, A, B, and F), suggesting a preferential selection for generalist species with broad niches in regions of strong environmental variability and elevated nutrient supply.

Competitive exclusion among species mediated by high temporal environmental variability (*26*) is an additional mechanism that may reduce richness at temperate latitudes. Strong seasonality in upper ocean stratification and associated variations in nutrients and light induce strong seasonality in productivity at mid- to high latitudes (Fig. 3C), with phytoplankton blooms dominated by relatively few species that monopolize local resources (*18, 27, 28*). While our presence-only raw data cannot reveal the dominance hierarchy of species in terms of abundance, our analysis suggests that low monthly richness at mid-latitudes is linked to rapid species turnover. These results provide the first confirmatory evidence to global mechanistic model results that proposed a negative influence of high temporal variability of the environment on phytoplankton richness, mediated

via competitive exclusion (*26*) and transient blooms of opportunists (*29*). However, unlike our results, mechanistic simulations proposed lowest phytoplankton richness at highest (*26*) or subtropical (*29*) latitudes, with the exact location of diversity minima depending on the parameterization of trophic interactions.

Our modeling approach cannot disentangle the relative importance of different mechanisms in generating the latitudinal richness gradient identified. However, our results suggest that environmental variability strongly modifies global gradients in phytoplankton diversity (Fig. 1, C and D), in line with recent findings related to marine bacteria (*30*). Rapoport's rule predicts an increase of species range size with increasing latitude (*31*) and has been used to explain a monotonically decreasing species richness from the equator to poles by an increase in seasonal environmental variability, selecting for wide-ranging species. We confirm a weak trend of increasing range size with increasing latitude (Fig. 2C) and a negative relationship between intra-annual variability and richness. Yet, the variability relevant for short-lived phytoplankton shows distinct mid-latitude peaks (Figs. 1D and 3C), which may explain the emergence of a latitudinal richness gradient that does not conform with the pattern expected by Rapoport's rule.

We do not exclude the possibility that factors besides temperature (table S2) and environmental variability affect richness. Spatial trends in body size and abundance may lead to departures of richness from the slope predicted by metabolic theory (*12*) as recently discussed for a similar nonlinear pattern in freshwater phytoplankton (*32*). Furthermore, richness responses to temperature may level off in the nutrient-poor tropical sea due to a slowdown of metabolic rates under nutrient scarcity, despite high temperatures (*33*). However, tropical richness gradients emerge in our analysis in the absence of clear nutrient gradients (*34*), and continuously increasing nutrient concentrations below ~17°C (*34*) cannot explain the marked change in the richness trend at ~11°C (Fig. 2A). While the effect of environmental variability thus remains a valid hypothesis for this change, finer-scale data are needed to assess the relative roles of nutrients, variability, and temperature in structuring global phytoplankton richness.

In conclusion, synthesizing results across key taxa reveals a first global pattern of phytoplankton species richness in line with a strong role of temperature on the evolutionary outcome of species' habitats and large-scale biodiversity (*12, 13, 22*). In addition, we demonstrate the critical role of environmental variability (*26, 27*) for species turnover and diversity through time. Together, these mechanisms shape a nonmonotonic latitudinal gradient in monthly phytoplankton species richness that is distinctly different from its terrestrial autotrophic counterpart (*35*) and also from most other marine taxa (*8–11*). Climate change is expected to modify both the variability and annual averages in environmental conditions, including temperature and ocean stratification. Our study proposes a link of phytoplankton richness with both temperature and ocean variability; therefore, responses of global patterns in marine phytoplankton diversity to climate change may be more complex than hitherto anticipated, with possible impacts on higher trophic organisms, productivity, and ecosystem function.

MATERIALS AND METHODS
To explore how phytoplankton diversity varies along environmental and spatial gradients, we compiled global phytoplankton presence data and oceanographic predictors. Phytoplankton presence data

were used to calibrate SDMs, designed to compensate for data bias and sample scarceness. Direct analyses of these raw data served as a robustness test for the model results. Furthermore, we used independent phytoplankton data for validation of our results. All analyses were conducted using R language.

Phytoplankton field data
We compiled phytoplankton data from the Global Biodiversity Information Facility (GBIF; https://www.gbif.org, retrieved on 7 December 2015), the Ocean Biogeographic Information System (OBIS; https://www.obis.org, retrieved on 5 December 2015), Villar *et al.* (*36*), and the MAREDAT initiative (table S3) (*37*). The final dataset contained 1,056,363 presence observations from 1298 species and two genera (collectively termed "species"), which were recorded at an average depth of 5.41 ± 6.95 m (mean ± SD) at 182,392 locations in space and time. Observation densities were spatially biased, with 49% of total observations originating from the north Atlantic and only 0.9% originating from the south Atlantic. Methods involved in original data collection included filters (*38*), microscopy (*36, 39*), and flow cytometers (*37*), among others. We retrieved all species observations for seven phyla: *Cyanobacteria*, *Chlorophyta* (excluding macroalgae), *Cryptophyta*, *Myzozoa*, *Haptophyta*, *Ochrophyta*, and *Euglenozoa*. More specifically, among the *Ochrophyta*, we included the classes *Bacillariophyceae*, *Chrysophyceae*, *Pelagophyceae*, and *Raphidophyceae*. Among the *Myzozoa*, we considered the class *Dinophyceae*. Among the *Euglenozoa*, we considered the class *Euglenoidea*. In addition, we compiled observations of *Prochlorococcus* and *Synechococcus* from the data sources. These genera are globally abundant but rarely identified to a species level. We excluded records (i) if they were listed as "fossil specimen" or "preserved specimen," (ii) if they were associated with year of collection >2015 or <1800, (iii) if they were associated with negative depths, or (iv) if they were associated with nonsensible coordinates. We removed data below the monthly climatological mixed layer based on the temperature criterion (*40*), as data at depth were insufficient to develop species models. However, for "mixed-layer species" (i.e., species recorded in the mixed layer), we assumed that data without depth indication stem from the mixed layer as well. Species names in the original data were harmonized following expert opinion (see Supplementary Materials and Methods). The dataset spanned all major phytoplankton taxa of the marine realm (*41*) and reflected, within the bounds of uncertainty, roughly similar factions of the total species known among key taxa (table S4).

Independent data (<2.9% of the observations overlap with the above dataset) (*39*), primarily based on Atlantic transect cruises, served for validation of results obtained from the main dataset. These independent data were collected on the basis of a consistent methodology by the same taxonomist, containing 4217 presence observations from 303 phytoplankton species within the mixed layer (*39*).

Open-ocean definition
To reduce confounding influences of more complex and fertile coastal environments on our open ocean phytoplankton analyses, we excluded data from seas shallower than 200 m (*42*) and from seas with surface salinities below 20 (*43*).

Environmental data
We compiled data on 10 environmental variables that represent key dimensions of phytoplankton ecological niches (*44–46*), which

shape species' distributions via effects on physiology, growth, and species competition (17, 25, 47). These variables served as candidate predictors for SDMs and as single predictors in species richness models. Variables were aggregated at a monthly ($n = 12$) climatological and globally gridded resolution (1° latitude × 1° longitude), as this was the best available resolution shared among datasets. Sea surface temperature (T; °C), salinity (S), nitrate (NO_3-; μM), phosphate (PO_4^{3-}; μM), and silicic acid [$Si(OH)_4$; μM] were obtained from World Ocean Atlas 2013 (1955–2012) fields (34, 43, 48). MLD (m) was included using the temperature criterion (40). Photosynthetically active radiation (PAR; μmol m^{-2} s^{-1}) and chlorophyll (Chl; μg liter^{-1}) were derived from the Sea-viewing Wide Field-of-view Sensor, using data from 1997 to 2007 (https://oceancolor.gsfc.nasa.gov). Sea surface wind stress (m s^{-1}) was derived from the Cross-Calibrated Multi-Platform (49) using data from 1987 to 2011 (https://podaac.jpl.nasa.gov). Data on carbon dioxide partial pressure in the surface sea (pCO_2; μatm) stem from (50).

We derived further predictors from these 10 variables: Photosynthetically available radiation over the MLD (MLPAR; μmol m^{-2} s^{-1}) was computed from T, MLD, and Chl (44). We also used the excess concentration of NO_3^-, relative to the Redfield ratio of 16:1 (μM; N-star), computed as $[NO_3^-] - 16[PO_4^{3-}]$. The use of N-star, rather than NO_3^-, avoids strong global correlations between NO_3^- and T (Spearman's ρ = −0.71). Si-star, the ratio of $[Si(OH)_4]$ to $[NO_3^-]$, was included as a predictor particularly relevant for the *Bacillariophyceae* (45). We also considered the temporal trends of T (dT/dt), NO_3^- (dNO_3^-/dt), PO_4^{3-} (dPO_4^{3-}/dt), $Si(OH)_4$ [$dSi(OH)_4$/dt], and MLD (dMLD/dt), calculated as the centered mean difference of the data of each month with its neighboring months. Logarithmic MLD, Chl, and nutrient levels to the base of 10 were used in addition to their original forms. Sea level height anomaly (m) (https://www.aviso.altimetry.fr/es/data/products/sea-surface-height-products/global/index.html) and nutricline depth (m), defined as the first depth at which nutrient levels exceeded a certain threshold (0.05 μM for NO_3^- and 0.05:16 μM for PO_4^{3-}), were also tested, yet these two variables were discarded due to poor skill in single-predictor model tests. Net primary production (NPP; mg C m^{-2} day^{-1}) was used for correlative analyses only, using data from 1998 to 2007 from the standard vertically generalized production model (http://www.science.oregonstate.edu/ocean.productivity).

Species distribution models

SDMs fit statistical associations between species' observed presences and environmental variables; i.e., they estimate a species' realized ecological niche (44). SDMs provide a useful framework to explore large-scale distributions of phytoplankton species, based on two major assumptions: (i) Species are not dispersal limited in the open ocean (51, 52) [but see (53)], a trait consistent with the generally wide geographic ranges of the species in our data; (ii) species are primarily controlled by environmental factors in their global distribution (36, 47, 52) and rapidly proliferate when conditions become suitable (18, 27). Applying SDMs to study marine phytoplankton distributions has emerged relatively recently (45). Since distribution patterns of phytoplankton species might change seasonally (30), reflecting the generally short generational cycles of phytoplankton owing to their largely microbial nature, we used a monthly matchup between species' presences and environmental variables to calibrate the niches. We then projected niches onto global environmental data fields at 1° and monthly resolution to obtain maps of species'

presence. We developed SDMs that address three principal sources of uncertainty: (i) biases in sampling effort, (ii) predictor selection, and (iii) algorithm choice. All steps used to build SDMs are described herein.

Data binning

We used phytoplankton presence data, rather than abundance data, as the former are less sensitive to differences in sampling methods and are more widely available. We binned species' presence observations into monthly 1° latitude × 1° longitude resolution to match the resolution of environmental predictors. Multiple observations per species and 1° cell that stemmed from the same month, but potentially from different years, thus counted as a single presence, resulting in a total of 245,322 species presences. The monthly data binning may have removed signals of temporal changes in species' distributions throughout the years. However, since data originated predominantly from a few decades between 1950 and 2000 (1984 ± 17; mean ± SD) and since climatic changes during this period were much smaller relative to current global amplitudes of environmental factors (for example, sea surface temperature spans ~ −1.8 to ~32°C) (48), we expect such changes to have only a minor impact on global SDM projections.

Environmental background data

Since absence data for phytoplankton are unreliable on the basis of traditional sampling methods (20) but required by our presence-absence SDMs, we selected background data (also termed pseudo-absences) for each species, using a so-called target-group approach (54). This approach addresses spatial and temporal sampling biases in field-based presence data of species via the selection of the number and location of pseudoabsences. We defined large groups of species as target groups, assuming that variation in sampling effort applied to the entire target group reflected variation in sampling applied to each species within the target group (54). The sampling of species' background data from the target group served two purposes: (i) Background sampling followed a sampling scheme similar to that of the species' presence data (and thus received similar bias), thereby balancing presence data bias when fitting SDMs; (ii) extensive ocean areas, which lacked sampling, were not misclassified as areas of species' absences. We used the *Bacillariophyceae*, *Dinoflagellata*, and *Haptophyta* separately to define "group-specific target groups" for their constituent species, as these taxa had different global sampling schemes. For the remaining taxa, the number of species was insufficient to build group-specific target groups. For these taxa, we used the total species as the target group, excluding *Bacillariophyceae*, as presence data of the latter were strongly north-south imbalanced.

In parallel, we used the total species as target group to sample the background for each species, which we refer to as "total target group" approach. We found that richness results were robust to the use of total versus group-specific target groups (fig. S3B).

We sampled background data in a stratified manner from the target group, dividing both the T and MLD gradient (spanned by the target group) into nine equally spaced intervals, yielding 81 strata (T × MLD combinations). Sampling data from each stratum separately assured that the breadth of these two key environmental factors was reflected in the backgrounds of species. The target group's presence data were gridded at monthly 1° resolution, before sampling backgrounds from it. We tested whether the density of these monthly 1° cells of the target group reflected original sampling efforts (approximated by the number of samples in the raw data) and found that the two measures were highly correlated (Spearman's

$\rho = 0.94$ for latitude; Spearman's $\rho = 0.99$ for longitude; binning data at 1° latitude or 1° longitude, respectively). For each species, we sampled 10 times more background data than the species had presences (55). Within each of the 81 strata, background data were randomly sampled. The amount of background data sampled from a specific stratum was proportional to the number of monthly 1° cells provided by the target group in this stratum, thereby reflecting original sampling efforts.

Statistical complexity

Statistical algorithm choice represents a key source of uncertainty in SDMs (56). We constructed SDMs based on either GLM (using the R package *stats*), GAM (R package *mgcv*), or RF (R package *randomForest*), as three algorithms of increasing statistical response shape complexity (57). We considered the GAM as our standard algorithm because of its intermediate complexity. We used comparably few predictors ($n = 4$) in models and fitted simple response shapes to account for the relatively few presences of most phytoplankton species (57). GLM included linear and quadratic terms and a stepwise bidirectional predictor selection procedure. GAM used smoothing terms with five basis dimensions, estimated by penalized regression splines without penalization to zero for single variables. To equalize the overall weight of presences versus background data per species, background data in GAM and GLM were weighted by the ratio of species' presence to background data points. RFs included 4000 trees, simple terms, and single end node size. The weighting of data in individual RF trees was balanced by randomly subsampling same amounts of background data as the species had presences.

Single predictor skill tests

In addition to algorithm choice, predictor choice represents a major source of uncertainty in phytoplankton SDMs, as these organisms are not well studied regarding their most important niche factors. To select powerful predictors for SDMs, we assessed the individual skill of an extensive number of candidate predictors ($n = 25$) in discriminating species' presences versus background data. The results of this test also served to identify the key environmental drivers of species' distributions independently of the SDM analysis. We fitted single-factor GLM, GAM, and RF models to the presences versus background data of each species, for each candidate predictor. The species ($n = 567$) considered for predictor analyses generally contained a minimum of 24 presences as was used as a lower threshold for species in SDMs. Model explanatory skill was evaluated using the adjusted D^2 (for GLM and GAM) and the out-of-bag error (for RF) statistic. For each species, predictors were ranked according to these statistics, and the mean variable ranks obtained across GLM, GAM, and RF served as a basis for predictor selection. We performed several sensitivity tests to evaluate the robustness of the predictor ranking. We compared rarely versus more frequently sampled species (i.e., ≥ 15, ≥ 24, and ≥ 50 presences), used different variables for the stratification of background sampling, and applied spatial thinning of species' presences to a distance ≥ 300 or ≥ 600 km (using the R package *spThin*), which reduces potentially confounding effects of autocorrelation. None of these modifications changed the result that temperature was the top-ranked predictor across total species. However, the rank of predictors other than temperature tended to vary between setups.

Predictor choice for models

To capture predictor-based uncertainties, we fitted five member models, each using a different set of four predictors, for building an SDM ensemble. The species ($n = 567$) considered for modeling contained ≥ 24 presences, which corresponds to a presence-to-predictor ratio ≥ 6 per species. We used a randomization approach to select the four predictors per member model, using the test-based predictor ranking (see above) of each species as a basis. For the first member model, we selected four predictors at random, without replacement from those predictors that ranked among the 10 most powerful predictors per species. We omitted Spearman's rank correlations between predictors greater than $|0.7|$ in each predictor set (computed from the predictor data at global monthly 1° resolution). Predictors of the four other members were composed by the same criterion. Yet, we allowed each predictor to be selected only up to twice among the five members to omit biases due to overrepresentation of individual predictors in SDMs. If sampling among the top 10 predictors did not provide a sufficient number of predictors for the 5 sets × 4 predictors (given the correlation criterion), candidate predictors that ranked >10 were selected. Predictors were equally used in GAM, GLM, and RF.

Monte Carlo simulation

We used a Monte Carlo simulation to quantify uncertainty in our results emerging from the choice of different predictor sets. For each species modeled, we randomly selected one of the five predictor sets prepared (see above) to fit the SDM of the species and then calculated richness and turnover of total species. This procedure was repeated ($n = 1000$ runs) and we present the SD across the runs (Figs. 1, C and D, and 3B).

Evaluation of model members and ensemble construction

For each species, we evaluated the predictive skill of each member model based on a repeated (4×) split-sample cross-validation test. In this test, the species' presences and background data are randomly split into four parts. The SDM member model is iteratively trained on the basis of three parts (75%) of the data and used to predict the remaining one-fourth (25%) of the data. The predicted values are then compared against the true values. We calculated the true skill statistic (TSS) (58) of this test. TSS ranges from −1 to +1 with values greater than zero indicating models performing better than at random. We retained member models with a TSS score of at least 0.35 for the construction of our SDM ensembles. Successful member models were then projected globally onto monthly ($n = 12$ months) environmental data fields, yielding probabilistic maps of species' presence. Presence probabilities were generally higher for high-latitude species than for lower-latitude species, as ecological niches at high latitudes were readily captured by SDMs. To avoid spatial biases in multispecies analyses, we therefore binarized the projected probabilities to presence-absence from thresholds maximizing the TSS (package *presenceAbsence*). For each month, we averaged the successful member models of each particular species to obtain monthly ensemble mean projections. Each species thus obtained a value between zero and one per monthly 1° cell. We did not further binarize the ensemble projections to presence-absence, as binarization tends to overestimate the species' presence toward the edges of the projected presence area, relative to its center. We hence argue that our ensemble mean projections characterize species' distribution patterns at a higher level of detail compared to 0/1 projections and are better suited for multispecies analyses, in line with previous work showing that the sum of overlapping 0/1 projections tended to overestimate species richness (59).

Global species richness map

We used the summation of monthly SDM ensemble projections of the successfully modeled species (536 species for GAM, 529 species

for GLM, and 538 species for RF) to obtain 12 monthly estimates of phytoplankton species richness (data file S1). The global species richness map (Fig. 1A) represents the average of these monthly estimates. We examined the robustness of the global species richness pattern with respect to balancing the relative representation of major taxa in our model analysis (fig. S3A). For each major taxon (table S4), we divided the number of species included in SDMs by the number of its totally known species, and we weighted the individual species inversely to this ratio in the corrected global richness pattern (fig. S3A).

Global species turnover map

Species turnover is a measure of the difference in species composition between two ecological communities. We assessed species turnover through time, analyzing the fraction of species whose identities changed between months ("species replacement"). We derived species communities at monthly 1° resolution from the species' monthly ensemble mean projections (for each monthly 1° cell, species with values larger than 0.5 counted as present). We calculated species turnover at 1° resolution between consecutive months ($n = 12$ monthly pairs) and averaged the results of the monthly pairs over the full year to obtain the species turnover map (Fig. 1B). We used the turnover component of the Jaccard dissimilarity as implemented in R's *betapart* package.

Explanatory power of single variables for global richness patterns

We tested whether richness itself could be modeled successfully using environmental variables. We fitted species richness using GLMs with a single predictor, thus assessing their univariate predictive skill (table S2). We fitted GLMs both to the species richness obtained from SDMs (aggregated from GAM, at monthly 1° resolution) and to the species richness observed at a sample level in the raw data (table S2). Each local sample was defined by a unique combination of latitude, longitude, depth, year, month, and day of collection in the phytoplankton raw data. Environmental variables were matched with richness values at 1° resolution, using monthly climatological data. The quality of linear model fits between single variables and GAM-based richness (at monthly 1° resolution) presented in the text is indicated by the R^2 statistic.

Global richness patterns in aggregate observational data

The number of detected species varied strongly between samples (fig. S1, B and C). To infer robust species richness patterns directly from raw observations (figs. S4 and S5), we split raw data into coarse environmental or geographic strata and repeatedly selected a certain number of samples at random, from each stratum. Determining the integral richness from the pooled samples, per stratum, served to depict richness patterns in the observational data. This approach, in part, ameliorates the problem of severe undersampling of richness by traditionally small sampling volumes (*20*).

Sensitivity of richness patterns to species detection

Our analysis included an extensive number of phytoplankton species and spanned all major phytoplankton taxa (table S4). However, results are limited to species for which in situ data were available. Patterns thus reflect the subset of detectable, and likely common, species, while potentially overlooking an unknown number of low-abundance species. We tested whether richness patterns in our data were robust to the omission of species that were relatively more rarely observed (figs. S4 and S5). We found that patterns were largely robust to their exclusion, in line with previous findings showing that global richness patterns are best approximated by subsets of common (rather than by rare) species (*60*).

Species range analysis

Species' observed ranges for key environmental factors were defined by the maximum and minimum value obtained from matching up the species' raw observations with sea surface temperature (*48*), sea surface wind stress (*49*), MLD (*40*), and additional factors, at monthly 1° resolution (Fig. 2B and fig. S6). Species' latitudinal ranges corresponded to the maximum and minimum latitude of raw observations (Fig. 2C). We assessed the expected maximum species richness over the full range of each environmental variable in the ocean based on species' observed presence ranges. At each environmental value (Fig. 2, B and C, and fig. S6), maximum richness is defined as the sum of the species' ranges that overlapped, and range size is defined as the species' mean range size.

To denote parts of the environmental gradients where edge effects likely distorted species' observed range patterns, we created a null model of the environmental response in expected maximum richness and average range size. To this end, we randomized the species' presences in the observational data ($n = 100$ runs) and derived species' range limits and expected richness/mean range size based on these randomized data. Thus, our null model represents the expected maximum richness/mean range size in absence of any environmental sorting of species' ranges. We identified areas with likely edge effects as those areas where the null model values deviated significantly (using the 95% confidence interval range) from the null model value obtained in the center (i.e., not affected by edge effects) of the environmental gradient studied. In Fig. 2 (B and C) and fig. S6, we only show estimates of maximum richness (red line) and range size (blue line) for areas unaffected by edge effects.

SUPPLEMENTARY MATERIALS

REFERENCES AND NOTES

1. C. B. Field, M. J. Behrenfeld, J. T. Randerson, P. Falkowski, Primary production of the biosphere: Integrating terrestrial and oceanic components. *Science* **281**, 237–240 (1998).
2. L. Guidi, S. Chaffron, L. Bittner, D. Eveillard, A. Larhlimi, S. Roux, Y. Darzi, S. Audic, L. Berline, J. R. Brum, L. P. Coelho, J. C. Ignacio-Espinoza, S. Malviya, S. Sunagawa, C. Dimier, S. Kandels-Lewis, M. Picheral, J. Poulain, S. Searson; Tara Oceans Coordinators, L. Stemmann, F. Not, P. Hingamp, S. Speich, M. Follows, L. Karp-Boss, E. Boss, H. Ogata, S. Pesant, J. Weissenbach, P. Wincker, S. G. Acinas, P. Bork, C. de Vargas, D. Iudicone, M. B. Sullivan, J. Raes, E. Karsenti, C. Bowler, G. Gorsky, Plankton networks driving carbon export in the oligotrophic ocean. *Nature* **532**, 465–470 (2016).
3. G. E. Hutchinson, The paradox of the plankton. *Am. Nat.* **95**, 137–145 (1961).

4. T. Rodríguez-Ramos, E. Marañón, P. Cermeño, Marine nano- and microphytoplankton diversity: Redrawing global patterns from sampling-standardized data. *Glob. Ecol. Biogeogr.* **24**, 527–538 (2015).

5. P. Cermeño, S. Dutkiewicz, R. P. Harris, M. Follows, O. Schofield, P. G. Falkowski, The role of nutricline depth in regulating the ocean carbon cycle. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 20344–20349 (2008).

6. R. Ptacnik, A. G. Solimini, T. Andersen, T. Tamminen, P. Brettum, L. Lepistö, E. Willén, S. Rekolainen, Diversity predicts stability and resource use efficiency in natural phytoplankton communities. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 5134–5138 (2008).

7. J. Liang, T. W. Crowther, N. Picard, S. Wiser, M. Zhou, G. Alberti, E. D. Schulze, A. D. McGuire, F. Bozzato, H. Pretzsch, S. de-Miguel, A. Paquette, B. Hérault, M. Scherer-Lorenzen, C. B. Barrett, H. B. Glick, G. M. Hengeveld, G. J. Nabuurs, S. Pfautsch, H. Viana, A. C. Vibrans, C. Ammer, P. Schall, D. Verbyla, N. Tchebakova, M. Fischer, J. V. Watson, H. Y. H. Chen, X. Lei, M. J. Schelhaas, H. Lu, D. Gianelle, E. I. Parfenova, C. Salas, E. Lee, B. Lee, H. S. Kim, H. Bruelheide, D. A. Coomes, D. Piotto, T. Sunderland, B. Schmid, S. Gourlet-Fleury, B. Sonké, R. Tavani, J. Zhu, S. Brandl, J. Vayreda, F. Kitahara, E. B. Searle, V. J. Neldner, M. R. Ngugi, C. Baraloto, L. Frizzera, R. Balazy, J. Oleksyn, T. Zawiła-Niedźwiecki, O. Bouriaud, F. Bussotti, L. Finér, B. Jaroszewicz, T. Jucker, F. Valladares, A. M. Jagodzinski, P. L. Peri, C. Gonmadje, W. Marthy, T. O'Brien, E. H. Martin, A. R. Marshall, F. Rovero, R. Bitariho, P. A. Niklaus, R. Alvarez-Loayza, N. Chamuya, R. Valencia, F. Mortier, V. Wortel, N. L. Engone-Obiang, L. V. Ferreira, D. E. Odeke, R. M. Vasquez, S. L. Lewis, P. B. Reich, Positive biodiversity-productivity relationship predominant in global forests. *Science* **354**, aaf8957 (2016).

8. D. P. Tittensor, C. Mora, W. Jetz, H. K. Lotze, D. Ricard, E. V. Berghe, B. Worm, Global patterns and predictors of marine biodiversity across taxa. *Nature* **466**, 1098–1101 (2010).

9. S. Rutherford, S. D'Hondt, W. Prell, Environmental controls on the geographic distribution of zooplankton diversity. *Nature* **400**, 749–753 (1999).

10. I. Rombouts, G. Beaugrand, F. Ibañez, S. Chiba, L. Legendre, Marine copepod diversity patterns and the metabolic theory of ecology. *Oecologia* **166**, 349–355 (2011).

11. M. Yasuhara, G. Hunt, H. J. Dowsett, M. M. Robinson, D. K. Stoll, Latitudinal species diversity gradient of marine zooplankton for the last three million years. *Ecol. Lett.* **15**, 1174–1179 (2012).

12. J. H. Brown, J. F. Gillooly, A. P. Allen, V. M. Savage, G. B. West, Toward a metabolic theory of ecology. *Ecology* **85**, 1771–1789 (2004).

13. A. P. Allen, J. H. Brown, J. F. Gillooly, Global biodiversity, biochemical kinetics, and the energetic-equivalence rule. *Science* **297**, 1545–1548 (2002).

14. A. P. Allen, J. F. Gillooly, V. M. Savage, J. H. Brown, Kinetic effects of temperature on rates of genetic divergence and speciation. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 9130–9135 (2006).

15. S. Malviya, E. Scalco, S. Audic, F. Vincent, A. Veluchamy, J. Poulain, P. Wincker, D. Iudicone, C. de Vargas, L. Bittner, A. Zingone, C. Bowler, Insights into global diatom distribution and diversity in the world's ocean. *Proc. Natl. Acad. Sci. U.S.A.* **113**, E1516–E1525 (2016).

16. S. Honjo, H. Okada, Community structure of coccolithophores in the photic layer of the mid-Pacific. *Micropaleontology* **20**, 209–230 (1974).

17. C. J. O'Brien, M. Vogt, N. Gruber, Global coccolithophore diversity: Drivers and future change. *Prog. Oceanogr.* **140**, 27–42 (2016).

18. S. Lehtinen, T. Tamminen, R. Ptacnik, T. Andersen, Phytoplankton species richness, evenness, and production in relation to nutrient availability and imbalance. *Limnol. Oceanogr.* **62**, 1393–1408 (2017).

19. H. Hillebrand, A. I. Azovsky, Body size determines the strength of the latitudinal diversity gradient. *Ecography* **24**, 251–256 (2001).

20. P. Cermeño, I. G. Teixeira, M. Branco, F. G. Figueiras, E. Marañón, Sampling the limits of species richness in marine phytoplankton communities. *J. Plankton Res.* **36**, 1135–1139 (2014).

21. H. Claustre, D. Antoine, L. Boehme, E. Boss, F. D'Ortenzio, O. F. D'Andon, C. Guinet, N. Gruber, N. O. Handegard, M. Hood, K. Johnson, A. Körtzinger, R. Lampitt, P. Y. LeTraon, C. Le Quéré, M. Lewis, M. J. Perry, T. Platt, D. Roemmich, P. Testor, S. Sathyendranth, U. Send, J. Yoder, Guidelines towards an integrated ocean observation system for ecosystems and biogeochemical cycles, in *Proceedings of OceanObs'09: Sustained Ocean Observations and Information for Society*, 1, Venice, Italy, 21 September to 25, ESA Publications WPP-306, 2010, 19 pp.

22. Á. López-Urrutia, E. San Martin, R. P. Harris, X. Irigoien, Scaling the metabolic balance of the oceans. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 8739–8744 (2006).

23. C. T. Kremer, M. K. Thomas, E. Litchman, Temperature- and size-scaling of phytoplankton population growth rates: Reconciling the Eppley curve and the metabolic theory of ecology. *Limnol. Oceanogr.* **62**, 1658–1670 (2017).

24. D. J. Currie, G. G. Mittelbach, H. V. Cornell, R. Field, J.-F. Guégan, B. A. Hawkins, D. M. Kaufman, J. T. Kerr, T. Oberdorff, E. O'Brien, J. R. G. Turner, Predictions and tests of climate-based hypotheses of broad-scale variation in taxonomic richness. *Ecol. Lett.* **7**, 1121–1134 (2004).

25. R. Margalef, Life-forms of phytoplankton as survival alternatives in an unstable environment. *Oceanol. Acta* **1**, 493–509 (1978).

26. A. D. Barton, S. Dutkiewicz, G. Flierl, J. Bragg, M. J. Follows, Patterns of diversity in marine phytoplankton. *Science* **327**, 1509–1511 (2010).

27. P. Cermeño, P. Chouciño, B. Fernández-Castro, F. G. Figueiras, E. Marañón, C. Marrasé, B. Mouriño-Carballido, M. Pérez-Lorenzo, T. Rodríguez-Ramos, I. G. Teixeira, S. M. Vallina, Marine primary productivity is driven by a selection effect. *Front. Mar. Sci.* **3**, 173 (2016).

28. X. Irigoien, J. Huisman, R. P. Harris, Global biodiversity patterns of marine phytoplankton and zooplankton. *Nature* **429**, 863–867 (2004).

29. S. M. Vallina, M. J. Follows, S. Dutkiewicz, J. M. Montoya, P. Cermeño, M. Loreau, Global relationship between phytoplankton diversity and productivity in the ocean. *Nat. Commun.* **5**, 4299 (2014).

30. J. Ladau, T. J. Sharpton, M. M. Finucane, G. Jospin, S. W. Kembel, J. O'Dwyer, A. F. Koeppel, J. L. Green, K. S. Pollard, Global marine bacterial diversity peaks at high latitudes in winter. *ISME J.* **7**, 1669–1677 (2013).

31. G. C. Stevens, The latitudinal gradient in geographical range: How so many species coexist in the tropics. *Am. Nat.* **133**, 240–256 (1989).

32. A. M. Segura, D. Calliari, C. Kruk, H. Fort, I. Izaguirre, J. F. Saad, M. Arim, Metabolic dependence of phytoplankton species richness. *Glob. Ecol. Biogeogr.* **24**, 472–482 (2015).

33. E. Marañón, M. P. Lorenzo, P. Cermeño, B. Mouriño-Carballido, Nutrient limitation suppresses the temperature dependence of phytoplankton metabolic rates. *ISME J.* **12**, 1836–1845 (2018).

34. H. E. Garcia, R. A. Locarnini, T. P. Boyer, J. I. Antonov, O. K. Baranova, M. M. Zweng, J. R. Reagan, D. R. Johnson, *World Ocean Atlas 2013, Volume 4: Dissolved Inorganic Nutrients (Phosphate, Nitrate, Silicate)*, S. Levitus, A. Mishonov, Eds. (NOAA, 2013), 25 pp.

35. H. Kreft, W. Jetz, Global patterns and determinants of vascular plant diversity. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 5925–5930 (2007).

36. E. Villar, G. K. Farrant, M. Follows, L. Garczarek, S. Speich, S. Audic, L. Bittner, B. Blanke, J. R. Brum, C. Brunet, R. Casotti, A. Chase, J. R. Dolan, F. D'Ortenzio, J. P. Gattuso, N. Grima, L. Guidi, C. N. Hill, O. Jahn, J. L. Jamet, H. Le Goff, C. Lepoivre, S. Malviya, E. Pelletier, J. B. Romagnan, S. Roux, S. Santini, E. Scalco, S. M. Schwenck, A. Tanaka, P. Testor, T. Vannier, F. Vincent, A. Zingone, C. Dimier, M. Picheral, S. Searson, S. Kandels-Lewis; Tara Oceans Coordinators, S. G. Acinas, P. Bork, E. Boss, C. de Vargas, G. Gorsky, H. Ogata, S. Pesant, M. B. Sullivan, P. Wincker, E. Karsenti, C. Bowler, F. Not, P. Hingamp, D. Iudicone, Environmental characteristics of Agulhas rings affect interocean plankton transport. *Science* **348**, 1261447 (2015).

37. E. T. Buitenhuis, M. Vogt, R. Moriarty, N. Bednaršek, S. C. Doney, K. Leblanc, C. Le Quéré, Y.-W. Luo, C. J. O'Brien, T. O'Brien, J. Peloquin, R. Schiebel, C. Swan, MAREDAT: Towards a world atlas of MARine Ecosystem DATa. *Earth Syst. Sci. Data* **5**, 227–239 (2013).

38. A. J. Richardson, A. W. Walnea, A. W. G. John, T. D. Jonas, J. A. Lindley, D. W. Sims, D. Stevens, M. Witt, Using continuous plankton recorder data. *Prog. Oceanogr.* **68**, 27–74 (2006).

39. S. Sal, Á. López-Urrutia, X. Irigoien, D. S. Harbour, R. P. Harris, Marine microplankton diversity database. *Ecology* **94**, 1658 (2013).

40. C. de Boyer Montégut, G. Madec, A. S. Fischer, A. Lazar, D. Iudicone, Mixed layer depth over the global ocean: An examination of profile data and a profile-based climatology. *J. Geophys. Res.* **109**, C12 (2004).

41. C. de Vargas, S. Audic, N. Henry, J. Decelle, F. Mahé, R. Logares, E. Lara, C. Berney, N. Le Bescot, I. Probert, M. Carmichael, J. Poulain, S. Romac, S. Colin, J. M. Aury, L. Bittner, S. Chaffron, M. Dunthorn, S. Engelen, O. Flegontova, L. Guidi, A. Horák, O. Jaillon, G. Lima-Mendez, J. Lukeš, S. Malviya, R. Morard, M. Mulot, E. Scalco, R. Siano, F. Vincent, A. Zingone, C. Dimier, M. Picheral, S. Searson, S. Kandels-Lewis, T. Oceans Coordinators, S. G. Acinas, P. Bork, C. Bowler, G. Gorsky, N. Grimsley, P. Hingamp, D. Iudicone, F. Not, H. Ogata, S. Pesant, J. Raes, M. E. Sieracki, S. Speich, L. Stemmann, S. Sunagawa, J. Weissenbach, P. Wincker, E. Karsenti, Eukaryotic plankton diversity in the sunlit ocean. *Science* **348**, 1261605 (2015).

42. C. Amante, B. W. Eakins, *ETOPO1 1 Arc-Minute Global Relief Model: Procedures, Data Sources and Analysis* (NOAA Technical Memorandum NESDIS NGDC-24, National Geophysical Data Center, NOAA, 2009).

43. M. M. Zweng, J. R. Reagan, J. I. Antonov, R. A. Locarnini, A. V. Mishonov, T. P. Boyer, H. E. Garcia, O. K. Baranova, D. R. Johnson, D. Seidov, M. M. Biddle, *World Ocean Atlas 2013, Volume 2: Salinity*, S. Levitus, A. Mishonov, Eds. (NOAA, 2013), 39 pp.

44. P. Brun, M. Vogt, M. R. Payne, N. Gruber, C. J. O'Brien, E. T. Buitenhuis, C. Le Quéré, K. Leblanc, Y. W. Luo, Ecological niches of open ocean phytoplankton taxa. *Limnol. Oceanogr.* **60**, 1020–1038 (2015).

45. S. Pinkernell, B. Beszteri, Potential effects of climate change on the distribution range of the main silicate sinker of the Southern Ocean. *Ecol. Evol.* **4**, 3147–3161 (2014).

46. A. D. Barton, A. J. Irwin, Z. V. Finkel, C. A. Stock, Anthropogenic climate change drives shift and shuffle in North Atlantic phytoplankton communities. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 2964–2969 (2016).

47. P. W. Boyd, R. Strzepek, F. Fu, D. A. Hutchins, Environmental control of open-ocean phytoplankton groups: Now and in the future. *Limnol. Oceanogr.* **55**, 1353–1376 (2010).

48. R. A. Locarnini, A. V. Mishonov, J. I. Antonov, T. P. Boyer, H. E. Garcia, O. K. Baranova, M. M. Zweng, C. R. Paver, J. R. Reagan, D. R. Johnson, M. Hamilton, D. Seidov, *World Ocean Atlas 2013, Volume 1: Temperature*, S. Levitus, A. Mishonov, Eds. (NOAA, 2013), 40 pp.

49. R. Atlas, R. N. Hoffman, J. Ardizzone, S. M. Leidner, J. C. Jusem, D. K. Smith, D. Gombos, A cross-calibrated, multiplatform ocean surface wind velocity product for

meteorological and oceanographic applications. *Bull. Am. Meteorol. Soc.* **92**, 157–174 (2011).

50. P. Landschützer, N. Gruber, D. C. E. Bakker, A 30 years observation-based global monthly gridded sea surface pCO2 product from 1982 through 2011 (NCEI Accession 0160558). Version 2.2. NOAA National Centers for Environmental Information. Dataset. doi: 10.3334/cdiac/otg.spco2_1982_2011_eth_somffn.

51. K. A. Whittaker, T. A. Rynearson, Evidence for environmental and ecological selection in a microbe with no geographic limits to gene flow. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 2651–2656 (2017).

52. P. Cermeño, P. G. Falkowski, Controls on diatom biogeography in the ocean. *Science* **325**, 1539–1541 (2009).

53. G. Chust, X. Irigoien, J. Chave, R. P. Harris, Latitudinal phytoplankton distribution and the neutral theory of biodiversity. *Glob. Ecol. Biogeogr.* **22**, 531–543 (2013).

54. S. J. Phillips, M. Dudík, J. Elith, C. H. Graham, A. Lehmann, J. Leathwick, S. Ferrier, Sample selection bias and presence-only distribution models: Implications for background and pseudo-absence data. *Ecol. Appl.* **19**, 181–197 (2009).

55. M. Barbet-Massin, F. Jiguet, C. H. Albert, W. Thuiller, Selecting pseudo-absences for species distribution models: How, where and how many? *Methods Ecol. Evol.* **3**, 327–338 (2012).

56. L. Buisson, W. Thuiller, N. Casajus, S. Lek, G. Grenouillet, Uncertainty in ensemble forecasting of species distribution. *Glob. Change Biol.* **16**, 1145–1157 (2010).

57. C. Merow, M. J. Smith, T. C. Edwards Jr., A. Guisan, S. M. McMahon, S. Normand, W. Thuiller, R. O. Wüest, N. E. Zimmermann, J. Elith, What do we gain from simplicity versus complexity in species distribution models. *Ecography* **37**, 1267–1281 (2014).

58. O. Allouche, A. Tsoar, R. Kadmon, Assessing the accuracy of species distribution models: Prevalence, kappa and the true skill statistic (TSS). *J. Appl. Ecol.* **43**, 1223–1232 (2006).

59. J. M. Calabrese, G. Certain, C. Kraan, C. F. Dormann, Stacking species distribution models and adjusting bias by linking them to macroecological models. *Glob. Ecol. Biogeogr.* **23**, 99–112 (2014).

60. W. Jetz, C. Rahbek, Geographic range size and determinants of avian species richness. *Science* **297**, 1548–1551 (2002).

61. R Development Core Team, *R: A Language and Environment for Statistical Computing v.3.2.1* (R Foundation for Statistical Computing, Vienna, Austria, 2015).

62. GBIF Secretariat, *GBIF Backbone Taxonomy*, Checklist Dataset https://doi.org/10.15468/39omei, accessed via https://www.gbif.org on 14 July 2015.

63. M. D. Guiry, G. M. Guiry, *AlgaeBase*, World-wide electronic publication, National University of Ireland, Galway, http://www.algaebase.org, accessed on 7 August 2017.

64. P. G. Falkowski, M. E. Katz, A. H. Knoll, A. Quigg, J. A. Raven, O. Schofield, F. J. R. Taylor, The evolution of modern eukaryotic phytoplankton. *Science* **305**, 354–360 (2004).

65. R. W. Jordan, L. Cros, J. R. Young, A revised classification scheme for living haptophytes. *Micropaleontology* **50**, 55–79 (2004).

Citation: D. Righetti, M. Vogt, N. Gruber, A. Psomas, N. E. Zimmermann, Global pattern of phytoplankton diversity driven by temperature and environmental variability. *Sci. Adv.* **5**, eaau6253 (2019).