LIMNOLOGY
and
OCEANOGRAPHY

ASLO

Limnol. Oceanogr. 60, 2015, 1020–1038

doi: 10.1002/lno.10074

Ecological niches of open ocean phytoplankton taxa

Philipp Brun,*[1,2] Meike Vogt,[1] Mark R. Payne,[1,2] Nicolas Gruber,[1] Colleen J. O'Brien,[1] Erik T. Buitenhuis,[3] Corinne Le Quéré,[3] Karine Leblanc,[4] Ya-Wei Luo[5]

[1]Environmental Physics Group, Institute of Biogeochemistry and Pollutant Dynamics, ETH Zürich, Universitätstrasse 16, 8092 Zürich, Switzerland

[2]Centre for Ocean Life, National Institute of Aquatic Resources (DTU-Aqua), Technical University of Denmark, Kavalergården 6, 2920, Charlottenlund, Denmark

[3]Tyndall Center for Climate Change Research and School of Environmental Sciences, University of East Anglia, Norwich Research Park, Norwich NR4 7TJ, United Kingdom

[4]Aix-Marseille Université, Université du Sud Toulon-Var, CNRS/INSU, IRD, Mediterranean Institute of Oceanography (MIO), UM110, 13288, Marseille, Cedex 09, France

[5]State Key Laboratory of Marine Environmental Science, Xiamen University, Xiamen 361102, China

Abstract

We characterize the realized ecological niches of 133 phytoplankton taxa in the open ocean based on observations from the MAREDAT initiative and a statistical species distribution model (MaxEnt). The models find that the physical conditions (mixed layer depth, temperature, light) govern large-scale patterns in phytoplankton biogeography over nutrient availability. Strongest differences in the realized niche centers were found between diatoms and coccolithophores. Diatoms (87 species) occur in habitats with significantly lower temperatures, light intensity and salinity, with deeper mixed layers, and with higher nitrate and silicate concentrations than coccolithophores (40 species). However, we could not statistically separate the realized niches of coccolithophores from those of diazotrophs (two genera) and picophytoplankton (two genera). *Phaeocystis* (two species) niches only clearly differed from diatom niches for temperature. While the realized niches of diatoms cover the majority of niche space, the niches of picophytoplankton and coccolithophores spread across an intermediate fraction and diazotroph and colonial *Phaeocystis* niches only occur within a relatively confined range of environmental conditions in the open ocean. Our estimates of the realized niches roughly match the predictions of Reynolds' C-S-R model for the global ocean, namely that taxa classified as nutrient stress tolerant have niches at lower nutrient and higher irradiance conditions than light stress tolerant taxa. Yet, there is considerable within-class variability in niche centers, and many taxa occupy broad niches, suggesting that more complex approaches may be necessary to capture all aspects of phytoplankton ecology.

Introduction

Marine phytoplankton ubiquitously inhabit the illuminated upper layers of the world's oceans, but their fate has consequences that reach far beyond the local mixed layer. Marine phytoplankton generate roughly half of the Earth's net primary production (Field et al. 1998) and make significant contributions to the global biogeochemical cycles of many biologically relevant elements such as carbon, nitrogen, phosphorus, and silicon (Falkowski 1994; Boyd and Doney 2003; Sarmiento and Gruber 2006). Phytoplankton are highly diverse, with representatives playing various roles in biogeochemistry (Falkowski et al. 2004; Le Quéré et al. 2005) and showing different distribution patterns (Boyd et al. 2010). Understanding the underlying mechanisms which shape phytoplankton biogeography and cause its current changes (e.g., Poloczanska et al. 2013) is, therefore, an important question but a complex one.

The ecological niche is a concept that has contributed greatly to our comprehension of patterns in the large-scale biogeography of taxa (Colwell and Rangel 2009). The fundamental ecological niche is the environment that permits sustained growth of a species (Hutchinson 1957), that is, it is a hypervolume in environmental space defined by favorable conditions in critical physical and chemical factors such as temperature, nutrients, and light. A species' realized niche (the conditions under which it can be observed) is usually a subset of its fundamental niche. The realized niche may be restricted by dispersal ability, interspecific competition or

*Correspondence: pgbr@aqua.dtu.dk

predation. One important strength of the concept of the ecological niche is that the favorable ranges in environmental conditions can be quantified based on laboratory experiments and field observations (Colwell and Rangel 2009).

In the past decades, a multitude of methods to quantify realized ecological niches based on field data have been developed. Phytoplankton ecological niches have been investigated with ordinal methods, such as the Outlying Mean Index, which compare habitat conditions of a species to mean habitat conditions of a sampling area (Dolédec et al. 2000; Grüner et al. 2011). However, these methods typically rely on abundance data from systematically sampled surveys—a condition rarely met in the open ocean.

Species distribution models (SDMs) have been shown to represent a powerful alternative with lower requirements to the extent of observational data. SDMs represent a family of statistical tools that are used to analyze and predict geographical ranges of species occurrence based on approximations of the realized ecological niche (Guisan and Zimmermann 2000; Elith and Leathwick 2009). Most SDM-based studies have been performed in terrestrial systems and only recently have SDMs been used in the marine realm (Robinson et al. 2011). In marine systems phytoplankton belong to the groups that received the least attention with the notable exception of Irwin et al. (2012) who used an SDM approach to characterize the realized ecological niches of diatoms and dinoflagellates in the North Atlantic.

In the field, nutrients (nitrate, phosphate, silicate, and iron), irradiance intensity (i.e., light), and temperature have been shown to be critical factors for phytoplankton ecological niches (reviewed in Boyd et al. 2010). However, the degree to which these factors control plankton biogeography is still poorly understood on the global scale (Buitenhuis et al. 2013; Luo et al. 2014). Observational data of phytoplankton distribution and abundance are very limited in the open ocean and often restricted to a few key species (Hood et al. 2006; Buitenhuis et al. 2013).

Different concepts exist to aggregate the roughly 20,000 species of phytoplankton (Falkowski et al. 2004) into few "manageable" groups. Two well established concepts are (i) plankton functional types (PFTs) (Iglesias-Rodríguez et al. 2002; Le Quéré et al. 2005; Hood et al. 2006), and the Reynolds' C-S-R model (Reynolds 1988, 2006), which builds on Margalef's "mandala" model (Margalef 1978).

The concept of PFTs involves the aggregation of plankton species into groups that share the ability to perform specific biogeochemical functions, such as silicification (diatoms) or calcification (coccolithophores), or fill a similar role due to their size, contribution to primary production, or trophic level (e.g., picophytoplankton, microzooplankton, mesozooplankton, and macrozooplankton) (Iglesias-Rodríguez et al. 2002; Le Quéré et al. 2005). This concept has been adopted widely in the marine modeling community, especially to represent lower trophic levels of marine ecosystems to investigate the impact of climate change on marine ecosystems, and to quantify potential feedbacks of marine ecosystems to global biogeochemical cycles (e.g., Moore et al. 2002; Anderson et al. 2010; Buitenhuis et al. 2010).

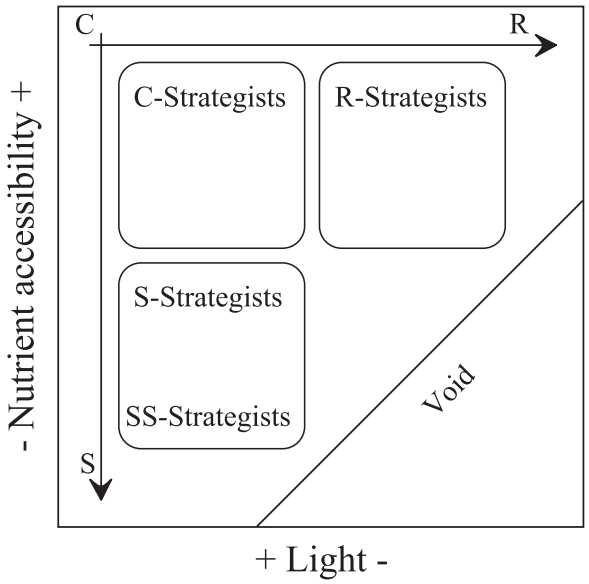

Fig. 1. Predicted distribution of phytoplankton groups in nutrient accessibility versus light availability space after Smayda and Reynolds (2001) and Reynolds (2006).

In the concept of Reynolds' C-S-R model the occurrence of phytoplankton is linked to two environmental factors, nutrient accessibility, and light availability (Fig. 1), which are assumed to be the major dimensions of their ecological niches. Three major survival strategies for phytoplankton are defined based on a set of physiological and morphological traits, and these survival strategies are linked to distinct ecological niches. Concretely, the following groups are distinguished: Colonist taxa (C-strategists) are fast growing, round, and rather small. C-strategists include many forms of nanoplanktonic flagellates but also some diatom genera. They mainly grow in coastal areas, where both factors, nutrients and light, are richly available. Nutrient stress tolerant species (S-strategists) are abundant in the open ocean. They consist of rather large species that grow slowly and have a high nutrient affinity (Fig. 1). Some S-strategists have additional adaptations to nutrient-poor conditions, for example, diazotrophy or mixotropy. S-strategists comprise the nitrogen-fixing cyanobacteria *Trichodesmium*, the coccolithophores and some diatoms but also the subgroup of chronic nutrient stress tolerant species (SS) such as the picoplanktonic cyanobacteria *Prochlorococcus*. Finally, light stress tolerant organisms (R-strategists) are groups of relatively large, but fast growing taxa. R-strategists typically have elongated body-shapes as a morphological adaptation to efficient light harvesting. Many diatom genera show R-strategist adaptations. Reynolds' C-S-R model has been used to study freshwater

Table 1. Observational data for different PFTs: Only taxa with more than 15 observations are included. Presence cells have been aggregated $1^\circ \times 1^\circ$ cells and masked for open ocean conditions and full environmental data coverage.

PFT	Coccos*	Diatoms	Diazos*	Phaeos*	Picos*
MAREDAT observations	11,702	90,648	3841	3527	40,946
Taxonomic resolution	Species	Species	Genus	Species	Genus
Grid cells with presence data	3771	4075	641	123	1609
Number of taxa	40	87	2	2	2
Grid cells per taxon	15-748	15-216	137-504	44-79	620-967
C-S-R classifications	40 S	4 S, 30 R	1 S	N/C†	1 SS
Reference	O'Brien et al. 2013	Leblanc et al. 2012	Luo et al. 2012	Vogt et al. 2012	Buitenhuis et al. 2013

*Coccos, coccolithophores; Diazos, diazotrophs; Phaeos, *Phaeocystis*; Picos, picophytoplankton.
†N/C, not classified.

phytoplankton (Reynolds 1988) and harmful algal blooms (Smayda and Reynolds 2001) but a large number of open ocean phytoplankton taxa have also been classified (Reynolds 2006). However, this classification is mainly derived from laboratory investigations and qualitative information about distributions (Smayda and Reynolds 2001).

Whether or not Reynolds C-S-R model has skill in predicting ecological niches of the classified taxa in the open ocean has never been thoroughly validated in the field. The legitimation of upscaling Reynolds' model to the open ocean may in fact be doubted because the relationship between environmental drivers and species occurrence is scale dependent (Elith and Leathwick 2009). Identifying light and nutrients as key drivers for phytoplankton succession in freshwater or regional marine systems does not mean they also govern global-scale patterns of species distributions. The importance of other limiting abiotic factors (Boyd et al. 2010) as well as biotic interactions such as grazing pressure may render the use of Reynolds' C-S-R model marginal on global scales.

In this article, we aim to challenge these categorization concepts by exploiting a novel database that contains observations for diverse plankton taxa on the global scale. The recently published MARine Ecosystem DATa (MAREDAT) initiative contains approximately 500,000 georeferenced global plankton abundance and biomass measurements collected from numerous field studies for 11 PFTs (Buitenhuis et al. 2013). While most observations are resolved to the species level, phytoplankton data in MAREDAT are grouped into five PFTs: coccolithophores, diatoms, diazotrophs, *Phaeocystis*, and picophytoplankton (Le Quéré et al. 2005). We use the MAREDAT database to characterize the realized niches of 133 phytoplankton species and genera in the open ocean by applying the SDM (MaxEnt) (Phillips et al. 2004, 2006). We investigate two overarching questions: First, we analyze which environmental factors show the highest predictive power for large-scale phytoplankton distributions. Second, we assess whether PFTs and a priori classified S- and R-strategists encompass taxa with similar ecological niches.

Methods

Presence data

The MAREDAT initiative includes observational biomass and abundance data for 11 PFTs (Buitenhuis et al. 2013). Five of them describe autotrophic organisms and are investigated in this study. These include coccolithophores (O'Brien et al. 2013), diatoms (Leblanc et al. 2012), diazotrophs (Luo et al. 2012), *Phaeocystis* (Vogt et al. 2012), and picophytoplankton (Buitenhuis et al. 2012). The number of raw observations within the investigated PFTs ranges from 3527 for *Phaeocystis* to 90,648 for the diatoms (Table 1). Most phytoplankton taxa in the MAREDAT dataset lack confirmed absence observations (diatoms, picophytoplankton). To have consistent data we remove all zero abundance observations. We furthermore discard biomass information and consider presence-only as the most comparable and robust piece of information in the dataset. Taxonomic information is not available for all PFTs to the same degree: full species-level taxonomic resolution is available for some coccolithophore, diatom, and *Phaeocystis* observations (68%, 75%, and 71%, respectively). Diazotrophs and picophytoplankton are resolved to the genus level at best (100% and 99%, respectively).

Environmental variables

Phytoplankton presence cells are matched up with environmental variables that have either been shown to be of physiological or ecological importance in the field or laboratory (e.g., Boyd et al. 2010). Since only a very limited number of colocated observations of such variables exist in MAREDAT, we matched phytoplankton observations with monthly data from climatological datasets. The environmental variables considered include mixed layer depth (MLD, m) (de Boyer Montégut 2004), as well as temperature (T, °C), salinity (S) and the concentrations of nitrate (NO_3^-), phosphate (PO_4^{3-}), and silicate ($Si(OH)_4$; all nutrients in units of μmoles L^{-1}) from the World Ocean Atlas 2009 (Antonov et al. 2010; Garcia et al. 2010; Locarnini et al. 2010). We average monthly climatological values of T, S, and nutrients over the climatological mixed layer. Furthermore, we use

modeled iron fields from the Community Earth System Model (Hurrell et al. 2013) and the Pelagic Interaction Scheme for Carbon and Ecosystem Studies (Aumont and Bopp 2006).

Finally, we consider mean photosynthetically active radiation in the mixed layer (MLPAR) estimated as

$$\mathrm{MLPAR}=\frac{1}{\mathrm{MLD}}\int_{0}^{z=\mathrm{MLD}}\mathrm{PAR}\times e^{-K_{\mathrm{ext}}}dz=\frac{\mathrm{PAR}}{K_{\mathrm{ext}}\mathrm{MLD}}(1-e^{-K_{\mathrm{ext}}\mathrm{MLD}})$$

The integration variable z represents depth (m), and PAR is photosynthetically active radiation at the surface (μmoles m^{-2} s^{-1}). The light attenuation coefficient K_{ext} (Morel 1988) in this equation is a function of chlorophyll concentration (CHL, μg L^{-1}) and includes the effect of sea water attenuation

$$K_{\mathrm{ext}}=0.121\times\mathrm{CHL}^{0.428},$$

where we assume CHL to be homogeneous within the mixed layer. Data for the variables CHL and PAR were obtained from climatologies derived from Sea-viewing Wide Field-of-view Sensor data (SeaWiFS; www.oceancolor.gsfc.nasa.gov).

Data treatment

Observations are aggregated into $1^\circ \times 1^\circ$ cells to reduce spatial biases, as numerous observations can occur within the same pixel. For each month and each taxon, cells with at least one (presence) observation within the local mixed layer are thereby defined as "presence cells" (22,138 cells in total). Presence cells from the same taxon can thus overlap geographically if observations were made in the same area but during different months. We aggregate data within the local MLD, which we assume to be of homogenous physico-chemical conditions as a result of physical mixing.

We restrict our study of phytoplankton ecological niches to the "open ocean" to avoid coastal effects that cannot be resolved by the coarse environmental data, we use to construct the statistical models. The open ocean is defined as the region having a water depth of more than 200 m as indicated by the ETOPO1 Global Relief Model (Amante and Eakins, 2009) and having a minimum sea surface salinity of 30. We furthermore confine the investigated area to cells that have corresponding data for all environmental variables used in our models. Missing data is primarily a problem for MLD values in the high latitudes. These restrictions lead to the omission of 43% of data-containing grid cells, predominantly in coastal areas.

Taxa that are present in fewer than 15 grid cells are removed to ensure a good performance of the MaxEnt models. A lower number would lead to a loss of model skill, while 15 presence cells have been shown to produce distribution predictions of a useful accuracy in terrestrial systems (Hernandez et al. 2006). The final number of grid cells used in this study is 11,212.

Final size of analyzed categories

The final sets of taxa resolved vary from 87 for diatoms to only two for diazotrophs, *Phaeocystis*, and picophytoplankton (*see* Tables 1, 2). Of the investigated 133 taxa, 73 were classified previously as S-strategists or R-strategists by Reynolds (2006) according to his C-S-R model. In our data, S-strategists represent the larger group including the 40 coccolithophore species, four diatom species and the diazotroph genus *Trichodesmium* while the related group of SS-strategists is represented by the picophytoplankton genus *Prochlorococcus*. R-strategists are represented by 27 diatom species (Table 1). A detailed list of the investigated taxa is given in Table 2. We do not test C-strategists here, since they only play a minor role in the open ocean.

Selection of investigated environmental variables

We explored the explanatory potential of all considered predictor variables in a preliminary run of our analysis. This led to the exclusion of iron from an in depth consideration here, despite its potentially important role in limiting phytoplankton growth (Boyd et al. 2010). The main reason for exclusion is the poor data coverage on the global scale (*see* also Palacz et al. 2013). Modeled iron concentrations predict phytoplankton distributions well, but differences between the two modeled iron fields are large (Pearson correlation coefficient (r) of 0.75 across the investigated area).

Similarly, a range of spatial and temporal gradients of the different variables were tested but did not improve the statistical model, therefore, none were included in our final set of predictor variables. We tested the spatial gradients of NO_3^-, salinity, and temperature at the surface and in 200 m depth and the temporal change of surface temperature, NO_3^-, PO_4^{3-}, and MLD. The obtained univariate model performances as measured by the area under the receiver operating characteristic curve (*see* section "Model performance" below) was between 0.66 (spatial gradients of NO_3^- and temperature at the surface) and 0.56 (temporal gradient of PO_4^{3-}) for all investigated taxa.

A correlation analysis applied to the remaining variables reveals strong correlations between NO_3^- and PO_4^{3-} ($r = 0.96$). Strong correlations can lead to misleading attribution of variable importance in SDM (Phillips et al. 2004; Dormann et al. 2012; Irwin et al. 2012). Therefore, we excluded PO_4^{3-}, but included P^*, a variable that reflects the excess (or deficiency) of PO_4^{3-} versus NO_3^- (Deutsch et al. 2007), defined as

$$P^*=[\mathrm{PO}_4^{3-}]-\frac{1}{16}\times[\mathrm{NO}_3^-]$$

P^* is closely related to N^* (Gruber and Sarmiento 1997), and is potentially important as it could define the habitats of

Table 2. Taxa included in the analyses, and their C-S-R classification according to Reynolds (2006).

Coccolithophores		
strategists		
Acanthoica acanthifera	*Acanthoica quattrospina*	*Algirosphaera robusta*
Anacanthoica acanthos	*Calcidiscus leptoporus*	*Calciopappus rigidus*
Calciosolenia brasiliensis	*Calciosolenia murrayi*	*Coccolithus pelagicus*
Coronosphaera mediterranea	*Discosphaera tubifera*	*Emiliania huxleyi*
Florisphaera profunda var. profunda	*Gephyrocapsa ericsonii*	*Gephyrocapsa oceanica*
Gephyrocapsa ornata	*Gladiolithus flabellatus*	*Helicosphaera carteri*
Helladosphaera cornifera	*Holococcolithophora sphaeroidea*	*Michaelsarsia adriaticus*
Michaelsarsia elegans	*Oolithotus antillarum*	*Oolithotus fragilis*
Ophiaster hydroideus	*Palusphaera vandelii*	*Reticulofenestra parvula*
Reticulofenestra sessilis	*Rhabdosphaera clavigera*	*Rhabdosphaera hispida*
Rhabdosphaera xiphos	*Syracosphaera molischii*	*Syracosphaera prolongata*
Syracosphaera pulchra	*Syracosphaera pulchra holococcolithophore*	*Turrilithus latericioides*
Umbellosphaera irregularis	*Umbellosphaera tenuis*	*Umbilicosphaera hulburtiana*
Umbilicosphaera sibogae		
Diatoms		
C-strategists		
Cerataulina pelagica	*Coscinodiscus oculus-iridis*	*Coscinodiscus radiatus*
R-strategists		
Chaetoceros affinis	*Chaetoceros atlanticus*	*Chaetoceros bulbosum*
Chaetoceros compressus	*Chaetoceros concavicornis*	*Chaetoceros convolutus*
Chaetoceros curvisetus	*Chaetoceros dadayi*	*Chaetoceros debilis*
Chaetoceros decipiens	*Chaetoceros dichaeta*	*Chaetoceros didymus*
Chaetoceros hyalochaetae	*Chaetoceros lorenzianus*	*Chaetoceros peruvianus*
Chaetoceros phaeoceros	*Chaetoceros socialis*	*Chaetoceros tetrastichon*
Thalassiosira angulata	*Thalassiosira anguste-lineata*	*Thalassiosira gravida*
Thalassiosira rotula	*Thalassiosira subtilis*	*Leptocylindrus danicus*
Leptocylindrus mediterraneus	*Leptocylindrus minimus*	*Skeletonema costatum*
S-strategists		
Hemiaulus hauckii	*Hemiaulus sinensis*	*Pseudosolenia calcar-avis*
Rhizosolenia styliformis		
Unclassified diatom species		
Asterionellopsis glacialis	*Bacteriastrum delicatulum*	*Bacteriastrum furcatum*
Climacodium frauenfeldianum	*Corethron criophilum*	*Cylindrotheca closterium*
Dactyliosolen antarcticus	*Dactyliosolen fragilissimus*	*Detonula pumila*
Ditylum brightwellii	*Eucampia antarctica*	*Eucampia cornuta*
Eucampia zodiacus	*Fragilariopsis kerguelensis*	*Fragilariopsis obliquecostata*
Fragilariopsis oceanica	*Fragilariopsis rhombica*	*Guinardia cylindrus*
Guinardia delicatula	*Guinardia flaccida*	*Guinardia striata*
Lauderia annulata	*Lioloma delicatulum*	*Lioloma pacificum*
Meuniera membranacea	*Navicula planamembranacea*	*Nitzschia bicapitata*
Nitzschia closterium	*Nitzschia delicatissima*	*Nitzschia longissima*
Nitzschia seriata	*Nitzschia tenuirostris*	*Paralia sulcata*
Planktoniella sol	*Proboscia alata*	*Proboscia gracillima*
Proboscia indica	*Pseudo-nitzschia delicatissima*	*Pseudo-nitzschia heimii*
Pseudo-nitzschia pungens	*Pseudo-nitzschia seriata*	*Rhizosolenia bergonii*
Rhizosolenia chunii	*Rhizosolenia delicatula*	*Rhizosolenia hebetata f. hebetata*
Rhizosolenia hebetata f. semispina	*Rhizosolenia imbricata*	*Rhizosolenia setigera*
Roperia tesselata	*Thalassionema bacillare*	*Thalassionema frauenfeldii*
Thalassionema nitzschioides	*Thalassiothrix longissima*	

TABLE 2. Continued

Diazotrophs	
S-strategists	
Trichodesmium	
Unclassified diazotroph genera	
Richelia	
Phaeocystis	
Unclassified Phaeocystis species	
Phaeocystis antarctica	*Phaeocystis pouchetii*
Picophytoplankton	
SS-strategists	
Prochlorococcus	
Unclassified picophytoplankton genera	
Synechococcus	

diazotrophic organisms (Deutsch et al. 2007). P^* is only very weakly correlated with NO_3^- ($r = 0.07$).

We chose to keep NO_3^-, $Si(OH)_4$ and T, despite NO_3^- and $Si(OH)_4$ showing a correlation coefficient of $r = 0.74$ and NO_3^- and T showing an r of -0.85 which is higher than the frequently used correlation threshold of $|r| = 0.7$ (Dormann et al. 2012). However, these variables are of high ecological importance for phytoplankton growth and distribution (Boyd et al. 2010) and we use the models only to analyze ecological niches and not for applications where high correlations are particularly problematic such as making predictions for other geographic areas or time periods (Dormann et al. 2012). We, therefore, believe this exception is justified. Furthermore, a sensitivity analysis revealed that the high level of correlation for these variables did not lead to distorted permutation importance estimates: the relative permutation importance of the remaining variables did not change when either of the variables was left out of the model (data not shown).

The SDM MaxEnt

The MaxEnt SDM (Phillips et al. 2004, 2006) is one of the most widely used species distribution/environmental niche models with well over 1000 published applications (Merow et al. 2013). Species distributions are derived from presence-only data, by contrasting environmental conditions that prevail at all locations where a given species was observed with those conditions typical for the entire global ocean ("background" conditions). Strengths of MaxEnt are its high performance in comparison to other SDMs (Elith et al. 2006), its ability to characterize species' distributions and their realized niches comparably well even when the observational data are very limited (Hernandez et al. 2006) and its ability to generate response curves of an ecologically realistic shape (i.e., not overly complex) (Irwin et al. 2012). MaxEnt is a probabilistic method that approximates the probability of presence of a species as a function of environmental conditions. Probability functions that depend on the range of the input variables (e.g., temperature, nutrient concentrations, etc.) are then used to predict species distributions. We use MaxEnt software 3.3.3e (www.cs.princeton.edu/~schapire/maxent/) to fit MaxEnt models and to obtain model-related statistics. We disable threshold features to restrict the fitted functions to a moderate level of complexity, and we represent background conditions by 1000 random grid cells in the open ocean for each month (12,000 in total).

Model performance

We evaluate the receiver operating characteristic (ROC) metric (Swets 1988; Manel et al. 2001) to quantify the predictive quality of MaxEnt models. The ROC metric is based on a large number of probability thresholds that are used to divide the continuous probabilistic output of MaxEnt models into binary presence/absence predictions. In an ROC plot, the proportions of correctly predicted presences are plotted against the proportion of wrongly predicted presences for each of the threshold-dependent presence/absence predictions. A good model maximizes the proportion of true presences for all thresholds, while minimizing the proportion of false presences. The ROC plot can be summarized by the "AUC," the area under the ROC curve. An AUC of 1 means perfect discrimination and an AUC of 0.5 is the value expected by an unbiased random presence/absence generator. For each taxon, the AUC is reported as the mean of five cross-validation replicates, both for a MaxEnt model including all variables (multivariate model) as well as for MaxEnt models that only consider one variable at a time (univariate models). Additionally we perform Student's t-tests to determine whether the five AUC replicates perform significantly better than random (AUC = 0.5) at the 95% significance level, which we use as quality criterion for estimated univariate niches (*see* below).

Table 3. Average AUC for fitted univariate and multivariate MaxEnt models. Percentages in brackets show the ratio of species for which fitted models predict significantly better than random. The first column ("All") represents averages over all investigated taxa.

	All	Coccos	Diatoms	Diazos	Phaeos	Picos
Multivariate models						
	0.91	0.9	0.91	0.89	0.99	0.83
Univariate models						
MLD	0.73	0.77	0.71	0.73	0.67	0.64
	(95%)	(100%)	(92%)	(100%)	(100%)	(100%)
Temperature (T)	0.73	0.69	0.74	0.74	0.96	0.62
	(92%)	(95%)	(90%)	(100%)	(100%)	(100%)
NO_3^-	0.71	0.69	0.73	0.74	0.88	0.62
	(90%)	(92%)	(89%)	(100%)	(100%)	(100%)
Average light (MLPAR)	0.7	0.72	0.68	0.7	0.78	0.56
	(88%)	(92%)	(85%)	(100%)	(100%)	(100%)
Salinity (S)	0.69	0.68	0.69	0.72	0.85	0.73
	(83%)	(78%)	(85%)	(100%)	(100%)	(100%)
$Si(OH)_4$	0.67	0.71	0.65	0.62	0.86	0.64
	(84%)	(93%)	(79%)	(100%)	(100%)	(100%)
P^*	0.66	0.68	0.65	0.73	0.67	0.61
	(72%)	(72%)	(70%)	(100%)	(100%)	(100%)

Variable importance

The permutation importance as implemented in the MaxEnt software allows for the quantification of the contribution of a variable to the performance of a multivariate model. It is estimated by randomly permuting the values of one variable and measuring how much the performance of the fitted model drops compared to a model fitted with the original values. A variable with a high permutation importance causes a relatively high drop in model quality when its values are permuted and thus contributes much to the quality of the multivariate model. We calculate the permutation importance for each taxon as the average of a fivefold cross-validation.

Univariate niche parametrization

We characterize realized niches based on the response curves derived from univariate models. Response curves are diagrams showing the probabilistic output of MaxEnt as a function of an environmental variable. Multivariate MaxEnt models are not appropriate for response curve analysis because MaxEnt discards a fraction of parameters in the initial model design based on statistical performance as a measure to avoid over fitting. This can lead to models not considering certain input variables and corresponding flat response curves despite existing patterns.

We follow Irwin et al. (2012) and summarize the response curves by two parameters, that is, the univariate niche center (median) and the univariate niche breadth (interquartile range). The latter gives an estimate of the tolerance range of the taxon for a particular factor. Specialist taxa will have small niche breadth, whereas the niches of generalist taxa are broad. Note that this estimate of niche breadth is mainly suitable as a relative measure to compare different taxa. The estimates of the latter two parameters depend on the considered range of environmental conditions. To optimize the accuracy of the parameter estimates, the integration range of the response curves was defined by the 0.05 and the 99.5 percentiles of the open ocean conditions. Five hundred replicates of model estimates for each univariate niche center and breadth are created by resampling with the bootstrapping method as implemented in the MaxEnt software. A preliminary sensitivity analysis (data not shown) showed that 500 bootstrapping replicates allow a reliable estimation of 90% confidence intervals for the derived niche parameters. We only analyze niche center and breadth for univariate models that perform better than random, that is, which have an AUC significantly larger than 0.5 as indicated by a Student's *t*-test applied to the five cross-validation replicates.

Results

Model performance

Model performance

Based on the AUC scores, the multivariate MaxEnt models estimate the realized niches of marine phytoplankton very well using the available set of predictor variables (Table 3). The AUC for the multivariate models ranges from 0.78 to 1.00 for 132 investigated taxa (only the model for the diatom *Nitzschia bicapitata* has a lower AUC of 0.70). The mean AUC is near 0.90 for all PFTs except picophytoplankton, for

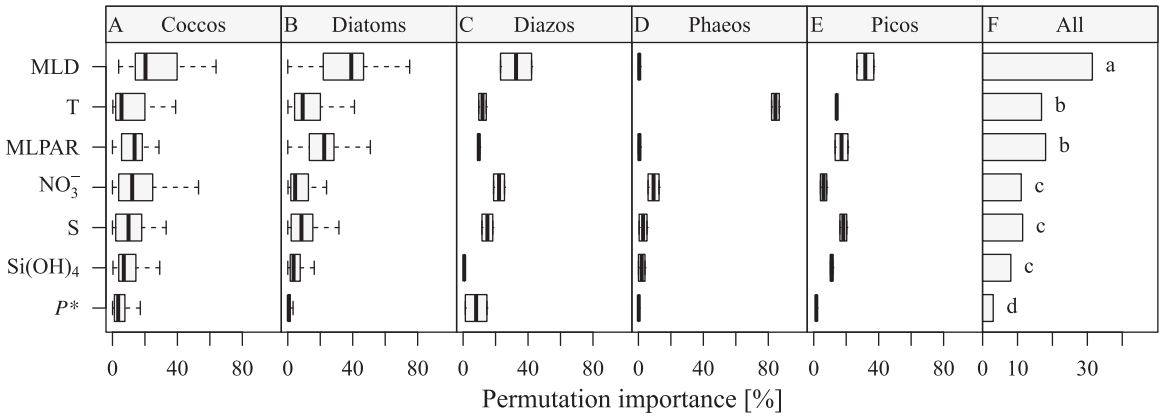

Fig. 2. Boxplots showing permutation importance of environmental variables for each PFT (A–E), central vertical lines indicate median values, boxes indicate interquartile ranges and error bars indicate 5th and 95th percentiles; barplot indicating mean permutation importance for all taxa (F). Permutation importance of variables with different letters are significantly different ($p \leq 0.05$) based on a Tukey HSD test.

which it is lower (0.83), and *Phaeocystis* for which very accurate predictions are possible (0.99).

Useful univariate models are found for between 72% and 95% of taxa, depending on the environmental factor (*see* values in brackets in Table 3). The best model performances are found for univariate models for T and MLD, with an average AUC of 0.73 in both cases (Table 3). The lowest AUC scores are obtained for *P** only models (0.66), which show lowest performance scores for all PFTs except diazotrophs.

Variable performance

The relative permutation importance measure, that is, the ranking of the predictive performance of the different environmental variables, reveals MLD is the most important variable for predicting the presence of the different phytoplankton taxa, significantly superior to all other variables (Tukey honest significant difference [HSD] test, Fig. 2, Panel F). The climate variables temperature and MLPAR are of similar permutation importance, forming a group of second best predictors ahead of the group containing NO_3^-, $Si(OH)_4$, and salinity. Similar to the performance of the corresponding univariate models (Table 3), the permutation importance of *P** is low.

When the permutation importance is averaged over PFT subsets, the patterns are more variable. MLD has the highest importance for coccolithophores, diatoms, diazotrophs, and picophytoplankton (approximately 30%). The second most important variables are less consistent among the investigated PFTs. In the case of coccolithophores and diazotrophs it is NO_3^- (16% and 22%), while for diatoms it is MLPAR (21%), and for picophytoplankton salinity (18%). A completely different pattern is found for *Phaeocystis*, for which temperature plays by far the most important role with a permutation importance of 84% followed by NO_3^- with 9%.

Univariate niches of phytoplankton

Most taxa investigated here have a niche center at temperatures between 10°C and 20°C (Fig. 3, Panel A). The midpoint of this range roughly corresponds to the mean temperature in the open ocean. The two *Phaeocystis* species considered have the lowest temperature niche centers (1°C on average), significantly lower than for any other PFT ($p \leq 0.01$; Table 4). Diatom niche centers occur across a wide range of temperatures, between 1°C and 26°C, but are significantly lower than those of coccolithophores and diazotrophs ($p \leq 0.001$ and $p \leq 0.01$, respectively), which have average niche centers of 21°C and 24°C, respectively. Picophytoplankton occupy broad temperature niches with an average niche center of 17°C.

The majority of the MLD niche centers are located at rather shallow mixed layers of less than 100 m, corresponding to the prevailing MLDs in the open ocean (Fig. 3, Panel B). Diazotrophs and coccolithophores occur at the shallowest mixed layers, on average, with average MLD niche centers of 39 m and 50 m, respectively. Mean MLD niche centers of both diatoms and *Phaeocystis* are significantly deeper than those of coccolithophores (Table 4) but vary between 23 m and 250 m. Approximately half of the taxa considered for those PFTs, as well as the two picophytoplankton taxa, have MLD niche centers between 100 m and 250 m depth, conditions which occur infrequently in the open ocean (Fig. 3, Panel B).

In terms of NO_3^-, most taxa have niche centers at elevated levels of 5 μmoles L^{-1} to 15 μmoles L^{-1}, conditions which are quite rare in the open ocean (Fig. 3, Panel C). The nitrate niches of diazotrophs and coccolithophores are closest to the oligotrophic conditions that prevail in the open ocean, with average niche centers of 2 μmoles L^{-1} and 5 μmoles L^{-1}, respectively. The niche centers of these two PFTs are distinct from those of diatoms ($p \leq 0.1$ and $p \leq 0.001$, respectively)

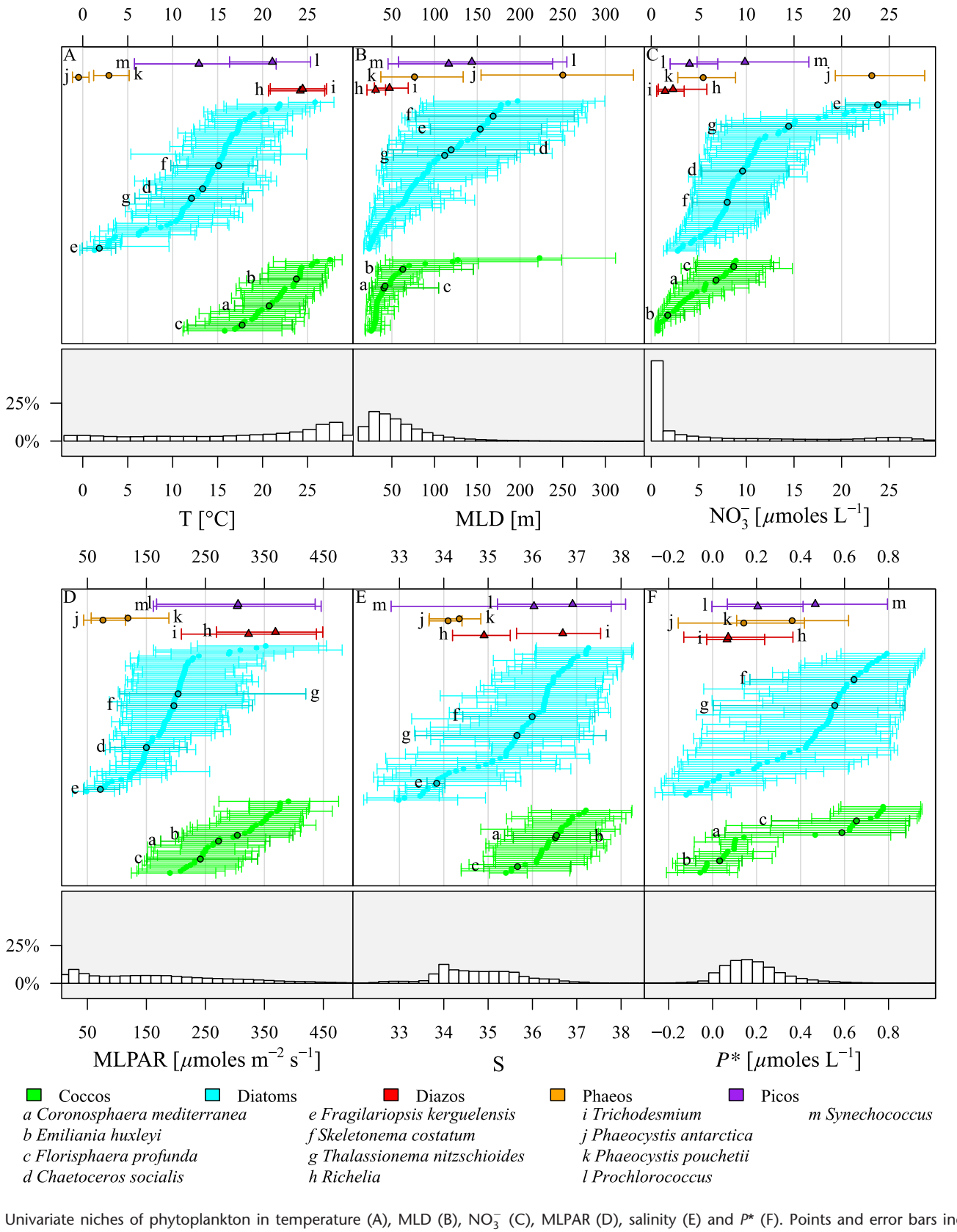

Fig. 3. Univariate niches of phytoplankton in temperature (A), MLD (B), NO_3^- (C), MLPAR (D), salinity (E) and *P** (F). Points and error bars indicate univariate niche centers (median) and breadths (inter quartile range), respectively. Points labeled by letters represent niches of well-known taxa. Only univariate niches of models that perform significantly better than random are illustrated, hence not all labeled taxa are shown in each panel. The histograms in the bottom plots indicate the frequency distribution of values of the environmental factor in the open ocean. Corresponding frequency distributions are estimated from the background points in the fitting procedure of MaxEnt models.

Table 4. Differences in univariate niche centers for PFTs and between S- and R-strategists. Depicted are significance levels of *p*-values for statistical tests comparing the niche centers in the environmental factors MLD, temperature, MLPAR, NO_3^-, salinity and *P**. No significance is symbolized with ns, "." means significance at a level of $p \leq 0.1$, * means significance at a level of $p \leq 0.05$, ** means significance at a level of $p \leq 0.01$ and *** means significance at a level of $p \leq 0.001$. Tukey HSD tests were used to compare PFTs while *t*-tests were used for the comparison of S-strategists with R-strategists.

	μ_{MLD}	μ_T	μ_{MLPAR}	μ_{NO3}	μ_S	$\mu_{Si(OH)4}$	μ_P*
Diatoms - Coccos	***	***	***	***	**	***	ns
Diazos - Coccos	ns	ns	ns	ns	ns	ns	ns
Phaeos - Coccos	**	***	***	*	*	**	ns
Picos - Coccos	ns	ns	ns	ns	ns	ns	ns
Diazos - Diatoms	ns	**	**	.	ns	ns	ns
Phaeos - Diatoms	ns	**	ns	ns	ns	ns	ns
Picos - Diatoms	ns	ns	*	ns	ns	ns	ns
Phaeos - Diazos	.	***	***	*	ns	ns	ns
Picos - Diazos	ns	ns	ns	ns	ns	ns	ns
Picos - Phaeos	ns	**	**	ns	ns	ns	ns
S – R	**	***	***	***	**	**	ns

and *Phaeocystis* ($p \leq 0.05$; Table 4) which have average niche centers of 11 μmoles L^{-1} and 14 μmoles L^{-1}, respectively. The latter PFTs, however, show high within-PFT variability in niche centers (ranges are 3-24 μmoles L^{-1} and 5-23 μmoles L^{-1}, respectively). Picophytoplankton occur at intermediate conditions in the statistical model (average niche center = 7 μmoles L^{-1}).

The niche centers in MLPAR occur predominantly between 150 μmoles m^{-2} s^{-1} and 300 μmoles m^{-2} s^{-1}, which is slightly biased toward high values compared to the average conditions in the open ocean. MLPAR niches of PFTs can be separated into two significantly distinct groups (Table 4). The niches of *Phaeocystis* and diatoms occur at low irradiance intensities (average niche centers at 98 μmoles m^{-2} s^{-1} and 186 μmoles m^{-2} s^{-1}, respectively) while coccolithophores, diazotrophs, and picophytoplankton have niche centers at comparatively high irradiance intensities (averages are 294 μmoles m^{-2} s^{-1}, 346 μmoles m^{-2} s^{-1}, and 305 μmoles m^{-2} s^{-1}, respectively).

Interestingly, most taxa have niche centers at rather high salinities relative to the frequency of open ocean conditions, with niche centers between 35 and 37 (Fig. 3, Panel E). *Phaeocystis* have niche centers at comparably low salinities (average is 34.2) whereas the realized niches of diatoms cover the whole range of possible salinities in the open ocean. Coccolithophore niche centers occur at significantly higher salinities, with an average of 36.2 ($p \leq 0.01$; Table 4).

More than 50% of the open ocean area has a $Si(OH)_4$ concentration below 5 μmoles L^{-1} but niche centers for most taxa are at elevated levels of 5 μmoles L^{-1} to 15 μmoles L^{-1}. Coccolithophore niche centers are found at significantly lower $Si(OH)_4$ concentrations than those of diatoms and *Phaeocystis* ($p \leq 0.01$; Table 4).

Finally, the diagnosed niche centers for *P** are biased toward high values of 0.3−0.7 μmoles L^{-1}, representing rare conditions that indicate clear surpluses of PO_4^{3-} compared to NO_3^- concentration with respect to the Redfield ratio (values are larger than zero; Fig. 3, Panel F). No significant differences are apparent between PFTs for *P** niche centers (Table 4; $p > 0.05$).

Realized niches in most environmental variables tend to be broader when the niche center is located at moderate conditions, whereas niches at the high or low end of the range of possible conditions tend to be narrower (Fig. 4). For instance, the temperature niches of *Phaeocystis* and diazotrophs are both rather narrow, while being located at the lower and upper end of the range of possible temperatures in the open ocean, respectively (Fig. 4, Panel A). Conversely, the temperature niches of picophytoplankton are broad and centered at moderate temperatures. MLD niches, however, do not become narrower if centers are located in deep mixed layers (Fig. 4, Panel B). This indicates that among the investigated taxa no "deep mixed layer specialists" exist. Instead, taxa occurring in deep mixed layers are rather generalists with a high tolerance of changing MLD. Differences in niche breadths also exist between PFTs: integrated over all investigated niche dimensions, diatoms and picophytoplankton tend be generalists with broad niches, whereas coccolithophores, diazotrophs, and *Phaeocystis* have narrower niches typical of specialist taxa.

Nitrate-light niches

Of the 133 phytoplankton taxa investigated in this study, 73 have previously been classified by Reynolds (2006) either as S-, SS-, or R-strategists based on morphological and physiological traits. This classification also includes an assumption about the ecological niches of taxa in terms of nutrient availability and light level. Here, we use our observation-based, univariate niche estimates to test these predicted

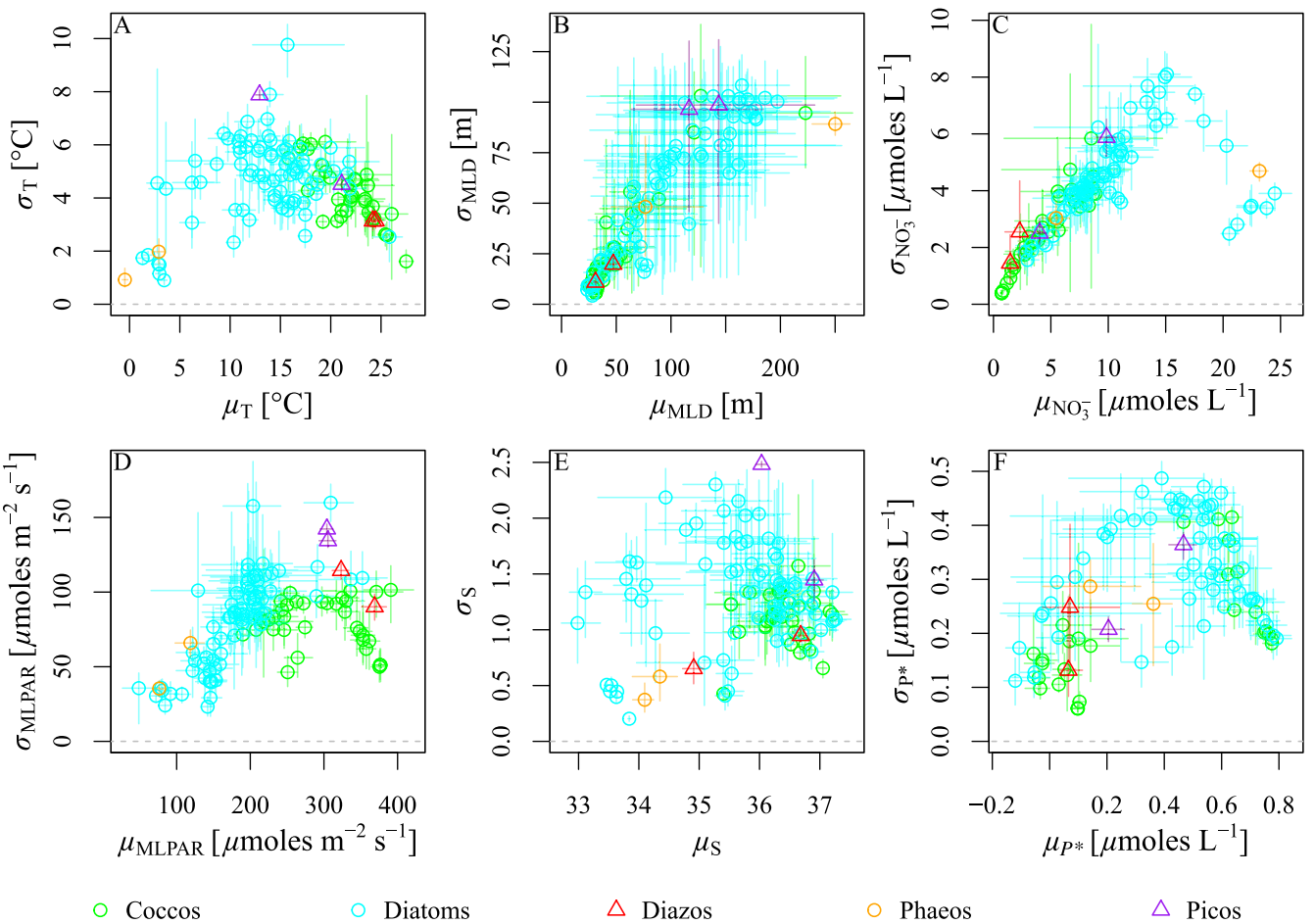

Fig. 4. centers (μs) and breadths (σs) of univariate niches for temperature (A), MLD (B), NO_3^- (C), MLPAR (D), S (E), and *P** (F). Horizontal and vertical lines associated with points indicate 90% confidence intervals based on bootstrap resampling. Univariate models which did not perform significantly better than random (AUC = 0.5) are not shown.

ecological niches. We use NO_3^- concentration to simulate general nutrient availability, as this factor is strongly correlated with other important nutrients such as PO_4^{3-} and $Si(OH)_4$ (*see* section "Selection of investigated environmental variables" above). Light level is represented by MLPAR (Fig. 5).

A quick comparison of Fig. 5, Panel A with the idealized scheme in the introduction (Fig. 1) indicates that our data do not include taxa with niches located in the upper left corner of the scheme, where conditions are simultaneously rich in nutrients and light availability. Such conditions are mainly expected to occur in coastal areas and are absent in the environmental data used to describe open ocean conditions in this study (frequency of considered open ocean conditions is illustrated by shading of background in Fig. 5, Panel A). Reynolds' scheme considers the area in the lower right corner, which is characterized by both poor nutrient and irradiance availability, to be unfavorable for phytoplankton growth ("void" area in Fig. 1). In contrast, we identify niches for some taxa (diatoms and *Phaeocystis*) under relatively poor light and nutrient availability (Fig. 5, Panel A). The very lower right corner in Fig. 5 Panel A, however, is not occupied by phytoplankton niches, and the relatively dark colored background there indicates that such conditions are frequent in the open ocean. These conditions are mostly located in temperate latitudes during winter months, and do not sustain phytoplankton growth due to low light and nutrient availability.

In agreement with Reynolds' predictions, we generally find taxa categorized as S-strategists to occur at lower NO_3^- concentrations and higher MLPAR intensities than those categorized as R-strategists ($p \leq 0.001$; Fig. 5, Panel B; Table 4). In addition, our results indicate that the S-strategists are significantly different from R-strategists, with the S-strategists having niches at shallower mixed layers, at higher temperatures, salinities and silicate concentration (Table 4).

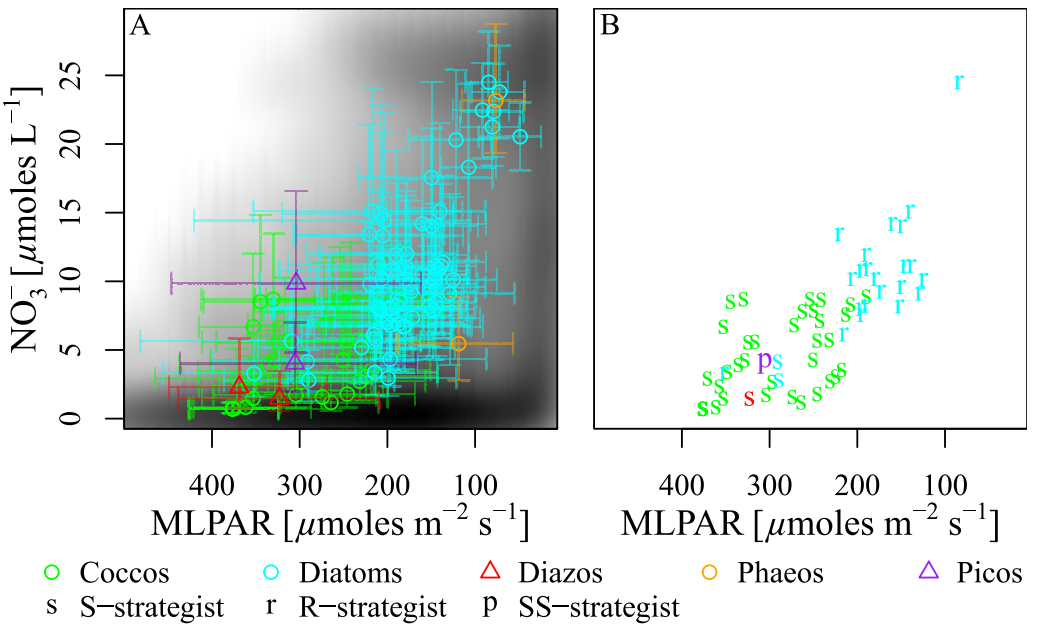

Fig. 5. (A) Niche centers (median) for NO_3^- and MLPAR for each taxon for which meaningful niches were found in both factors. In the background a kernel density estimate of open ocean conditions is shown: black indicates very frequent conditions in the open ocean whereas white colored combinations of NO_3^- concentration and MLPAR do not occur. Error bars indicate niche breadths (interquartile range). Colors represent different PFTs; green: coccolithophores; blue: diatoms; red: diazotrophs; orange: *Phaeocystis*; purple: picophytoplankton. (B) Niche centers labeled with predicted niche position according to Reynolds (2006). The same color code as in Panel A is used to represent PFTs in Panel B. The axes of MLPAR are reversed to be consistent with the axis in Reynolds scheme shown in Fig. 1.

However, there are also taxa for which the predictions do not correspond to our findings. The diatom *Chaetoceros dadayi*, for instance, is classified as R-strategist but we find it to have an S-strategist niche with a niche center of 3 μmoles L^{-1} in nitrate and 352 μmoles m^{-2} s^{-1} in MLPAR. Furthermore, Reynolds predicted the picophytoplankton *Prochlorococcus* to be a specialist for chronic nutrient stress (SS-strategist, Reynolds 2006). Our results suggest these organisms to have a relatively broad niche (breadth ranges from 167 μmoles m^{-2} s^{-1} to 436 μmoles m^{-2} s^{-1} for MLPAR and from 2 μmoles L^{-1} to 7 μmoles L^{-1} for NO_3^- concentration), and we cannot distinguish it from the regular S-strategists characterized in this study (purple "p" in Fig. 5, Panel B and corresponding purple cross in Fig. 5, Panel A).

Discussion

Importance of environmental variables

We find MLD to be the environmental variable with the highest permutation importance. This is in agreement with previous findings for diatoms and dinoflagellates in the North Atlantic (Irwin et al. 2012) and indicates that MLD may also be a key driver for the distributions of phytoplankton taxa in the global open ocean. The high importance of MLD for phytoplankton occurrence is established in various other contexts such as seasonal succession on the local scale (e.g., Margalef 1978) or productivity on the global scale (e.g., Boyce et al. 2010). MLD regimes and associated conditions appear to impose a major pressure for phytoplankton to which they have to adapt. The importance of MLD may stem from it being an ideal proxy for the joint influence of numerous important processes driving phytoplankton growth and distribution patterns: among these are grazing pressure, entrainment of nutrient-rich deep water, light availability, temperature, and short-term environmental variability (Evans and Parslow 1985; Behrenfeld 2010; Irwin et al. 2012).

The climatic variables temperature and MLPAR were of second-highest importance, with a permutation importance of 17-18%. Temperature and irradiance intensity are key variables that have been shown to control the distributions of phytoplankton in many ocean areas (reviewed in Boyd et al. 2010). The physiological effects of these factors are well studied for numerous phytoplankton groups and described by temperature and light sensitivities and optima (Litchman and Klausmeier 2008; Thomas et al. 2012). Furthermore, climatic variables tend to vary along large-scale gradients and on the global scale such variables have been shown to be the best predictors of species distributions in terrestrial systems (Elith and Leathwick 2009). The combination of the direct physiological effect on phytoplankton and the large-scale patterns of variation are likely to be responsible for the high

predictive performance of temperature and MLPAR in this study.

The solute variables NO_3^-, $Si(OH)_4$, and salinity have been found to be significantly less important than most of the other tested variables (permutation importance of 11%, 11%, and 8%, respectively). The relatively low importance of NO_3^- is surprising, since it is perhaps the single most important variable controlling phytoplankton growth in the ocean, either directly (Boyd et al. 2010) or indirectly through the promotion or inhibition of competitors as in the case of diazotrophs (Weber and Deutsch 2010).

One explanation for this conundrum is that in contrast to the other controlling variables, phytoplankton also strongly influence the NO_3^- and, in the case of diatoms, $Si(OH)_4$ concentration in the mixed layer through rapid consumption of these resources. Resupply of nutrients happens through vertical mixing processes at small scales and mesoscales which bring up nutrient-rich deeper waters (Falkowski et al. 1991; McGillicuddy et al. 2007; Lévy 2008). The interplay between mixing and consumption largely determines the conditions experienced by phytoplankton and leads to habitat variations at small scales (d'Ovidio et al. 2010), which cannot be resolved by the available climatological nutrient data. The identified nitrate and silicate niches, therefore, mainly reflect coarse-scale nutrient regimes which are correlated with temperature, in particular in the case of NO_3^- ($r = -0.86$). The case is similar for the predictive quality of PO_4^{3-} climatologies (not shown sensitivity analysis of this study; Irwin et al. 2012). Despite the unarguably major impact of nutrient availability on phytoplankton growth, nutrient variables that are averaged over hundreds of kilometers and months are of limited predictive quality for phytoplankton distribution, as the interaction between phytoplankton and nutrients is likely to happen on spatiotemporal scales that cannot be resolved in our analysis. In essence, many phytoplankton taxa are present in relatively low-nutrient regimes and wait for a nutrient pulse to enhance their growth and to become abundant.

Salinity affects phytoplankton cells directly via differences in osmotic potential. In the salinity range of the open ocean (30-40) the direct effect of salinity on phytoplankton reproduction rate, however, is relatively small (Brand 1984). The moderate predictive quality of salinity is, therefore, likely to arise from indirect effects such as salinity-induced stratification (Beaugrand et al. 2013), correlated evaporation/precipitation regimes (Hosoda et al. 2009), or the influence of rivers or sea ice in the high latitudes.

The relative excess of PO_4^{3-} in comparison to NO_3^-, P^*, was the variable with the lowest permutation importance in this study (3% on average; Fig. 2). This is not surprising, as the ratio of nutrients matters only in areas where a colimitation of both nutrients exists for phytoplankton. For diazotrophs and coccolithophores, which primarily occur in low nutrient areas, the importance of P^* was highest, although still relatively low (8.5% and 7.1%, respectively). In addition to only being relevant in limited ocean areas, this factor is affected by the combined uncertainty of two nutrient variables, which further constrains its accuracy and makes P^* a relatively poor predictor variable.

In conclusion, large-scale patterns of phytoplankton distribution can be predicted well and are most clearly linked to climate regimes. The tight link between climatic conditions and phytoplankton distributions further emphasizes the susceptibility of phytoplankton biogeography to climate change (Poloczanska et al. 2013). Based on these findings, we see two major promising directions for further research involving SDM of phytoplankton.

First, the characterized realized ecological niches could be used as a quantitative basis to estimate how distributions of different phytoplankton taxa may change in the future. Phytoplankton have several properties that correspond the assumptions of correlative SDMs comparably well: phytoplankton are far less limited by dispersal ability than terrestrial organisms. Shifts of phytoplankton distribution occur with rates that are high enough to track observed changes in sea surface temperature (Poloczanska et al. 2013). Moreover, phytoplankton are immediately affected by the state of the environment since they almost exclusively rely on abiotic resources, their body temperature, and metabolic rates are directly determined by the surrounding temperature and they have limited behavioral capabilities to evade the conditions they are exposed to (Robinson et al. 2011; McManus and Woodson 2012). However, some phytoplankton groups also have properties that hamper the identification of one single realized ecological niche for all of its life forms. For example, the ubiquitous distribution of single *Phaeocystis* cells contrasts with the occurrence patterns of *Phaeocystis* colonies, which were investigated in this study (Vogt et al. 2012). Furthermore, our results indicate that it is not currently possible to characterize the whole complexity of phytoplankton ecological niches due to the limited resolution of in situ environmental variables, and the scarcity of biological observations (discussed in further detail below). Similarly, it is currently unclear to which degree phytoplankton is able to physiologically and genetically adapt to changing climatic conditions (Colwell and Rangel 2009). Thus, further investigations are needed to determine the reliability of SDM-based predictions of phytoplankton biogeography.

Second, SDMs could be used to improve the monitoring of phytoplankton biogeography. Other indirect variables similar to MLD, such as ambient chlorophyll *a* concentration, may be explored as potential proxies to predict and monitor phytoplankton biogeography on the global scale. Chlorophyll in combination with other environmental variables derived from remote sensing (sea surface temperature and salinity) could potentially lead to better resolved estimates of phytoplankton biogeography. In addition to monitoring, such estimates could be used for the evaluation of

satellite algorithms which estimate the distribution of phytoplankton groups based on optical properties of the sea surface (e.g., Brewin et al. 2011; Hirata et al. 2011).

Plankton functional types

For the PFTs with relatively large numbers of investigated taxa, diatoms, and coccolithophores (87 taxa and 40 taxa, respectively), we found the strongest distinction of realized ecological niches (Table 4). Compared to diatoms, coccolithophores occurred in significantly warmer and saltier areas with shallower mixed layers, a higher irradiance, and lower nitrate and silicate concentration ($p \leq 0.01$). These findings are similar to the conclusions of extensive review studies (Balch 2004; Boyd et al. 2010). However, for the remaining PFTs investigated in this study differences in univariate niche centers were not as clearly distinct. Our results indicated that the simulated niche centers allow us to distinguish between two major groupings of PFTs: coccolithophores, diazotrophs, and picophytoplankton occur at higher MLPAR and temperatures than diatoms and *Phaeocystis*. Furthermore, coccolithophores and diazotrophs occur at significantly lower NO_3^- concentrations than diatoms and *Phaeocystis* ($p \leq 0.05$, except for the marginally significant difference between diatoms and diazotrophs of $p \leq 0.1$) and coccolithophores inhabit significantly more saline areas than diatoms and *Phaeocystis* ($p \leq 0.05$; Table 4). The niche centers within these groupings could not be statistically separated except for the significantly lower MLPAR niche centers of *Phaeocystis* in comparison to those of diatoms ($p \leq 0.05$). A part of the lack in statistically significant separation of the PFTs contained within the two groupings could result from the smaller number of taxa included for the diazotrophs, picophytoplankton and *Phaeocystis*: each of these PFTs was only represented by two taxa. Thus, for those groups absolute differences in univariate niche centers had to be larger to achieve statistical significance in a Tukey HSD test.

To which degree a similar analysis based on a larger number of taxa would lead to more distinct differences between the PFTs, however, is unclear. Whereas we expect additional diazotroph taxa to occur under habitat conditions similar to those of the taxa investigated here (Luo et al. 2012), we believe that the preferred habitat conditions of the *Phaeocystis* PFT are more diverse than our results suggest: the third major bloom forming *Phaeocystis* species, *Phaeocystis globosa*, was not included in this study due to scarcity of observational data, but is known to also occur in tropical areas (Schoemann et al. 2005).

In conclusion, we have shown that there are systematic differences in phytoplankton realized niches between PFTs, but at the same time we found a broad range of preferred habitat conditions within many PFTs. These results may act as cornerstones for the validation of marine ecosystem models and remote sensing approaches which currently show considerable uncertainties about the distribution and dominance patterns of different PFTs in the global ocean (e.g., Brewin et al. 2011; M. Vogt pers. comm.). Yet, comparisons of estimated niches and distributions have to be made with caution: marine ecosystem models are designed to capture the majority of the biomass and hence productivity of a PFT. Based on our approach, we are not able to make any inference as to how abundant the PFTs are in the different parts of their occupied niche space. However, although we only included two taxa for several PFTs, the most prevalent taxa of all PFTs are represented and, therefore, our results can be seen as an indication that some PFTs may be more easily parameterized using large scale environmental variables in global marine ecosystem models, as they group species with similar biogeographic and niche characteristics (e.g., diazotrophs, colonial *Phaeocystis*) while other groups consist of species with more diverse environmental preferences, and may be more challenging to implement as a single PFT in coarse-scale global marine biogeochemistry and climate models (e.g., diatoms).

Niches in Reynolds' space

We used our approximations of realized niche centers and breadths to test whether the predictions Reynolds' C-S-R model are meaningful when upscaled to the global ocean. The expected niches of R-strategists and S-strategists could be reproduced, but the simulated niches were much broader and more overlapping than suggested by the Reynolds model (Smayda and Reynolds 2001). In agreement with the theory, niche centers of S-strategists prevailed in nutrient-poor conditions, whereas the niche centers of the R-strategists were predominantly located in areas with low light intensity. However, our data also showed that phytoplankton taxa are often not restricted to a narrow window of suitable conditions in light and nutrient concentrations but can be found across a large fraction of the available open ocean conditions of the two parameters. *Prochlorococcus*—the only representative of organisms tolerant to chronic nutrient scarcity (SS-strategists; Reynolds 2006)—, for instance, was found to occur in nutrient scarce conditions with high light intensity, in agreement with its demonstrated biomass dominance in nutrient-deficient areas such as the subtropical gyres (Hirata et al. 2011). But its simulated niche also encompasses elevated nutrient concentrations and lower light intensity, in agreement with the findings of Johnson et al. (2006) who found *Prochlorococcus* at marginal light levels at depths of 150 m or more, far below the local mixed layer which formed the vertical boundary in this study. In contrast to the assumption in Reynolds' model (Smayda and Reynolds 2001), the shapes of our simulated niches, therefore, indicate that on the global scale phytoplankton occurrence is not only determined by the accessibility of major nutrient concentrations and light intensity. Other factors including temperature, micronutrients such as iron, carbon dioxide, or grazing pressure may play an important role as well (Boyd et al. 2010).

The results of the variable importance analysis strengthen this case. On the global scale our proxies for nutrient accessibility and light availability (NO_3^- and MLPAR) did not prove to be the major niche dimensions for phytoplankton. For 16% of the preclassified taxa, no preferences could be found in at least one of the factors NO_3^- and MLPAR based on the analyzed distributional patterns (no meaningful MaxEnt model could be fitted). Furthermore, averaged over all taxa, we found the latter two variables to be of intermediate permutation importance compared to the other investigated environmental factors, especially MLD. Nevertheless, the univariate model performance of both factors is relatively high (0.71 and 0.7, respectively) and MLD, according to our results the most important variable on the global scale, is related to nutrient supply rates and irradiance intensity. Therefore, across all open ocean areas and on monthly time-scales, we confirm the significance of nutrient accessibility and light availability, but we cannot capture their predicted overarching importance (Smayda and Reynolds 2001) by the NO_3^- and MLPAR climatologies used in this study. However, better resolved light and in particular nutrient data may lead to different conclusions about the relative importance of these two factors for phytoplankton ecological niches.

Reynolds' C-S-R classification comprises key traits for phytoplankton survival, but not in all cases do the assumed impacts and importance of the traits match the views of the more recent literature. According to Reynolds (2006), the main traits that separate R-strategists from S-strategists are their higher growth rate, lower nutrient affinity and more elongated body shapes. Furthermore, R-strategists tend to be smaller than S-strategists. Growth rate and nutrient affinity are broadly recognized as key traits for the competitive success of phytoplankton (Litchman and Klausmeier 2008; Edwards et al. 2012*a*) and are used to parameterize phytoplankton in many ecosystem models (e.g., Buitenhuis et al. 2010). Organism size is recognized to strongly influence the competitive success of phytoplankton under different environmental conditions (Litchman and Klausmeier 2008), but many studies indicate that nutrient affinity decreases with body size (e.g., Edwards et al. 2012*b*) suggesting that nutrient specialists should be small, as opposed to the predicted large size of S-strategists by Reynolds. The reason for this seeming contradiction is that many S-strategists occur in colonies, for example *Trichodesmium*, with rather large aggregate sizes (Reynolds 2006). The role of body shape, and in particular elongation, as a trait determining light affinity has been demonstrated in multiple studies (Reynolds 2006; Naselli-Flores et al. 2007). However, the effect is somewhat distorted due to the high plasticity of the body shape of many phytoplankton taxa (Litchman and Klausmeier 2008).

Finally, the link between phytoplankton traits and realized ecological niches, as defined by Reynolds, is of descriptive nature and assigns broad ranges of possible trait values to the different strategists without the definition of distinct thresholds. It is thus not possible to assign a clear strategy to phytoplankton with certain trait combinations. For instance slow growing phytoplankton cells with a volume of 10^3-10^4 μm^3 could be either S-strategists or R-strategists unless they are perfectly round or have strongly elongated body shapes (Reynolds 2006).

In summary, we find that the predicted niches of S- and R-strategists in the Reynolds model are supported by global-scale observations. Yet, the relatively broad and overlapping niches and the moderate importance of our proxies for nutrients and light suggest that phytoplankton realized niches, as many aspects in ecology, may be influenced by several factors and can only partly be explained by a simple model. Future extensions may link traits in a more extensive and mechanistic way to realized ecological niches, to produce more accurate models of the relationship between phytoplankton traits and realized ecological niches.

Strengths and weaknesses of data and method

In this study, we investigated the realized ecological niches of marine phytoplankton on a global scale, based on an extensive set of field data. We statistically modeled the niches of 133 taxa and consistently achieved high performance scores and realistic niche distributions. However, there are also a number of limitations associated with this study, which will be discussed below.

The available phytoplankton observations were biased toward high abundances as a result of more frequent sampling in high biomass regions and seasons, while there were many areas that were severely undersampled (Buitenhuis et al. 2013). We attempted to reduce the impact of an overrepresentation of frequently sampled areas on our results by binning the data to monthly $1^\circ \times 1^\circ$ cells, but there are many regions without any observations (e.g., South Pacific). A sensitivity analysis showed that a random reduction of sample size down to our limit of 15 presence cells does not greatly reduce AUC, but leaving out observations from whole ocean basins can cause distinct drops in model performance. Furthermore, estimates of niche centers may be biased and niche breadths may be estimated too narrow. However, currently we are not able to assess the degree of such limitations as we lack independent validation datasets with high spatial coverage.

One possibility to increase the size of the dataset would be to include richly available observations from coastal areas. Such observations were left out in this study to avoid biases in ecological niche estimates due to river input, atmospheric deposition of nutrients, nutrient resupply due to sediment interactions, or anthropogenic sources of pollution. Further studies are necessary to assess the optimal trade-off between

more representative niche simulations due to the addition of further observations and biases due to coastal effects.

Furthermore, the taxonomic resolution of the data is biased toward certain groups and in some cases only genus level classifications were available. Diatoms represent 65% of the investigated taxa and coccolithophores 30%. This does not represent the diversity of phytoplankton in the open ocean, and even of the latter two PFTs we considered only a minor fraction of the known diversity (Falkowski et al. 2004). The classification of diazotrophs and picophytoplankton to the genus level is coarse and could lead to realized niche estimates of limited accuracy, for instance in the case of *Prochlorococcus* with its numerous distinct ecotypes (Johnson et al. 2006). Additionally, the investigated genus *Richelia* represents heterocystous diazotrophs (Luo et al. 2012). Heterocystous diazotrophs live in association with diatom species and their ecological niches may, thus, be influenced by these tight interactions. Our results show, however, that the niches of the two investigated diazotroph taxa are very similar in all niche dimensions except for salinity, indicating that no obvious expansion or contraction of *Richelia* niches results from this association. This observation suggests that the diazotrophs profit in some other way from the association, for example, through enhanced growth rate (Foster et al. 2011).

More generally, phytoplankton taxonomy is a constantly changing field, which makes working with historical datasets challenging. Phytoplankton classifications are predominantly based on morphological characteristics. With the increasing use of molecular methods, cryptic speciation (i.e., the evolution of genetically distinct species with very similar morphology) has been found to occur within many phytoplankton taxa (e.g., Smayda 2011; Degerlund et al. 2012). Thus, species previously considered to be cosmopolitan were found to comprise multiple genetically distinct cryptic species, each occupying a narrower ecological niche (e.g., Degerlund et al. 2012). An updated and more extensive set of phytoplankton observations will become available with the next release of MAREDAT, hopefully improving some of these issues. For an optimal performance of SDMs, a revised version of MAREDAT furthermore should include both presence–absence and biomass information, and aim to resolve all included groups to the species level.

Additionally, the environmental variables used to model the realized ecological niches do not include all critical factors. We characterize the realized ecological niches of the phytoplankton only using bottom-up factors and do not explicitly include biotic interactions. Grazing rates or grazer abundance, for instance, may act as a strong control on phytoplankton communities (Litchman and Klausmeier 2008; Prowe et al. 2012). Grazing, however, mainly controls phytoplankton biomass but by itself it is unlikely to lead to the complete exclusion of a phytoplankton taxon from a certain area. The impact of grazing on the shape of the realized niches may, thus, be limited. Furthermore, by including MLD we account for part of the effects of grazing indirectly (Behrenfeld 2010). Further abiotic factors such as iron concentration have previously been shown to be important for phytoplankton niches, in particular in high-nitrate, low-chlorophyll (HNLC) regions (Boyd et al. 2010). This could not be further investigated in this study due to the lack of existing climatologies.

Last but not least, the spatial and temporal resolution of the environmental data we used is relatively coarse in comparison to nutrient consumption rates (Litchman and Klausmeier 2008) and to the patchiness of phytoplankton distribution (d'Ovidio et al. 2010). Match-ups of phytoplankton observations and environmental data will thus deviate from the conditions that the organisms actually experienced at the time of sampling. Our findings are, therefore, limited to the effects of broad gradients in environmental conditions of phytoplankton distribution. However, the high model performance scores indicate that our explanatory variables explain a large fraction of the spatial patterns of the investigated observations. This is in agreement with the notion of Robinson et al. (2011) that climate variables and coarse-scale SDMs are suitable for widely distributed species in environments which show comparatively small fine-scale variations, such as the pelagic realm.

Despite these limitations, our approach represents an important missing link between observation-based studies (e.g., Irwin et al. 2012), which are typically of limited spatial extent, and global-scale studies using ecosystem models and remote sensing (e.g., Buitenhuis et al. 2010, Hirata et al. 2011) which have a coarse resolution of phytoplankton diversity. This allowed us to test how well the realized niches of phytoplankton taxa can be summarized using idealized categories such as PFTs and Reynolds' C-S-R model. We characterized the realized niches of five PFTs based on rich quantitative data, in particular for the diatoms and coccolithophores. These results may help to improve the parameterization of PFTs in marine ecosystem models. Moreover, we tested open ocean predictions of Reynolds' C-S-R model for numerous taxa and found that taxa described as S- and R-strategists indeed occupy different light and nutrient niches. Yet, the variability of the characterized niches and the moderate permutation importance of our proxies for the environmental variables in Reynolds' model suggest that the relationship between phytoplankton traits and ecological niches may be more complex than assumed by this model. Finally, we found temperature, light and particularly MLD to be more important environmental drivers of the distribution of the investigated phytoplankton taxa than large-scale patterns in nutrient concentrations.

References

Amante, C., and B. W. Eakins. 2009. ETOPO1 1 arc-minute global relief model: Procedures, data sources and analysis, NOAA Technical Memorandum NESDIS NGDC-24. National Geophysical Data Center.

Anderson, T. R., W. C. Gentleman, and B. Sinha. 2010. Influence of grazing formulations on the emergent properties of a complex ecosystem model in a global ocean general circulation model. Prog. Oceanogr. **87**: 201-213. doi:10.1016/j.pocean.2010.06.003.

Antonov, J. I., and others. 2010. World Ocean Atlas 2009, Volume **2**: Salinity, p. 184. In S. Levitus [ed.], NOAA Atlas NESDIS 69. U.S. Government Printing Office.

Aumont, O., and L. Bopp. 2006. Globalizing results from ocean in situ iron fertilization studies. Global Biogeochem. Cycles **20**: GB2017. doi:10.1029/2005GB002591.

Balch, W. M. 2004. Coccolithophores: From molecular processes to global impact, p. 165-190. In H. R. Thierstein and J. R. Young [eds.], Coccolithophores: From molecular processes to global impact. Springer.

Beaugrand, G., D. Mackas, and E. Goberville. 2013. Applying the concept of the ecological niche and a macroecological approach to understand how climate influences zooplankton: Advantages, assumptions, limitations and requirements. Prog. Oceanogr. **111**: 75–90. doi:10.1016/j.pocean.2012.11.002.

Behrenfeld, M. J. 2010. Abandoning sverdrup's critical depth hypothesis on phytoplankton blooms. Ecology **91**: 977–989. doi:10.1890/09-1207.1.

Boyce, D. G., M. R. Lewis, and B. Worm. 2010. Global phytoplankton decline over the past century. Nature **466**: 591–596. doi:10.1038/nature09268.

Boyd, P., and S. Doney. 2003. The impact of climate change and feedback processes on the ocean carbon cycle, p. 157–193. In M.R. Fasham [ed.], Ocean biogeochemistry SE - 8. Springer Berlin Heidelberg.

Boyd, P. W., R. Strzepek, F. Fu, and D. A. Hutchins. 2010. Environmental control of open-ocean phytoplankton groups: Now and in the future. Limnol. Oceanogr. **55**: 1353–1376. doi:10.4319/lo.2010.55.3.1353.

Brand, L. E. 1984. The salinity tolerance of forty-six marine phytoplankton isolates. Estuar. Coast. Shelf Sci. **18**: 543–556. doi:10.1016/0272-7714(84)90089-1.

Brewin, R. J. W., and others. 2011. An intercomparison of bio-optical techniques for detecting dominant phytoplankton size class from satellite remote sensing. Remote Sens. Environ. **115**: 325–339. doi:10.1016/j.rse.2010.09.004.

Buitenhuis, E. T., and others. 2012. Picophytoplankton biomass distribution in the global ocean. Earth Syst. Sci. Data **4**: 37–46. doi:10.5194/essd-4-37-2012.

Buitenhuis, E. T., R. B. Rivkin, S. Sailley, and C. Le Quéré. 2010. Biogeochemical fluxes through microzooplankton. Global Biogeochem. Cycles **24**: GB4015. doi:10.1029/2009GB003601.

Buitenhuis, E. T., and others. 2013. MAREDAT: Towards a world atlas of MARine Ecosystem DATa. Earth Syst. Sci. Data **5**: 227–239. doi:10.5194/essd-5-227-2013.

Colwell, R. K., and T. F. Rangel. 2009. Hutchinson's duality: The once and future niche. Proc. Natl. Acad. Sci. USA (Suppl.) **106**: 19651–19658. doi:10.1073/pnas.0901650106.

d'Ovidio, F., S. De Monte, S. Alvain, Y. Dandonneau, and M. Lévy. 2010. Fluid dynamical niches of phytoplankton types. Proc. Natl. Acad. Sci. USA **107**: 18366–18370. doi:10.1073/pnas.1004620107.

de Boyer Montégut, C. 2004. Mixed layer depth over the global ocean: An examination of profile data and a profile-based climatology. J. Geophys. Res. **109**: C12003. doi:10.1029/2004JC002378.

Degerlund, M., S. Huseby, A. Zingone, D. Sarno, and B. Landfald. 2012. Functional diversity in cryptic species of Chaetoceros socialis Lauder (Bacillariophyceae). J. Plankton Res. **34**: 416–431. doi:10.1093/plankt/fbs004.

Deutsch, C., J. L. Sarmiento, D. M. Sigman, N. Gruber, and J. P. Dunne. 2007. Spatial coupling of nitrogen inputs and losses in the ocean. Nature **445**: 163–167. doi:10.1038/nature05392.

Dolédec, S., D. Chessel, C. Gimaret-Carpentier, V. Cedex, and L. De Biome. 2000. Niche separation in community analysis: A new method. Ecology **81**: 2914–2927. doi:10.1890/0012-9658(2000)081[2914:NSICAA]2.0.CO;2.

Dormann, C. F., and others. 2012. Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. Ecography **36**: 27–46. doi:10.1111/j.1600-0587.2012.07348.x.

Edwards, K. F., E. Litchman, and C. A. Klausmeier. 2012a. Functional traits explain phytoplankton community structure and seasonal dynamics in a marine ecosystem. Ecol. Lett. **16**: 56–63. doi:10.1111/ele.12012.

Edwards, K. F., M. K. Thomas, C. A. Klausmeier, and E. Litchman. 2012b. Allometric scaling and taxonomic variation in nutrient utilization traits and maximum growth rate of phytoplankton. Limnol. Oceanogr. **57**: 554–566. doi:10.4319/lo.2012.57.2.0554.

Elith, J., and others. 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography **29**: 129–151. doi:10.1111/j.2006.0906-7590.04596.x.

Elith, J., and J. R. Leathwick. 2009. Species distribution models: Ecological explanation and prediction across space and time. Annu. Rev. Ecol. Evol. Syst. **40**: 677–697. doi:10.1146/annurev.ecolsys.110308.120159.

Evans, G. T., and J. S. Parslow. 1985. A model of annual plankton cycles. Biol. Oceanogr. **3**: 327–347. doi:10.1080/01965581.1985.10749478.

Falkowski, P. G. 1994. The role of phytoplankton photosynthesis in global biogeochemical cycles. Photosynth. Res. **39**: 235–258. doi:10.1007/BF00014586.

Falkowski, P. G., M. E. Katz, A. H. Knoll, A. Quigg, J. A. Raven, O. Schofield, and F. J. R. Taylor. 2004. The evolution of modern eukaryotic phytoplankton. Science **305**: 354–360. doi:10.1126/science.1095964.

Falkowski, P. G., D. Ziemann, Z. Kolber, and P. K. Bienfang. 1991. Role of eddy pumping in enhancing primary production in the ocean. Nature **352**: 55–58. doi:10.1038/352055a0.

Field, C., M. Behrenfeld, J. Randerson, and P. Falkowski. 1998. Primary production of the biosphere: Integrating terrestrial and oceanic components. Science **281**: 237–240.doi:10.1126/science.281.5374.237.

Foster, R. A., M. M. M. Kuypers, T. Vagner, R. W. Paerl, N. Musat, and J. P. Zehr. 2011. Nitrogen fixation and transfer in open ocean diatom-cyanobacterial symbioses. ISME J. **5**: 1484–1493. doi:10.1038/ismej.2011.26.

Garcia, H. E., R. A. Locarnini, T. P. Boyer, J. I. Antonov, M. M. Zweng, O. K. Baranova, and D. R. Johnson. 2010. World Ocean Atlas 2009, Volume **4**: Nutrients (phosphate, nitrate, silicate), p. 398. In S. Levitus [ed.], NOAA Atlas NESDIS 71. U.S. Government Printing Office.

Gruber, N., and J. L. Sarmiento. 1997. Global patterns of marine nitrogen fixation and denitrification. Global Biogeochem. Cycles **11**: 235–266. doi:10.1029/97GB00077.

Grüner, N., C. Gebühr, M. Boersma, U. Feudel, K. H. Wiltshire, and J. A. Freund. 2011. Reconstructing the realized niche of phytoplankton species from environmental data: Fitness versus abundance approach. Limnol. Oceanogr. Methods **9**: 432–442. doi:10.4319/lom.2011.9.432.

Guisan, A., and N. E. Zimmermann. 2000. Predictive habitat distribution models in ecology. Ecol. Modell. **135**: 147–186. doi:10.1016/S0304-3800(00)00354-9.

Hernandez, P. A., C. H. Graham, L. L. Master, and D. L. Albert. 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. Ecography **29**: 773–785. doi:10.1111/j.0906-7590.2006.04700.x.

Hirata, T., and others. 2011. Synoptic relationships between surface Chlorophyll-a and diagnostic pigments specific to phytoplankton functional types. Biogeosciences **8**: 311–327. doi:10.5194/bg-8-311-2011.

Hood, R. R., and others. 2006. Pelagic functional group modeling: Progress, challenges and prospects. Deep Sea Res. Part II Top. Stud. Oceanogr. **53**: 459–512. doi:10.1016/j.dsr2.2006.01.025.

Hosoda, S., T. Suga, N. Shikama, and K. Mizuno. 2009. Global surface layer salinity change detected by Argo and its implication for hydrological cycle intensification. J. Oceanogr. **65**: 579–586. doi:10.1007/s10872-009-0049-1.

Hurrell, J. W., and others. 2013. The community earth system model: A framework for collaborative research. Bull. Am. Meteorol. Soc. **94**: 1339–1360. doi:10.1175/BAMS-D-12-00121.1.

Hutchinson, G. E. 1957. Concluding remarks. Cold Spring Harb. Symp. Quant. Biol. **22**: 415–427.doi:10.1101/SQB.1957.022.01.039.

Iglesias-Rodríguez, M. D., C. W. Brown, S. C. Doney, J. Kleypas, D. Kolber, Z. Kolber, P. K. Hayes, and P. G. Falkowski. 2002. Representing key phytoplankton functional groups in ocean carbon cycle models: Coccolithophorids. Global Biogeochem. Cycles **16**: 47-1–47-20. doi:10.1029/2001GB001454.

Irwin, A. J., A. M. Nelles, and Z. V. Finkel. 2012. Phytoplankton niches estimated from field data. Limnol. Oceanogr. **57**: 787–797. doi:10.4319/lo.2012.57.3.0787.

Johnson, Z. I., E. R. Zinser, A. Coe, N. P. McNulty, E. M. S. Woodward, and S. W. Chisholm. 2006. Niche partitioning among Prochlorococcus ecotypes along ocean-scale environmental gradients. Science **311**: 1737–1740. doi:10.1126/science.1118052.

Le Quéré, C., and others. 2005. Ecosystem dynamics based on plankton functional types for global ocean biogeochemistry models. Glob. Chang. Biol. **11**: 2016–2040. doi:10.1111/j.1365-2486.2005.1004.x.

Leblanc, K., and others. 2012. A global diatom database—abundance, biovolume and biomass in the world ocean. Earth Syst. Sci. Data **4**: 149–165. doi:10.5194/essd-4-149-2012.

Lévy, M. 2008. The modulation of biological production by oceanic mesoscale turbulence, p. 219–261. In J. Weiss and A. Provenzale [eds.], Transport and mixing in geophysical flows SE - 9. Springer Berlin Heidelberg.

Litchman, E., and C. A. Klausmeier. 2008. Trait-based community ecology of phytoplankton. Annu. Rev. Ecol. Evol. Syst. **39**: 615–639. doi:10.1146/annurev.ecolsys.39.110707.173549.

Locarnini, R. A., and others. 2010. World Ocean Atlas 2009, Volume **1**: Temperature, p. 184. In S. Levitus [ed.], NOAA Atlas NESDIS 68. U.S. Government Printing Office.

Luo, Y.-W., and others. 2012. Database of diazotrophs in global ocean: Abundance, biomass and nitrogen fixation rates. Earth Syst. Sci. Data **4**: 47–73. doi:10.5194/essd-4-47-2012.

Luo, Y.-W., I. D. Lima, D. M. Karl, C. A. Deutsch, and S. C. Doney. 2014. Data-based assessment of environmental controls on global marine nitrogen fixation. Biogeosciences **11**: 691–708. doi:10.5194/bg-11-691-2014.

Manel, S., H. C. Williams, and S. J. Ormerod. 2001. Evaluating presence-absence models in ecology: The need to account for prevalence. J. Appl. Ecol. **38**: 921–931. doi:10.1046/j.1365-2664.2001.00647.x.

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

McGillicuddy, D. J., and others. 2007. Eddy/wind interactions stimulate extraordinary mid-ocean plankton blooms. Science **316**: 1021–1026. doi:10.1126/science.1136256.

McManus, M. A., and C. B. Woodson. 2012. Plankton distribution and ocean dispersal. J. Exp. Biol. **215**: 1008–1016. doi:10.1242/jeb.059014.

Merow, C., M. J. Smith, and J. A. Silander. 2013. A practical guide to MaxEnt for modeling species' distributions: What it does, and why inputs and settings matter. Ecography **36**: 1058–1069. doi:10.1111/j.1600-0587.2013.07872.x.

Moore, J. K., S. C. Doney, J. A. Kleypas, D. M. Glover, and I. Y. Fung. 2002. An intermediate complexity marine ecosystem model for the global domain. Deep Sea Res. Part II Top. Stud. Oceanogr. **49**: 403–462. doi:10.1016/S0967-0645(01)00108-4.

Morel, A. 1988. Optical modeling of the upper ocean in relation to its biogenous matter content (case I waters). J. Geophys. Res. Ocean. **93**: 10749–10768. doi:10.1029/JC093iC09p10749.

Naselli-Flores, L., J. Padisák, and M. Albay. 2007. Shape and size in phytoplankton ecology: Do they matter? Hydrobiologia **578**: 157–161. doi:10.1007/s10750-006-2815-z.

O'Brien, C. J., and others. 2013. Global marine plankton functional type biomass distributions: Coccolithophores. Earth Syst. Sci. Data **5**: 259–276. doi:10.5194/essd-5-259-2013.

Palacz, A. P., M. A. St. John, R. J. W. Brewin, T. Hirata, and W. W. Gregg. 2013. Distribution of phytoplankton functional types in high-nitrate, low-chlorophyll waters in a new diagnostic ecological indicator model. Biogeosciences **10**: 7553–7574. doi:10.5194/bg-10-7553-2013.

Phillips, S. J., R. P. Anderson, and R. E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. Ecol. Modell. **190**: 231–259. doi:10.1016/j.ecolmodel.2005.03.026.

Phillips, S. J., P. Avenue, F. Park, M. Dudík, and R. E. Schapire. 2004. A maximum entropy approach to species distribution modeling. Proceedings of the 21st International Conference on Machine Learning. ACM Press. 83.

Poloczanska, E. S., and others. 2013. Global imprint of climate change on marine life. Nat. Clim. Change **3**: 919–925. doi:10.1038/nclimate1958.

Prowe, A. E. F., M. Pahlow, S. Dutkiewicz, M. Follows, and A. Oschlies. 2012. Top-down control of marine phytoplankton diversity in a global ecosystem model. Prog. Oceanogr. **101**: 1–13. doi:10.1016/j.pocean.2011.11.016.

Reynolds, C. S. 1988. Functional morphology and the adaptive strategies of freshwater phytoplankton, p. 338–433. In C. D. Sandgren [ed.], Growth and reproductive strategies of freshwater phytoplankton. Cambridge Univ. Press.

Reynolds, C. S. 2006. The ecology of phytoplankton. Cambridge Univ. Press.

Robinson, L. M., J. Elith, A. J. Hobday, R. G. Pearson, B. E. Kendall, H. P. Possingham, and A. J. Richardson. 2011. Pushing the limits in marine species distribution modelling: Lessons from the land present challenges and opportunities. Glob. Ecol. Biogeogr. **20**: 789–802. doi:10.1111/j.1466-8238.2010.00636.x.

Sarmiento, J. L., and N. Gruber. 2006. Ocean biogeochemical dynamics. Princeton Univ. Press.

Schoemann, V., S. Becquevort, J. Stefels, V. Rousseau, and C. Lancelot. 2005. Phaeocystis blooms in the global ocean and their controlling mechanisms: A review. J. Sea Res. **53**: 43–66. doi:10.1016/j.seares.2004.01.008.

Smayda, T. J. 2011. Cryptic planktonic diatom challenges phytoplankton ecologists. Proc. Natl. Acad. Sci. USA **108**: 4269–4270. doi:10.1073/pnas.1100997108.

Smayda, T. J., and C. S. Reynolds. 2001. Community assembly in marine phytoplankton: Application of recent models to harmful dinoflagellate blooms. J. Plankton Res. **23**: 447–461. doi:10.1093/plankt/23.5.447.

Swets, J. A. 1988. Measuring the accuracy of diagnostic systems. Science **240**: 1285–1293. doi:10.1126/science.3287615.

Thomas, M. K., C. T. Kremer, C. A. Klausmeier, and E. Litchman. 2012. A global pattern of thermal adaptation in marine phytoplankton. Science **338**: 1085–1088. doi:10.1126/science.1224836.

Vogt, M., and others. 2012. Global marine plankton functional type biomass distributions: Phaeocystis spp. Earth Syst. Sci. Data **4**: 107–120. doi:10.5194/essd-4-107-2012.

Weber, T. S., and C. Deutsch. 2010. Ocean nutrient ratios governed by plankton biogeography. Nature **467**: 550–554. doi:10.1038/nature09403.

Acknowledgments

We thank L. Bopp, C. Laufkötter and the MAREMIP CMIP5 project for the modelled iron data. Furthermore, we thank Dr. Niklaus Zimmermann for his support in technical aspects of species distribution modeling. P.B., M.V., and N.G. acknowledge funding from ETH Zurich. C.O.B and M.P. have received funding from the European Community's Seventh Framework Programme (FP7 2007-2013) under grant agreement no. 238366 (Greencycles II) and 308299 (NACLIM), respectively. P.B. and M.P. acknowledge the Villum foundation that funds the Center for Ocean Life. Y.-W.L. is supported by the Ministry of Education of China 985 Phase III (south China marine science center). We thank all original data originators for their contribution to the MAREDAT global plankton atlas (http://lgmacweb.env.uea.ac.uk/green_ocean/data/index.shtml?d1#biomass), as this work would not have been possible without their help.

Submitted 14 August 2014
Revised 2 February 2015
Accepted 26 January 2015

Associate editor: Heidi M. Sosik