A Journal of Macroecology

Global Ecology and Biogeography

RESEARCH PAPER

Marine nano- and microphytoplankton diversity: redrawing global patterns from sampling-standardized data

Tamara Rodríguez-Ramos[1]*, Emilio Marañón[1] and Pedro Cermeño[2]

[1]Departamento de Ecoloxía e Bioloxía Animal, Universidade de Vigo, Campus As Lagoas-Marcosende, 36310, Vigo, Spain, [2]Instituo de Ciencias del Mar, Consejo Superior de Investigaciones Científicas, Passeig Marítim de la Barceloneta, 37-49 E-08003, Barcelona, Spain

*Correspondence: Tamara Rodríguez-Ramos, Departamento de Ecoloxía e Bioloxía Animal, Universidade de Vigo, Campus As Lagoas-Marcosende, 36310, Vigo, Spain.
E-mail: tamara@uvigo.es

ABSTRACT

Aim We analysed marine phytoplankton diversity data as a function of latitude, temperature, primary production and several environmental and biological variables to ascertain whether large-scale variability in the diversity of marine nano- and microphytoplankton (including diatoms, dinoflagellates and coccolithophores) follows similar patterns to those observed for macroorganisms. For the first time we explored these relationships after correcting the observed patterns of species richness by sampling effort.

Location The global ocean.

Methods To standardize the estimates of species richness by sampling effort we used interpolation and extrapolation based on Hill numbers and shareholder quorum subsampling (SQS) methods. Then, we fitted linear and quadratic models to species richness data to explore their variability with latitude, inverse temperature and biomass. These relationships were compared with the patterns obtained from non-standardized data. In addition, we used a stepwise multiple linear regression model to explain the variability of species richness as the combined effect of multiple drivers acting together.

Results Marine phytoplankton diversity was weakly correlated with latitude, temperature or biomass. The hotspots of species richness at intermediate latitudes largely vanished after standardization for sampling effort. Neither latitude, temperature, primary production (as diagnostics of energy supply) nor any other variable or combination of variables, explained the patterns of phytoplankton species richness.

Main conclusions None of the hypotheses tested explained a significant amount of the variability in species richness. The patterns observed for microorganisms in previous studies may have resulted at least partially from differences in sampling effort along productivity gradients and systematic undersampling of species. We conclude that large-scale processes such as passive dispersal and recurrent habitat recolonization dominate the distribution of species. Sampling protocols and data analyses must be improved in order to obtain estimates of diversity that are comparable across ecosystems.

Keywords
Dispersal, latitudinal diversity gradient, marine phytoplankton, productivity–diversity relationship, sampling effort standardization, temperature.

DOI: 10.1111/geb.12274
http://wileyonlinelibrary.com/journal/geb

INTRODUCTION

The latitudinal diversity gradient (LDG), which predicts a decline in species richness from the equator to the poles, has been reported for macro- and microorganisms in aquatic (freshwater and marine) and terrestrial systems (Pianka, 1966; Fuhrman *et al.*, 2008; Stomp *et al.*, 2011). The mechanisms that lead to the LDG have long been discussed (Pianka, 1966; Hillebrand, 2004) and several ecological, historical and evolutionary explanations have been proposed (Willig *et al.*, 2003; Mittelbach *et al.*, 2007). The 'species–energy' hypothesis states that the LDG is generated and maintained by an equatorward increase in the amount of incoming energy. According to this hypothesis, the biological diversity of an ecosystem is related to the availability of kinetic and potential forms of energy in the environment, expressed as temperature and net primary production, respectively (Allen *et al.*, 2007). Founded on the concept of the metabolic theory of ecology (MTE) it has been suggested that temperature could control the probability of mutation and speciation (Allen *et al.*, 2007) by adjusting the kinetics of enzymatic reactions. At low latitudes, higher incident solar radiation and consequently higher temperatures can thus regulate the latitudinal distribution of species through regional variations in the rate of diversification. Primary production, the rate of energy capture and carbon fixation by primary producers, has been also proposed as an important factor controlling the number of species present in the community, but the mechanisms underlying the productivity–diversity relationship (PDR) are not clear and its shape has been found to depend on the type of organism and the scale of observation (Mittelbach *et al.*, 2001; Chase & Leibold, 2002). For aquatic microorganisms, this relationship has been proposed to be 'hump-shaped' (Horner-Devine *et al.*, 2003; Irigoien *et al.*, 2004; Stomp *et al.*, 2011), such that diversity increases with the amount of resources but decreases in highly productive environments, where a few dominant species account for the bulk of community biomass. However, differences in sampling effort across ecosystems and/or time periods could bias the intercomparison of data sets, distorting the patterns of species richness reported for these marine microorganisms (Cermeño *et al.*, 2013).

The growth of marine phytoplankton depends on their ability to acquire and use resources, and hence the geographical variability of light regimes and nutrient supply might be involved in controlling the large-scale distribution of phytoplankton species in the ocean. However, microorganisms have huge populations, broad dispersal ranges and short generation times which decrease the probability of local extinction. These features of microbial populations led to the idea that broad-scale processes, such as dispersal and recurrent habitat recolonization, reduce the effect of environmental factors in shaping the patterns of microbial diversity, maintaining a large pool of coexisting species within local communities (Finlay, 2002; Gibbons *et al.*, 2013).

The scarcity of field data, the systematic undersampling of marine phytoplankton communities (Cermeño *et al.*, 2013; Rodríguez-Ramos *et al.*, 2014) and the idiosyncrasy of different observers in taxonomic assignation (Irigoien *et al.*, 2004; Cermeño *et al.*, 2013; Chust *et al.*, 2013) limit our ability to obtain accurate estimates of species richness, delineate broad-scale patterns and identify the underlying mechanisms. Here, we use two extensive data sets, including species composition and several biotic and abiotic variables, to investigate the patterns of marine phytoplankton diversity in the global ocean. A novel contribution of this work is the careful selection of samples and data analysis in order to avoid some of the main shortcomings of previous studies. Here: (1) cell counts and taxonomic identifications were conducted by a single expert (D. S. Harbour, Plymouth Marine Laboratory, Prospect Place, West Hoe, Plymouth PL1 3DH, UK) (Sal *et al.*, 2013); (2) we compiled data collected during all seasons over several years and at different regions around the worlds' oceans, including a large variety of habitats, to avoid spatiotemporal dependence in our results; and (3) we explored for the first time numerous potential explanatory variables and mechanisms while controlling for the effect of differential sampling effort across data sets. Our goals were: (1) to test whether the predictions of the LDG and the MTE apply for marine nano- and microphytoplankton; (2) to investigate the shape of the PDR at different scales of observation (from local to global); and (3) to identify the combination of biological and environmental variables which best explain the patterns of species richness at a global scale.

METHODS

Data acquisition

A complete list of the variables included in each data set, their units, the number of observations and a brief definition/description is available in Table 1.

The global database

A very valuable data set, the global database (GD) was compiled by Sal *et al.* (2013), comprising data collected world-wide on microplankton species composition and hydrographic, environmental and biological variables, covering a great variety of marine ecosystems from coastal to open-ocean areas and from polar to equatorial latitudes (see Sal *et al.*, 2013, for further details on methodologies). Within the original array of species, we selected those corresponding to phytoplankton taxa (diatoms, dinoflagellates and coccolithophores). Our final data set includes $S = 651$ species and $n = 744$ sampling sites. As the volume settled per sample is unknown, we assumed that cell counting was always done on 50-ml seawater samples, which is commonly used as the 'standard' volume. Then, abundances were extrapolated from the reported cell densities to a total abundance in 50 ml. Although it is possible that in high-biomass conditions the volume settled was smaller, the number of potentially biased cases is low enough [in only 15% of the samples was the chlorophyll *a* (Chl*a*) concentration was higher than 3 mg m^{-3}] to have any significant influence over our results. Having the total number of individuals counted per sample allows the standardization of raw species richness as a function

Table 1 List of variables included in the study, their units, the number of observations in each data set (Atlantic meridional transects (n_{AMT}) and the global (n_{GD}) data sets) and a brief definition.

Variable	Units	n_{AMT}	n_{GD}	Definition
Lat	–	98	744	Latitude
Sal	–	95	–	Salinity at surface
ND	m	80	–	Nitracline depth, at which $[NO_3] > 0.05$ µmol l^{-1}
T	°C	95	698	Temperature, at surface
PP2–20 µm	mg C m^{-3} h^{-1}	70	–	Primary production rate at surface, size fraction 2–20 µm (nanoplankton)
MLD	m	72	744	Mixed layer depth, at which $\Delta T > 0.5$° C
PP0.2–2 µm	mg C m^{-3} h^{-1}	70	–	Primary production rate at surface, size fraction 0.2–2 µm (picoplankton)
Chl*a*2–20 µm	mg Chl*a* m^{-3}	68	–	Chl*a* concentration at surface, size fraction 2–20 µm (nanoplankton)
TotPP	mg C m^{-3} h^{-1}	93	–	Total (all size fractions) primary production rate, at surface
TotChl*a*	mg Chl*a* m^{-3}	92	744	Total (all size fractions) Chl*a* concentration, at surface
ELD	m	71	–	Euphotic layer depth, at which light intensity = 1% of superficial intensity
Chl*a*0.2-2 µm	mg Chl*a* m^{-3}	68	–	Chl*a* concentration at surface, size fraction 0.2–2 µm (picophytoplankton)
Chl*a*≥20 µm	mg Chl*a* m^{-3}	67	–	Chl*a* concentration at surface, size fraction ≥ 20 µm (microphytoplankton)
PP≥20 µm	mg C m^{-3} h^{-1}	70	–	Primary production rate at surface, fraction ≥ 20 µm (microphytoplankton)
Chl*a*≥2µmInt	mg Chl*a* m^{-2}	62	–	Chl*a* concentration integrated in the euphotic layer, size fraction ≥ 2 µm (nano- and microphytoplankton)
TotAbd	No. of individuals	98	744	Total (all size fractions) abundance per volume of sample settled, at surface
PP≥2µmInt	mg C m^{-2} h^{-1}	62	–	Primary production rate integrated in the euphotic layer, size fraction ≥ 2 µm (nano- and microphytoplankton)
PP < 2µmInt	mg C m^{-2} h^{-1}	62	–	Primary production rate integrated in the euphotic layer, size fraction < 2 µm (picophytoplankton)
S_{SQS}	species	97	744	Number of species standardized by shareholder quorum subsampling (SQS) for $q = 0.7$ (70% subsampled)
PPTotInt	mg C m^{-2} h^{-1}	95	–	Total (all size fractions) primary production rate, integrated in the euphotic layer
Chl*a*TotInt	mg Chl*a* m^{-2}	97	–	Total (all size-fractions) Chl*a* concentration, integrated in the euphotic layer
Chl*a*<2µmInt	mg Chl*a* m^{-2}	62	–	Chl*a* concentration, integrated in the euphotic layer, size fraction < 2 µm (picophytoplankton)
TotBio	mg C m^{-3}	98	744	Total (all size fractions) biomass, at surface
S_{obs}	No. of species	98	744	Observed species richness, at surface
S_H	–	97	744	Shannon's or standard information index of diversity
S_{iNext}	No. of species	93	744	Number of species estimated by interpolation and extrapolation methodology (iNext) for $n = 10000$ individuals
PP/Chl*a*	mg C mg Chla^{-1} h^{-1}	95	–	Ratio PPTotInt : Chl*a*TotInt
PAR	mol photons m^{-2} day^{-1}	–	744	Photosynthetically active radiation, at surface
Nit	µmol l^{-1}	–	744	Nitrate concentration, at surface
Phosp	µmol l^{-1}	–	744	Phosphate concentration, at surface
Silic	µmol l^{-1}	–	744	Silicate concentration, at surface

of sampling effort (number of individuals counted per sample). Finally, we selected only near-surface samples (depth < 20 m) to avoid any potential interference between light availability and diversity.

Atlantic meridional transects

We compiled data on physical, chemical and biological variables concurrently determined during four Atlantic meridional transects (AMTs) carried out on board RRS *James Clark Ross* during September and October 1995 (AMT-1), April and May 1996 (AMT-2), September and October 1996 (AMT-3), and April and May 1997 (AMT-4), crossing temperate, subtropical and tropical regions in the North and South Atlantic Ocean (Marañón *et al.*, 2000). Seawater samples were collected at 25 sampling stations distributed along each latitudinal transect, from five to ten depths in the upper 200 m of the water column. Sampling depths were selected according to the vertical distribution of fluorescence, covering the entire euphotic layer.

The entire data set comprises 28 variables describing environmental conditions (e.g. temperature), total and size-fractioned primary production, local hydrodynamic conditions, resource availability (e.g. the depth of the nitracline) (Cermeño *et al.*, 2008a) and several measurements of phytoplankton standing stocks (total and size-fractionated Chl*a* concentration, total biomass and abundance). The methodologies used to obtain the hydrographic structure of the water column as well as to measure the biological and environmental variables included in the AMT data set, are explained in detail in the Appendix S1 in the Supporting Information (and references therein).

In both data sets, carbon biomass (TotBio) was estimated from cell numbers (TotAbd) by using biovolume estimates and empirical relationships between biovolume and cell carbon reported by Holligan *et al.* (1984).

Data analyses and statistics

Standardization of the observed species richness and diversity metrics

Our calculations are based on raw estimates of species richness or non-standardized data (S_{obs}). To obtain comparable estimates of species richness from areas with different levels of productivity and cell densities we used two different methods of standardization. Firstly, we applied the interpolation and extrapolation with Hill number methodology (Chao *et al.*, 2014) by using the package iNext (Hsieh *et al.*, 2014) implemented in R (R Core Team, 2013). This tool uses individual-based rarefaction analyses (Gotelli & Colwell, 2001) or extrapolation, depending on whether the total abundance of individuals in the sample, TotAbd, is higher or lower than the number of individuals at which we want to compare the samples (n), respectively. We defined S_{iNext} as the number of species expected when subsampling or extrapolating to $n = 10{,}000$ individuals. Secondly, we used the shareholder quorum subsampling (SQS) method (Alroy, 2010) that estimates the expected number of species by sampling a given coverage of the underlying species abundance distribution. The more diverse the sample, the higher the number of individuals that must be sampled to detect a similar proportion of the species present in the community, and thus a non-uniform sampling effort is needed. To define S_{SQS}, we sampled 70% (0.7 quorum subsampling) of the species abundance distribution on each community after eliminating the most dominant taxa (the species with the highest abundance in the sample). The SQS routine was performed by using the function *sqs* in R (R Core Team, 2013). In both cases, choosing a different number of individuals or a different percentage of the species abundance distribution did not change significantly the results obtained from standardized data.

Finally, we also calculated the Shannon diversity index, S_H, which computes diversity as a dual variable taking into account species richness and evenness (i.e. the equitability in the distribution of abundance among species) (Hurlbert, 1971).

Objective 1: the latitudinal diversity gradient

The LDG hypothesis predicts a decline in diversity with increasing latitude from the equator to the poles. To test this hypothesis we represented the latitudinal distribution of four diversity metrics (data from the GD) as a function of absolute latitude. Then, we fitted ordinary least squares (OLS) linear regression models to data.

Objective 2: the metabolic theory of ecology and biodiversity

The MTE predicts that species richness increases exponentially with temperature (T), such that the natural logarithm of species richness is linearly related to $1/kT$ (k is the Boltzmann constant), with a slope E (activation energy) ranging between 0.60 and 0.70 eV (Brown *et al.*, 2004). To test this hypothesis, we fitted linear and quadratic models (first- and second-order polynomial functions, respectively) to the relationship between the broad-scale variation of phytoplankton (ln-transformed) species richness and the inverse of temperature. The Akaike information criterion (AIC) was used to select the best model (Burnham & Anderson, 2002).

Objective 3: biomass–diversity relationship

The PDR is one of the most explored but least clear patterns in ecology. We represented the relationship between species richness (raw and standardized) and total biomass (mg C m^{-3}) per sampling site. Total biomass is used here as a surrogate for primary production, which is not available in the GD data set. Moreover, we investigated the PDR at different regions and seasons. To do so, we divided the GD into SubpolarN (> 49° N), TemperateN (35 to 49° N), OligoN (20 to 34° N), Equat+Upwell (9° S to 19° N), OligoS (10 to 34° S) and TemperateS (35 to 48° S). The SubpolarS (> 48° S) region contained only four sampling stations and was discarded from the analysis. The GD was also divided into different seasons, including Spring (March–May), Summer (June–August), Autumn (September–November) and Winter (December–February) for the Northern Hemisphere and vice versa for the Southern Hemisphere. We then fitted linear and quadratic models to data and selected the best models according to the AIC (Burnham & Anderson, 2002).

Objective 4: single drivers of diversity

We searched for correlation between pairwise variables to quantify the variation of marine phytoplankton diversity in relation to latitude and several environmental and biological factors. We used the function *MINE* (Reshef *et al.*, 2011) in R (R Core Team, 2013), which computes the maximal information coefficient (MIC), a measure of dependence for two-variable relationships, such that the higher the value of MIC, the stronger the correlation between the variables under consideration. The main advantage of using this function is its capacity to identify a broad range of relationships other than the linear correlation, traditionally computed as the Pearson product–moment correlation coefficient. The P-values or significance levels corresponding to the obtained MIC scores were pre-computed for different numbers of observations, and are available at http://www.exploredata.net/downloads/p-value-tables. The MIC threshold, indicating a significant correlation between the pair of variables considered, was defined as a P-value < 0.001. This correlation analysis enables to detect correlations among geographical and environmental variables in order to avoid redundant information in the following analysis (see Objective 5).

Objective 5: multiple drivers of diversity

To test whether the patterns of marine phytoplankton diversity (or its absence) are driven by the combined effect of multiple

factors acting together, we used a stepwise multiple linear regression analysis with backward selection of variables. This analysis allows us to explore the relative performance of different combinations of variables as explanatory factors of species richness (S_{obs} or S_{iNext} as the dependent variable). Within each data set, we selected only the stations for which all the variables included in the analysis were measured ($n = 51$ and $n = 677$ stations, for AMT and GD, respectively). Variables found to be *strongly* or *very strongly* correlated with each other in the previous analysis (see Appendix S2) were not allowed to be simultaneously included in the regression analyses to avoid redundant information. The analysis was computed in R (R Core Team, 2013) using the function *step*, which calculates the AIC for a model describing the dependent variable as a function of the number of explanatory variables (k). Then, it calculates the AIC for all possible models derived from the initial one taken into account $k - 1$ explanatory variables. The variable which most improves the fit (giving rise to the lowest AIC) with respect to the initial model is removed from the subset of potentially explanatory variables. This routine is repeated until the fit cannot be further improved by removing any of the remaining variables. The method uses a maximum of $k = 7$ variables per model, so we selected different subsets of seven variables until we found the best model. Then, we used the function *moran.test* implemented in the package *spdep* in R (R Core Team, 2013) to check if the residuals from the multiple linear regression models were spatially autocorrelated.

RESULTS

Patterns in species occurrence and abundance

In the GD, 17% of the species were very infrequent (found in only one or two samples), 13% were infrequent (present in less than 10% of the samples) and only 0.6% were widespread (found in at least half of the samples, distributed in both hemispheres); no cosmopolitan species (present in more than 90% of the samples) were found. In the AMTs, 33% of the species were very infrequent, 41% were infrequent, 5% were widespread and < 1% were cosmopolitan. However, in both data sets, the occurrence of a species, defined as the number of samples in which the species is present, was correlated positively with its total abundance in the corresponding data set (Appendix S3). This result suggests that infrequent and very infrequent species, typically present in very low abundance, could be systematically undersampled owing to their low population densities. Hence, in addition to raw species richness (S_{obs}), we included three additional descriptors of diversity in our analyses: S_{iNext}, S_{SQS} and S_H (see Methods and Table 1).

Objective 1: LDG

The hotspots of species richness found at intermediate latitudes largely vanished after standardization for sampling effort (Fig. 1). All the diversity metrics showed a slight tendency to decrease linearly with increasing latitude. However, the coefficients of determination (R^2) were very low (ranging from 0.005 for S_H to 0.07 for S_{iNext}), suggesting that latitudinal gradients had little power to account for the patterns of nano- and microphytoplankton species richness in the ocean.

Objective 2: MTE

The shape of the temperature–species richness relationship was dependent on the expression of species richness used (Fig. 2). The parameters of the models fitted to data are available in Table 2. For S_{obs} and S_{iNext}, the AIC showed that this relationship was best described by a curvilinear relationship. On the

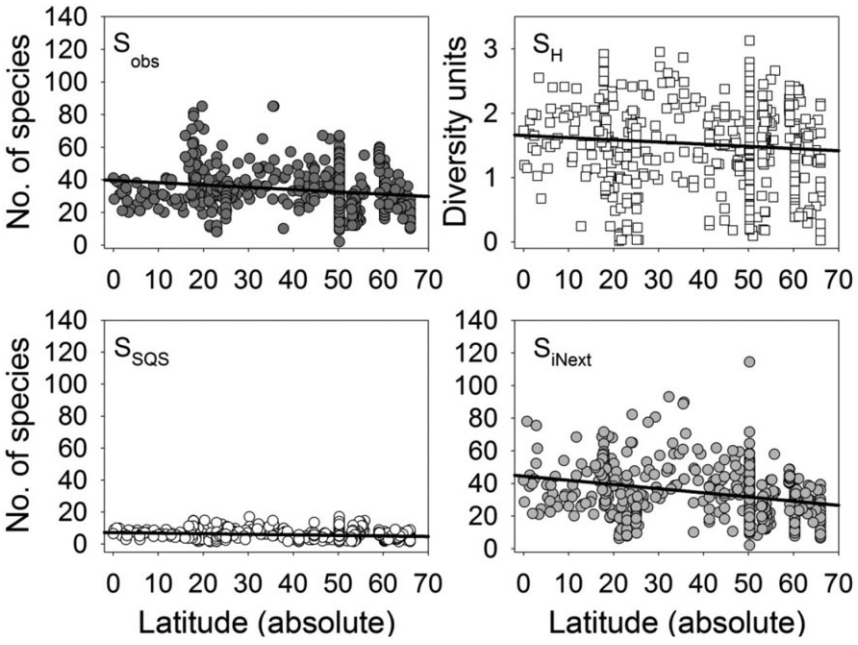

Figure 1 Global latitudinal distribution of superficial nano- and microphytoplankton species richness and diversity. Solid lines represent the ordinary least squares regression fitted to diversity data as a function of absolute latitude (Lat). S_{obs}, observed species richness ($S_{obs} = 39.65 - 14.10 \times$ Lat, $P < 0.0001$, $R^2 = 0.028$); S_{iNext}, species richness standardized by Chao's method for $n = 10{,}000$ individuals ($S_{iNext} = 44.47 - 0.26 \times$ Lat, $P < 0.0001$, $R^2 = 0.07$); S_{SQS}, species richness standardized by Alroy's SQS method after sampling 70% of the abundance distribution, $q = 0.7$ ($S_{SQS} = 7.08 - 0.03 \times$ Lat, $P < 0.0001$, $R^2 = 0.031$); and S_H, Shannon diversity index ($S_H = 1.65 - 0.003$ * Lat, $P < 0.05$, $R^2 = 0.005$).

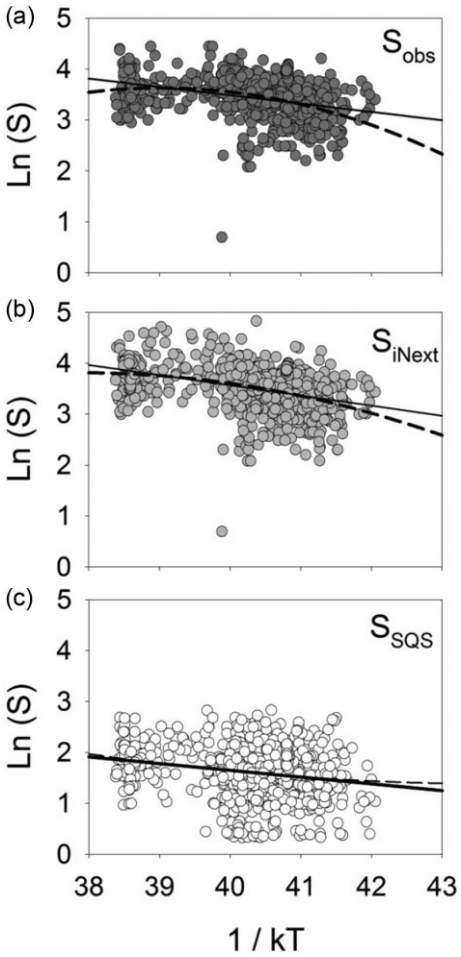

Figure 2 Relationship between temperature and species richness. The natural logarithm of species richness (a) observed (S_{obs}) and standardized by sampling effort (b) by Chao's method for $n = 10{,}000$ individuals (S_{iNext}) and (c) by Alroy's SQS method for $q = 0.7$ (70% resampled) (S_{SQS}) is expressed as a function of the inverse, absolute temperature at the surface, $1/kT$ (k, Boltzmann's constant), using data from the global data set (GD, $n = 698$). The parameters of linear (solid line) and quadratic (dashed line) functions fitted to the data are available in Table 2. The best fitted model, with the lowest value of the Akaike information criterion, is highlighted with a thicker line.

contrary, S_{SQS} declined linearly with decreasing temperature as predicted by the MTE. However, in all cases, temperature explained only a minor percentage of the variability in species richness. Moreover, the regression slopes (absolute values) of the relationship between ln-transformed species richness and temperature were much lower (ranging from 0.13 for S_{SQS} to 0.19 for S_{iNext}) than the range predicted by the MTE, regardless of the expression used for species richness (Table 2).

Objective 3: PDR

The global relationship between S_{obs} and community biomass (Fig. 3, upper panel) was best described by a quadratic function (see Appendix S4), with diversity peaking at intermediate levels of community biomass. Biomass explained *c.* 22% of the variability in species richness (S_{obs}). Although this unimodal pattern was consistent for all the expressions of diversity, the coefficient of determination was very low after standardizing species richness by sampling effort or when using the Shannon diversity index (i.e. biomass explained only *c.* 9% of variability for S_{iNext}, *c.* 4% for S_{SQS} and *c.* 3% for S_H).

Our analysis showed poor relationships across geographical regions and seasons. Only in the northern oligotrophic region (OligoN) and in the low-latitude region (Equat+Upwell) did the biomass explain > 50% and *c.* 40% of variability in S_{obs}, respectively (Fig. 3). Again, the patterns observed with raw data mostly disappeared or changed after standardizing species richness by sampling effort. For the different seasons (Fig. 4), raw species richness exhibited a 'hump-shaped' relationship with biomass. This relationship vanished or turned linear and negative when standardizing by sampling effort, except for the summer subset (see Appendices S4 & S5 for further details on linear and quadratic models fitted to data, for different regions and seasons, respectively).

Objective 4: single drivers of diversity

We performed analyses of correlation between pairwise variables. The results are summarized in correlation (semi-) matrices (Appendix S2 for the AMT (A) and GD (B) data sets, respectively). Significant correlations are categorized into different classes of 'strength' (weak, moderate, strong and very strong), defined as ranges of maximum information coefficient (MIC) scores (see the legends and figures in Appendix S2).

AMT data set

Among all the diversity expressions, we found only a significant correlation with latitude for S_{SQS} (Appendix S2). Likewise, none of the environmental variables which were strongly correlated with latitude (Sal, ND, *T*, MLD or ELD) were significantly correlated with any of the four expression of diversity. Temperature was not correlated with any diversity metric, reinforcing our results about the minor role of temperature on species richness distribution. TotChl*a* was not correlated with any of the diversity metrics (Appendix S2).

GD data set

Environmental and resource-related variables showed the strongest correlation with latitude. TotAbd, TotChl*a* and TotBio showed a significant correlation with increasing latitude. The estimates of species richness showed different patterns: S_{obs} was significantly (although weakly, see Appendix S2) correlated with all the variables included in the analysis. S_{iNext}, S_{SQS}, and S_H were significantly (although weakly, see Appendix S2) correlated with latitude.

In summary, for both the AMT and GD data sets, none of the variables included in the correlation analysis were strongly

Table 2 Parameters of the linear and quadratic models fitted to the relationship between the natural logarithm of species richness (observed, S_{obs}, or standardized by sampling effort by Chao's method for $n = 10{,}000$ individuals (S_{iNext}) and by Alroy's SQS method for $q = 0.7$ (70% resampled) S_{SQS}) and the inverse temperature (independent variable).

Dependent variable	Model	Param. 1	Param. 2	Param. 3	R^2	AIC	*P*-value
$\ln(S_{obs})$	L	10.0 (± 0.7)	−0.16 (± 0.02)		0.104	693.7	< 0.0001
	Q	−121.5 (± 29.0)	6.41 (± 1.43)	−0.082 (± 0.02)	0.129	**674.9**	< 0.0001
$\ln(S_{iNext})$	L	11.0 (± 0.8)	−0.19 (± 0.02)		0.12	780.8	< 0.0001
	Q	−37.1 (± 30.9)	2.22 (± 1.55)	−0.03 (± 0.02)	0.12	**780.3**	< 0.0001
$\ln(S_{SQS})$	L	7.0 (± 1.0)	−0.13 (± 0.02)		0.04	**1101**	< 0.0001
	Q	34.8 (± 39.0)	−1.53 (± 1.95)	0.02 (± 0.71)	0.039	1102.8	< 0.0001

The *P*-value indicates the level of significance of the fit. R^2 is the coefficient of determination. AIC, Akaike information criterion for each fitted model. The model with lowest AIC (in bold) was the best fitted to data.
Models:
L (linear): $\ln(S) = \text{Param. 1 (± SE)} + \text{Param. 2 (± SE)} \times (1/kT)$
where Param.1 = intercept; Param.2 = slope.
Q (quadratic): $\ln(S) = \text{Param.1 (± SE)} + \text{Param. 2 (± SE)} \times (1/kT) + \text{Param. 3 (± SE)} \times (1/kT)^2$
where Param.1 = free term; Param.2 = linear term; Param.3 = quadratic term; k (Boltzmann constant) = 1.38×10^{-23} m^2 kg s^{-2} K^{-1}.

correlated with S_{obs}, S_{iNext}, S_{SQS} or S_H. Consequently, the distribution of nano- and microphytoplankton species richness in the oceans cannot be explained by a single controlling factor.

Objective 5: multiple determinants of diversity

A stepwise multiple linear regression analysis with backward selection of variables was implemented as an attempt to forecasting the patterns in S_{obs} or S_{iNext} as a function of biotic and abiotic factors. For the AMT data set, the analysis showed that a model combining three variables (T, TotBio and TotPP) explained 34% of the variability in S_{obs} along the Atlantic Ocean, while for S_{iNext} it decreased to 21% and temperature was discarded as a driver of species richness (see Appendix S6 for further details on model fitting). For the GD, with a smaller number of potential independent factors but a higher number of observations per variable, a model combining four drivers (T and TotBio, as well as incident PAR and MLD; see Table 1 for a definition of the terms) was able to explain only 17% of the variability in S_{obs} and 18% of the variability in S_{iNext} (Appendix S6).We calculated Moran's I statistic for each model's residuals (the number of species observed minus the number of species predicted by the model) to test for spatial autocorrelation. In both cases the result was not significant at the $\alpha = 0.05$ level ($P > 0.05$ in both cases and for both data sets), which means that we cannot reject the null hypothesis of a random distribution of the residuals and thus we can discard spatial autocorrelation in our data (Dormann *et al.*, 2007).

DISCUSSION

The latitudinal diversity gradient

We have found a weak latitudinal diversity gradient regardless of the diversity metric used. The hotspots of species richness that are commonly observed at intermediate latitudes largely vanished after standardizing the number of species by sampling effort. The observation of a weak latitudinal diversity gradient is in accordance with previous findings for marine planktonic microorganisms (Hillebrand & Azovsky, 2001; Cermeño *et al.*, 2008b) and compatible with the bipolar distribution of some specific marine bacteria (Sul *et al.*, 2013) and a cyst-forming dinoflagellate species (Montresor *et al.*, 2003). As suggested by Pedrós-Alió (2006), everything would be likely to be everywhere if rare taxa forming the seed bank of species were detectable by the existing methods.

Alternatively, some authors have proposed the existence of a strong latitudinal diversity pattern for marine microorganisms. For instance, Pommier *et al.* (2007) and Fuhrman *et al.* (2008) found that the species richness of marine bacterioplankton was significantly correlated with latitude and temperature. However, these results might be biased by methodological issues such as a limited sampling coverage and systematic undersampling. In the first case, the analysis is based on nine sampling sites spatially separated by wide distances, and thus the results are not strictly representative of a global pattern. In the second study, the authors report a substantial proportion of unexplained variation in diversity, limiting potential interpretations about the nature of a LDG for bacterioplankton. As stated above, these studies were potentially biased by differences in sampling effort and thus disparate probabilities of detection of rare species (Rodríguez-Ramos *et al.*, 2014). Using a global ocean ecosystem model, Barton *et al.* (2010) predict a decrease in phytoplankton diversity with increasing latitude and identify hotspots of diversity at tropical and subtropical latitudes. However, the robustness of their model results has been questioned, with some arguing that minor deviations from their assumption of neutrality would result in very different predictions (Huisman, 2010). Recently, Stomp *et al.* (2011) reported on a LDG for freshwater phytoplankton despite their data set only covering a narrow latitudinal range. However, the patterns of microbial diversity might obey different rules in freshwater ecosystems,

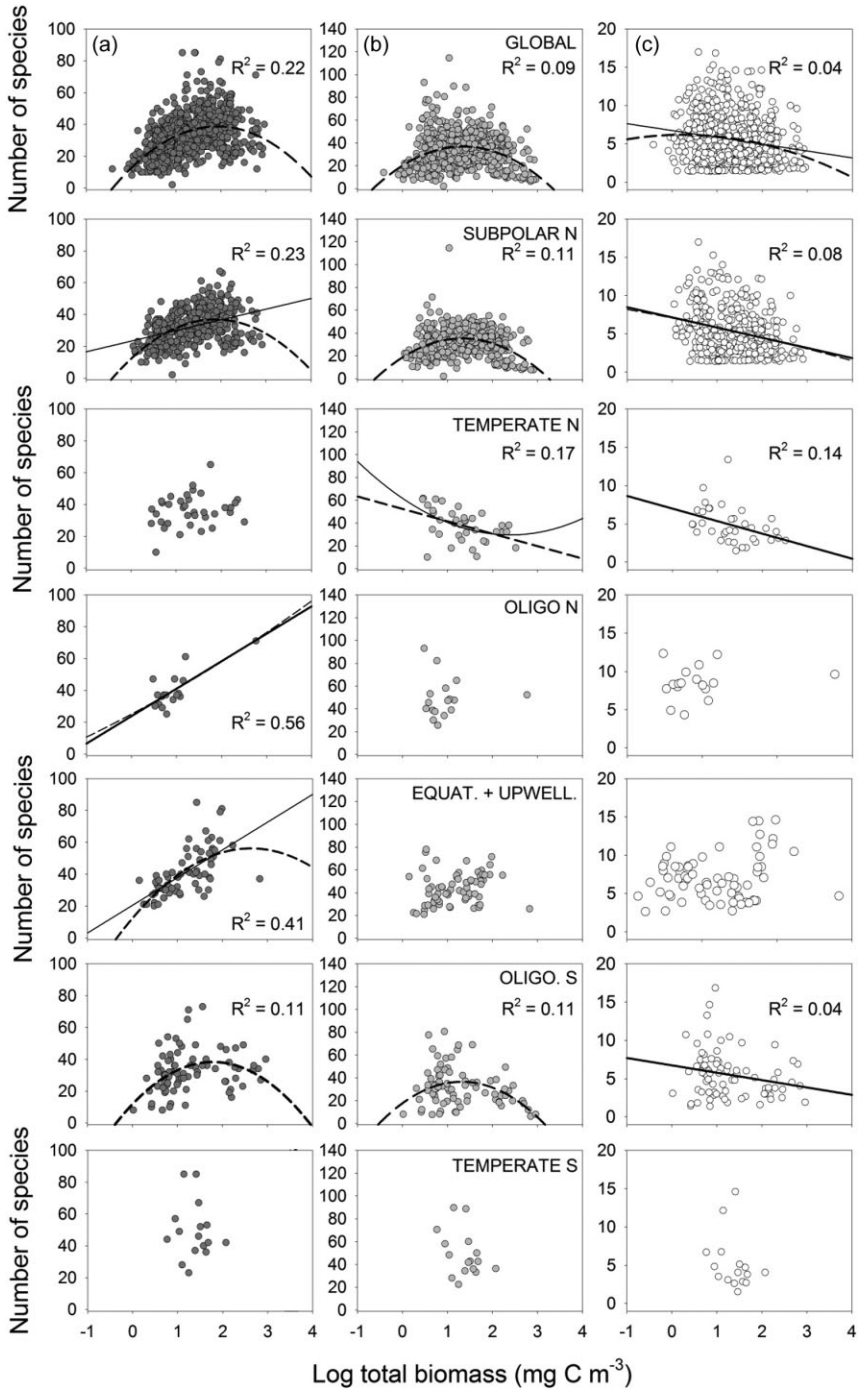

Figure 3 Productivity–diversity relationship at different spatial scales. Superficial species richness (a) observed (S_{obs}) and standardized by sampling effort (b) by Chao's method for $n = 10{,}000$ individuals (S_{iNext}) and (c) by Alroy's SQS method for $q = 0.7$ (70% resampled) (S_{SQS}) is expressed as a function of total biomass. Data correspond to the global data set (GD, $n = 744$). Solid and dashed lines represent linear and quadratic functions fitted to data, respectively. Only significant fits are shown. The parameters describing each fitted model are shown in Appendix S4. The best fitted model, with the lowest value of the Akaike information criterion, is highlighted with a thicker line.

where species populations are strongly limited by dispersal. This reduced habitat connectivity would have increased the probability of geographical isolation and hence the rate of diversification under favourable conditions for phytoplankton growth.

Testing the prediction of the MTE

The relationship between S_{SQS} and temperature was best described by a linear regression model, while the relationships between S_{obs} and S_{iNext} with temperature were best described by a quadratic model, in agreement with previous results for marine planktonic foraminifera (Rutherford *et al*., 1999) and several terrestrial taxa (Algar *et al*., 2007). In all cases, the inverse of temperature explained only a low percentage of variability in species richness. Besides, the slopes of the linear regressions were significantly lower than those predicted by the MTE (Brown *et al*., 2004) (Table 2). These results argue against an exponential association between temperature and species richness as predicted by the MTE, and add marine phytoplankton to the list of taxonomic guilds that do not support

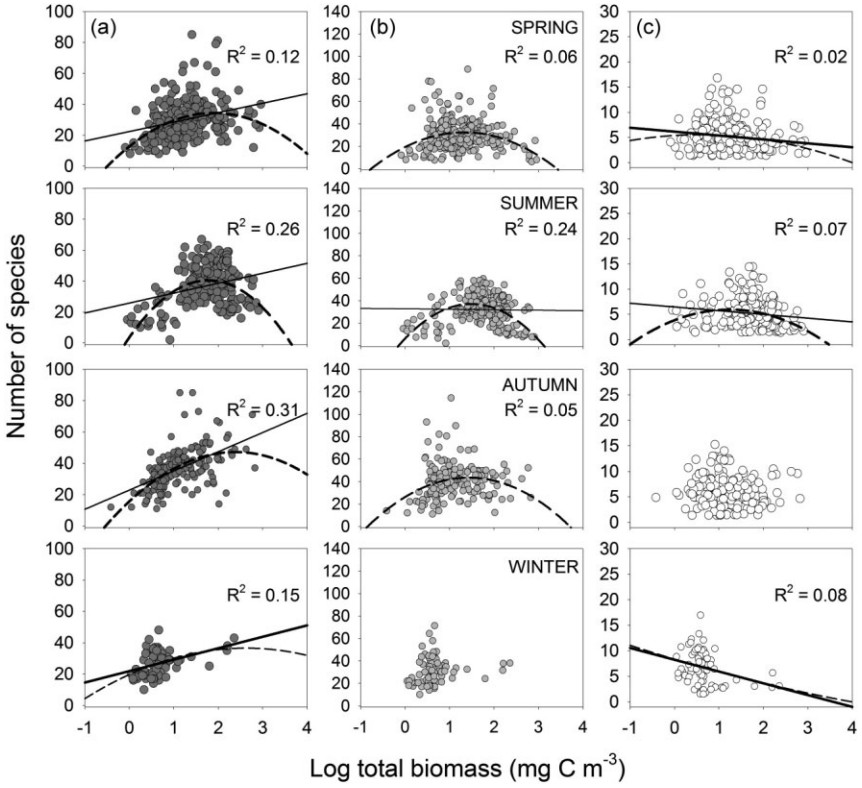

Figure 4 Productivity–diversity relationships for different seasons. Superficial species richness (a) observed (S_{obs}) and standardized by sampling effort (b) by Chao's method for $n = 10{,}000$ individuals (S_{iNext}) and (c) by Alroy's SQS method for $q = 0.7$ (70% resampled) (S_{SQS}) is expressed as a function of total biomass. Data correspond to the global data set (GD, $n = 744$). Solid and dashed lines represent the fitted linear and quadratic models, respectively. Only significant fits are shown. The parameters describing each fitted model are shown in Appendix S5. The best fitted model, with the lowest value of the Akaike information criterion, is highlighted with a thicker line.

this relationship as a cause of variability in species richness (Hawkins *et al.*, 2007).

Traditional sampling protocols severely underestimate the number of species with low population densities (Rodríguez-Ramos *et al.*, 2014). Hence, the number of species detected is expected to be higher where specific populations attain higher cell densities (i.e. in productive ecosystems). This assertion is supported by the significant positive relationship found between the number of observations per species and its total abundance per data set (Appendix S3). Recently, Marañón *et al.* (2014) found that resource limitation attenuates the temperature dependence of metabolic rates, and showed that phytoplankton growth rate (per day) is adjusted by changes in resource supply, with seawater temperature playing only a minor role. Their results could imply (omitting the effect of processes of biomass loss) a similar lack of dependence between the number of species and temperature, in agreement with our conclusions.

The productivity–diversity pattern

We performed an extensive analysis to investigate the shape and nature of the PDR over regional and seasonal scales. Globally, the pattern resulting from raw data supported, although with a large scatter of data, the long-standing idea of a hump-shaped relationship between diversity and biomass (Irigoien *et al.*, 2004). However, changes in community biomass were unable to explain a significant amount of variability in S_{obs}, S_{iNext}, S_{SQS} or S_H, regardless of the geographical region (Appendix S4) or seasonal period (Appendix S5). These results support the conclusions of Adler *et al.* (2011) and Cermeño *et al.* (2013), who suggested that diversity and biomass (or primary production) in communities of terrestrial plants and marine phytoplankton are not linked mechanistically.

Factors explaining diversity patterns

The pairwise correlation analysis confirmed the absence of a relationship between diversity and primary production or temperature, and revealed that no other single environmental or biological variable was able to explain per se a significant amount of variability in species richness. We thus used a stepwise multiple regression analysis to test for the combined effect of several factors upon diversity patterns, a method which has previously been satisfactorily used for a variety of organisms. For example, freshwater phytoplankton diversity is significantly affected by temperature, Chl*a* concentration and lake area and depth, explaining more than 50% of the variability in species richness (Stomp *et al.*, 2011). Recently, Azovsky & Mazei (2013) determined that the species richness of marine benthic ciliates was highly dependent on salinity and investigation effort, which together explained about 90% of variability in species richness. In the present study, however, the multiple-drivers model (including standing stock descriptors and temperature but no other environmental variables) explained only a relatively low percentage of variability in species richness, both for observed and standardized estimates (Appendix S6). Our results suggest that environmental regression is a poor method

for the prediction of diversity patterns for marine nano- and microphytoplankton, although some previous studies have shown its utility. For instance, Ladau *et al.* (2013) used a species distribution modelling approach to generate global maps of marine bacterial diversity by regressing observations of diversity on environmental conditions and then projecting the regression into geographical space. They found a seasonal pattern in the latitudinal distribution of marine bacterial diversity, with peaks at high latitudes in winter. However, this result could be partially influenced by the observation that species are easily detected during that season because communities exhibit typically more even distributions (Caporaso *et al.*, 2012).

We suggest two potential, non-exclusive explanations for these results. First, conventional sampling methods systematically overlook the number of rare species (Rodríguez-Ramos *et al.*, 2014), which limits our ability to delineate the patterns of species richness and the underlying mechanisms (Cermeño *et al.*, 2013). Second, niche and neutral mechanisms may act simultaneously to adjust the taxonomic composition of local communities. In this scenario, the identity of dominant species would be determined primarily by local environmental selection, while neutral processes such as individuals' dispersal and demographic stochasticity would account for the large pool of rare taxa present in these communities.

The results presented here suggest that environmental conditions, seawater temperature and ecosystem productivity are poor predictors of phytoplankton species richness variability in the world ocean. Our analyses are based on morphologically defined species, easily recognized under the microscope, but obviate small-sized picophytoplankton species which account for much of the community biomass and primary production in low-latitude open-ocean regions. We have shown for variability of species richness large-sized species may be controlled to first order by broad-scale processes such as the dispersal of individuals. Assuming that these mechanisms also influence the distribution of small picophytoplankton taxonomic units, we would expect similar diversity patterns to those observed for the nano- and microphytoplankton groups.

Conclusions

Marine nano- and microphytoplankton exhibit a weak latitudinal diversity gradient. Temperature and productivity were unable to explain the variability in species richness across ecosystems. We conclude that the use of non-uniform sampling effort along productivity gradients distorted the patterns of species richness reported previously. Sampling biases and insufficient data are key issues affecting the estimates of species richness and previously reported patterns of microbial plankton biogeography.

Our results agree with the view that the patterns of species richness in the microbial world differ from those of macroorganisms (Hillebrand & Azovsky, 2001; Finlay, 2002; Fenchel & Finlay, 2004). Emergent hypotheses (Holt, 2006; Vergnon *et al.*, 2009; Segura *et al.*, 2011) suggest that neutrality can arise from ecological and evolutionary interactions among species, giving rise to groups of ecologically similar coexisting species. According to our results, niche- and neutral-based processes might not be mutually exclusive but could be concurrent drivers of diversity patterns and community composition. Passive dispersal with marine currents gives rise to a seed bank of species, from which only a few species are environmentally selected over short time-scales.

New surveys applying increased sampling efforts and standardization techniques are required in order to obtain meaningful and comparable estimates of phytoplankton species richness and community composition. These are key objectives for a better understanding of the spatial and temporal distribution of microbial plankton species, and the linkage between community structure and ecosystem functioning.

ACKNOWLEDGEMENTS

We thank all people who contributed to the collection of data and sample analyses. We also thank J. L. Otero-Ferrer for his valuable advice and help on spatial statistical analyses and two anonymous referees for their helpful comments.

This research was funded by the Spanish Ministry of Economy and Competitiveness (MINECO) through the project DISTRAL (CTM2011-25035). This study is a contribution to the international IMBER project and was supported by the UK Natural Environment Research Council National Capability funding to the Plymouth Marine Laboratory and the National Oceanography Centre, Southampton. This is contribution number 258 of the AMT programme. T.R.-R. acknowledges the receipt of a FPI Predoctoral Fellowship from MINECO, through the project 'Macroecological patterns in marine phytoplankton' (CTM2008-03699). P.C. was supported by a 'Ramón y Cajal' contract from MINECO.

REFERENCES

Adler, P.B., Seabloom, E.W., Borer, E.T. *et al.* (2011) Productivity is a poor predictor of plant species richness. *Science*, **333**, 1750–1753.

Algar, A.C., Kerr, J.T. & Currie, D.J. (2007) A test of metabolic theory as the mechanism underlying broad-scale species-richness gradients. *Global Ecology and Biogeography*, **16**, 170–178.

Allen, A.P., Gillooly, J.F. & Brown, J.H. (2007) Recasting the species–energy hypothesis: the different roles of kinetic and potential energy in regulating biodiversity. *Scaling biodiversity* (ed. by D. Storch, P.A. Marquet and J.H. Brown), pp. 283–299. Cambridge University Press, Cambridge.

Alroy, J. (2010) Geographical, environmental and intrinsic biotic controls on Phanerozoic marine diversification. *Palaeontology*, **53**, 1211–1235.

Azovsky, A. & Mazei, Y. (2013) Do microbes have macroecology? Large-scale patterns in the diversity and distribution of marine benthic ciliates. *Global Ecology and Biogeography*, **22**, 163–172.

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

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

Burnham, K.P. & Anderson, D.R. (2002) *Model selection and multi-model inference: a practical information-theoretic approach*. Springer-Verlag, New York.

Caporaso, J.G., Paszkiewicz, K., Field, D., Knight, R. & Gilbert, J.A. (2012) The western English Channel contains a persistent microbial seed bank. *ISME Journal*, **6**, 1089–1093.

Cermeño, P., Dutkiewicz, S., Harris, R.P., Follows, M., Schofield, O. & Falkowski, P.G. (2008a) The role of nutricline depth in regulating the ocean carbon cycle. *Proceedings of the National Academy of Sciences USA*, **105**, 20344–20349.

Cermeño, P., Maranón, E., Harbour, D., Figueiras, F.G., Crespo, B.G., Huete-Ortega, M., Varela, M. & Harris, R. (2008b) Resource levels, allometric scaling of population abundance, and marine phytoplankton diversity. *Limnology and Oceanography*, **53**, 312–318.

Cermeño, P., Rodríguez-Ramos, T., Dornelas, M., Figueiras, F.G., Marañón, E., Teixeira, I.G. & Vallina, S.M. (2013) Species richness in marine phytoplankton communities is not correlated to ecosystem productivity. *Marine Ecology Progress Series*, **488**, 1–9.

Chao, A., Gotelli, N.J., Hsieh, T., Sander, E.L., Ma, K., Colwell, R.K. & Ellison, A.M. (2014) Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies. *Ecological Monographs*, **84**, 45–67.

Chase, J.M. & Leibold, M.A. (2002) Spatial scale dictates the productivity–biodiversity relationship. *Nature*, **416**, 427–430.

Chust, G., Irigoien, X., Chave, J. & Harris, R.P. (2013) Latitudinal phytoplankton distribution and the neutral theory of biodiversity. *Global Ecology and Biogeography*, **22**, 531–543.

Dormann, F.C., McPherson, M.J., Araújo, B.M., Bivand, R., Bolliger, J., Carl, G., Davies, G.R., Hirzel, A., Jetz, W., Kissling, D.W., Kühn, I., Ohlemüller, R., Peres-Neto, P.R., Reineking, B., Schröder, B., Schurr, F.M. & Wilson, R. (2007) Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography*, **30**, 609–628.

Fenchel, T. & Finlay, B.J. (2004) The ubiquity of small species: patterns of local and global diversity. *Bioscience*, **54**, 777–784.

Finlay, B.J. (2002) Global dispersal of free-living microbial eukaryote species. *Science*, **296**, 1061–1063.

Fuhrman, J.A., Steele, J.A., Hewson, I., Schwalbach, M.S., Brown, M.V., Green, J.L. & Brown, J.H. (2008) A latitudinal diversity gradient in planktonic marine bacteria. *Proceedings of the National Academy of Sciences USA*, **105**, 7774–7778.

Gibbons, S.M., Caporaso, J.G., Pirrung, M., Field, D., Knight, R. & Gilbert, J.A. (2013) Evidence for a persistent microbial seed bank throughout the global ocean. *Proceedings of the National Academy of Sciences USA*, **110**, 4651–4655.

Gotelli, N.J. & Colwell, R.K. (2001) Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters*, **4**, 379–391.

Hawkins, B.A., Albuquerque, F.S., Araujo, M.B., Beck, J., Bini, L.M., Cabrero-Sanudo, F.J., Castro-Parga, I., Diniz-Filho, J.A.F., Ferrer-Castan, D. & Field, R. (2007) A global evaluation of metabolic theory as an explanation for terrestrial species richness gradients. *Ecology*, **88**, 1877–1888.

Hillebrand, H. (2004) On the generality of the latitudinal diversity gradient. *The American Naturalist*, **163**, 192–211.

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

Holligan, P.M., Williams, P.J., Purdie, D. & Harris, R.P. (1984) Photosynthesis, respiration and nitrogen supply of plankton populations in stratified, frontal and tidally mixed shelf waters. *Marine Ecology Progress Series*, **17**, 201–213.

Holt, R.D. (2006) Emergent neutrality. *Trends in Ecology and Evolution*, **21**, 531–533.

Horner-Devine, M.C., Leibold, M.A., Smith, V.H. & Bohannan, B.J.M. (2003) Bacterial diversity patterns along a gradient of primary productivity. *Ecology Letters*, **6**, 613–622.

Hsieh, T.C., Ma, K.H. & Chao, A. (2014) iNEXT: an R package for interpolation and extrapolation in measuring species diversity. Available at: http://chao.stat.nthu.edu.tw/blog/software-download/inext-online/

Huisman, J. (2010) Comment on 'Patterns of diversity in marine phytoplankton'. *Science*, **329**, 512–512.

Hurlbert, S.H. (1971) The nonconcept of species diversity: a critique and alternative parameters. *Ecology*, **52**, 577–586.

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

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

Marañón, E., Holligan, P.M., Varela, M., Mouriño, B. & Bale, A.J. (2000) Basin-scale variability of phytoplankton biomass, production and growth in the Atlantic Ocean. *Deep Sea Research Part I: Oceanographic Research Papers*, **47**, 825–857.

Marañón, E., Cermeño, P., Huete-Ortega, M., López-Sandoval, D.C., Mouriño-Carballido, B. & Rodríguez-Ramos, T. (2014) Resource supply overrides temperature as a controlling factor of marine phytoplankton growth. *PLoS ONE*, **9**, 6, e99312.

Mittelbach, G.G., Steiner, C.F., Scheiner, S.M., Gross, K.L., Reynolds, H.L., Waide, R.B., Willig, M.R., Dodson, S.I. & Gough, L. (2001) What is the observed relationship between species richness and productivity? *Ecology*, **82**, 2381–2396.

Mittelbach, G.G., Schemske, D.W., Cornell, H.V., Allen, A.P., Brown, J.M., Bush, M.B., Harrison, S.P., Hurlbert, A.H., Knowlton, N. & Lessios, H.A. (2007) Evolution and the latitudinal diversity gradient: speciation, extinction and biogeography. *Ecology Letters*, **10**, 315–331.

Montresor, M., Lovejoy, C., Orsini, L., Procaccini, G. & Roy, S. (2003) Bipolar distribution of the cyst-forming dinoflagellate *Polarella glacialis*. *Polar Biology*, **26**, 186–194.

Pedrós-Alió, C. (2006) Marine microbial diversity: can it be determined? *Trends in Microbiology*, **14**, 257–263.

Pianka, E.R. (1966) Latitudinal gradients in species diversity: a review of concepts. *The American Naturalist*, **100**, 33–46.

Pommier, T., Canbäck, B., Riemann, L., Boström, K.H., Simu, K., Lundberg, P., Tunlid, A. & Hagström, Å. (2007) Global patterns of diversity and community structure in marine bacterioplankton. *Molecular Ecology*, **16**, 867–880.

R Core Team (2013) R: a language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria. Available at: http://www.R-project.org/.

Reshef, D.N., Reshef, Y.A., Finucane, H.K., Grossman, S.R., McVean, G., Turnbaugh, P.J., Lander, E.S., Mitzenmacher, M. & Sabeti, P. (2011) Detecting novel associations in large data sets. *Science*, **334**, 1518–1524.

Rodríguez-Ramos, T., Dornelas, M., Marañón, E. & Cermeño, P. (2014) Conventional sampling methods severely underestimate phytoplankton species richness. *Journal of Plankton Research*, **36**, 334–343.

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

Sal, S., López-Urrutia, Á., Irigoien, X., Harbour, D.S. & Harris, R.P. (2013) Marine microplankton diversity database: ecological archives E094-149. *Ecology*, **94**, 1658.

Segura, A.M., Calliari, D., Kruk, C., Conde, D., Bonilla, S. & Fort, H. (2011) Emergent neutrality drives phytoplankton species coexistence. *Proceedings of the Royal Society B: Biological Sciences*, **278**, 2355–2361.

Stomp, M., Huisman, J., Mittelbach, G.G., Litchman, E. & Klausmeier, C.A. (2011) Large-scale biodiversity patterns in freshwater phytoplankton. *Ecology*, **92**, 2096–2107.

Sul, W.J., Oliver, T.A., Ducklow, H.W., Amaral-Zettlera, L.A. & Sogin, M.L. (2013) Marine bacteria exhibit a bipolar distribution. *Proceedings of the National Academy of Sciences USA*, **110**, 2342–2347.

Vergnon, R., Dulvy, N.K. & Freckleton, R.P. (2009) Niche versus neutrality: uncovering the drivers of diversity in a species-rich community. *Ecology Letters*, **12**, 1079–1090.

Willig, M.R., Kaufman, D.M. & Stevens, R.D. (2003) Latitudinal gradients of biodiversity: pattern, process, scale, and synthesis. *Annual Review of Ecology, Evolution, and Systematics*, **34**, 273–309.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at the publisher's web-site.

Appendix S1 Methodology employed for measuring the variables included in the AMT data set.

Appendix S2 (Semi-)matrices of correlation between pairwise variables for each data set.

Appendix S3 Occurrence of each species as a function of its total abundance in each data set.

Appendix S4 Details of the models fitted to diversity–biomass relationships for different regions.

Appendix S5 Details on the models fitted to diversity–biomass relationships for different seasons.

Appendix S6 Parameters of the multiple linear regression models explaining observed species richness (S_{obs}) and species richness standardized by Chao's method (S_{iNext}), for each data set.

BIOSKETCHES

Tamara Rodríguez-Ramos is a biological oceanographer interested in the physiology and macroecology of marine phytoplankton. She combines empirical and theoretical approaches as well as modelling and data analysis tools to better understand the mechanisms and processes responsible for community structure and species assemblage, as well as the relationship between metabolism, cell size and environment.

Emilio Marañón is a biological oceanographer interested in the ecology and biogeochemical role of phytoplankton. He obtained his PhD from the University of Oviedo (1995) and was a post-doctoral researcher at the Southampton Oceanography Centre (1996–98). Since 1999 he has been a senior lecturer in ecology at the University of Vigo.

Pedro Cermeño is research scientist at the Marine Science Institute located in Barcelona and belonging to the National Environmental Research Council. His research focuses on understanding the patterns of microbial species distribution and their impact on global biogeochemical cycles using field data, experimental analyses and modelling work.

Editor: Michael Rex