



UvA-DARE (Digital Academic Repository)

Large-scale variation in phytoplankton community composition of >1000 lakes across the USA

Verspagen, J.M.H.; Ji, X.; Liu, Q.-X.; Huisman, J.

DOI

[10.1088/2752-664X/ac788c](https://doi.org/10.1088/2752-664X/ac788c)

Publication date

2022

Document Version

Final published version

Published in

Environmental Research: Ecology

License

CC BY

[Link to publication](#)

Citation for published version (APA):

Verspagen, J. M. H., Ji, X., Liu, Q.-X., & Huisman, J. (2022). Large-scale variation in phytoplankton community composition of >1000 lakes across the USA. *Environmental Research: Ecology*, 1(1), Article 015001 . <https://doi.org/10.1088/2752-664X/ac788c>

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (<https://dare.uva.nl>)

PAPER • OPEN ACCESS

Large-scale variation in phytoplankton community composition of >1000 lakes across the USA

To cite this article: Jolanda M H Verspagen *et al* 2022 *Environ. Res.: Ecology* 1 015001

View the [article online](#) for updates and enhancements.

You may also like

- [Phosphorus-only fertilization rapidly initiates large nitrogen-fixing cyanobacteria blooms in two oligotrophic lakes](#)
Lewis A Molot, Scott N Higgins, Sherry L Schiff *et al.*
- [Analyzing the algal bloom risk and its relationship with environmental variables in urban landscape water](#)
Kun Luo, Conglin Wu, HeZhen Zheng *et al.*
- [Canonical correspondence analysis of relationship between characteristics of phytoplankton community and environmental factors in Wolong Lake](#)
Shanshan Ma, Xingchen Lv, Shicong Geng *et al.*

ENVIRONMENTAL RESEARCH ECOLOGY



PAPER

Large-scale variation in phytoplankton community composition of >1000 lakes across the USA

OPEN ACCESS

RECEIVED

11 March 2022

REVISED

29 May 2022

ACCEPTED FOR PUBLICATION

14 June 2022

PUBLISHED

30 August 2022

Jolanda M H Verspagen^{1,3,*} , Xing Ji^{1,2,3} , Quan-Xing Liu²  and Jef Huisman¹ 

¹ Department of Freshwater and Marine Ecology, Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, Amsterdam, The Netherlands

² Research Center for Global Change and Complex Ecosystems, School of Ecological and Environmental Science, East China Normal University, Shanghai, People's Republic of China

³ These authors contributed equally to this work.

* Author to whom any correspondence should be addressed.

E-mail: j.m.h.verspagen@uva.nl

Keywords: biodiversity, eutrophication, temperature, cyanobacteria, diatoms, chlorophytes, dinoflagellates

Supplementary material for this article is available [online](#)

Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](#).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.



Abstract

Although environmental impacts on the biodiversity and species composition of lakes have been studied in great detail at local and regional scales, unraveling the big picture of how lake communities respond to environmental variation across large spatial scales has received less attention. We performed a comprehensive analysis to assess how the phytoplankton community composition varies among >1000 lakes across the conterminous United States of America. Our results show that lake-to-lake similarity in species composition was low even at the local scale, and slightly decreased with geographical distance. Analysis of the compositional data by Dirichlet regression revealed that the geographical variation in phytoplankton community composition was best explained by total phosphorus (TP), water temperature, pH, and lake size. High TP concentrations were associated with high relative abundances of cyanobacteria and euglenophytes at the expense of other phytoplankton groups. High lake temperatures stimulated cyanobacteria, dinoflagellates, desmids and euglenophytes, whereas cryptophytes, golden algae and diatoms were relatively more abundant in colder lakes. Low lake pH correlated with high dissolved CO₂ concentrations, which may explain why it benefitted phytoplankton groups with inefficient carbon concentrating mechanisms such as golden algae and euglenophytes. Conversely, the relative abundance of cyanobacteria showed a pronounced increase with lake pH. Large lakes showed higher relative abundances of cyanobacteria and diatoms, whereas small lakes showed higher relative abundances of chlorophytes, desmids and euglenophytes. Biodiversity increased with lake temperature, but decreased at high TP concentrations and pH. The key environmental variables identified by our study (high phosphorus loads, warm temperature, low pH) are associated with anthropogenic pressures such as eutrophication, global warming and rising atmospheric CO₂ concentration. Hence, our results provide a comprehensive illustration of the major impact of these anthropogenic pressures on the biodiversity and taxonomic composition of lake phytoplankton communities.

1. Introduction

One of the major challenges in the aquatic sciences is to understand and predict how the species composition of lakes will respond to environmental changes induced by eutrophication, global warming and other anthropogenic pressures. A range of methods are available to address this challenge, including the development of mathematical models (Elliott 2010, Brauer *et al* 2012, Janssen *et al* 2019), analysis of the

genetic and physiological traits of species using sophisticated molecular tools and laboratory experiments (Steffen *et al* 2014, Sandrini *et al* 2016), dedicated lake experiments (Schindler 1974, Huisman *et al* 2004), and extensive monitoring studies (Nöges *et al* 2008, Wagner and Adrian 2009, Pollard *et al* 2018, Zhang *et al* 2018). Each of these approaches has its own advantages and disadvantages, and integration of multiple approaches will be needed to obtain the full picture.

For example, laboratory studies have greatly contributed to our understanding of the physiological and ecological traits of different phytoplankton taxa, such as their temperature responses (Dauta *et al* 1990, Butterwick *et al* 2005, Lüring *et al* 2013, Paerl and Otten 2013), and nutritional requirements (Rhee and Gotham 1980, Tilman *et al* 1981, Sommer 1985, Spijkerman *et al* 2005, Passarge *et al* 2006, Ji *et al* 2017). However, laboratory studies typically investigate only a limited number of species grown under controlled conditions. It is less clear to what extent results obtained from laboratory isolates can be extrapolated to complex natural plankton communities with a high diversity of interacting species exposed to a wide range of natural and human-induced environmental pressures (Burford *et al* 2020).

An approach that can be very fruitful in this respect is the analysis of large-scale lake surveys (Mantzouki *et al* 2018, Pollard *et al* 2018). In these large-scale lake surveys, sampling, sample analysis and data processing follow the same standards. This approach has several interesting characteristics. (a) As species responses can deviate strongly between lakes in different settings, surveys that cover a large geographical range with many different types of lakes are well suited to identify important environmental drivers. (b) Large-scale surveys allow space-for-time substitution, where instead of a long-term temporal study in a single lake, many lakes are studied across large spatial scales (Pickett 1989). For instance, lakes in warmer climates may provide a proxy for the impact of global warming on lakes in colder climates. (c) Large-scale lake surveys provide an important benchmark for the applicability of lab findings and the predictive capacity of model studies.

Prominent examples of large-scale lake surveys include the European REBECCA dataset (Ptacnik *et al* 2008a), the European Multi Lake Survey (Mantzouki *et al* 2018), a survey of Iowa lakes in the U.S.A. (Filstrup *et al* 2016), as well as other lake surveys in South America (Kosten *et al* 2012), Canada (Vogt *et al* 2017, Hansson *et al* 2019) and China (Tong *et al* 2017, Tao *et al* 2020). The most extensive survey is probably the National Lakes Assessment (NLA) of the United States Environmental Protection Agency (US EPA), conducted in the summers of 2007, 2012 and 2017 across the conterminous USA. In each of these years, biological, chemical and physical measurements and samples were taken in approximately 1000 lakes, and the datasets have been used for a wide range of limnological studies (Pollard *et al* 2018).

Together, these large-scale lake surveys have provided new insights and broadened our knowledge on aquatic ecosystems. For example, these studies have identified nutrients and temperature as the most important environmental drivers of phytoplankton community composition (Watson *et al* 1997, Ptacnik *et al* 2008a, Filstrup *et al* 2016, Sodr e *et al* 2020). High nutrient loads generally increase phytoplankton biomass (Vollenweider *et al* 1974) and favour cyanobacterial dominance (Watson *et al* 1997, Ptacnik *et al* 2008a, Beaulieu *et al* 2013, Taranu *et al* 2015, Filstrup *et al* 2016, Beaver *et al* 2018, Ho and Michalak 2020, Sodr e *et al* 2020). These large-scale findings align with numerous local and regional lake studies in which eutrophication promoted cyanobacterial abundance (Schindler 1974, Paerl *et al* 2011), often at the expense of chrysophytes (Ptacnik *et al* 2008a) and desmids (Coesel 1982). Such shifts in phytoplankton community composition with increasing nutrient loads may also impact phytoplankton biodiversity. Specifically, phytoplankton biodiversity tends to increase with increasing nutrient load, but may level off or even decline at very high nutrient loads (Jeppesen *et al* 2000, Stomp *et al* 2011).

Temperature impacts the thermal stratification of lakes, which suppresses vertical mixing. Thermal stratification is therefore expected to favor buoyant species including several bloom-forming cyanobacteria (Huisman *et al* 2004, Elliott 2010). Conversely, large diatom species may be at a disadvantage in stratified lakes, as they sink to the sediments due to their relatively heavy silica frustules (Huisman *et al* 2002, Winder *et al* 2009, R uhland *et al* 2015). Moreover, although there is considerable interspecific variation, cyanobacteria tend to have higher temperature optima than diatoms (Dauta *et al* 1990, Butterwick *et al* 2005, Paerl and Otten 2013). Consequently, many lake studies have reported increasing abundances of cyanobacteria (J ohnk *et al* 2008, Wagner and Adrian 2009, Kosten *et al* 2012, Beaulieu *et al* 2013) and decreasing abundances of large diatoms at higher temperatures (Winder *et al* 2009, R uhland *et al* 2015, Sodr e *et al* 2020).

Many of the large-scale lake surveys have focused on specific taxonomic groups, particularly on cyanobacteria (Kosten *et al* 2012, Beaulieu *et al* 2013, Rigosi *et al* 2014, Taranu *et al* 2015, Beaver *et al* 2018, Ho and Michalak 2020). The interest in cyanobacteria is understandable, because they can produce toxic blooms which have negative impacts on water quality, and hence are highly relevant in lake management (Chorus and Bartram 1999, Meriluoto *et al* 2017, Huisman *et al* 2018). However, in addition to cyanobacteria, other freshwater phytoplankton taxa, such as chrysophytes (Hiltunen *et al* 2012) and euglenophytes (Zimba *et al* 2017), are also capable of toxin production. Moreover, other phytoplankton

groups may play key roles in biogeochemical cycles, such as silica drawdown by diatoms (Smol and Stoermer 2010), or may have unexpected impacts on food web structure such as mixotrophic taxa capable of both photoautotrophic and heterotrophic growth (Flynn *et al* 2013, Hansson *et al* 2019). It is therefore important to gain a better understanding of how entire phytoplankton communities respond to environmental change, including phytoplankton groups that have received less attention in previous large-scale studies.

In this study, we analyze how phytoplankton community composition and biodiversity varies with major environmental drivers across the conterminous USA. The data are obtained from the NLA of the US EPA in the summer of 2012, which provides a unique data set in which environmental variables and phytoplankton community composition of >1000 lakes were sampled according to standardized protocols (Pollard *et al* 2018). Our analysis aims at ‘the big picture’, and will undoubtedly overlook details of relevance for the distribution of specific taxa or the community composition of specific lakes. We first investigate correlation patterns among the major environmental variables measured in the lakes, and assess which of these environmental variables best explain the observed variation in community composition. Next, we investigate how each of the selected environmental variables impacts the relative abundance of all major phytoplankton groups, and their most common genera, and how these variables affect phytoplankton biodiversity and dominance patterns.

2. Methods

2.1. Data collection

All data used for this study were obtained from the NLA 2012 of the US EPA (2016). A total of 1235 lakes from the conterminous USA with a size greater than 1 ha were sampled once or twice during summer (May–September) (US EPA 2011a). In this study, we investigated one sample per lake and therefore used only the data from the first visit for those lakes that were visited twice. At each lake, an extensive suite of abiotic and biotic parameters was measured. We describe the methods only briefly below, and refer to the US EPA reports for more extensive methodological details of the sampling and analyses (US EPA 2011b, 2012).

Depth profiles of water temperature, pH, and dissolved oxygen were measured with a water quality probe or multi-parameter sonde. We used these data to calculate the average temperature, pH, and dissolved oxygen of the surface water between 0 and 2 m depth. Furthermore, we calculated the maximum buoyancy frequency (N^2) as a proxy of the stability of lake stratification (Denman and Gargett 1983). Specifically, buoyancy frequency was calculated from depth profiles of water density as a function of temperature and salinity (Chen and Millero 1986), where salinity was calculated from conductivity (Lewis 1980).

Furthermore, water samples were collected from the surface water, at the location where the lake reached its deepest point, using a 2 m integrated sampler. If the euphotic zone (estimated from Secchi disk measurements) was less than 2 m deep, water samples were taken only within the euphotic zone (US EPA 2011b). Dissolved organic carbon (DOC) was measured using UV-promoted persulfate oxidation to CO₂ with infrared detection. Dissolved silica (DSi) was measured using flow injection analysis. Total nitrogen (TN) and total phosphorus (TP) were measured using flow injection analysis after persulfate digestion. Acid neutralizing capacity (ANC) was determined using automated acidimetric titration to pH ≤ 3.5 with modified Gran plot analysis. Conductivity was measured at 25 °C using Man-Tech TitraSip automated analysis (Man-Tech, Guelph, Canada) (US EPA 2012). Turbidity was measured using Man-Tech TitraSip analysis (Man-Tech, Guelph, Canada) or manual analysis using a HACH turbidimeter for high turbidity samples. Chlorophyll-*a* (Chl *a*) content was extracted with 90% acetone and analyzed by fluorometry. For details on measurement precision and quality control, see the Laboratory Operations Manual (US EPA 2012).

Dissolved CO₂ and HCO₃[−] concentrations were not measured in the lakes. Therefore, we calculated dissolved CO₂ and HCO₃[−] concentrations from pH and ANC (Stumm and Morgan 2012), and then transformed dissolved CO₂ concentrations to partial pressure (pCO₂, in ppm) based on Henry’s law using equilibrium constants corrected for salinity and temperature (Weiss 1974, Millero *et al* 2006).

Phytoplankton samples were collected from the upper 2 m of the water column (or from the euphotic zone if the euphotic zone was less than 2 m deep) using an integrated sampler, at the same location where the other parameters were measured (US EPA 2011b). The samples were preserved with Lugol’s iodine. Phytoplankton was identified to the lowest taxonomic level (usually the species or genus level), and counted with an inverted microscope using Utermöhl sedimentation chambers. A total of 400 natural algal units per sample was counted in 8–100 randomly selected fields. Phytoplankton biovolume of each taxon was calculated from cellular dimensions and geometric shapes measured by microscopic examination (Hillebrand *et al* 1999). A detailed description of the procedures can be found in US EPA (2012). For each lake, we calculated ‘relative abundances’ of the different taxa, defined as the relative contributions of these taxa to the total phytoplankton biovolume. We focused our analysis on the relative abundances of the eight

most abundant phytoplankton groups, which together comprised >98% of the total phytoplankton biovolume when averaged over all lakes.

We used the following classification: cyanobacteria, diatoms, chlorophytes, desmids, dinoflagellates, cryptophytes, euglenophytes and golden algae (which include chrysophytes and synurophytes). We also analyzed relative abundances at the genus level, for those phytoplankton genera that appeared in more than 100 of the lakes selected for regression analysis. We used the genus names as they were given in the NLA dataset.

2.2. Biodiversity and similarity indices

We calculated taxon richness and the Shannon–Wiener diversity index to assess phytoplankton biodiversity of each lake. The Shannon–Wiener index (H) was calculated as:

$$H = - \sum_{i=1}^R (q_i \ln q_i) \quad (1)$$

where taxon richness R is the total number of taxa in a lake, and q_i is the proportion of individuals belonging to the i th taxon in this lake.

We also calculated the similarity in phytoplankton community composition between each pair of lakes using the Sørensen and Bray–Curtis similarity indices. For this purpose, we used all phytoplankton taxa identified to at least the genus level, which comprised 90% of the counted taxa, and ~94% of the counted biovolume. The Sørensen similarity index only considers presence/absence data, whereas the Bray–Curtis similarity index takes phytoplankton abundance into account. Both similarity indices (S) are defined as:

$$S = \frac{2C}{A + B} \quad (2)$$

For the Sørensen similarity index, A and B are the number of taxa in lake A and lake B , and C is the number of taxa shared by the two lakes. For the Bray–Curtis similarity index, A and B represent the sum of the relative abundances of the phytoplankton taxa in lake A and in lake B , and C is the sum of the lesser relative abundances for those taxa shared by the two lakes. The similarity indices were calculated using the function *vegdist* from the *vegan* package in R. For comparison with the similarity data, we also quantified the environmental variation among lakes by calculating the (absolute) difference between the values of environmental variables for each pair of lakes. Furthermore, we calculated the geographical distance between each pair of lakes using the function *distVincentyEllipsoid* from the *geosphere* package in R (Karney 2013).

To test whether the similarity in community composition and the environmental variation changed with increasing distance between lakes, we performed two-sided Jonckheere–Terpstra tests (a non-parametric trend test) from R package *clinfun*, using 3000 permutations for the reference distribution.

2.3. Statistical analyses

We first tested whether all explanatory variables and their log-transformed values followed a normal distribution using the Shapiro–Wilks test. If log-transformation improved normality and/or homoscedasticity, log₁₀-transformed data were used for further analysis. We note that buoyancy frequency values were transformed as $\log_{10}(x + 1)$, and ANC values were transformed as $\log_{10}(x + 1000)$. Correlations between environmental variables were investigated using Pearson’s product-moment correlation, with p values corrected for multiple hypothesis testing using the false discovery rate (FDR: Benjamini and Hochberg 1995).

For our further analysis, we standardized all explanatory variables using z -transformation to obtain standardized effect sizes (Feld *et al* 2016). We calculated the variance inflation factor (VIF) to quantify multicollinearity of the explanatory variables, and selected variables with a VIF <5 in a stepwise manner using the R package *usdm* (Naimi *et al* 2014). We then assessed the relation between the relative abundances of the phytoplankton groups and the selected explanatory variables using multiple regression analysis.

Relative abundances of phytoplankton taxa represent compositional data, consisting of continuous proportions or percentages that sum to a total of one or 100% (Aitchison 1982, Greenacre 2021). Compositional data may show spurious correlations between taxa because the total sum of relative abundances is constrained. Furthermore, compositional data cannot be analyzed by standard linear regression models, because the data are not normally distributed and the variance is not homogeneous across the range from zero to one. These issues are often addressed by transforming the relative abundances using for instance a logit or log-ratio transformation (e.g. the centered log-ratio (clr) transform), which changes the data from values bounded between zero and one to unconstrained values. Subsequently, the transformed data can be analyzed by standard statistical techniques, such as ordinary least-squares regression. However,

Table 1. Classification of lakes into four categories according to total phosphorus concentration, water temperature, lake pH and lake area. The number of lakes (*n*) in each category is indicated between brackets.

Environmental variable	Low		←→	High	Units
Total phosphorus concentration	<20 (234)	20–40 (259)	40–100 (281)	>100 (267)	μg l ⁻¹
Water temperature	<19 (167)	19–24 (352)	24–29 (398)	>29 (124)	°C
Lake pH	<7.0 (163)	7.0–8.0 (257)	8.0–9.0 (504)	>9.0 (117)	—
Lake area	<10 (232)	10–30 (278)	30–100 (247)	>100 (284)	ha

data transformations have some limitations that may introduce biases or hamper the interpretation. For example, a logit transformation will strongly increase the error for relative abundances close to zero (many phytoplankton taxa in our dataset) or one. Furthermore, the clr transform lacks subcompositional coherence. This means that changes in only a subset of a multispecies community (e.g. competitive replacement of species A by species B) will alter the clr transform of all species in the community, even if the relative abundances of the other species were not affected. Therefore, instead of a data transformation, we preferred a regression analysis that uses the original scale of measurement. We chose for Dirichlet regression, which is specifically tailored for compositional data consisting of multiple categories (multiple taxa) (Acevedo-Trejos *et al* 2013, Douma and Weedon 2019).

To account for frequent zero-valued observations, the relative abundances were first adjusted according to the following equation (Douma and Weedon 2019):

$$p_i^* = p_i \left(1 - \frac{1}{n}\right) + \frac{1}{Gn} \quad (3)$$

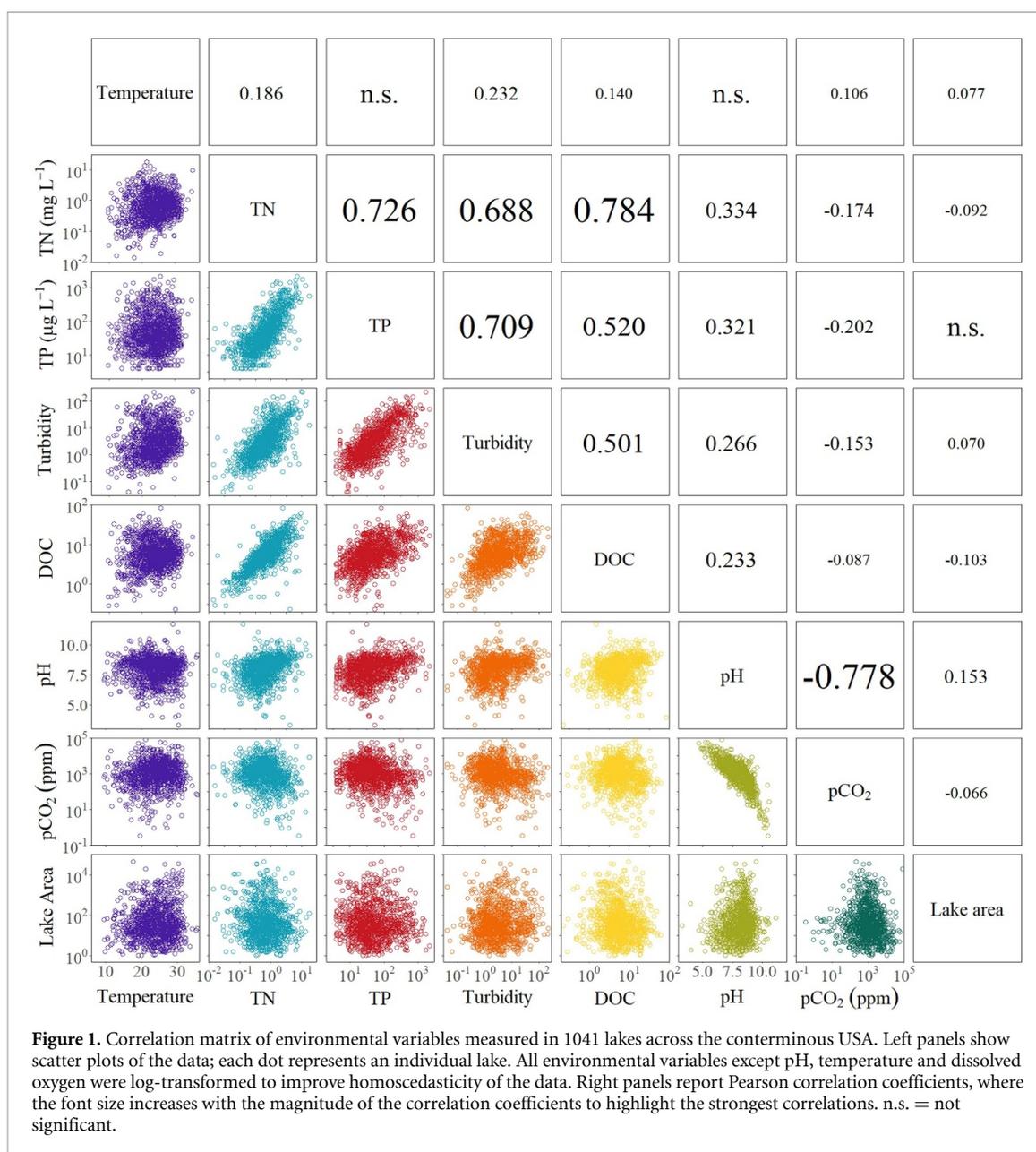
In our analysis, p_i is the relative abundance of phytoplankton group i , n is the total number of lakes used in the analysis ($n = 1041$), and G is the number of phytoplankton groups ($G = 8$). This adjustment implies that zero values are replaced by a very small value ($1/Gn$), while non-zero values are nearly identical to the original data because of the high number of lakes, and the total relative abundances still sum to one.

We performed a stepwise multiple Dirichlet regression including all phytoplankton groups, as function of the environmental variables using the R package *DirichletReg* (Maier 2014). We selected the environmental variables that contributed most to the explanatory power of the regression model based on the Bayesian information criterion (BIC: Schwarz 1978), where the smallest value indicates the best model fit. We chose BIC over other model selection criteria (e.g. Akaike information criterion), because BIC penalizes parameters more strongly and therefore tends to select the most parsimonious set of explanatory variables (Raftery 1995). The four environmental variables selected by the Dirichlet regression were log(TP), water temperature, pH and log(lake area).

Next, we investigated the relation between the relative abundances of the phytoplankton groups and these four environmental variables in further detail. For this purpose, the TP concentration, temperature, pH and area of the lakes were each classified into four different categories, where we ensured that each category comprised at least 100 lakes (table 1).

Relative abundances of the phytoplankton groups, the Shannon–Wiener diversity index and taxon richness were not normally distributed over the lakes, as confirmed by the Shapiro–Wilk test. We therefore used non-parametric Kruskal–Wallis one-way analysis of variance to assess whether relative abundances of phytoplankton groups, taxon richness and the Shannon–Wiener diversity index differed among lakes with different TP concentrations, water temperature, pH and lake area. The Kruskal–Wallis test was followed by Dunn’s post hoc pairwise comparisons, and the p values were corrected for multiple hypothesis testing using the FDR (Benjamini and Hochberg 1995).

Furthermore, we calculated the ‘probability of dominance’ of each phytoplankton group. The probability of dominance was defined as the percentage of lakes in which a given phytoplankton group comprised >75% of the total phytoplankton biovolume. We investigated whether the probability of dominance of the phytoplankton groups differed among lakes with different TP concentrations, temperature, pH and lake area using Chi-square tests. If Chi-square tests were significant, we used the two-proportions Z-test for post hoc pairwise comparisons of the probabilities of dominance. The p -values of the Chi-square tests and Z-tests were corrected for multiple hypothesis testing using the FDR (Benjamini and Hochberg 1995). All statistical analyses were performed in R (version 4.0.2; R Core Team 2020).

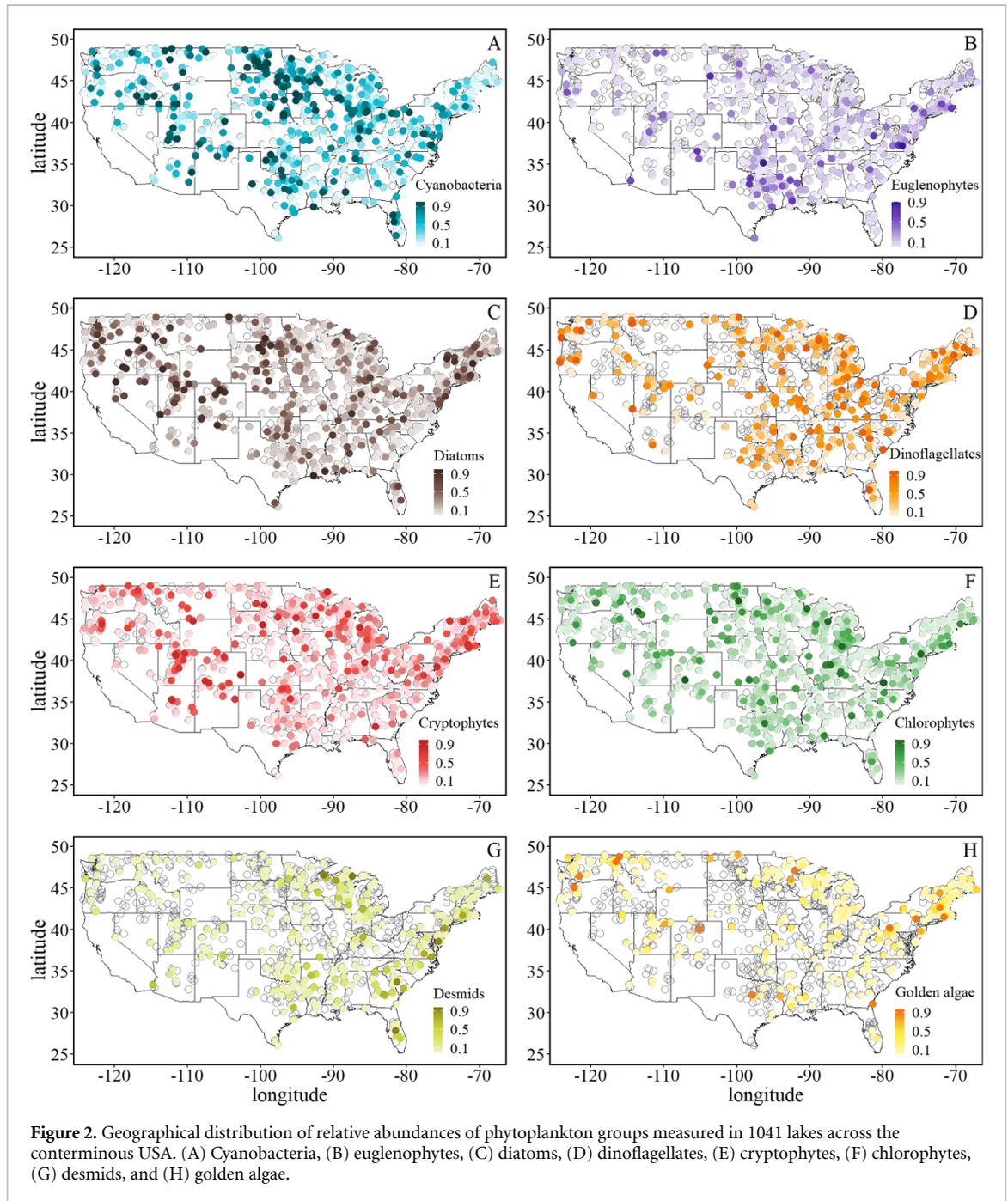


3. Results

3.1. Environmental variation among lakes

The lakes cover a wide range of environmental conditions, with summer water temperatures ranging from <10 to >35 °C, TN and TP concentrations spanning three orders of magnitude, and pH values ranging from 3 to 11 (figure 1). Several of the environmental variables measured in the lakes were strongly associated with each other (figure 1, table S1). TP, TN, turbidity, DOC and pH were all positively correlated, illustrating that lakes with high nutrient concentrations tend to have high levels of aquatic primary production, which increases turbidity, DOC production and pH. Lakes with high pH had low pCO₂ and high HCO₃⁻ concentrations, in agreement with the pH dependence of carbon speciation. Furthermore, pCO₂ showed weak negative correlations and pH showed weak positive correlations with TN and TP in this data set (figure 1). Some variables, such as water temperature and lake area, only showed very weak or insignificant correlations with most other variables (figure 1). Lake area correlated positively with lake depth, however, illustrating that large lakes tend to be deeper (Cael *et al* 2017).

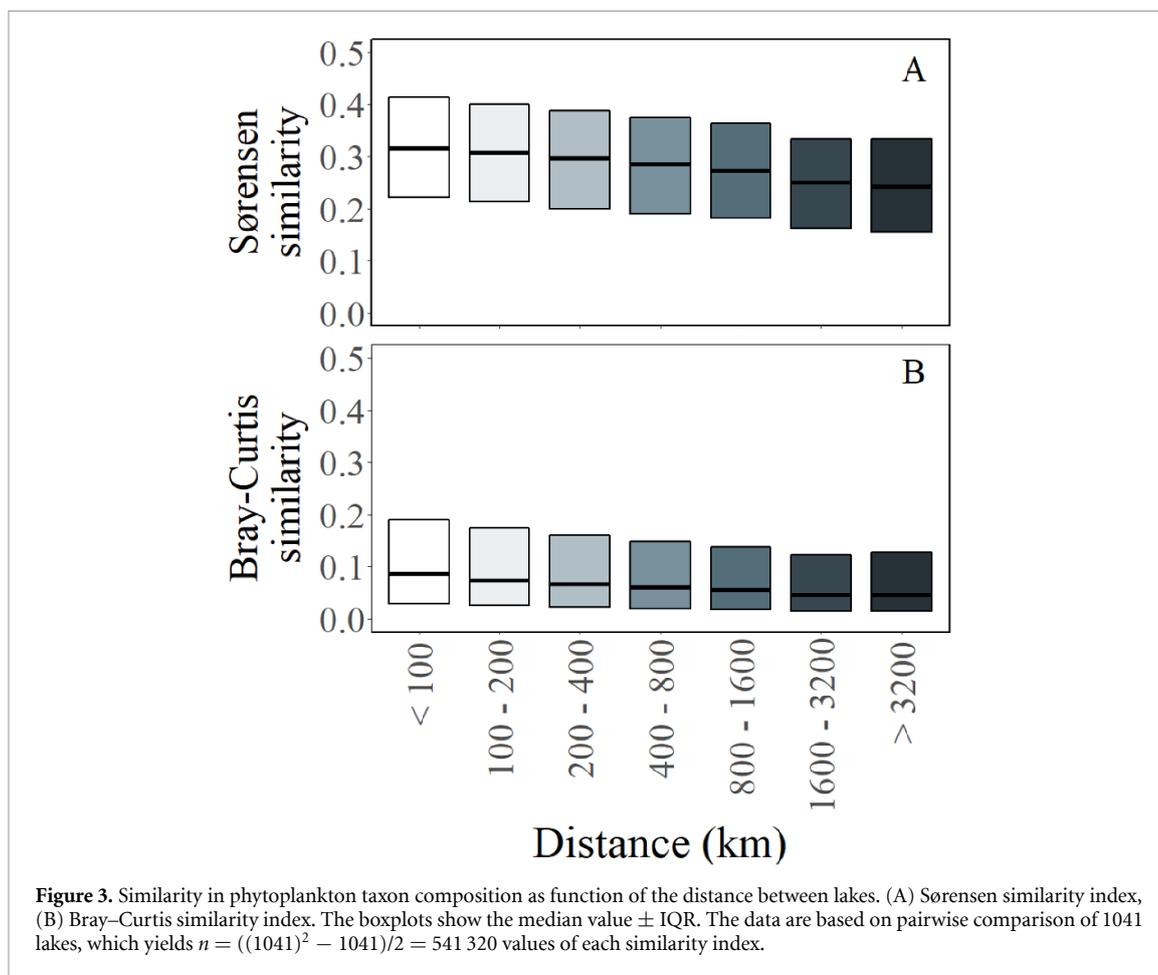
To remove the redundancy among lake variables, we calculated VIFs. This led to removal of the variables log(pCO₂) and log(HCO₃⁻) which had a high collinearity with pH and log(ANC), log(Secchi depth) which had a high collinearity with log(turbidity), and log(conductivity) which had a high collinearity with log(ANC). Although log(TN) and log(TP) also showed a high collinearity, we decided to maintain both



log(TN) and log(TP) in our analysis, because of the major effects of both N and P limitation on phytoplankton communities (Smith 1983, Elser *et al* 1990, Conley *et al* 2009). Hence, based on the VIFs, the explanatory variables selected for further analysis were dissolved oxygen, water temperature, pH, log(ANC), log(lake depth), log(lake area), log(DOC), log(TP), log(TN), log(DSi), log(turbidity), and log(N^2). After removing all missing observations from the dataset, a total of 1041 lakes remained that included all selected variables.

3.2. Geographical distribution of phytoplankton composition

All major phytoplankton groups occurred throughout the conterminous USA, yet some biogeographical patterns were apparent (figure 2). Cyanobacteria occurred in almost all lakes and often dominated in the Midwest (figure 2(A)). Euglenophytes had high relative abundances in the South and along the East Coast (figure 2(B)). Diatoms had high relative abundances in many lakes throughout the USA, and particularly in the Rocky Mountains and the Northeast (figure 2(C)). Dinoflagellates were less abundant in the Great Plains and Southwest than in other regions (figure 2(D)). Cryptophytes and chlorophytes occurred throughout the



conterminous USA (figures 2(E) and (F)). Desmids and golden algae mostly occurred in the East, Northeast and Pacific Northwest, but were largely absent in the central part of the USA (figures 2(G) and (H)).

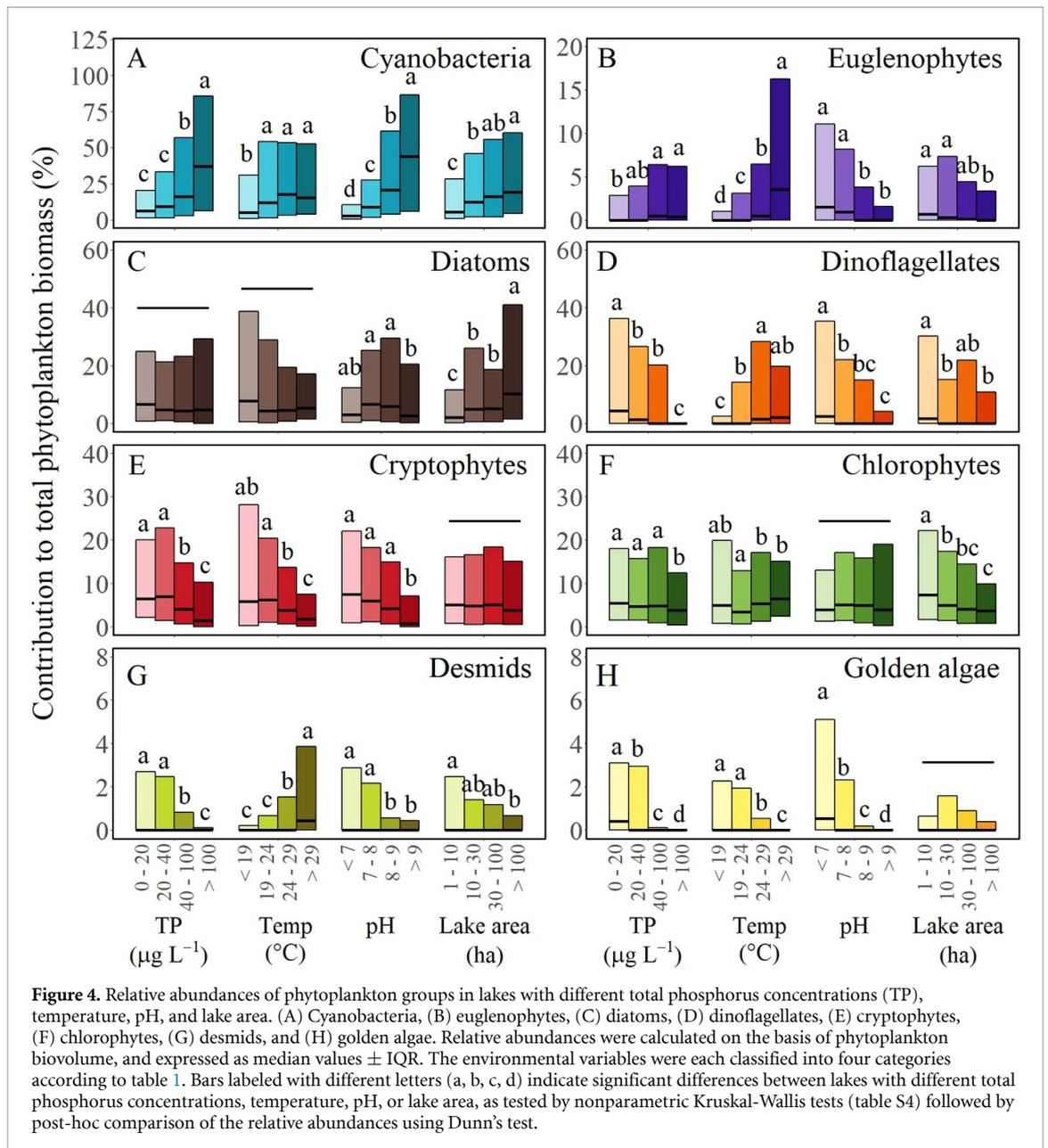
The phytoplankton community composition varied not only at the regional scale, but in many cases it also differed strongly between nearby lakes at local scales (figure 2). Consequently, Sørensen and Bray-Curtis similarity indices were relatively low even for lakes in the same area, and decreased only slightly but significantly with increasing geographical distance between lakes (figure 3, Sørensen similarity, two-sided Jonckheere-Terpstra test, $T_{JT} = 4.59 \times 10^{10}$, $p < 0.001$; Bray-Curtis similarity, two-sided Jonckheere-Terpstra test, $T_{JT} = 4.83 \times 10^{10}$, $p < 0.001$). The decrease in similarity in phytoplankton composition corresponded with a significant increase in variation in TP concentrations, water temperature and pH with increasing geographical distance between lakes (figure S1).

3.3. Environmental variables explaining phytoplankton composition

Subsequently, we analyzed which environmental variables provided the most parsimonious explanation for the observed variation in phytoplankton community composition. Stepwise selection of environmental variables by Dirichlet multiple regression showed that a model incorporating TP, water temperature, pH, and lake area best explained the relative abundances of the phytoplankton groups across the 1041 lakes (tables S2 and S3). Our further analysis therefore focused on these four explanatory variables.

3.3.1. TP

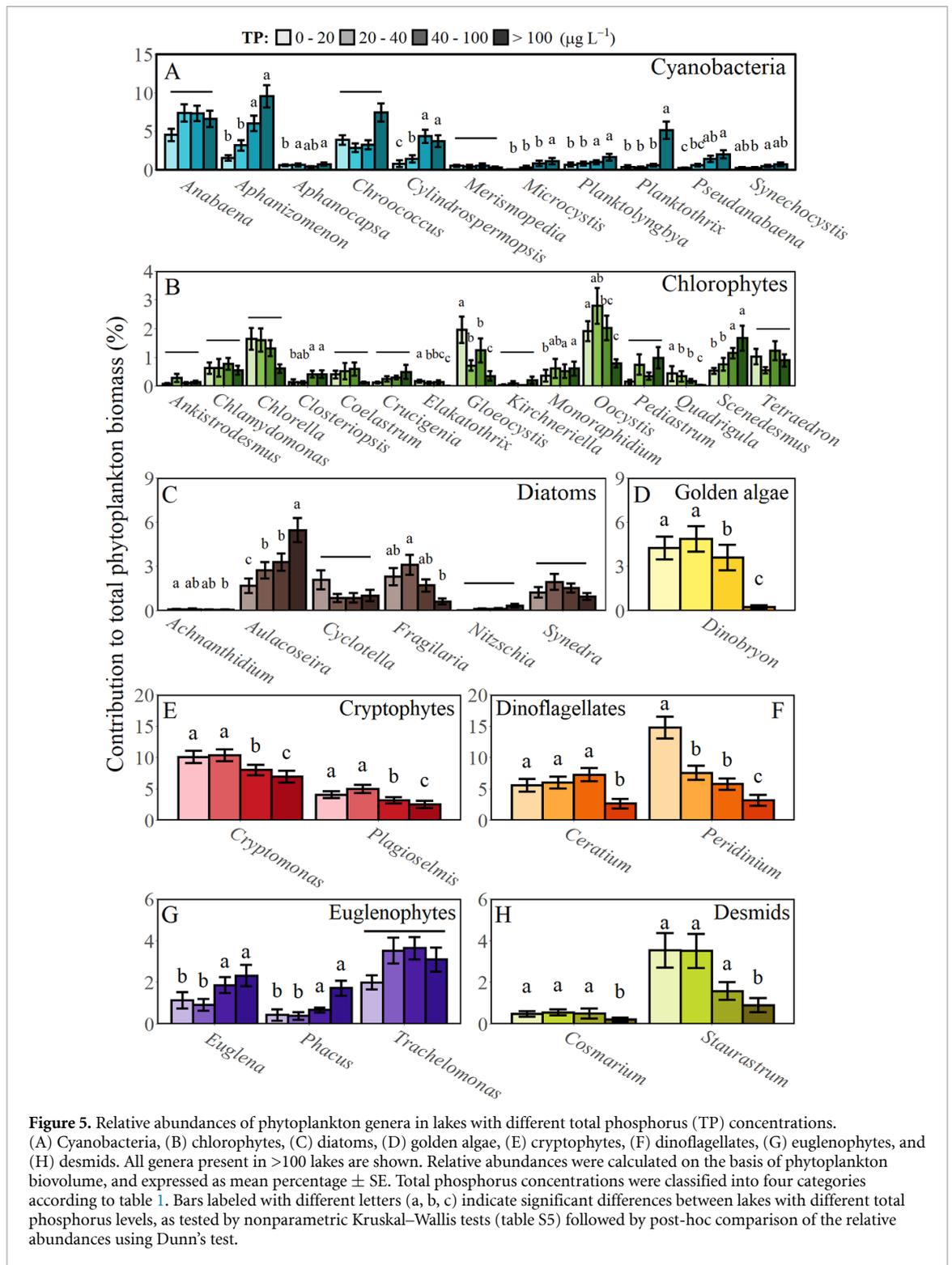
TP concentrations varied among lakes at both local and regional scales. Lakes in the Appalachian Highlands, Rocky Mountains, Great Lakes area and Pacific Northwest generally had the lowest TP concentrations, while many lakes in the Great Plains area had high TP concentrations (figure S2(A)). Phytoplankton community composition varied strongly with the TP concentration (figure 4, table S4). Relative abundance of cyanobacteria and euglenophytes were highest in phosphorus-rich lakes (figures 4(A) and (B)), whereas dinoflagellates, cryptophytes, desmids and golden algae reached their highest relative abundances in phosphorus-poor lakes (figures 4(D), (E), (G) and (H)). Relative abundances of diatoms and chlorophytes did not show major variation with the TP concentration (figures 4(C) and (F)).



At the genus level, the response to TP largely reflected that of the corresponding phytoplankton group, except for the diatoms and chlorophytes where some genera increased, others decreased, and again others showed no response to TP (figure 5, table S5). These opposite responses of individual genera may explain the lack of response of chlorophytes and diatoms at the phytoplankton group level.

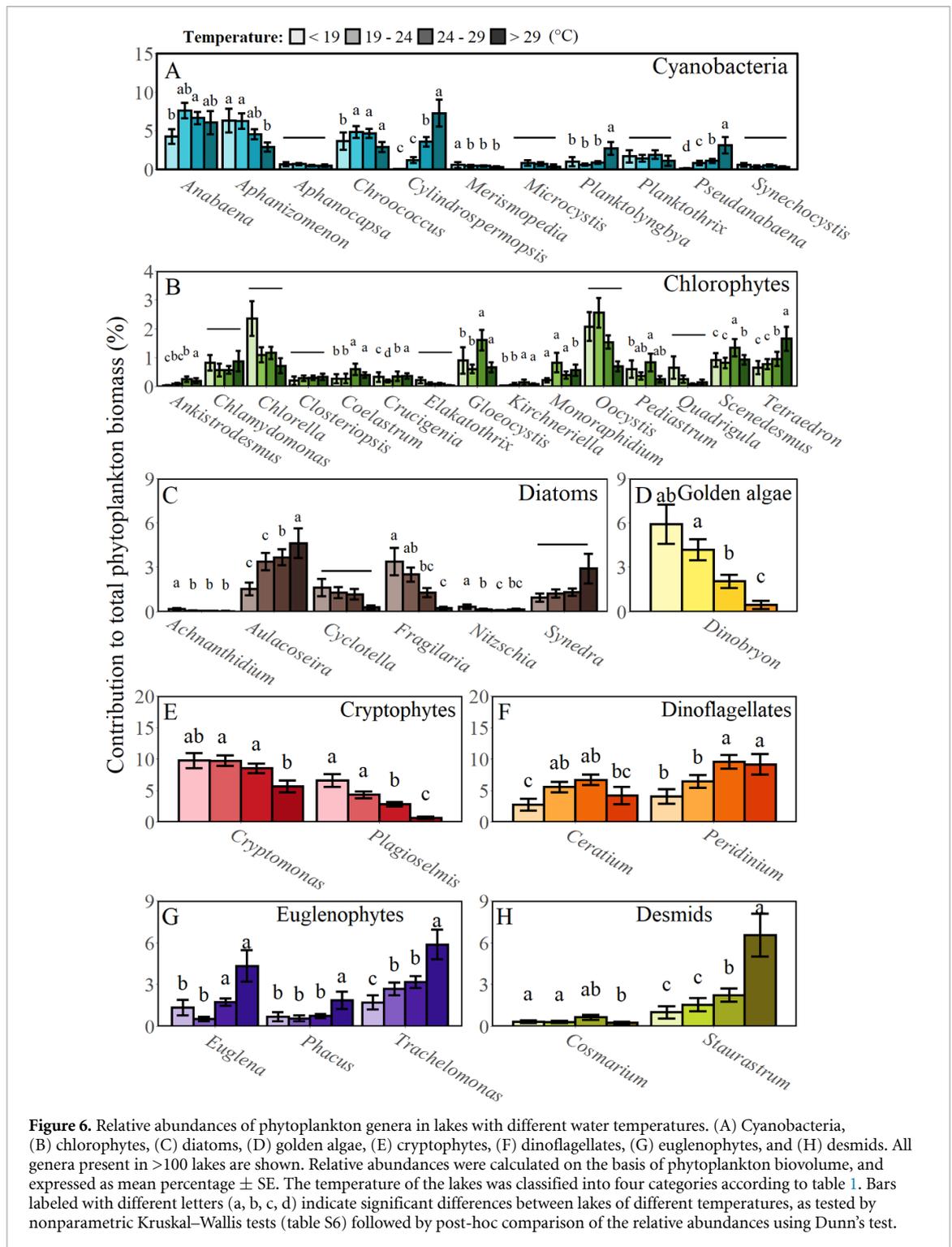
3.3.2. Lake temperature

In line with expectation, lakes were considerably warmer in the Southeast and Southern California than in the Rocky Mountains and Pacific Northwest (figure S2(B)). Phytoplankton community composition varied significantly with lake temperature (figure 4, table S4). Although cyanobacteria and dinoflagellates showed only a modest increase with temperature (figures 4(A) and (D)), relative abundances of desmids and euglenophytes were much higher in warm than in cold lakes (figures 4(B) and (G)). Conversely, relative abundances of cryptophytes and golden algae were highest in cold lakes (figures 4(E) and (H)). Relative abundances of diatoms and chlorophytes did not show major variation in response to lake temperature (figures 4(C) and (F)). For cyanobacteria, chlorophytes and diatoms, patterns at the genus level did not always reflect the temperature response of the corresponding phytoplankton group (figure 6, table S6). Among the cyanobacteria, for example, relative abundances of *Cylindrospermopsis* and *Pseudanabaena* increased but the relative abundance of *Aphanizomenon* decreased significantly with lake temperature (figure 6(A)).



3.3.3. Lake pH

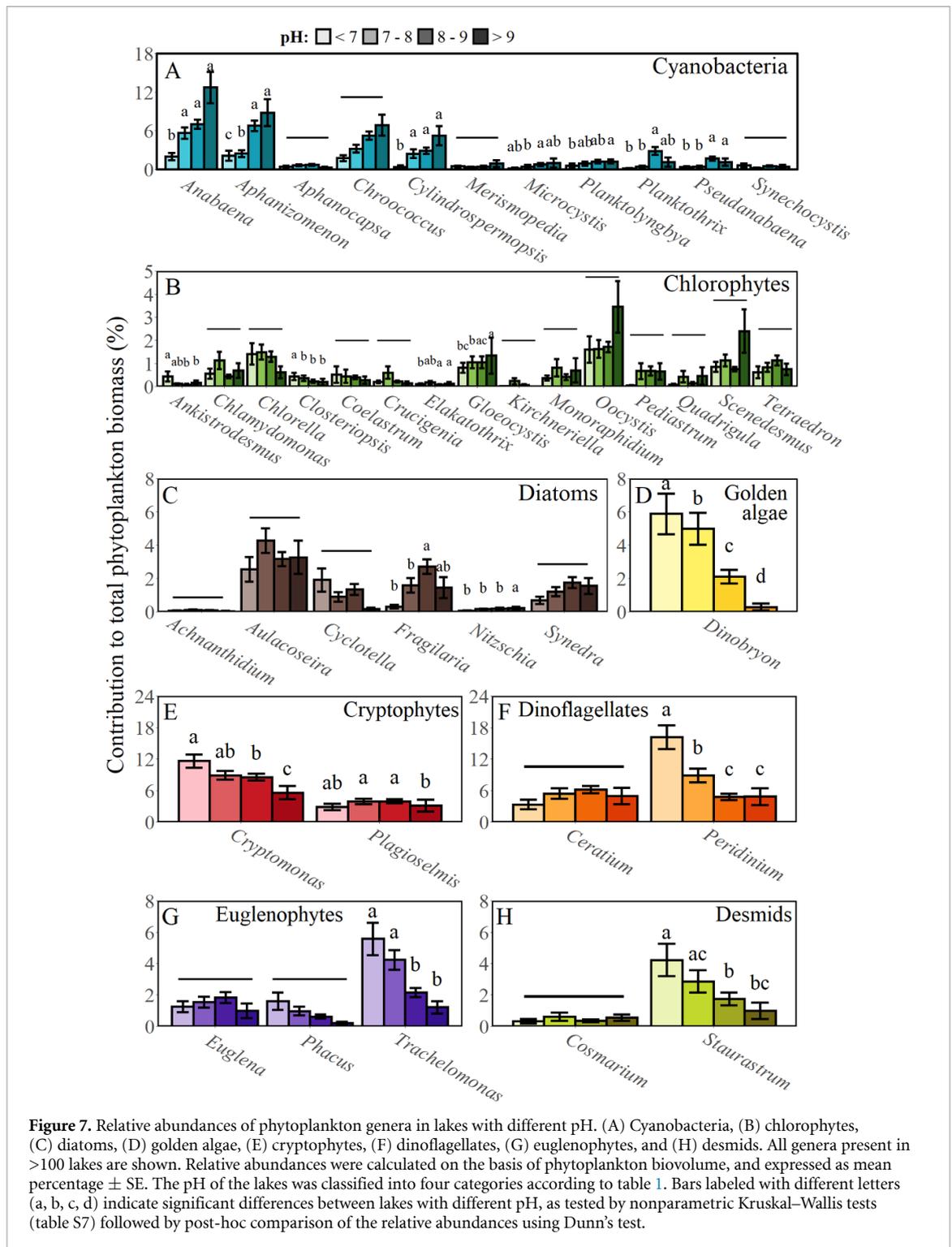
Many lakes along the East Coast, Great Lakes area and Pacific Northwest are moderately acidic, whereas $\text{pH} > 8$ in most lakes in the semi-arid regions of the Great Plains and Southwest (figure S2(C)). This pattern is largely associated with geographical variation in precipitation, with higher rainfall leaching alkaline elements from the soil causing lower pH (Slessarev *et al* 2016). Decay of organic matter and the production of humic acids in peatlands and forested areas may also contribute to low pH. Phytoplankton community composition varied significantly with lake pH (figure 4, table S4). Relative abundance of cyanobacteria was highest in alkaline lakes (figure 4(A)), whereas relative abundances of euglenophytes, dinoflagellates, cryptophytes, desmids and golden algae were all highest in acidic lakes (figures 4(B), (D), (E), (G) and (H)). Diatoms and chlorophytes did not show a clear pattern in response to lake pH (figures 4(C) and (F)). At the genus level,



many genera showed either a similar pH response as the corresponding phytoplankton group, or no clear response to pH (figure 7, table S7). Relative abundances of the filamentous nitrogen-fixing cyanobacteria *Anabaena*, *Aphanizomenon* and *Cylindrospermopsis* increased particularly strongly with pH (figure 7(A)).

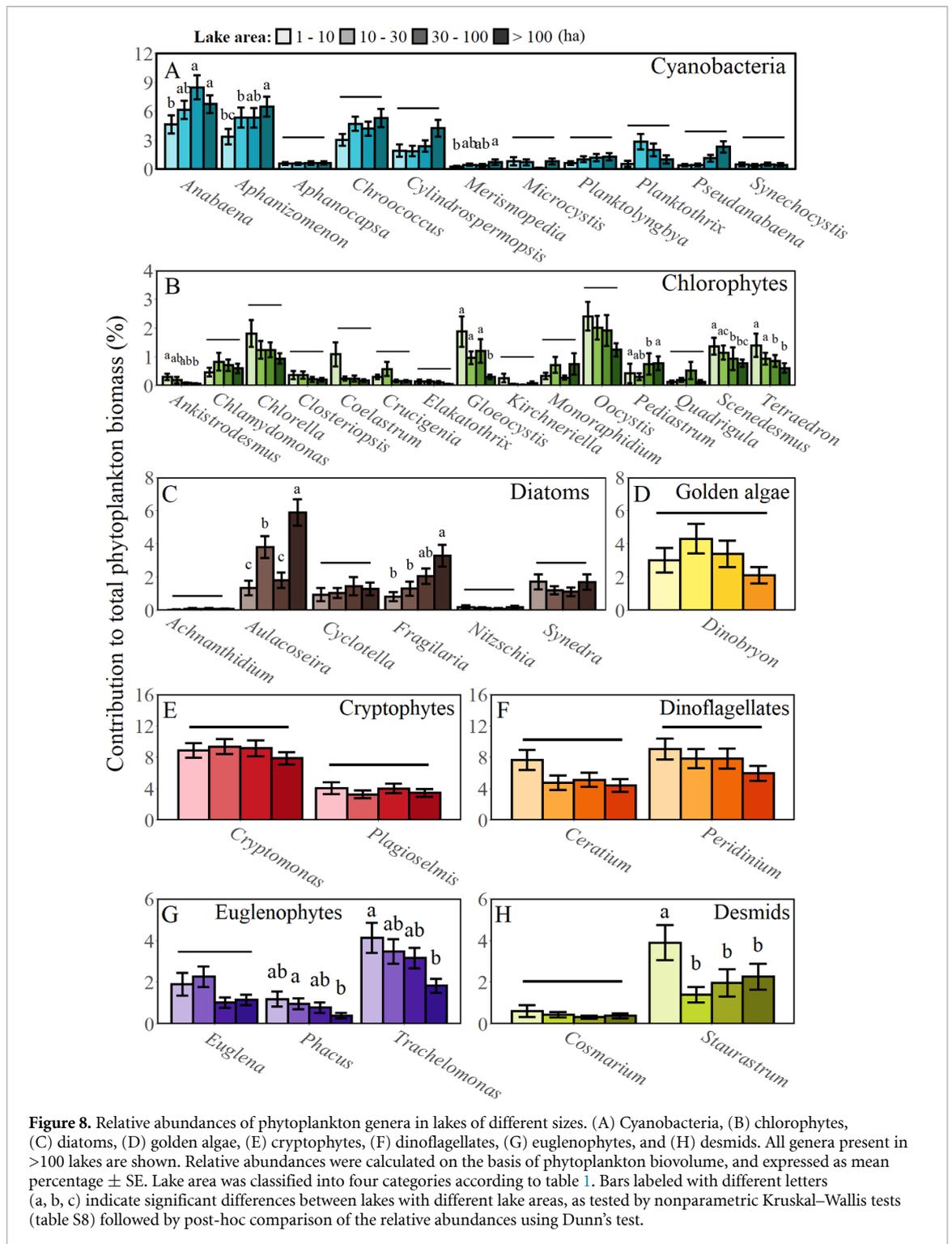
3.3.4. Lake area

Lake area varied over more than four orders of magnitude, but the geographical distribution of lake area did not show a distinct large-scale pattern in this dataset (figure S2(D)). Relative abundances of cyanobacteria and diatoms were significantly higher in large lakes (figures 4(A) and (C)), whereas relative abundances of euglenophytes and chlorophytes were significantly higher in small lakes (figures 4(B) and (F)). Patterns at the genus level largely reflected patterns at the level of the phytoplankton groups (figure 8, table S8).



3.4. Dominant phytoplankton groups

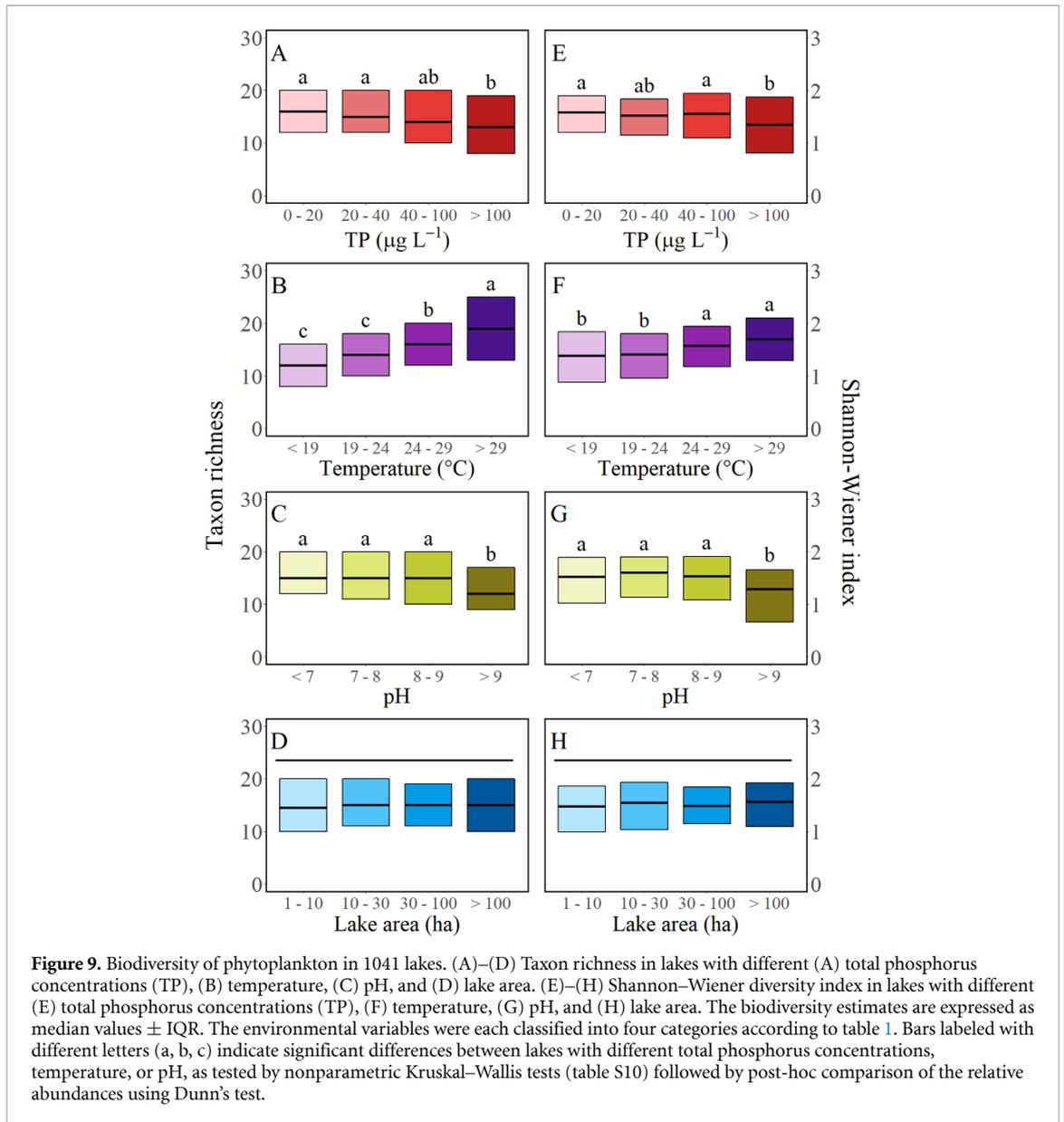
We also investigated under which environmental conditions specific phytoplankton groups became dominant. For this purpose, the ‘probability of dominance’ was calculated as the percentage of lakes in which a phytoplankton group comprised >75% of the total phytoplankton biomass. For cyanobacteria, the probability of dominance showed a strong increase with TP concentration and lake pH (figure S3(A); see table S9 for the statistics). Specifically, cyanobacteria were dominant in <5% of the lakes with a low TP concentration or a pH < 7, whereas they were dominant in >30% of the lakes with a high TP concentration or a pH > 9. The probability of dominance of diatoms was highest in cold waters (figure S3(C)), the probability of dominance of dinoflagellates and golden algae was highest in relatively acidic lakes



(figures S3(D) and (H)), and the probability of dominance of chlorophytes was highest in small ponds (figure S3(F)).

3.5. Biodiversity patterns

Phytoplankton biodiversity, expressed as taxon richness and by the Shannon–Wiener diversity index, varied among lakes at both local and regional scales (figure S4). Biodiversity was significantly lower in lakes with high TP concentrations and high pH, and increased significantly with lake temperature (figure 9; tables S10 and S11). Biodiversity did not vary significantly with lake area (figures 9(D) and (H); table S10). The same



lack of significance between biodiversity and lake area was also obtained if we used a linear regression analysis over all $n = 1041$ lakes (table S11), indicating that this result was robust.

4. Discussion

4.1. Large-scale patterns in community composition

The more than 1000 lakes analyzed in this study span a wide variety of geographical regions and environmental conditions, from oligotrophic to hypertrophic lakes, from cold mountain waters to warm subtropical lakes, and from acidic to highly alkaline lakes (figures 1 and 2). Phytoplankton community composition varied considerably between lakes, and not only at large spatial scales. Even phytoplankton communities of nearby lakes in the same local area often had a very different taxonomic composition. Yet, despite this local variation, large-scale patterns in phytoplankton community composition in lakes across the USA can also be discerned. Our analysis shows that the variation in phytoplankton community composition is best explained, in a statistical sense, by TP, temperature, pH and lake size. For example, cyanobacteria dominated the phytoplankton in many lakes with high phosphorus concentrations and high pH in the Midwest. Euglenophytes were abundant in several warm lakes in the South, diatoms preferred the colder lakes in, e.g. the Rocky Mountains, and dinoflagellates, desmids and golden algae tended to prefer the more oligotrophic and less alkaline lakes (figures 2 and 4).

The selection of TP, temperature, pH and lake size by our statistical analysis does not necessarily mean that these specific variables offer a causal explanation for the observed variation in phytoplankton community composition, since correlation does not necessarily reflect causation. For example, TP, TN and turbidity showed high collinearity (figure 1). TP was only a slightly better predictor than turbidity and TN (see the one-parameter models in table S2), and their high collinearity makes it difficult to disentangle effects of phosphorus, nitrogen and light on community composition from these lake data. This implies that TP represents a simple proxy of the overall trophic status of the lakes, including concomitant variation in nitrogen and light levels. Similarly, lake pH showed a strong negative correlation with pCO₂ and a strong positive correlation with the bicarbonate concentration (figure 1), and therefore variation in pH also represents variation in carbon speciation among the lakes. This implies that the lake data cannot distinguish whether shifts in community composition in response to the pH gradient represent direct pH effects on the species or indirect effects mediated by the availability and speciation of inorganic carbon. Hence, TP, temperature, pH and lake area can best be interpreted as the four major 'summary variables' that capture the underlying large-scale variation in a whole suite of environmental conditions governing phytoplankton community composition.

4.2. Variation in taxonomic composition

4.2.1. Cyanobacteria

Our results show that cyanobacteria reached their highest relative abundances in phosphorus-rich lakes with high pH, and showed a weaker relationship with lake temperature and lake area (figures 4 and S3). The increase of cyanobacteria with TP concentration is in agreement with many previous studies, which have shown that high nutrient loads stimulate the development of cyanobacterial blooms (Downing *et al* 2001, Paerl *et al* 2001, O'Neil *et al* 2012, Beaver *et al* 2018, Huisman *et al* 2018, Ho and Michalak 2020).

High relative abundances of cyanobacteria in lakes with high pH can be interpreted as a preference of cyanobacteria for alkaline waters with relatively high bicarbonate concentrations. Bicarbonate is the dominant inorganic carbon species at pH > 6.5 and cyanobacteria use a sophisticated CO₂ concentrating mechanism (CCM) including several bicarbonate uptake systems with different kinetic properties, which makes them highly effective in bicarbonate acquisition in waters at high pH (Price 2011, Sandrini *et al* 2014, Burnap *et al* 2015, Ji *et al* 2020). For example, Ji *et al* (2020) showed that in CO₂-depleted waters of high pH the cyanobacterium *Microcystis aeruginosa* strongly down-regulates its CO₂ uptake capacity and fully relies on bicarbonate uptake to cover its photosynthetic carbon fixation. Moreover, the photosynthetic activity of dense cyanobacterial blooms can be responsible for the drawdown of pCO₂, as indicated by the negative correlation between pCO₂ and chlorophyll *a* in lakes dominated by cyanobacteria (figure S5). This creates a positive feedback loop, where cyanobacterial dominance is not only favoured by high pH, but CO₂ depletion by dense cyanobacterial blooms in nutrient-rich waters also contributes to a high pH and a shift to bicarbonate as the dominant inorganic species (Verspagen *et al* 2014, Visser *et al* 2016).

The positive relation between relative cyanobacterial abundance and lake area appears to be a robust result (figures 4 and S3, table S3). Interestingly, this relationship is particularly pronounced for the nitrogen-fixing genera *Anabaena* (now known as *Dolichospermum*) and *Aphanizomenon* (figure 8(A)). Indeed, nitrogen-fixing cyanobacteria such as *Aphanizomenon*, *Anabaena* and *Nodularia* are also common genera in several other large waters, such as the Baltic Sea and Lake Peipsi (Stal *et al* 2003, Nöges *et al* 2008, Kahru *et al* 2020).

Although many studies have pointed out that cyanobacterial blooms will benefit from high temperatures (Paerl and Huisman 2008, Wagner and Adrian 2009, Kosten *et al* 2012, O'Neil *et al* 2012, Visser *et al* 2016, Huisman *et al* 2018, Ho and Michalak 2020), our results showed only a weak positive relationship of lake temperature with the relative abundance and no relationship with the dominance of cyanobacteria (figures 4 and S3). Previous analyses of the lake data of the US EPA collected in the summers of 2007 and 2012 did find positive effects of temperature on cyanobacteria (Beaulieu *et al* 2013, Rigosi *et al* 2014). Similar to our results, however, these studies also pointed out that nutrients had stronger effects on cyanobacterial abundance than temperature (Beaulieu *et al* 2013, Rigosi *et al* 2014). At the genus level, temperature affected the relative abundances of some cyanobacteria more strongly (figure 6(A)). In particular, relative abundances of the bloom-forming genera *Cylindrospermopsis*, *Planktolyngbya* and *Pseudanabaena* were significantly higher in warm than in cold lakes. *Cylindrospermopsis* is a genus generally found in warmer waters that has rapidly expanded from tropical to temperate zones in recent decades (Padisák 1997), presumably in response to global warming (Briand *et al* 2004). Also, *Pseudanabaena* is known to prefer relatively warm waters (Gao *et al* 2018). Other genera such as *Aphanizomenon* decreased with lake temperature, however (figure 6(A)). These opposite temperature effects on different cyanobacterial genera could potentially cancel out

temperature effects at higher levels of taxonomic aggregation, and hence may explain the weak temperature response for cyanobacteria as a whole.

4.2.2. Diatoms

The probability of dominance of diatoms was highest in lakes with low water temperature and low TP concentration (figure S3(C)). These results are in line with the common observation that, as a functional group, freshwater diatoms often dominate in cold and relatively nutrient-poor waters (Reynolds 2006, Edwards *et al* 2012). Furthermore, the relative abundance of diatoms increased with lake area (figure 4). This pattern was particularly pronounced for the genera *Aulacoseira* and *Fragilaria* (figure 8(C)), which consist of several large and heavily silicified taxa that require a considerable degree of turbulence to remain suspended in the water column (Sherman *et al* 1998, Ptacnik *et al* 2003, Rühland *et al* 2008, 2015). Hence, the positive association of these genera with lake area is in line with the observation that larger lakes are often more exposed to wind action and tend to experience deeper and more extensive vertical mixing (Fee *et al* 1996). Another possible explanation for the positive association between diatoms and lake area is that large lakes tended to have low DOC concentrations (figure 1), which results in relatively clear waters dominated by blue and green wavelengths that provide a suitable light environment for the fucoxanthin pigments of diatoms (Holtrop *et al* 2021).

At the aggregated level of the entire functional group, the relative abundance of diatoms did not show a clear relationship with TP and lake temperature (figure 4). However, cell size, temperature dependence and nutrient physiology vary considerably among diatom taxa, which may cause substantial changes in the taxonomic composition of diatom assemblages along temperature and nutrient gradients (Kilham *et al* 1996, Winder *et al* 2009, Rühland *et al* 2015). This is also illustrated by our results, which show large variation in temperature and nutrient responses among different diatom taxa (figures 5 and 6).

4.2.3. Chlorophytes and desmids

Chlorophytes did not display clear patterns in relation to TP, lake temperature and pH, but showed a preference for small lakes. Our results match previous studies reporting no significant response of relative chlorophyte abundance to TP (Watson *et al* 1997), although some studies report an increase (Jensen *et al* 1994). Similar to diatoms, the data show large variation in the response to TP, lake temperature and lake pH among different chlorophyte taxa (figures 5–7), which likely explains the lack of major trends at the level of the entire functional group. Many chlorophyte genera with high relative abundances in small lakes (e.g. *Gloeocystis*, *Oocystis*, *Scenedesmus*, *Tetraedron*; figure 8(B)) are common members of the periphyton (Eminson and Moss 1980, McCormick *et al* 1996), which may explain why chlorophytes as a group showed a higher relative abundance and higher probability of dominance in small lakes (figures 4 and S3).

Relative abundances of desmids were highest in small lakes with low TP concentration, low pH, and high temperature. These findings are in line with the observation that many desmid species inhabit the oligotrophic and relatively acidic waters of bogs and fens (Brook 1981, Coesel 1982), although some taxa prefer more eutrophic and alkaline waters (Spijkerman *et al* 2005). The preference of desmids for warm waters is in good agreement with laboratory experiments showing that a variety of desmid species isolated from temperate climates all had high temperature optima in the range of 25 °C–30 °C (Coesel and Wardenaar 1990, Stamenković and Hanelt 2013). Indeed, although desmids have a cosmopolitan distribution, their taxonomic diversity and quantitative abundance are generally highest in tropical regions (Coesel 1996, Barbosa and Padisák 2002), and most desmids in the temperate zone reach peak abundances during the summer months (Canter and Lund 1966, Coesel and Kooijman-Van Blokland 1994).

4.2.4. Dinoflagellates

Dinoflagellates reached their highest relative abundances in small- to medium-sized lakes with low TP concentration, low pH, and relatively high temperature. Vertical migration of dinoflagellates provides a major competitive advantage in stratified waters with a nutrient-depleted surface layer but higher nutrient availability at depth (Lieberman *et al* 1994, Doblin *et al* 2006, Peacock and Kudela 2014). However, the motility of dinoflagellates is sensitive to shear stress generated by high turbulence (Thomas and Gibson 1990, Berdalet *et al* 2007), although the sensitivity varies among taxa (Smayda 2002, Sullivan and Swift 2003). This likely explains why dinoflagellates are less common in large lakes that are more exposed to wind stress and experience deeper vertical mixing (Fee *et al* 1996). High temperatures strengthen the thermal stratification of lakes and enable a faster swimming speed of dinoflagellates, which may further improve conditions for the motile behavior of flagellated phytoplankton species (Kamykowski and McCollum 1986, Huisman *et al* 1999, Winder and Sommer 2012). Hence, although dinoflagellates display considerable physiological, morphological and behavioral variation, our results align with the common view that as a group they are

generally well adapted to thermally stratified waters with a relatively low nutrient availability in the epilimnion.

In our study, the relative abundance and probability of dominance of dinoflagellates decreased strongly with pH (figures 4 and S3). One possible explanation for this pronounced pattern is that many dinoflagellate species can grow mixotrophically, combining photosynthesis and bacterivory (Stoecker 1999, Flynn *et al* 2013). Several studies have indicated that mixotrophic taxa are especially abundant in humic waters with high DOC concentrations and bacterial abundances (Jansson *et al* 1996, Hansson *et al* 2019). Humic waters also tend to have a relatively low pH, which may explain the prevalence of dinoflagellates in lakes with pH < 7.

4.2.5. Golden algae, cryptophytes and euglenophytes

According to the phycological literature, golden algae (including chrysophytes and synurophytes) are typically associated with oligotrophic or mesotrophic waters with a slightly acidic to neutral pH, cryptophytes can be found in a wide range of aquatic ecosystems, and euglenophytes are particularly abundant in nutrient-rich waters and also in acidic environments (De Huszar and Caraco 1998, Reynolds 2006, Wehr *et al* 2015). This is in good agreement with the distribution patterns observed in our study.

In general, less attention has been given to the CCMs of golden algae, cryptophytes and euglenophytes than to the CCMs of other phytoplankton groups. For cryptophytes, in particular, the existence of a CCM has not yet been well investigated (Kroth 2015). Chrysophytes can take up CO₂ but not bicarbonate, and hence appear to lack a CCM (Maberly *et al* 2009). Euglenophytes have either no CCM or a relatively ineffective CCM, depending on the species (Colman and Balkos 2005). Hence, golden algae and euglenophytes are likely to benefit from high CO₂ availability at low pH values. This is confirmed by our results, where relative abundances of golden algae and euglenophytes were highest in acidic lakes (figure 4), but remained low in alkaline lakes where bicarbonate is the predominant inorganic carbon source and CO₂ concentrations are low.

The temperature response of golden algae, cryptophytes and euglenophytes has also received relatively little attention in the literature. Yet, all three groups showed a pronounced temperature response in the lake data. Specifically, golden algae and cryptophytes had higher relative abundances in cold lakes, whereas euglenophytes reached high relative abundance in warm lakes (figure 4). Indeed, several experimental and field studies have shown that golden algae prefer relatively cold waters and can be sensitive to high temperatures (Sandgren *et al* 1995, Butterwick *et al* 2005, Bergkemper *et al* 2018). Cryptophytes are common in temperate, boreal and polar lakes (McKnight *et al* 2000, Lepistö and Holopainen 2003), where they often dominate during or shortly after ice cover (Wiedner and Nixdorf 1998, Vehmaa and Salonen 2009). Conversely, blooms of euglenophytes are quite common in (sub)tropical climates and several other studies also found a positive relationship between euglenophyte abundances and water temperature (Chattopadhyay and Banerjee 2007, Rahman *et al* 2007).

4.3. Biodiversity patterns

Biodiversity is generally considered to be an indication of ecosystem functioning and resilience (Loreau *et al* 2001, Folke *et al* 2004, Ptacnik *et al* 2008b, Tilman *et al* 2014, Bestion *et al* 2021). Consistent with other lake studies, we found that phytoplankton biodiversity increased with temperature (Stomp *et al* 2011, Segura *et al* 2015), which supports the common observation of a latitudinal diversity gradient with highest species richness at low latitudes (Willig *et al* 2003, Hillebrand 2004, Pontarp *et al* 2019). Furthermore, our analysis showed a significantly lower taxon richness and Shannon–Wiener diversity index in lakes with high TP concentrations, which is in line with many other observations that the biodiversity of aquatic and terrestrial primary producers often declines at high nutrient levels, resulting in a low biodiversity of eutrophic ecosystems (Leibold 1999, Mittelbach *et al* 2001, Stevens *et al* 2004, Hautier *et al* 2009, Fraser *et al* 2015).

A classic prediction derived from the theory of island biogeography is that biodiversity will increase with habitat size (MacArthur and Wilson 1967). Since lakes represent ‘aquatic islands’ in a terrestrial landscape, this prediction may also apply to lake size. Indeed, a study on phytoplankton species richness across aquatic systems that spanned >15 orders of magnitude in spatial extent found that phytoplankton biodiversity increased with habitat area (Smith *et al* 2005). However, a study on phytoplankton species richness across USA lakes spanning a narrower spatial extent similar to our study (4 orders of magnitude) only found a very weak species-area relationship (Stomp *et al* 2011). In our analysis of >1000 lakes, neither phytoplankton taxon richness nor the Shannon–Wiener diversity index increased with lake area. Hence, at this continental scale, water temperature and nutrient status seemed to be the key determinants of phytoplankton biodiversity.

4.4. Caveats

As any study, this study suffers from a number of caveats. For example, the lake samples collected by the US EPA and analysed in our study were restricted to one point in time, and only considered one location in each

lake. Furthermore, samples were taken from the top 2 m of the surface water, which may underrepresent species that prefer deeper water layers. That said, however, the elegant simplicity of this sampling protocol has enabled the collection of data from >1000 lakes spread across the continent and all using the exact same procedures, which is essential for a coherent analysis of large-scale patterns in community composition.

Another caveat is that microscopic analysis of lake samples reveals larger phytoplankton taxa, while picophytoplankton is often overlooked. Picophytoplankton are common in lakes, and their diversity and abundance can vary strongly with environmental variables (Callieri and Stockner 2002, Stomp *et al* 2007, Schiaffino *et al* 2013). The use of dedicated techniques such as flow cytometric analysis (Schiaffino *et al* 2013, Mojica *et al* 2015) and high-throughput sequencing of 16S and 18S ribosomal RNA (Pierella Karlusich *et al* 2020, Vuorio *et al* 2020) may help to identify picophytoplankton in future lake surveys and elucidate their contribution to the phytoplankton community composition across large spatial scales.

5. Conclusions

Overall, our analysis of this large-scale data set shows that TP, temperature, pH and lake size are major determinants of the phytoplankton community composition in lakes. Interestingly, three of these four environmental parameters are associated with major human impacts on aquatic ecosystems, i.e. eutrophication results in higher phosphorus inputs into lakes, global warming is predicted to increase lake temperature, and rising atmospheric CO₂ concentrations in combination with enhanced DOC inputs from the surrounding watershed may lower the pH. Our results are based on correlations, which do not necessarily reflect causal relationships, as phytoplankton communities may be influenced by TP, temperature and pH through a variety of pathways and mechanisms. Moreover, the observed lake-to-lake variation in taxonomic composition is substantial and not all of this variation can be captured by just four environmental variables. Despite these cautionary notes, the comprehensive patterns in this extensive data set nicely illustrate how phytoplankton community structure varies across large spatial scales, and provide a tentative indication of which phytoplankton taxa are likely to be among the winners and losers in response to anthropogenic pressures such as eutrophication and global warming.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: www.epa.gov/national-aquatic-resource-surveys/data-national-aquatic-resource-surveys. (US Environmental Protection Agency 2016)

Acknowledgments

We gratefully acknowledge the efforts of dedicated field crews, laboratory staff, data management and quality control staff, analysts and many others from US EPA, states, tribes, federal agencies, universities, and other organizations in collecting the data of the National Lakes Assessment 2012. We thank the anonymous reviewers for their helpful comments. This research was supported by the Division of Earth and Life Sciences (ALW) of the Netherlands Organization for Scientific Research (NWO). X J and Q X L were also supported by the National Key R&D Program of China (2017YFC0506001) and National Natural Science Foundation of China (32071609).

ORCID iDs

Jolanda M H Verspagen  <https://orcid.org/0000-0001-9295-4183>

Xing Ji  <https://orcid.org/0000-0003-2539-8097>

Quan-Xing Liu  <https://orcid.org/0000-0002-8602-0154>

Jef Huisman  <https://orcid.org/0000-0001-9598-3211>

References

- Acevedo-Trejos E, Brandt G, Merico A and Smith S L 2013 Biogeographical patterns of phytoplankton community size structure in the oceans: phytoplankton biogeographical patterns *Glob. Ecol. Biogeogr.* **22** 1060–70
- Aitchison J 1982 The statistical analysis of compositional data *J. R. Stat. Soc. B* **44** 139–60
- Barbosa F A R and Padisák J 2002 The forgotten lake stratification pattern: atelomixis, and its ecological importance *Verh. Int. Ver. Theor. Angew. Limnol.* **28** 1385–95
- Beaulieu M, Pick F and Gregory-Eaves I 2013 Nutrients and water temperature are significant predictors of cyanobacterial biomass in a 1147 lakes data set *Limnol. Oceanogr.* **58** 1736–46

- Beaver J R, Tausz C E, Scotese K C, Pollard A I and Mitchell R M 2018 Environmental factors influencing the quantitative distribution of microcystin and common potentially toxigenic cyanobacteria in U.S. lakes and reservoirs *Harmful Algae* **78** 118–28
- Benjamini Y and Hochberg Y 1995 Controlling the false discovery rate: a practical and powerful approach to multiple testing *J. R. Stat. Soc. B* **57** 289–300
- Berdalet E, Peters F, Koumandou V L, Roldán C, Guadayol Ò and Estrada M 2007 Species-specific physiological response of dinoflagellates to quantified small-scale turbulence *J. Phycol.* **43** 965–77
- Bergkemper V, Stadler P and Weisse T 2018 Moderate weather extremes alter phytoplankton diversity—a microcosm study *Freshw. Biol.* **63** 1211–24
- Bestion E, Haegeman B, Codesal S A, Garreau A, Huet M, Barton S and Montoya J M 2021 Phytoplankton biodiversity is more important for ecosystem functioning in highly variable thermal environments *Proc. Natl Acad. Sci. USA* **118** e2019591118
- Brauer V S, Stomp M and Huisman J 2012 The nutrient-load hypothesis: patterns of resource limitation and community structure driven by competition for nutrients and light *Am. Nat.* **179** 721–40
- Briand J-F, Lebourlanger C, Humbert J-F, Bernard C and Dufour P 2004 *Cylindrospermopsis raciborskii* (Cyanobacteria) invasion at mid-latitudes: selection, wide physiological tolerance, or global warming? *J. Phycol.* **40** 231–8
- Brook A J 1981 *The Biology of Desmids* (Berkeley, CA: University of California Press)
- Burford M A, Carey C C, Hamilton D P, Huisman J, Paerl H W, Wood S A and Wulff A 2020 Perspective: advancing the research agenda for improving understanding of cyanobacteria in a future of global change *Harmful Algae* **91** 101601
- Burnap R, Hagemann M and Kaplan A 2015 Regulation of CO₂ concentrating mechanism in cyanobacteria *Life* **5** 348–71
- Butterwick C, Heaney S I and Talling J F 2005 Diversity in the influence of temperature on the growth rates of freshwater algae, and its ecological relevance *Freshw. Biol.* **50** 291–300
- Cael B B, Heathcote A J and Seekell D A 2017 The volume and mean depth of Earth's lakes *Geophys. Res. Lett.* **44** 209–18
- Callieri C and Stockner J G 2002 Freshwater autotrophic picoplankton: a review *J. Limnol.* **61** 1
- Canter H M and Lund J W G 1966 The periodicity of planktonic desmids in Windermere, England *Verh. Int. Ver. Theor. Angew. Limnol.* **16** 163–72
- Chattopadhyay C and Banerjee T C 2007 Temporal changes in environmental characteristics and diversity of net phytoplankton in a freshwater lake *Turk. J. Bot.* **31** 287–96
- Chen C-T A and Millero F J 1986 Thermodynamic properties for natural waters covering only the limnological range *Limnol. Oceanogr.* **31** 657–62
- Chorus I and Bartram J 1999 *Toxic Cyanobacteria in Water: A Guide to Their Public Health Consequences, Monitoring, and Management* (London, New York: E & FN Spon)
- Coesel P F M 1982 Structural characteristics and adaptations of desmid communities *J. Ecol.* **70** 163
- Coesel P F M 1996 Biogeography of desmids *Hydrobiologia* **336** 41–53
- Coesel P F M and Wardenaar K 1990 Growth responses of planktonic desmid species in a temperature-light gradient *Freshw. Biol.* **23** 551–60
- Coesel P and Kooijman-Van Blokland H 1994 Distribution and seasonality of desmids in the maarsseveen lakes area *Neth. J. Aquat. Ecol.* **28** 19–24
- Colman B and Balkos K D 2005 Mechanisms of inorganic carbon acquisition in two *Euglena* species *Can. J. Bot.* **83** 865–71
- Conley D J, Paerl H W, Howarth R W, Boesch D F, Seitzinger S P, Havens K E, Lancelot C and Likens G E 2009 Controlling eutrophication: nitrogen and phosphorus *Science* **323** 1014–5
- Dauta A, Devaux J, Piquemal F and Boumnick L 1990 Growth rate of four freshwater algae in relation to light and temperature *Hydrobiologia* **207** 221–6
- De Huszar V L M and Caraco N F 1998 The relationship between phytoplankton composition and physical–chemical variables: a comparison of taxonomic and morphological–functional descriptors in six temperate lakes *Freshw. Biol.* **40** 679–96
- Denman K L and Gargett A E 1983 Time and space scales of vertical mixing and advection of phytoplankton in the upper ocean *Limnol. Oceanogr.* **28** 801–15
- Doblin M A, Thompson P A, Revill A T, Butler E C V, Blackburn S I and Hallegraeff G M 2006 Vertical migration of the toxic dinoflagellate *Gymnodinium catenatum* under different concentrations of nutrients and humic substances in culture *Harmful Algae* **5** 665–77
- Douma J C and Weedon J T 2019 Analysing continuous proportions in ecology and evolution: a practical introduction to beta and dirichlet regression *Methods Ecol. Evol.* **10** 1412–30
- Downing J A, Watson S B and McCauley E 2001 Predicting cyanobacteria dominance in lakes *Can. J. Fish. Aquat. Sci.* **58** 1905–8
- Edwards K F, Thomas M K, Klausmeier C A and Litchman E 2012 Allometric scaling and taxonomic variation in nutrient utilization traits and maximum growth rate of phytoplankton *Limnol. Oceanogr.* **57** 554–66
- Elliott J A 2010 The seasonal sensitivity of cyanobacteria and other phytoplankton to changes in flushing rate and water temperature *Glob. Change Biol.* **16** 864–76
- Elser J J, Marzolf E R and Goldman C R 1990 Phosphorus and nitrogen limitation of phytoplankton growth in the freshwaters of North America: a review and critique of experimental enrichments *Can. J. Fish. Aquat. Sci.* **47** 1468–77
- Eminson D and Moss B 1980 The composition and ecology of periphyton communities in freshwaters: 1 The influence of host type and external environment on community composition *Br. Phycol. J.* **15** 429–46
- Fee E J, Hecky R E, Kasian S E M and Cruikshank D R 1996 Effects of lake size, water clarity, and climatic variability on mixing depths in Canadian shield lakes *Limnol. Oceanogr.* **41** 912–20
- Feld C K, Segurado P and Gutiérrez-Cánovas C 2016 Analysing the impact of multiple stressors in aquatic biomonitoring data: a 'cookbook' with applications in R *Sci. Total Environ.* **573** 1320–39
- Filstrup C T, Heathcote A J, Kendall D L and Downing J A 2016 Phytoplankton taxonomic compositional shifts across nutrient and light gradients in temperate lakes *Inland Waters* **6** 234–49
- Flynn K J, Stoecker D K, Mitra A, Raven J A, Glibert P M, Hansen P J, Granéli E and Burkholder J M 2013 Misuse of the phytoplankton–zooplankton dichotomy: the need to assign organisms as mixotrophs within plankton functional types *J. Plankton Res.* **35** 3–11
- Folke C, Carpenter S, Walker B, Scheffer M, Elmqvist T, Gunderson L and Holling C S 2004 Regime shifts, resilience, and biodiversity in ecosystem management *Annu. Rev. Ecol. Syst.* **35** 557–81
- Fraser L H *et al* 2015 Worldwide evidence of a unimodal relationship between productivity and plant species richness *Science* **349** 302–5

- Gao J, Zhu J, Wang M and Dong W 2018 Dominance and growth factors of *Pseudanabaena* sp. in drinking water source reservoirs, Southern China *Sustainability* **10** 3936
- Greenacre M 2021 Compositional data analysis *Annu. Rev. Stat. Appl.* **8** 271–99
- Hansson T H, Grossart H, Giorgio P A, St-Gelais N F and Beisner B E 2019 Environmental drivers of mixotrophs in boreal lakes *Limnol. Oceanogr.* **64** 1688–705
- Hautier Y, Niklaus P A and Hector A 2009 Competition for light causes plant biodiversity loss after eutrophication *Science* **324** 636–8
- Hillebrand H 2004 On the generality of the latitudinal diversity gradient *Am. Nat.* **163** 192–211
- Hillebrand H, Dürselen C-D, Kirschtel D, Pollinger U and Zohary T 1999 Biovolume calculation for pelagic and benthic microalgae *J. Phycol.* **35** 403–24
- Hiltunen T, Barreiro A and Hairston N G 2012 Mixotrophy and the toxicity of *Ochromonas* in pelagic food webs *Freshw. Biol.* **57** 2262–71
- Ho J C and Michalak A M 2020 Exploring temperature and precipitation impacts on harmful algal blooms across continental U.S. lakes *Limnol. Oceanogr.* **65** 992–1009
- Holtrop T, Huisman J, Stomp M, Biersteker L, Aerts J, Grébert T, Partensky F, Garczarek L and van der Woerd H J 2021 Vibrational modes of water predict spectral niches for photosynthesis in lakes and oceans *Nat. Ecol. Evol.* **5** 55–66
- Huisman J, Arrayás M, Ebert U and Sommeijer B 2002 How do sinking phytoplankton species manage to persist? *Am. Nat.* **159** 245–54
- Huisman J, Codd G A, Paerl H W, Ibelings B W, Verspagen J M H and Visser P M 2018 Cyanobacterial blooms *Nat. Rev. Microbiol.* **16** 471–83
- Huisman J, Sharples J, Stroom J M, Visser P M, Kardinaal W E A, Verspagen J M H and Sommeijer B 2004 Changes in turbulent mixing shift competition for light between phytoplankton species *Ecology* **85** 2960–70
- Huisman J, van Oostveen P and Weissing F J 1999 Species dynamics in phytoplankton blooms: incomplete mixing and competition for light *Am. Nat.* **154** 46–68
- Janssen A B G, Teurlinckx S, Beusen A H W, Huijbregts M A J, Rost J, Schipper A M, Seelen L M S, Mooij W M and Janse J H 2019 PCLake+: a process-based ecological model to assess the trophic state of stratified and non-stratified freshwater lakes worldwide *Ecol. Modelling* **396** 23–32
- Jansson M, Blomqvist P, Jonsson A and Bergström A-K 1996 Nutrient limitation of bacterioplankton, autotrophic and mixotrophic phytoplankton, and heterotrophic nanoflagellates in lake Öträsket *Limnol. Oceanogr.* **41** 1552–9
- Jensen J P, Jeppesen E, Olrik K and Kristensen P 1994 Impact of nutrients and physical factors on the shift from cyanobacterial to chlorophyte dominance in shallow Danish lakes *Can. J. Fish. Aquat. Sci.* **51** 1692–9
- Jeppesen E, Jensen J P, Søndergaard M, Lauridsen T and Landkildehus F 2000 Trophic structure, species richness and biodiversity in Danish lakes: changes along a phosphorus gradient: a detailed study of Danish lakes along a phosphorus gradient *Freshw. Biol.* **45** 201–18
- Ji X, Verspagen J M H, Stomp M and Huisman J 2017 Competition between cyanobacteria and green algae at low versus elevated CO₂: who will win, and why? *J. Exp. Bot.* **68** 3815–28
- Ji X, Verspagen J M H, Van de Waal D B, Rost B and Huisman J 2020 Phenotypic plasticity of carbon fixation stimulates cyanobacterial blooms at elevated CO₂ *Sci. Adv.* **6** eaax2926
- Jöhnk K D, Huisman J, Sharples J, Sommeijer B, Visser P M and Stroom J M 2008 Summer heatwaves promote blooms of harmful cyanobacteria *Glob. Change Biol.* **14** 495–512
- Kahru M, Elmgren R, Kaiser J, Wasmund N and Savchuk O 2020 Cyanobacterial blooms in the Baltic Sea: correlations with environmental factors *Harmful Algae* **92** 101739
- Kamykowski D and McCollum S A 1986 The temperature acclimatized swimming speed of selected marine dinoflagellates *J. Plankton Res.* **8** 275–87
- Karney C F F 2013 Algorithms for geodesics *J. Geod.* **87** 43–55
- Kilham S S, Theriot E C and Fritz S C 1996 Linking planktonic diatoms and climate change in the large lakes of the yellowstone ecosystem using resource theory *Limnol. Oceanogr.* **41** 1052–62
- Kosten S *et al* 2012 Warmer climates boost cyanobacterial dominance in shallow lakes *Glob. Change Biol.* **18** 118–26
- Kroth P G 2015 The biodiversity of carbon assimilation *J. Plant Physiol.* **172** 76–81
- Leibold M A 1999 Biodiversity and nutrient enrichment in pond plankton communities *Evol. Ecol. Res.* **1** 73–95
- Lepistö L and Holopainen A-L 2003 Occurrence of cryptophyceae and katablepharids in boreal lakes *Hydrobiologia* **502** 307–14
- Lewis E 1980 The practical salinity scale 1978 and its antecedents *IEEE J. Ocean. Eng.* **5** 3–8
- Lieberman O S, Shilo M and van Rijn J 1994 The physiological ecology of a freshwater dinoflagellate bloom population: vertical migration, nitrogen limitation, and nutrient uptake kinetics *J. Phycol.* **30** 964–71
- Loreau M *et al* 2001 Biodiversity and ecosystem functioning: current knowledge and future challenges *Science* **294** 804–8
- Lürling M, Eshetu F, Faassen E J, Kosten S and Huszar V L M 2013 Comparison of cyanobacterial and green algal growth rates at different temperatures *Freshw. Biol.* **58** 552–9
- Maberly S C, Ball L A, Raven J A and Sültemeyer D 2009 Inorganic carbon acquisition by chrysophytes *J. Phycol.* **45** 1052–61
- MacArthur R H and Wilson E O 1967 *The Theory of Island Biogeography* (Princeton, NJ: Princeton University Press) (available at: <https://press.princeton.edu/books/paperback/9780691088365/the-theory-of-island-biogeography>)
- Maier M J 2014 DirichletReg: dirichlet regression for compositional data in R (WU Vienna University of Economics and Business) (available at: <https://epub.wu.ac.at/4077/>)
- Mantzouki E *et al* 2018 A European multi lake survey dataset of environmental variables, phytoplankton pigments and cyanotoxins *Sci. Data* **5** 180226
- McCormick P V, Rawlik P S, Lurding K, Smith E P and Sklar F H 1996 Periphyton-water quality relationships along a nutrient gradient in the Northern florida everglades *J. North Am. Benthol. Soc.* **15** 433–49
- McKnight D M, Howes B L, Taylor C D and Goehring D D 2000 Phytoplankton dynamics in a stably stratified Antarctic lake during winter darkness *J. Phycol.* **36** 852–61
- Meriluoto J, Spoof L and Codd G A 2017 *Handbook of Cyanobacterial Monitoring and Cyanotoxin Analysis* (New York: Wiley)
- Millero F J, Graham T B, Huang F, Bustos-Serrano H and Pierrot D 2006 Dissociation constants of carbonic acid in seawater as a function of salinity and temperature *Mar. Chem.* **100** 80–94
- Mittelbach G G, Steiner C F, Scheiner S M, Gross K L, Reynolds H L, Waide R B, Willig M R, Dodson S I and Gough L 2001 What is the observed relationship between species richness and productivity? *Ecology* **82** 2381–96

- Mojica K D A, van de Poll W H, Kehoe M, Huisman J, Timmermans K R, Buma A G J, van der Woerd H J, Hahn-Woernle L, Dijkstra H A and Brussaard C P D 2015 Phytoplankton community structure in relation to vertical stratification along a north-south gradient in the Northeast Atlantic Ocean *Limnol. Oceanogr.* **60** 1498–521
- Naimi B, Hamm N A S, Groen T A, Skidmore A K and Toxopeus A G 2014 Where is positional uncertainty a problem for species distribution modelling? *Ecography* **37** 191–203
- Nöges T, Laugaste R, Nöges P and Tönno I 2008 Critical N:P ratio for cyanobacteria and N₂-fixing species in the large shallow temperate lakes Peipsi and Võrtsjärv, North-East Europe *Hydrobiologia* **599** 77–86
- O’Neil J M, Davis T W, Burford M A and Gobler C J 2012 The rise of harmful cyanobacteria blooms: the potential roles of eutrophication and climate change *Harmful Algae* **14** 313–34
- Padisák J 1997 *Cylindrospermopsis raciborskii* (Woloszynska) Seenayya et Subba Raju, an expanding, highly adaptive cyanobacterium: worldwide distribution and review of its ecology *Arch. Hydrobiol. Suppl. Monogr. Beitr.* **107** 563–93
- Paerl H W, Fulton R S, Moisaner P H and Dyble J 2001 Harmful freshwater algal blooms, with an emphasis on cyanobacteria *Sci. World J.* **1** 76–113
- Paerl H W and Huisman J 2008 Blooms like it hot *Science* **320** 57–58
- Paerl H W and Otten T G 2013 Blooms bite the hand that feeds them *Science* **342** 433–4
- Paerl H W, Xu H, McCarthy M J, Zhu G, Qin B, Li Y and Gardner W S 2011 Controlling harmful cyanobacterial blooms in a hyper-eutrophic lake (Lake Taihu, China): the need for a dual nutrient (N & P) management strategy *Water Res.* **45** 1973–83
- Passarge J, Hol S, Escher M and Huisman J 2006 Competition for nutrients and light: stable coexistence, alternative stable states, or competitive exclusion? *Ecol. Monogr.* **76** 57–72
- Peacock M B and Kudela R M 2014 Evidence for active vertical migration by two dinoflagellates experiencing iron, nitrogen, and phosphorus limitation *Limnol. Oceanogr.* **59** 660–73
- Pickett S T A 1989 Space-for-time substitution as an alternative to long-term studies *Long-Term Studies in Ecology: Approaches and Alternatives* ed G E Likens (New York: Springer) pp 110–35
- Pierella Karlusich J J, Ibarbalz F M and Bowler C 2020 Phytoplankton in the Tara Ocean *Ann. Rev. Mar. Sci.* **12** 233–65
- Pollard A I, Hampton S E and Leech D M 2018 The promise and potential of continental-scale limnology using the U.S. Environmental protection agency’s National lakes assessment *Limnol. Oceanogr. Bull.* **27** 36–41
- Pontarp M *et al* 2019 The latitudinal diversity gradient: novel understanding through mechanistic eco-evolutionary models *Trends Ecol. Evol.* **34** 211–23
- Price G D 2011 Inorganic carbon transporters of the cyanobacterial CO₂ concentrating mechanism *Photosynth. Res.* **109** 47–57
- Ptácnik R, Diehl S and Berger S 2003 Performance of sinking and nonsinking phytoplankton taxa in a gradient of mixing depths *Limnol. Oceanogr.* **48** 1903–12
- Ptácnik R, Lepistö L, Willén E, Brettum P, Andersen T, Rekolainen S, Lyche Solheim A and Carvalho L 2008a Quantitative responses of lake phytoplankton to eutrophication in Northern Europe *Aquat. Ecol.* **42** 227–36
- Ptácnik R, Solimini A G, Andersen T, Tamminen T, Brettum P, Lepistö L, Willén E and Rekolainen S 2008b Diversity predicts stability and resource use efficiency in natural phytoplankton communities *Proc. Natl Acad. Sci.* **105** 5134–8
- R Core Team 2020 *R: A Language and Environment for Statistical Computing* (Vienna: R Foundation for Statistical Computing) (available at: www.R-project.org/)
- Raftery A E 1995 Bayesian model selection in social research *Sociol. Methodol.* **25** 111–63
- Rahman M M, Jewel M A S, Khan S and Haque M M 2007 Study of euglenophytes bloom and its impact on fish growth in Bangladesh *Algae* **22** 185–92
- Reynolds C S 2006 *The Ecology of Phytoplankton* (Cambridge: Cambridge University Press)
- Rhee G-Y and Gotham I J 1980 Optimum N:P ratios and coexistence of planktonic algae *J. Phycol.* **16** 486–9
- Rigosi A, Carey C C, Ibelings B W and Brookes J D 2014 The interaction between climate warming and eutrophication to promote cyanobacteria is dependent on trophic state and varies among taxa *Limnol. Oceanogr.* **59** 99–114
- Rühland K M, Paterson A M and Smol J P 2015 Lake diatom responses to warming: reviewing the evidence *J. Paleolimnol.* **54** 1–35
- Rühland K, Paterson A M and Smol J P 2008 Hemispheric-scale patterns of climate-related shifts in planktonic diatoms from North American and European lakes *Glob. Change Biol.* **14** 2740–54
- Sandgren C D, Smol J P and Kristiansen J 1995 *Chrysophyte Algae: Ecology, Phylogeny and Development* (Cambridge: Cambridge University Press)
- Sandrini G, Ji X, Verspagen J M H, Tann R P, Slot P C, Luimstra V M, Schuurmans J M, Matthijs H C P and Huisman J 2016 Rapid adaptation of harmful cyanobacteria to rising CO₂ *Proc. Natl Acad. Sci. USA* **113** 9315–20
- Sandrini G, Matthijs H C P, Verspagen J M H, Muyzer G and Huisman J 2014 Genetic diversity of inorganic carbon uptake systems causes variation in CO₂ response of the cyanobacterium *Microcystis* *ISME J.* **8** 589–600
- Schiaffino M R, Gasol J M, Izaguirre I and Unrein F 2013 Picoplankton abundance and cytometric group diversity along a trophic and latitudinal lake gradient *Aquat. Microb. Ecol.* **68** 231–50
- Schindler D W 1974 Eutrophication and recovery in experimental lakes: implications for lake management *Science* **184** 897–9
- Schwarz G 1978 Estimating the dimension of a model *Ann. Stat.* **6** 461–4
- Segura A M, Calliari D, Kruk C, Fort H, Izaguirre I, Saad J F and Arim M 2015 Metabolic dependence of phytoplankton species richness: phytoplankton large-scale richness and temperature *Glob. Ecol. Biogeogr.* **24** 472–82
- Sherman B S, Webster I T, Jones G J and Oliver R L 1998 Transitions between *Aulacoseira* and *Anabaena* dominance in a turbid river weir pool *Limnol. Oceanogr.* **43** 1902–15
- Slessarev E W, Lin Y, Bingham N L, Johnson J E, Dai Y, Schimel J P and Chadwick O A 2016 Water balance creates a threshold in soil pH at the global scale *Nature* **540** 567–9
- Smayda T J 2002 Turbulence, watermass stratification and harmful algal blooms: an alternative view and frontal zones as “pelagic seed banks” *Harmful Algae* **1** 95–112
- Smith V H 1983 Low nitrogen to phosphorus ratios favor dominance by blue-green algae in lake phytoplankton *Science* **221** 669–71
- Smith V H, Foster B L, Grover J P, Holt R D, Leibold M A and deNoyelles F 2005 Phytoplankton species richness scales consistently from laboratory microcosms to the world’s oceans *Proc. Natl Acad. Sci. USA* **102** 4393–6
- Smol J P and Stoermer E F 2010 *The Diatoms: Applications for the Environmental and Earth Sciences* (Cambridge: Cambridge University Press)
- Sodré E D O, Langlais-Bourassa A, Pollard A I and Beisner B E 2020 Functional and taxonomic biogeography of phytoplankton and zooplankton communities in relation to environmental variation across the contiguous USA *J. Plankton Res.* **42** 141–57

- Sommer U 1985 Comparison between steady state and non-steady state competition: experiments with natural phytoplankton *Limnol. Oceanogr.* **30** 335–46
- Spijkerman E, Maberly S C and Coesel P F 2005 Carbon acquisition mechanisms by planktonic desmids and their link to ecological distribution *Can. J. Bot.* **83** 850–8
- Stal L J, Albertano P, Bergman B, von Bröckel K, Gallon J R, Hayes P K, Sivonen K and Walsby A E 2003 BASIC: baltic Sea cyanobacteria. An investigation of the structure and dynamics of water blooms of cyanobacteria in the Baltic Sea-responses to a changing environment *Cont. Shelf Res.* **23** 1695–714
- Stamenković M and Hanelt D 2013 Adaptation of growth and photosynthesis to certain temperature regimes is an indicator for the geographical distribution of *Cosmarium* strains (Zygnematophyceae, Streptophyta) *Eur. J. Phycol.* **48** 116–27
- Steffen M M, Dearth S P, Dill B D, Li Z, Larsen K M, Campagna S R and Wilhelm S W 2014 Nutrients drive transcriptional changes that maintain metabolic homeostasis but alter genome architecture in *Microcystis* *ISME J.* **8** 2080–92
- Stevens C J, Dise N B, Mountford J O and Gowing D J 2004 Impact of nitrogen deposition on the species richness of grasslands *Science* **303** 1876–9
- Stoecker D K 1999 Mixotrophy among dinoflagellates *J. Eukaryot. Microbiol.* **46** 397–401
- Stomp M, Huisman J, Mittelbach G G, Litchman E and Klausmeier C A 2011 Large-scale biodiversity patterns in freshwater phytoplankton *Ecology* **92** 2096–107
- Stomp M, Huisman J, Vörös L, Pick F R, Laamanen M, Haverkamp T and Stal L J 2007 Colourful coexistence of red and green picocyanobacteria in lakes and seas *Ecol. Lett.* **10** 290–8
- Stumm W and Morgan J J 2012 *Aquatic Chemistry: Chemical Equilibria and Rates in Natural Waters* (New York: Wiley)
- Sullivan J M and Swift E 2003 Effects of small-scale turbulence on net growth rate and size of ten species of marine dinoflagellates *J. Phycol.* **39** 83–94
- Tao S *et al* 2020 Changes in China's lakes: climate and human impacts *Natl Sci. Rev.* **7** 132–40
- Taranu Z E *et al* 2015 Acceleration of cyanobacterial dominance in north temperate-subarctic lakes during the Anthropocene *Ecol. Lett.* **18** 375–84
- Thomas W H and Gibson C H 1990 Quantified small-scale turbulence inhibits a red tide dinoflagellate, *Gonyaulax polyedra* Stein *Deep Sea Res. A* **37** 1583–93
- Tilman D, Isbell F and Cowles J M 2014 Biodiversity and ecosystem functioning *Annu. Rev. Ecol. Evol. Syst.* **45** 471–93
- Tilman D, Mattson M and Langer S 1981 Competition and nutrient kinetics along a temperature gradient: an experimental test of a mechanistic approach to niche theory *Limnol. Oceanogr.* **26** 1020–33
- Tong Y *et al* 2017 Decline in Chinese lake phosphorus concentration accompanied by shift in sources since 2006 *Nat. Geosci.* **10** 507–11
- US Environmental Protection Agency 2011a 2012 National Lakes assessment *Site Evaluation Guidelines* EPA-841-B-11-005
- US Environmental Protection Agency 2011b 2012 National lakes assessment *Field Operations Manual* EPA-841-B-11-003
- US Environmental Protection Agency 2012 2012 National lakes assessment *Laboratory Operations Manual* EPA-841-B-11-004
- US Environmental Protection Agency 2016 National lakes assessment 2012 *National Aquatic Resource Surveys* (data and metadata files) (available at: www.epa.gov/national-aquatic-resource-surveys/data-national-aquatic-resource-surveys) (Accessed 30 March 2017)
- Vehmaa A and Salonen K 2009 Development of phytoplankton in Lake Pääjärvi (Finland) during under-ice convective mixing period *Aquat. Ecol.* **43** 693–705
- Verspagen J M H, Van de Waal D B, Finke J E, Visser P M, Van Donk E and Huisman J 2014 Rising CO₂ levels will intensify phytoplankton blooms in eutrophic and hypertrophic lakes *PLoS One* **9** e104325
- Visser P M, Verspagen J M H, Sandrini G, Stal L J, Matthijs H C P, Davis T W, Paelr H W and Huisman J 2016 How rising CO₂ and global warming may stimulate harmful cyanobacterial blooms *Harmful Algae* **54** 145–59
- Vogt R J, St-Gelais N F, Bogard M J, Beisner B E and Del Giorgio P A 2017 Surface water CO₂ concentration influences phytoplankton production but not community composition across boreal lakes *Ecol. Lett.* **20** 1395–404
- Vollenweider R A, Munawar M and Stadelmann P 1974 A comparative review of phytoplankton and primary production in the Laurentian Great Lakes *J. Fish. Res. Board Can.* **31** 739–62
- Vuorio K, Mäki A, Salmi P, Aalto S L and Tirola M 2020 Consistency of targeted metatranscriptomics and morphological characterization of phytoplankton communities *Front. Microbiol.* **11** 96
- Wagner C and Adrian R 2009 Cyanobacteria dominance: quantifying the effects of climate change *Limnol. Oceanogr.* **54** 2460–8
- Watson S B, McCauley E and Downing J A 1997 Patterns in phytoplankton taxonomic composition across temperate lakes of differing nutrient status *Limnol. Oceanogr.* **42** 487–95
- Wehr J D, Sheath R G and Kocielek J P 2015 *Freshwater Algae of North America: Ecology and Classification* (Amsterdam: Elsevier)
- Weiss R F 1974 Carbon dioxide in water and seawater: the solubility of a non-ideal gas *Mar. Chem.* **2** 203–15
- Wiedner C and Nixdorf B 1998 Success of chrysophytes, cryptophytes and dinoflagellates over blue-greens (cyanobacteria) during an extreme winter (1995/96) in eutrophic shallow lakes *Hydrobiologia* **369** 229–35
- Willig M R, Kaufman D M and Stevens R D 2003 Latitudinal gradients of biodiversity: pattern, process, scale, and synthesis *Annu. Rev. Ecol. Evol. Syst.* **34** 273–309
- Winder M, Reuter J E and Schladow S G 2009 Lake warming favours small-sized planktonic diatom species *Proc. R. Soc. B* **276** 427–35
- Winder M and Sommer U 2012 Phytoplankton response to a changing climate *Hydrobiologia* **698** 5–16
- Zhang M, Shi X, Yang Z, Yu Y, Shi L and Qin B 2018 Long-term dynamics and drivers of phytoplankton biomass in eutrophic Lake Taihu *Sci. Total Environ.* **645** 876–86
- Zimba P V, Huang I-S, Gutierrez D, Shin W, Bennett M S and Triemer R E 2017 Euglenophycin is produced in at least six species of euglenoid algae and six of seven strains of *Euglena Sanguinea* *Harmful Algae* **63** 79–84