

## ORIGINAL RESEARCH

# Lake regionalization and diatom metacommunity structuring in tropical South America

Xavier Benito<sup>1</sup>  | Sherilyn C. Fritz<sup>1</sup> | Miriam Steinitz-Kannan<sup>2</sup> |  
Maria I. Vélez<sup>3</sup> | Michael M. McGlue<sup>4</sup>

<sup>1</sup>Department of Earth and Atmospheric Sciences and School of Biological Sciences, University of Nebraska–Lincoln, Lincoln, Nebraska

<sup>2</sup>Department of Biological Sciences, Northern Kentucky University, Highland Heights, Kentucky

<sup>3</sup>Department of Geology, University of Regina, Regina, SK, Canada

<sup>4</sup>Department of Earth and Environmental Sciences, University of Kentucky, Lexington, Kentucky

**Correspondence**

Xavier Benito, National Socio-Environmental Synthesis Center (SESYNC), University of Maryland, Annapolis, MD, USA.  
Email: xbenitogranel@sesync.org

**Funding information**

Funded by NSF EAR-1338694, NSF-EAR1251678, NASA 15-BIODIV15-0013, and National Geographic #8672-09 grants to SF. The Colombia diatom database collection was funded from NWO-WOTRO grant WB 75368 to H. Hooghiemstra and MV, and Inter-American Institute for Global Change (IAI grant No. CRN3038)-NSF (grant GEO-1128040) to MV. Lowland sampling in Brazil, Argentina and parts of Bolivia was funded by National Geographic #9797-15 and NSF EAR 1541247 to MM. Ecuador sampling was supported by NSF Grants BSR-8021539 and BSR-8202658 to Paul Colinvaux and MSK. MSK thanks Michael C. Miller, Mark Nienaber and many NKU undergraduates who helped put the Ecuador diatom database together.

**Abstract**

Lakes and their topological distribution across Earth's surface impose ecological and evolutionary constraints on aquatic metacommunities. In this study, we group similar lake ecosystems as metacommunity units influencing diatom community structure. We assembled a database of 195 lakes from the tropical Andes and adjacent lowlands (8°N–30°S and 58–79°W) with associated environmental predictors to examine diatom metacommunity patterns at two different levels: taxon and functional (deconstructed species matrix by ecological guilds). We also derived spatial variables that inherently assessed the relative role of dispersal. Using complementary multivariate statistical techniques (principal component analysis, cluster analysis, nonmetric multidimensional scaling, Procrustes, variance partitioning), we examined diatom–environment relationships among different lake habitats (sediment surface, periphyton, and plankton) and partitioned community variation to evaluate the influence of niche- and dispersal-based assembly processes in diatom metacommunity structure across lake clusters. The results showed a significant association between geographic clusters of lakes based on gradients of climate and landscape configuration and diatom assemblages. Six lake clusters distributed along a latitudinal gradient were identified as functional metacommunity units for diatom communities. Variance partitioning revealed that dispersal mechanisms were a major contributor to diatom metacommunity structure, but in a highly context-dependent fashion across lake clusters. In the Andean Altiplano and adjacent lowlands of Bolivia, diatom metacommunities are niche assembled but constrained by either dispersal limitation or mass effects, resulting from area, environmental heterogeneity, and ecological guild relationships. Topographic heterogeneity played an important role in structuring planktic diatom metacommunities. We emphasize the value of a guild-based metacommunity model linked to dispersal for elucidating mechanisms underlying latitudinal gradients in distribution. Our findings reveal the importance of shifts in ecological drivers across climatic and physiographically distinct lake clusters, providing a basis for comparison of broad-scale community gradients in lake-rich regions elsewhere. This may

help guide future research to explore evolutionary constraints on the rich Neotropical benthic diatom species pool.

#### KEYWORDS

diatom guilds, lakes, latitudinal gradient, metacommunity, topographic heterogeneity

## 1 | INTRODUCTION

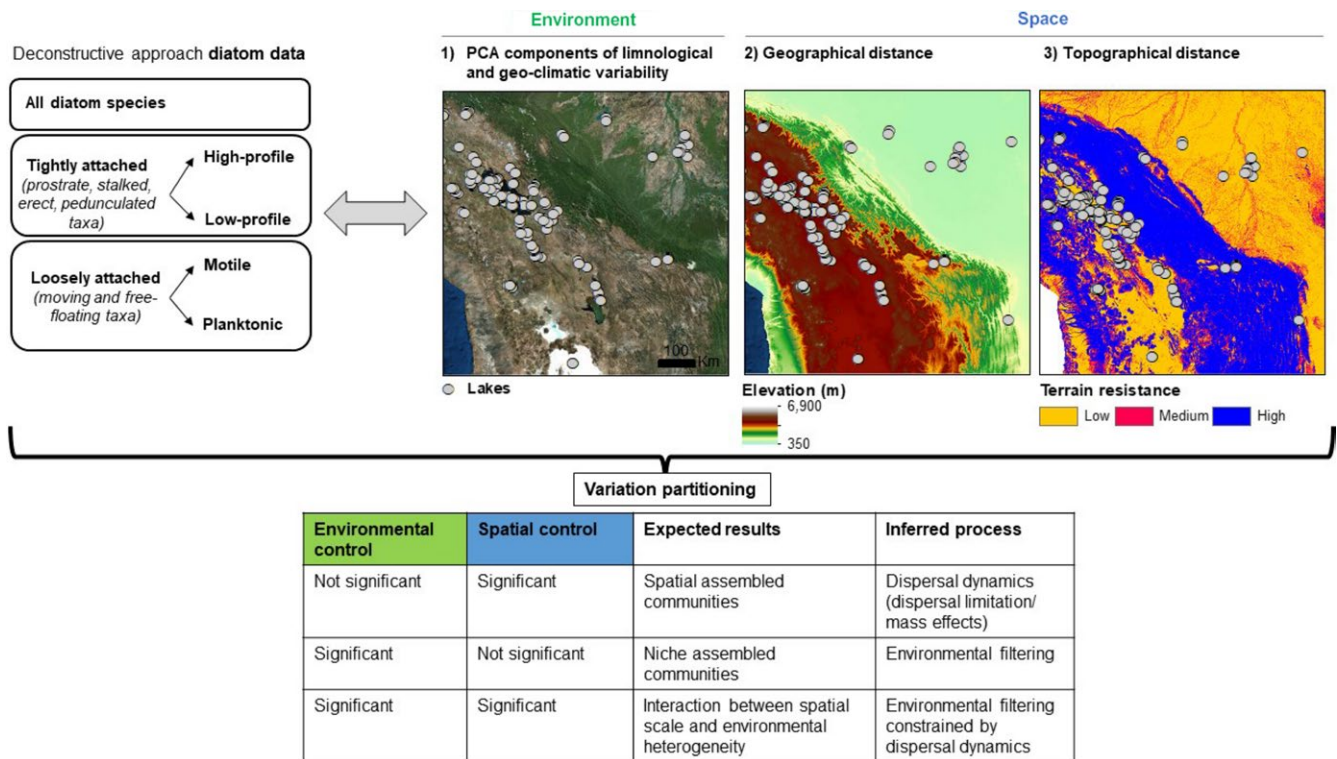
Lakes are one focal point in biogeography and community ecology, because they cover a small proportion of the Earth's surface but make disproportionate contributions to regional biodiversity, global biogeochemical cycles, and ecosystem services (Vitousek, Mooney, Lubchenco, & Melillo, 1997). The overlap between biogeography and community ecology is rapidly expanding, particularly for questions that examine the effects of spatiotemporal gradients and the influences of historical legacies (dispersal) on contemporary regional and local biotas (Jenkins & Ricklefs, 2011; Verleyen et al., 2009; Vyverman et al., 2007). In this context, metacommunity theory, which recognizes a set of local ecological communities that are connected by the dispersal of potentially interacting species, provides a flexible framework to integrate both disciplines, because it considers both regional (e.g., dispersal, climate) and local (e.g., biotic interactions, limnology) drivers of species composition (Gonçalves-Souza, Romero, & Cottenie, 2014; Leibold et al., 2004; Soininen, Jamoneau, Rosebery, & Passy, 2016; Viana et al., 2014). Combining biogeography and metacommunity theory is timely in light of global environmental change, because impacts from climate change (global processes) to habitat fragmentation and pollution (local processes) will likely affect lake ecosystems at intermediate spatial scales (Jenkins & Ricklefs, 2011).

Implicitly, metacommunity theory focuses on two nonexclusive paradigms to explain the composition of ecological communities: species sorting, with an emphasis on species autoecology and environmental gradients as the major influences (niche assembly rules), and dispersal (Heino et al., 2015). Species sorting requires a moderate dispersal rate to allow species to be sorted into their most suitable environmental habitats. When dispersal rates are high, communities are homogenized irrespective of environmental conditions via mass effects, thereby obscuring species sorting, particularly at small spatial scales. In contrast, low dispersal rates hinder a species ability to effectively track suitable environmental conditions, resulting in dispersal limited communities, especially at large spatial extents (Heino et al., 2015). Recent metacommunity studies propose that ecological drivers of community structure should fit within a framework defined by environmental heterogeneity, spatial scale, and the innate dispersal abilities of the organisms themselves (Brown, Sokol, Skelton, & Tornwall, 2017; Passy, 2017).

For organisms that disperse widely and passively, like microalgae or fungi, few studies have used a deconstructive approach that splits community data into different functional groups as surrogates for dispersal ability (Bie et al., 2012). Diatoms (unicellular siliceous algae)

are useful organisms to study, because they are a very species-rich group of algae, disperse widely, and have different growth morphologies for resource use and to resist physical disturbances (Passy, 2017). Passy (2007) and Rimet and Bouchez, (2011) classified diatom species into four ecological guilds: high-profile, low-profile, motile, and planktic, with all guilds supposedly connected to their dispersal abilities (Wetzel et al., 2012). Recent findings indicate that environmental and spatial mechanisms that structure species composition differ across guilds in lotic diatoms (Dong et al., 2016; Liu, Soininen, Han, & Declerck, 2013; Soininen et al., 2016; Wetzel et al., 2012), yet little evidence exists for process that structure diatom species composition in lakes. For instance, Vilmi, Tolonen, Karjalainen, and Heino (2017) hypothesized that diatom species tightly attached to the substratum, typically corresponding to the high-profile and low-profile guilds, are less likely to be dispersed than species that can move along a substrate or are floating in the water column, such as the motile and planktic guilds, respectively. However, no previous study has analyzed diatom metacommunity structuring across gradients of climate, physiography, and limnology in the tropics, and certainly not using a trait-based approach. Thus, we apply this approach to lakes of tropical South America, where very diverse benthic and planktic floras and many lake-rich regions are present (Rumrich, Lange-Bertalot, & Rumrich, 2000). Besides the strong provinciality observed in the Southern Hemisphere diatom flora (Vyverman et al., 2007), tropical lakes have neither been previously compared nor analyzed to explain possible mechanisms behind diversity gradients.

From a landscape perspective, lake ecosystems are units defined by the surrounding catchment, including topography, land cover, bedrock geology, and climate. Lakes are rarely isolated on the landscape; clusters or lake districts are the rule, not the exception (Catalan, Curtis, & Kernan, 2009). This makes lakes ideal systems to test macroecological effects of local and regional processes on assemblage composition (i.e., "biogeographic islands," Colinvax & Steinitz-Kannan, 1980). Disentangling the relative effects of pure environmental versus spatial factors on community species composition using empirical data is generally challenging, partly because of the presence of many direct and indirect connections among landscape processes operating at different spatial and temporal scales (Logue, Mouquet, Peter, & Hillebrand, 2011). Moreover, from the high mountains to lowlands, a diversity of conditions associated with topography (e.g., slope, aspect etc.) and continuum of aquatic systems impose a variety of dispersal pathways in organisms (e.g., geographic, network, landscape resistance) (Moritz et al., 2013). Thus, the ecological consequences of lake districts for metacommunities are not well understood (Catalan & Donato Rondon, 2016).



**FIGURE 1** Study design and working hypothesis to examine lacustrine diatom metacommunity patterns in tropical South America. Diatom data were analyzed using taxon and functional approaches (deconstructed species matrix by ecological guilds). We used three sets of predictors that represent environment, geographic, and topographic components to investigate the influence of niche and dispersal effects on diatom community structure. Variation partitioning analysis was used to quantify pure and shared proportions of variation on community composition explained by the three set of predictors. We summarized inferred processes according to variation partitioning results and the expected results of environmental and spatial controls

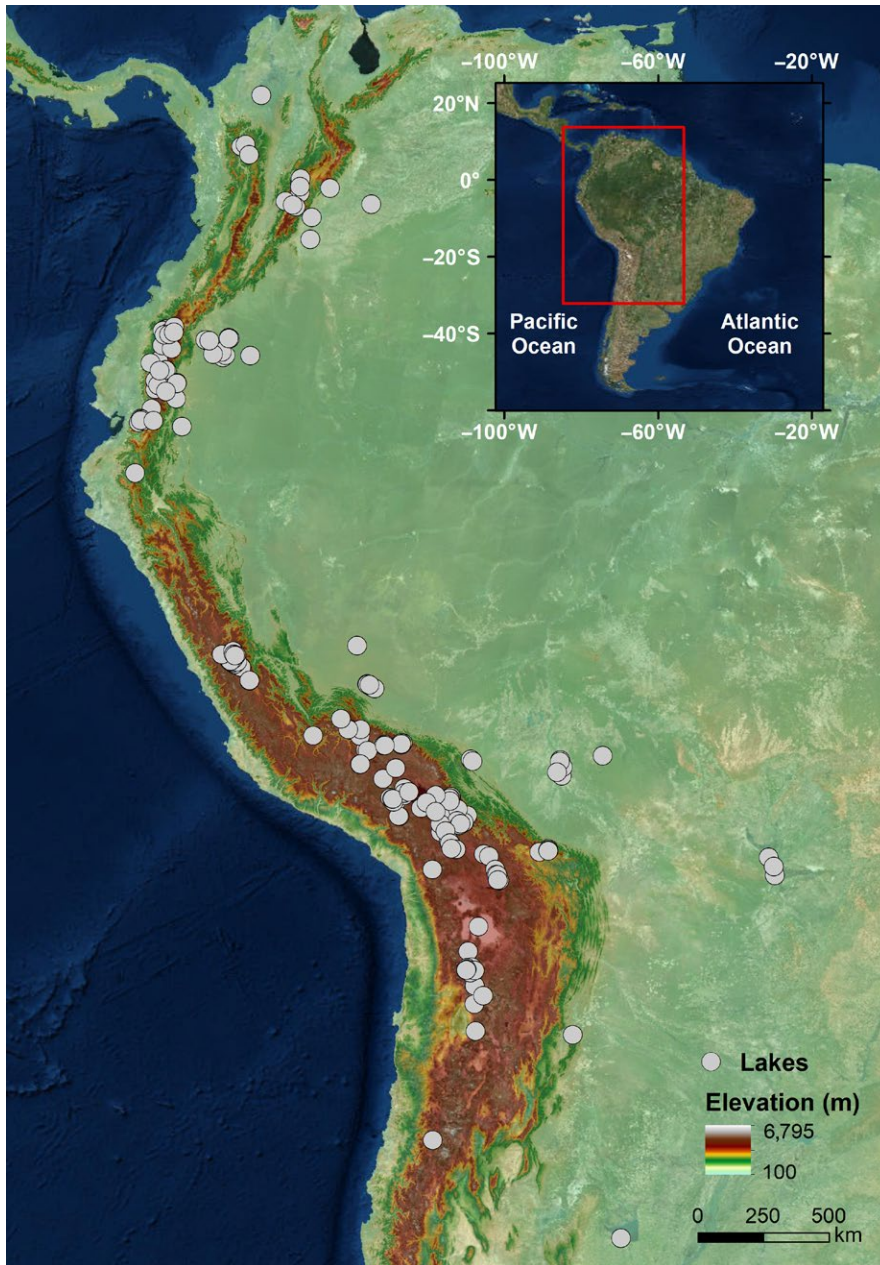
The overarching aim of this study is to classify groups of tropical lakes as functional units to examine diatom metacommunity structuring. First, we examined the extent to which limnological and geo-climatic environmental factors typify groups of lakes and how such classification correlates with diatom distribution. Second, we analyzed the relative importance of environmental and spatial structuring of diatom assemblages for each lake cluster individually. Third, we tested whether environmental and/or spatial processes differed between diatom guilds and lake clusters. Given the link between life forms and dispersal abilities of organisms in metacommunity structuring, we hypothesize that diatom guilds tightly attached to the substratum (high-profile and low-profile, “weak dispersers”) are strongly affected by spatial variables, whereas guilds that glide along substrates or are free floating (motile and planktic, “strong dispersers”) are strongly influenced by environmental variables (e.g., water chemistry, climate) (Figure 1).

## 2 | MATERIAL AND METHODS

### 2.1 | Study area

The study lakes are distributed across the tropical Andes and the Andean foreland plains (Figure 2), covering an altitudinal gradient from 220 to 5,070 m a.s.l. between 8°N–30°S and 58–79°W. Our

lake database encompasses a wide range of physiographic and climatic settings that produce diverse limnological conditions. Lowland sites are primarily lakes that occupy old river channels and floodplain wetlands, spread across Ecuador, Perú, Bolivia, and Brazil. In these lowland regions, some lake basins are seasonally connected to large rivers (e.g., Parana, Paraguay, Napo), which increases seasonal variability in limnological and hydrological conditions (McGlue et al., 2011). Colombia’s lowland lakes are distributed from the eastern savannas to the very wet western rain forests (Vélez, Wille, Hooghiemstral, & Metcalfe, 2005). Along the Andean cordillera, lakes occur at a range of high elevations (3,000–5,070 m a.s.l.) and lie in closed (endorheic) basins. The tropical Andes shows a predominant north–south landscape gradient with varied topographic heterogeneity that influences both local and regional climates (Valencia et al., 2016). Northern Andean lakes in Ecuador and Colombia lie in montane forests, inter-Andean valleys, and Páramo ecosystems. In the central Andean Cordillera of Perú and Bolivia, most of the study lakes are closed basins of glacial origin dominated by montane grass and shrubland. In the Altiplano plateau (central Andes), the northern region is characterized by cold and relatively humid conditions. Lakes are mainly freshwater and lie in extensive interconnected hinterland basins (Cohen et al., 2014). The southern Altiplano is drier, and most lakes are isolated and saline due to the basin geology and high evaporation rates (Sylvestre, Servant-Vildary, & Roux, 2001).



**FIGURE 2** Map of tropical South America showing the location of the study lakes ( $n = 195$ ). See Supporting Information Appendix S1–Table S1.1 for detailed information about the lakes

## 2.2 | Diatom database

Diatom data included 195 lakes with a total of 303 samples (Supporting Information Appendix S1–Table S1.1). Diatom samples comprised sediment surface, periphyton, and plankton, to ensure that the majority of diatom ecological guilds were collected. The samples were collected in the period 1977–2009 and fall well within the temporal window of the analyzed climatic variables for comparison. Diatom samples were analyzed separately for habitat type and sampling year, except for the Ecuador dataset, in which samples were composited. Preliminary analyses showed that little variation was seen between sampling dates. Diatom samples were cleaned using 30%  $H_2O_2$  and 37% HCl to remove organic material and carbonates, respectively. Cleaned diatoms were mounted in Naphrax (refractive index 1.74). All identifications were made to

the species level when possible, using South American diatom floras and regional studies (Manguin & Manguin, 1964; Metzeltin & Lange-Bertalot, 2007; Rumrich et al., 2000; Servant-Vildary, 1986) and taxonomic resources available at the diatom herbarium of the Philadelphia Academy of Natural Sciences. In about 90% of the samples, species relative abundance was enumerated by counting >300 diatom valves; the species data of the remaining 10% of the samples consisted of presence–absence counts. To better account for differences in the sampling methodology and counting methods, site-by-species abundance matrices were transformed to site-by-species presence–absence matrices prior to all analyses. Measures to ensure taxonomic consistency included aggregating varieties of species, scanning the data for taxonomic synonyms, and lumping species complexes (e.g., *Achnantheidium minutissimum*, *Sellaphora pupula*, *Discostella stelligera*, *Ulnaria ulna*) from the database entries.

Taxonomic harmonization was carried out to update diatom taxonomic nomenclature based on the most up-to-date classification (Guiry & Guiry, 2017).

### 2.3 | Ecological guilds

The diatom species matrix was split into ecological guilds to model the dispersal potential of diatoms taxa following Passy (2007) and Rimet and Bouchez (2011). The four guilds are high-profile, low-profile, motile, and planktic. The high- and low-profile guilds were considered to be “weak dispersers,” which includes species tightly attached to the substrate, such as adnate, prostrate, stalked, erect, and pedunculated taxa, such as *Achnanthes*, *Gomphonema*, or *Cocconeis*. High-profile and low-profile species are adapted to a certain degree to physical disturbances by growing closely to the substrate; high-profile species are adapted to high nutrient concentrations, and low-profile species are adapted to low nutrient concentrations. The motile and planktic guilds were considered to be “strong dispersers” and include species either loosely attached or with no obvious method of attachment, such as *Navicula*, *Nitzschia*, *Cyclostephanos*, and *Discostella*. All guild assignments were made at the genus level where possible (Supporting Information Appendix S3).

### 2.4 | Environmental and spatial variables

Different sets of explanatory variables were collected from several sources to characterize local and regional environmental gradients of the study lakes. Local environmental variables included limnological parameters that have been identified as important for lake diatoms in the tropical Andes in previous studies (Benito et al., 2018; Steinitz-Kannan, 1979; Sylvestre et al., 2001; Tapia, Fritz, Seltzer, Rodbell, & Metiever, 2006) and that were available for a large number of sites. These proximal (i.e., site-specific) environmental variables consisted of water temperature (°C), pH, and conductivity ( $\mu\text{S}/\text{cm}$ ). Regional environmental variables included geographical and climatic variables. A total of 11 geo-climatic variables were used, including: latitude/longitude, elevation (m), mean annual air temperature (MAT, °C), mean annual precipitation (MAP, mm), temperature seasonality (standard deviation; °C), precipitation seasonality (coefficient of variation; mm), % aquatic habitat, connectivity, Terrain Ruggedness Index (TRI), and lake area ( $\text{km}^2$ ). The STRM 90 m digital elevation model (Jarvis, Reuter, Nelson, & Guevara, 2008) was used to obtain elevation and calculate TRI values for each cell using the Raster Terrain Analysis Plugin in QGIS v.2.8.2 (QGIS Development Team, 2013). The TRI quantifies terrain heterogeneity (slope) by summarizing the change in elevation within a  $3 \times 3$  pixel cell grid (Riley, 1999). Climatic variables were obtained and extracted from the WorldClim 1.4 database. WorldClim contains averaged monthly climate data for the period 1950–2000 at a 1 km grid resolution (Hijmans, Cameron, Parra, Jones, & Jarvis, 2005). The Global Lakes and Wetlands Database (GLWD, Lehner & Döll, 2004) comprises lakes, rivers, and different wetland types with a surface area of  $>0.1 \text{ km}^2$  in the form of a global map of  $\sim 1$  km precision. Using the GLWD as a basemap,

equal grids of  $50 \text{ km}^2$  were created to extract the surface area occupied by freshwater in each grid as proxy of % aquatic habitat, and the density of water bodies in each grid as proxy of connectivity. The surface area of each lake was obtained by digitizing using the ESRI World Imagery layer as a basemap. All maps were manipulated using ArcGIS 10.4.1.

Two different distance matrices were calculated representing the physical distance (geography) and the resistance of the landscape to dispersal (topography) between pairs of sites to evaluate potential dispersal. Geographical distance was calculated using Euclidean distances, based on site coordinates. To obtain the distance matrix related to landscape resistance, TRI values for each cell were reclassified into three different levels of resistance by assigning values of 1 (low terrain rugosity), 50 (medium terrain rugosity), and 100 (high terrain rugosity) using ArcGis 10.4.1. These resistances were chosen assuming that lake basins with low rugosity are prone to landscape permeability, whereas less connectivity is expected in landscapes with complex topography (high terrain heterogeneity). Next, the resultant raster map was imported into the CIRCUITSCAPE program (McRae, 2006) to calculate pairwise landscape resistance distances to dispersal. CIRCUITSCAPE uses circuit theory to integrate dispersal through grid cells, allowing for multiple pathways between sites.

Spatial variables were generated for each of the two distance matrices (geographic and topographic) using distance-based Moran's Eigenvector Maps (Dray, Legendre, & Peres-Neto, 2006). These spatial db-MEM variables were obtained using principal coordinates of neighborhood matrix (PCNM). This method describes the spatial variability across study sites by generating eigenvector-based variables, which can be used as predictors in constrained ordination analysis. Only positive eigenvectors were employed as spatial predictors for posterior statistical analysis (variance partitioning). This analysis was performed using the *pcnm* function of the *vegan* package written in R (Oksanen et al., 2016; R Development Team, 2016).

### 2.5 | Statistical analyses

All statistical analyses were performed using the R software version 3.3.1 (R Development Team, 2016).

First, we summarized major patterns among limnological and geo-climatic variability to classify groups of lakes. Prior to running ordination analyses, all variables were transformed ( $\log_{10}[x + 1]$ ) to meet assumptions of linearity and homogeneity of variances (homoscedasticity). A principal component analysis (PCA) with the algorithm NIPALS (nonlinear estimation by iterative partial least squares) was performed using the *nipals* function of the *ade4* package (Dray & Dufour, 2007). The NIPALS algorithm allowed the computation of a PCA without deleting samples with missing data or estimating the missing values (Ibáñez et al., 2012). With this approach, we avoid eliminating sites with missing values, because of few widespread measurements of some limnologic variables (see Supporting Information Appendix S1—Table S1.2 for % data values). A previous study of the region, Benito et al. (2018) showed that water temperature and climatic variables (MAP, MAT) are correlated at the

spatial scale of the data used for this study. Also, other limnological variables (conductivity, pH, and nutrients) were outperformed by macroecological gradients associated with distinct climatic and topographic conditions, because local environmental conditions are temporally unstable compared with geo-climatic variables, such as elevation, catchment geology, and ecoregion in lakes of tropical Andes and adjacent lowlands (Benito et al., 2018). Considering these issues, the PCA axes act as composite variables of environmental drivers of lakes. The number of significant PCA axes, which indicate the nonrandom variability, were selected by broken stick model using the *evplot* function (Borcard, Gillet, & Legendre, 2011). In this case, we retained the first three orthogonal components explaining 72% of total variance (Supporting Information Appendix S4). Metrics for usefulness and sampling adequacy of the data for the PCA were obtained with the Kaiser–Meyer–Olkin index (KMO; critical value >0.70; Dziuban & Shirkey, 1974) and Bartlett's test of sphericity (Budaev, 2010), respectively. The PCA site scores of the first three axes were subsequently used as inputs for hierarchical cluster analysis based on Euclidean distances, with flexible beta as the linkage method using the *agnes* function of the *cluster* package (Maechle, 2012). Both methods (i.e., PCA and cluster analysis) are complementary and helped identify groups of lakes having similar environmental characteristics (cluster analysis) and provided information about the pattern of variation within and between groups in ordination space (PCA).

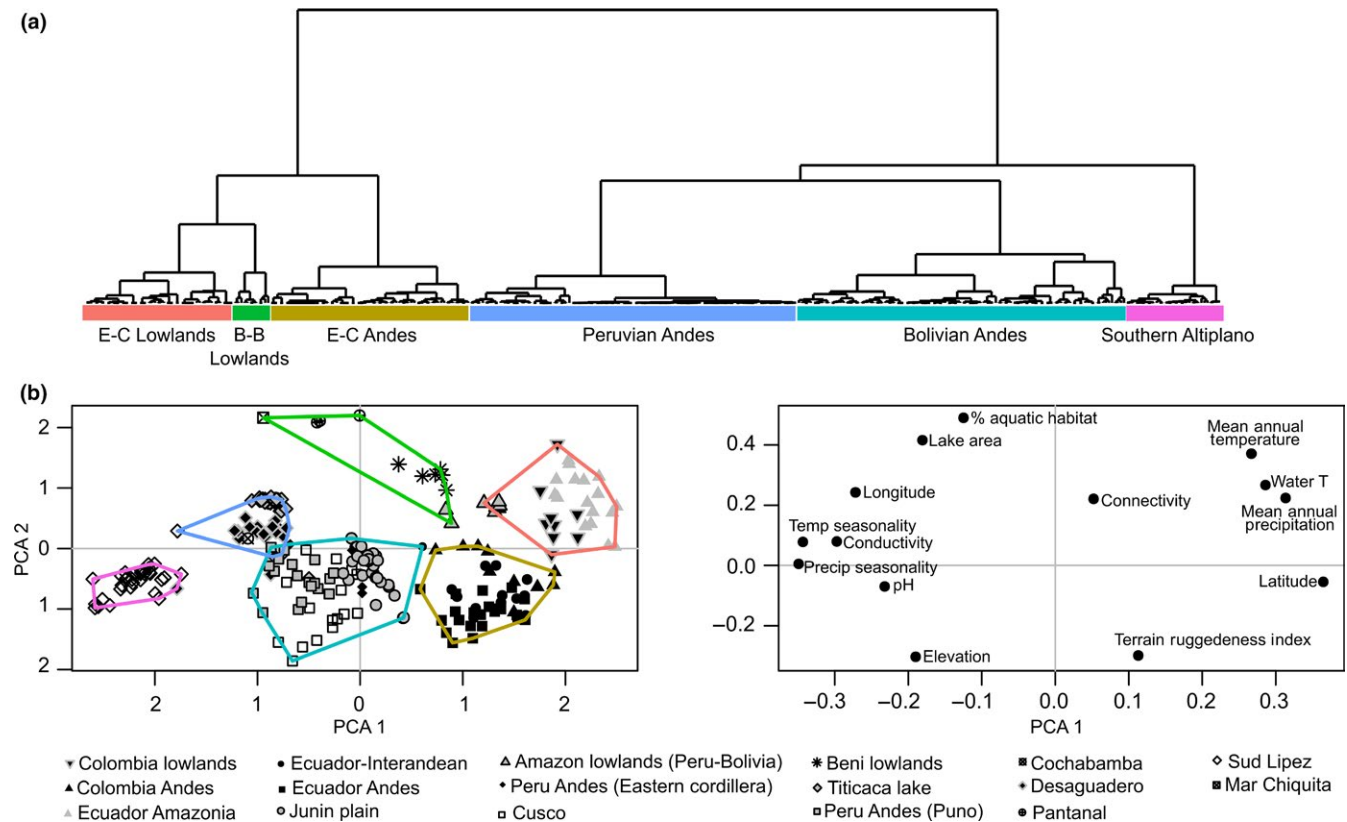
Second, we analyzed the major structure of the diatom data using a nonmetric multidimensional scaling (NMDS) with Chao distance measure on Hellinger transformed presence–absence matrix, in the *vegan* package. The Chao index was selected to account for unseen shared species and thus to reduce bias in sampling effort among study regions (Colwell, Mao, & Chang, 2004). All diatom taxa ( $n = 1,635$ ), including singletons, were included in the analyses. To aid interpretation of the NMDS axes without incorporating any environmental constraint, limnological and geo-climatic variables were fitted using the environmental fitting technique with the *envfit* function in *vegan*. Except for the Ecuador data set, additional NMDS analyses were performed for the sediment, periphyton, and plankton samples separately to assess variability of lake habitats and determine any relationship with environmental variables. If geo-climatic variables are the most important factors for structuring diatom habitat samples, the group division of lakes should be related to the regions, showing clear breakpoints in assemblage composition. In contrast, if limnological variables are more important, site groupings should not be related to the regions, and hence no clear breakpoints in assemblage composition should be observed (e.g., Heino, Soininen, Alahuhta, Lappalainen, & Virtanen, 2016). To account for the possible influence of temporal effects due to the time span of diatom collection, sampling year was included as a variable in the environmental fitting procedure.

Third, we characterized each group of lakes identified by PCA and cluster analysis in terms of their environmental heterogeneity and spatial extent following Tonkin, Death, Muotka, Astorga, and Lytle (2018). Environmental heterogeneity was estimated by

calculating the mean distance of each site to the corresponding group centroid using the *betadisper* function of the *vegan* package. For spatial extent, the function *ordihull* in *vegan* was used to enclose all sites that form each cluster, and the relative area was then estimated using the *polygon* function. Differences in environmental heterogeneity among clusters were tested using the *adonis2* function in *vegan* (which uses a permutational multivariate analysis of variance with 999 permutations) (PERMANOVA, Anderson, 2005).

Fourth, to explore the relationships among diatom community distribution and environmental variables and to test the significance of any correlation found, we used the Procrustes and analysis of variance following Lisboa et al. (2014). Procrustes is a correlative multivariate method that assesses species–environment relationships obtained from different unconstrained ordinations. The Procrustean analysis was performed between the NMDS (diatom data) and PCA (environmental data) matrices with the three-first components of each ordination using the *procrustes* and *protest* functions in *vegan*. The degree of concordance is given by the  $m^2$  statistic and associated  $p$  values with 999 permutation tests. Lower values of  $m^2$ , which vary from 0 to 1, indicate greater concordance (Peres-Neto & Jackson, 2001). If both ordinations (NMDS and PCA) are correlated, diatom distributions should be associated with lake clustering, suggesting regionally distinct communities. To further test this idea, we extracted the residual vector of association between diatom and environmental data, the so-called procrustean association metric (PAM), using the function *residuals*. Therefore, the statistical difference in PAMs across lake clusters was assessed through an analysis of variance (ANOVA) with the function *aov* at significance level  $\alpha = 0.05$ .

At last, the relative role of environmental and spatial components on diatom community composition was determined with redundancy analysis (RDA) and variance partitioning analysis. Variance partitioning quantifies pure and shared proportions of variation in community composition explained by different set of predictors (Peres-Neto, Legendre, Dray, & Borcard, 2006). We used eigenvector-related variables as predictors: the three-first PCA axes were used as variables related to niche-based factors but excluding spatial variables (latitude, longitude, connectivity and TRI), and the two sets of eigenvectors extracted from geographic and topographic distances using the PCNM analysis were used as spatial predictors. The diatom community structure was first regressed onto each set of predictors (environmental, geographic and topographic), including all eigenvectors using global RDA models individually, with the *rda* function in *vegan*, and tested for significance. If the global model was significant, a forward selection procedure was performed using the two stop criteria with the *ordiR2step* function of *vegan*. The forward-selected variables were posteriorly used for the variance partitioning as explanatory variables using the *varpart* function in *vegan*. The pure effects accounted for by the environmental, geographical, and topographical components were tested using the ANOVA function. We primarily based our results on the adjusted  $R^2$  values rather than significance alone, because we were interested in quantifying the effect sizes. Variance partitioning analyses were performed for the



**FIGURE 3** Combined approach of cluster analysis using principal component analysis (PCA) site scores as inputs (a) and PCA ordination with the analyzed limnological and geo-climatic variables (b). Bottom left: site scores labeled by regions. Bottom right: factor scores of the variables. B-B: Bolivian-Brazilian; E-C: Ecuadorian-Colombian

entire diatom species matrix and separately for each ecological guild across all lake clusters.

## 3 | RESULTS

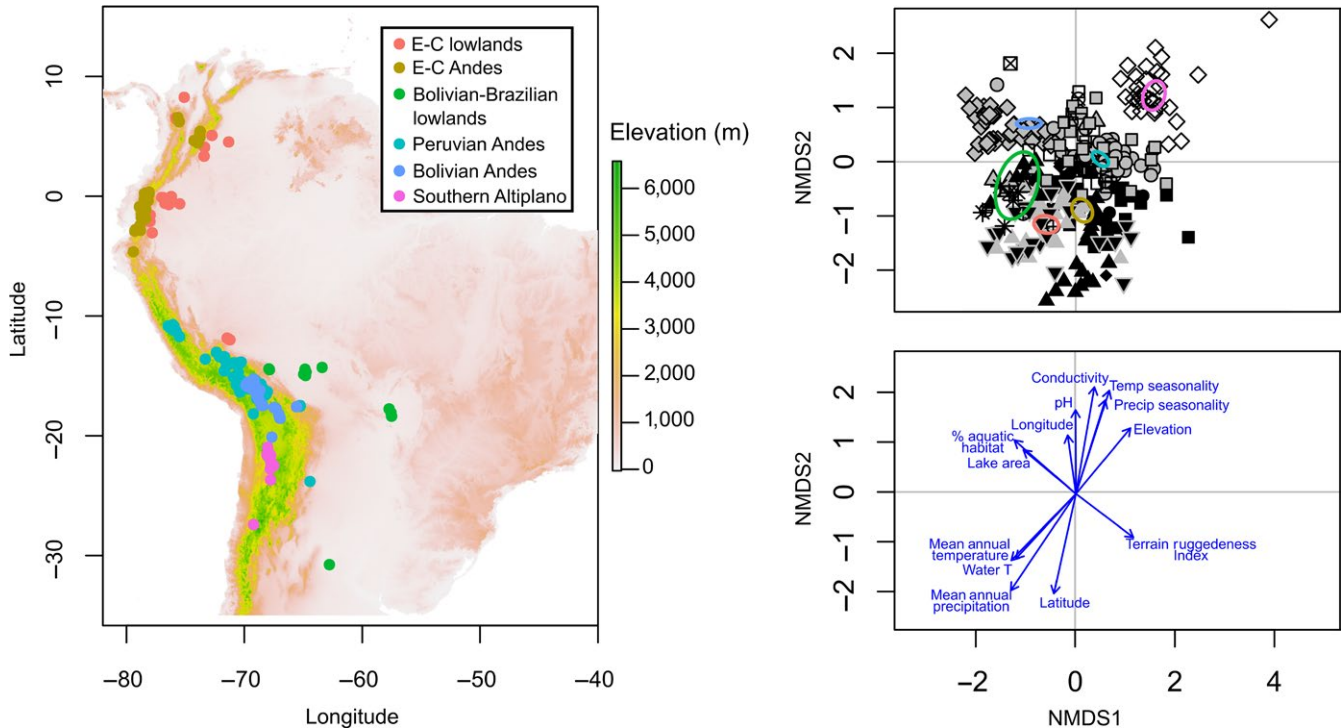
### 3.1 | Regionalization of lakes

The PCA adequately summarizes the limnological and geo-climatic variability of the data, as indicated by the KMO's measure of adequacy (0.71) and Bartlett's test of sphericity ( $p < 0.001$ ). According to the broken stick model, the amount of non-random variability is explained by the three-first principal components, explaining 46.8%, 16.6%, and 10.2%, respectively (Supporting Information Appendix S4); the PCA ordination plot illustrates axes 1 and 2 for graphical purposes (Figure 3b). The first PCA axis is associated with climatic and limnological variability; positive scores occur for MAT, MAP, and water temperature, whereas negative scores were associated with seasonality in precipitation and temperature, pH, and conductivity. The first component arranged lakes from the Amazonian lowlands of Ecuador and Colombia to the central Andes of Bolivia. PCA axis 2 was associated with a lake's landscape configuration (Figure 3b); positive loadings occur for % aquatic habitat, lake area, connectivity, and low elevation, whereas negative scores occurred for high elevation, high terrain heterogeneity, and more isolated conditions (lower % aquatic habitat and connectivity). Thus, the second component separated all lowland sites from the high-elevation Andean lakes.

Cluster analyses resulted in six lake groups that are arranged along the first PCA axis and exhibit a clear latitudinal structure (Figure 3a and Supporting Information Appendix S2). Environmental heterogeneity differed among the clusters (PERMANOVA:  $F = 40.537$ ;  $p < 0.01$ ; Supporting Information Appendix S2—Figure S2.2). Clusters 1 and 2 comprised all lakes near the equator; cluster 1 grouped all lowland lakes (Ecuadorian-Colombian lowlands, Figure 3a), and cluster 2 grouped Andean lakes from Ecuador and Colombia (Ecuadorian-Colombian Andes, Figure 3a). Cluster 3 was composed of lakes in the Amazonian lowlands of Bolivia and the Pantanal in Brazil (Bolivian-Brazilian lowlands, Figure 3a). This cluster showed the highest environmental heterogeneity (Supporting Information Appendix S2—Figure S2.2). Clusters 4, 5, and 6 consisted of high-elevation Andean lakes across Perú and Bolivia. The Peruvian Andes cluster (cluster 4, Figure 3a) showed a combination of the greatest spatial extent and wide environmental heterogeneity, whereas the Bolivian Andes (cluster 5, Figure 1a) and Southern Altiplano clusters (cluster 6, Figure 3a) had the lowest environmental heterogeneity and spatial extents (Supporting Information Appendix S2—Figure S2).

### 3.2 | Diatom metacommunities

Diatom species composition of sediment surface and periphyton samples showed clear regional differences in the NMDS ordination,



**FIGURE 4** Results of nonmetric multidimensional scaling (NMDS) with Chao distance measure on Hellinger transformed presence-absence diatom data for the entire species matrix (2D stress = 0.20). Left: Lake distribution labeled by clusters. Top right: site scores labeled by regions (see Figure 3 for study region labels); ellipses represent 95% confidence level of each lake's cluster centroid identified through PCA and cluster analyses. Bottom right: environmental variable fitting showing the selected variables; the length of each vector is proportional to the correlation between variables and NMDS axes. E-C: Ecuadorian-Colombian

suggesting that species composition variability is mainly related to among-region differences (Supporting Information Appendix S5—Figure S5.1–2). In contrast, plankton samples showed less clear separation between regions, suggesting a higher effect of local environmental variables (Supporting Information Appendix S5—Figure S5.3). Based on  $R^2$  values, geo-climatic variables had the strongest ( $R^2 > 0.60$ ,  $p < 0.05$ ) relationship on surface sediment and periphyton species composition, whereas conductivity and pH had the strongest ( $R^2 > 0.65$ ,  $p < 0.05$ ) relationship on planktic species composition. The relationship between the variable “year” and diatom species composition was statistically significant for the sediment surface ( $R^2 > 0.16$ ;  $p < 0.05$ ), periphyton ( $R^2 > 0.68$ ;  $p < 0.05$ ), and plankton ( $R^2 > 0.44$ ;  $p < 0.05$ ) samples.

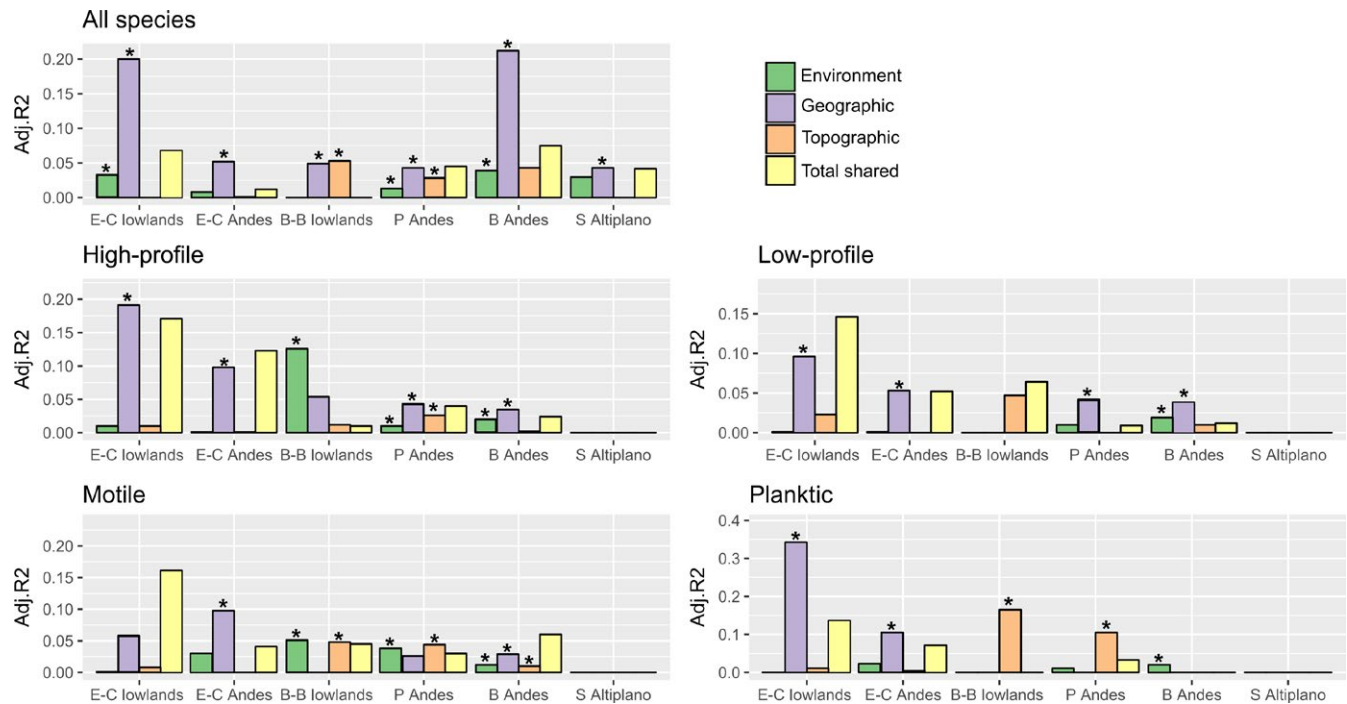
When considering the entire diatom species matrix, limnological and geo-climatic variables, except for connectivity, had significant relationships with diatom data ( $R^2 > 0.20$ ;  $p < 0.05$ ), as indicated by the NMDS with environmental fitting (Figure 4). The significant strength of concordance between the NMDS and PCA ordinations (Protest  $m^2$  statistic = 0.58;  $p < 0.001$ ) and differences across lake clusters in diatom–environment relationships (ANOVA's PAM:  $F = 122.3$ ;  $p < 0.01$ ) supported the cluster division of lakes with distinct diatom communities associated with environmental characteristics of groups of lakes (Figure 4).

The first NMDS axis arranged diatom metacommunities from the lowlands (Ecuadorian-Colombian lowlands and

Bolivian-Brazilian lowlands, Figure 4) to high-elevation Andes (Ecuadorian-Colombian Andes, Peruvian Andes, Bolivian Andes, and Southern Altiplano, Figure 4) following gradients of % aquatic habitat, lake area, and terrain heterogeneity; thus, NMDS axis 1 is primarily a measure of a lake's landscape configuration, similar to PCA axis 2. NMDS axis 2 separated diatom metacommunities from higher latitudes with greater seasonality in precipitation and temperature and higher pH and conductivity (Bolivian-Brazilian lowlands, Peruvian Andes, Bolivian Andes, and Southern Altiplano, Figure 4), from sites near the equator, where diatom metacommunities were associated with higher MAT and MAP and lower pH and conductivities (Ecuadorian-Colombian lowlands and Ecuadorian-Colombian Andes, Figure 5). NMDS axis 2 is thus associated with gradients of climatic and limnological variability, as is PCA axis 1. The variable “year” was significantly correlated with NMDS axis 2 scores (Pearson's  $r = 0.33$ ;  $n = 370$ ;  $p < 0.001$ ), but NMDS axis 1 scores were not significantly correlated with this variable (Pearson's  $r = -0.11$ ;  $n = 303$ ;  $p = 0.19$ ).

The RDA results with forward-selected variables are summarized in Supporting Information Appendix S6. When the whole diatom species matrix is used, variance partitioning revealed that the relative effects of the spatial component (geographic and topographic factors) outperformed the effect of the environmental component in all diatom metacommunities, and equally significant environmental and spatial effects were found in diatom metacommunities





**FIGURE 5** Variance partitioning results of environment, geographic, and topographic variables on all diatom species data and for each guild individually across the six lake clusters identified using PCA and cluster analyses. Asterisks denote the statistical significance ( $p < 0.05$ ) of the pure fractions of environment, geographic or topographic variables. Lake clusters are arranged from low to high latitude. B: Bolivia; B-B: Bolivian-Brazilian; E-C: Ecuadorian-Colombian; P: Perú; S: Southern. See Supporting Information Appendix S6—Table S6.1 and Table S6.2 for detailed RDA results and forwarded-selected variables

for certain high-latitude lake clusters (Bolivian-Brazilian lowlands, Peruvian Andes, Bolivian Andes and Southern Altiplano; Figure 5).

If diatom guilds are analyzed individually, varied effects of environmental and spatial components emerged across lake clusters: (a) for the high-profile and low-profile guilds (“weak dispersers”), the spatial component overrode the effect of environmental factors in most lake clusters and showed a decreasing trend of the explained variability with latitude (Figure 5); (b) for the motile and planktic guilds (“strong dispersers”), the environmental and spatial components were equally significantly associated with diatom meta-community structuring in certain lake clusters (Bolivian-Brazilian lowlands, Peruvian Andes and Bolivian Andes; Figure 5). This was particularly evident for the motile guild (Figure 5), especially for the Bolivian-Brazilian lowlands, but not for the planktic guild; and (c) for the motile and planktic guilds, the spatial effects on diatom community composition were associated with the topographic component only for certain high-latitude meta-communities (Bolivian-Brazilian lowlands and Peruvian Andes), whereas for the low-latitude meta-communities (Ecuadorian-Colombian lowlands and Andes), it was associated with the geographic component (Figure 5). The Southern Altiplano showed no environmental or spatial structure for any of the diatom guilds. Of all the diatom guilds, the unique fractions explained by environmental and spatial components were relatively small, and the total shared effect between environmental, geographic, and topographic factors was highest in low-latitude lake clusters (Figure 5; Supporting Information Appendix S6—Table S6.2).

## 4 | DISCUSSION

Our results indicate that lacustrine diatoms in tropical Andean mountains and the adjacent lowlands form ecologically meaningful clusters following gradients of local and regional environmental conditions. The PCA identified two independent driving factors in the formation of lake clusters, namely climatic and limnological variability (PCA1) and landscape configuration (PCA2), which resulted in six different geographically distinct clusters in the way that these environmental factors combine. Covering large environmental gradients and including different within-lake habitats necessarily crosses multiple species pools that show different response to regional and local environmental factors (Heino et al., 2016). From a diatom community perspective, planktic communities more closely track limnological conditions compared with sediment surface and periphyton communities, suggesting a stronger influence of regional-scale environmental gradients on benthic diatom species (Winogardner, Beisner, Legendre, & Gregory-Eaves, 2015). In addition, our study provides new evidence that regional diatom communities can be treated as unique meta-communities, because of the distribution of species in environmentally similar lake clusters, as indicated by the significant association between PCA and NMDS. From a diatom meta-community perspective, distinctive context-dependent processes emerged. We argue these processes originate from the relationships between spatial extent, environmental heterogeneity, and ecological guilds (as surrogate of dispersal abilities).

As expected, lake clusters differed in terms of spatial extent and environmental heterogeneity. These two factors are essential for distinguishing among species sorting, dispersal limitation, and mass effects in lake metacommunities (Heino et al., 2015; Vilmi et al., 2017). For instance, a higher environmental heterogeneity within-lake clusters would lead to stronger species sorting, likely because the clusters have a greater diversity of ecological gradients that can be occupied by species with different niches (Leibold et al., 2004). Yet, lake clusters that showed the highest environmental heterogeneity (e.g., Bolivian-Brazilian lowlands and Peruvian Andes) did not reveal stronger niche-based mechanisms relative to dispersal-based ones (Figure 5). Moreover, the positive relationship between environmental filtering and habitat heterogeneity might depend on spatial extent, because dispersal processes (i.e., dispersal limitation or mass effects) may reduce species sorting at either large or small extents (Zorzal-Almeida, Soininen, Bini, & Bicudo, 2017). In the Bolivian Andes and Southern Altiplano lake clusters, both spatial and environmental effects correlated with diatom species composition, indicating that diatom metacommunities are niche assembled but constrained either by dispersal limitation or mass effects, or both. While our results are correlative, and thus we cannot unequivocally provide causal evidence for distinguishing between dispersal limitation and mass effects, we associated the significant spatial role to mass effects due to the relatively small spatial extent of these two lake clusters (Supporting Information Appendix S2), according to the expectations of Heino et al. (2015) and Tonkin et al. (2016). In contrast, at the largest spatial extent (e.g., Peruvian Andes), dispersal limitation would be primary in structuring diatom metacommunities.

The structure of lake metacommunities has been often analyzed using exclusively local environmental and spatial effects (Declerck, Winter, Shurin, Suttle, & Matthews, 2013; Heino et al., 2016). Yet climatic effects have not been evaluated as thoroughly, although they may affect environmental filtering (Alahuhta & Heino, 2013; Loewen, 2017). Some latent environmental predictors (e.g., catchment productivity) are a function of climatic and several lake/catchment features (e.g., precipitation, topography) across large scale in tropical regions. For instance, Steinitz-Kannan, Colinvaux, and Kannan (1983) found that nutrient levels are related to altitude in Andean lakes of Ecuador. Benito et al. (2018) show that geoclimatic variables might be partially manifested via local limnological variables in Andean lakes of Peru and Ecuador. Here, we used PCA site scores to capture latent environmental variables and likely those local variables (e.g., nutrients) that are spatially patterned and outperformed by macroecological variables in most groups of our study lakes (Benito et al., 2018), and hence use environmental drivers of lakes' clustering to infer niche-based assembly processes (e.g., Steinitz-Kannan et al., 1983). Nonetheless, the relative influence of space over environmental factors, as indicated by variance partitioning (Figure 5), provided limited evidence for niche assembled diatom communities at a regional metacommunity scale. At the biogeographic scale (continental), we found that spatial factors outperformed environmental factors, as well (Benito et al., 2018), suggesting that dispersal-based processes predominantly control the

structuring of lake diatom assemblages in tropical South America at different spatial scales. Rather than being mutually exclusive, species sorting and dispersal dynamics may jointly drive diatom community composition with varying effects mediated by ecological guilds, as discussed below.

Evidence from studies of small spatial scales indicates that diatom guilds can not only track environmental gradients but also spatial factors due to the relationship between life forms and dispersal abilities (Dong et al., 2016; Liu et al., 2013; Riera, Magnuson, Kratz, & Webster, 2000; Vilmi et al., 2017). At continental spatial scales, previous studies demonstrated strong responses of diatom guild distributions in streams to both environmental and spatial factors (Passy, 2017; Soininen et al., 2016). Our results in groups of tropical lakes indicated that high-profile and low-profile guilds ("weak dispersers") were driven by spatial variables along a latitudinal gradient (in terms of uniquely explained variation by geographic component decreasing with latitude; Figure 5). Hence, the high-profile and low-profile guilds were less spatially structured in high-latitude lake clusters, which have more variable climatic conditions (higher seasonality) than equatorial latitudes. Climatic stability with latitude has been suggested to drive a dispersal-ecological specialization trade-off at metacommunity level (Zaharescu, Hooda, Burghlea, & Palanca-Soler, 2016). However, environmental variables can be spatially structured, thereby resulting in shared effects, as revealed in our study (Supporting Information Appendix S6—Table S6.2). These effects may lead to spurious interpretations of spatial effects as proxies of dispersal dynamics. Nonetheless, the total shared effect among environment, geographic, and topographic components also showed a clear decreasing latitudinal pattern (Figure 5). This suggests that prevalence of ecological guild variation in response to higher regionally structured environment toward the equator might explain the formation of the latitudinal diatom metacommunity gradient.

The detection of significant environmental effects on the motile guild supported the importance of niche-based processes in affecting loosely attached diatoms ("strong dispersers") (e.g. Jocque, Field, Brendonck, & Meester, 2010; Vilmi et al., 2017). This is particularly evident in the Bolivian-Brazilian lowlands, where the strength of the environmental component outperformed the effect of spatial factors on diatom community composition, likely due to the high environmental heterogeneity of tropical floodplain lakes, which is evident in the Pantanal lakes (McGlue et al., 2011). These hydrologically pulsing environments favor the dispersal of aquatic microorganisms among habitats (Cardoso et al., 2017; Devercelli, Scarabotti, Mayora, Schneider, & Giri, 2016; Dias et al., 2016). Although local environmental variables appeared to exert a major control on planktic communities (Supporting Information Appendix S5—Figure S5.3), variance partitioning did not provide support for niche-based assembly processes for the planktic guild among-lake clusters, in which pure spatial controls were mostly detected. Dispersal-related mechanisms in diatom metacommunities have also been suggested in mountainous areas where constrained aerial dispersed is likely due to high elevational gradients and step valleys (Dong et al., 2016;

Jamoneau, Passy, Soininen, Lebourer, & Tison-Rosebery, 2018). Nonetheless, we cannot exclude the possibility that the nonsignificant environmental structure at metacommunity level is a product of the low abundance of planktic taxa in our database (10% of the total taxa), or the noninclusion of ecologically important processes for phytoplankton communities (e.g., biotic interactions, trophic state) (Nabou, Siqueira, Bini, & de Nogueira, 2009).

Landscape features mediate the importance of regional processes in shaping metacommunities by either promoting or limiting dispersal (Badgley et al., 2017). In our study, a significant proportion of community composition in all diatom guilds was explained by geographic distances. Similar results have been found by Zorzal-Almeida et al. (2017) in a set of tropical reservoirs. However, topographic constraints, such as mountain barriers, may yield more ecologically informative relationships among sites than straight-line distances (Dong et al., 2016). Topographic distances had only a significant relationship with diatom guilds loosely attached to the substrate (motile and planktic) in certain lake clusters (Bolivian-Brazilian lowlands and Peruvian Andes). These two regions mostly correspond to lake clusters characterized by a combination of rugged topography and relatively low % aquatic habitat (Supporting Information Appendix S2), suggesting that complex topography in isolated lake systems indeed exerts constraints for motile and planktic guild distribution. Unexpectedly, in the topographically complex lake clusters of the Andes and lowlands of Ecuador and Colombia, spatial effects due to rugged topography did not appear to exert influence on the community composition of these two diatom guilds, as indicated by the stronger influence of geographical distance. Subtle differences in topographic heterogeneity and hydrological connections between lakes through riverine network (not accounted for in this study) might explain the varied spatial control in diatom guilds. These differences may further contribute to the development of regionally distinct diatom metacommunities in tropical South America.

Our results indicating shifts in ecological drivers on diatom metacommunity structuring across climatic and physiographically distinct lake regions would not have been detected without a functional approach. These findings are in accord with those of Vilmi et al. (2017) in lakes and Jamoneau et al. (2018) in streams, who found different responses to environmental and spatial factors when diatom ecological guilds are analyzed individually, rather than entire assemblages. Nonetheless, it is hard to conclude that large-scale ecological patterns can be explained by environmental and/or spatial variables, given the complex spatial-temporal mechanisms of community assembly for a certain metacommunity (Brown et al., 2017; Padiol et al., 2014). For the present study, two independent lines of evidence supported this statement. First, percentages of explained variation among models for the whole diatom community and for each ecological guild separately were low (adjusted  $R^2 = 1\%$ – $20\%$ ). Among others, this can be attributed to intraguild interactions that alter fundamental ecological niches via competition-facilitation effects (Passy, 2017), the noninclusion of explicit measures of nutrients, which are a major driver of diatom community composition over large-scale

surveys (Soininen et al., 2016; Verleyen et al., 2009; Vyverman et al., 2007; Winegardner et al., 2015), and methodological difficulties in measuring multiple turnovers in community composition (Heino et al., 2016). Although such low values are comparable with other large-scale studies that analyze aquatic metacommunities (adjusted  $R^2 = 1\%$ – $35\%$ ) (Bie et al., 2012; Hájek et al., 2011; Heino et al., 2016; Padiol et al., 2014; Pandit, Kolasa, & Cottenie, 2009; Soininen et al., 2016; Zorzal-Almeida et al., 2017), our study indicates that tropical lake diatom metacommunities are highly dynamic and that any inference from one-off snapshot sampling may be misleading. This is shown by significant temporal effects found in the diatom data from the analyses of sediment surface, periphyton, and plankton communities, as well as the entire diatom species matrix. Second, the effects of environmental and spatial variables were not significant in the Southern Altiplano metacommunity (Figure 5). Previous studies in lentic systems suggest that null results can result from unmeasured fine-scale environmental variability (Nabou et al., 2009). This fact, coupled with the extreme environmental conditions found in the shallow hypersaline lakes ("salares") of the Southern Altiplano (Sylvestre et al., 2001), could further explain these null results.

Our study revealed broadly similar patterns to those that were generated from the analyses of diatom communities when considering all the species versus separating species from sediment surface and periphyton habitats. This suggests that the rich benthic species pool closely tracks different environmental and spatial influences on lacustrine community composition, and thus reinforces the guild-specific metacommunity model to account for biogeographic variation (i.e., latitudinal gradient). Combining two disciplines at disparate spatial scales (metacommunity and biogeography) might help guide future diatom research, for example, in exploring the evolutionary origin of the marked interhemispheric differences in species composition in response to past tectonic and climatic events and their role in the strong latitudinal gradient of diatom diversity (Verleyen et al., 2009).

## ACKNOWLEDGMENTS

The authors are grateful to colleagues who provided samples and published and unpublished diatom and associated environmental data for assembling the database: P. Tapia, N. Michelutti, W. Gosling, F. Mayle, B. Valencia, and J.P. Bradbury. Funded by NSF EAR-1338694, NSF-EAR1251678, NASA 15-BIODIV15-0013, and National Geographic #8672-09 grants to SF. The Colombia diatom database collection was funded from NWO-WOTRO grant WB 75368 to H. Hooghiemstra and MV, and Inter-American Institute for Global Change (IAI grant No. CRN3038)-NSF (grant GEO-1128040) to MV. Lowland sampling in Brazil, Argentina and parts of Bolivia was funded by National Geographic #9797-15 and NSF EAR 1541247 to MM. Ecuador sampling was supported by NSF Grants BSR-8021539 and BSR-8202658 to Paul Colinvaux and MSK. MSK thanks Michael C. Miller, Mark Nienaber and many NKU undergraduates who helped put the Ecuador diatom database together. The authors acknowledge

constructive comments received from two anonymous referees and the associate editor, which improved the manuscript substantially.

## CONFLICT OF INTEREST

None declared.

## AUTHOR CONTRIBUTIONS

XB and SF conceived the ideas; XB, SF, MK, MV, and MM analyzed and processed individual diatom data sets; XB analyzed the data and lead the writing of the manuscript with contributions of all the authors.

## DATA ACCESSIBILITY

Diatom community data, water chemistry, and geo-climatic variables of lakes are available at Dryad Digital Repository <http://doi.org/doi:10.5061/dryad.6jk8h77>.

## ORCID

Xavier Benito  <http://orcid.org/0000-0003-0792-2625>

## REFERENCES

- Alahuhta, J., & Heino, J. (2013). Spatial extent, regional specificity and metacommunity structuring in lake macrophytes. *Journal of Biogeography*, *40*, 1572–1582. <https://doi.org/10.1111/jbi.12089>
- Anderson, M. J. (2005). *Permutational multivariate analysis of variance*. Auckland, New Zealand: Department of Statistics, University of Auckland. Retrieved from [http://img2.timg.co.il/forums/1\\_124959686.pdf](http://img2.timg.co.il/forums/1_124959686.pdf)
- Badgley, C., Smiley, T. M., Terry, R., Davis, E. B., DeSantis, L. R., Fox, D. L., ... Yanites, B. J. (2017). Biodiversity and topographic complexity: Modern and geohistorical perspectives. *Trends in Ecology & Evolution*, *32*, 211–226. <https://doi.org/10.1016/j.tree.2016.12.010>
- Benito, X., Fritz, S. C., Steinitz-Kannan, M., Tapia, P., Kelly, M. A., & Lowell, T. V. (2018). Geo-climatic factors drive diatom community distribution in tropical South American freshwaters. *Journal of Ecology*, *106*, 1660–1672. <https://doi.org/10.1111/1365-2745.12934>
- Bie, T., Meester, L., Brendonck, L., Martens, K., Goddeeris, B., Ercken, D., ... Declerck, S. A. (2012). Body size and dispersal mode as key traits determining metacommunity structure of aquatic organisms. Hillebrand H, editor. *Ecology Letters*, *15*, 740–747. <https://doi.org/10.1111/j.1461-0248.2012.01794.x>
- Borcard, D., Gillet, F., & Legendre, P. (2011). *Numerical ecology with R*. Springer Science & Business Media: Springer Science & Business Media. <https://doi.org/10.1007/978-1-4419-7976-6>
- Brown, B. L., Sokol, E. R., Skelton, J., & Tornwall, B. (2017). Making sense of metacommunities: Dispelling the mythology of a metacommunity typology. *Oecologia*, *183*, 643–652. <https://doi.org/10.1007/s00442-016-3792-1>
- Budaev, S. (2010). Using principal components and factor analysis in animal behaviour research: Caveats and guidelines. *Ethology*, *116*, 472–480. <https://doi.org/10.1111/j.1439-0310.2010.01758.x>
- Cardoso, S. J., Nabout, J. C., Farjalla, V. F., Lopes, P. M., Bozelli, R. L., Huszar, V. L., & Roland, F. (2017). Environmental factors driving phytoplankton taxonomic and functional diversity in Amazonian floodplain lakes. *Hydrobiologia*, *802*, 115–130. <https://doi.org/10.1007/s10750-017-3244-x>
- Catalan, J., Curtis, C. J., & Kernan, M. (2009). Remote European mountain lake ecosystems: Regionalisation and ecological status. *Freshwater Biology*, *54*, 2419–2432. <https://doi.org/10.1111/j.1365-2427.2009.02326.x>
- Catalan, J., & Donato Rondon, J. C. (2016). Perspectives for an integrated understanding of tropical and temperate high-mountain lakes. *Journal of Limnology*, *75*, 215–234.
- Cohen, A., McGlue, M. M., Ellis, G. S., Zani, H., Swarzenski, P. W., Assine, M. L., & Silva, A. (2014). Lake formation, characteristics, and evolution in retroarc deposystems: A synthesis of the modern Andean orogen and its associated basins. *Geological Society of America Memoirs*, *212*, 212–216.
- Colinvaux, P., & Steinitz-Kannan, M. (1980). Species richness and area in Galapagos and Andean lakes: Equilibrium phytoplankton communities and a paradox of the zooplankton. In University Press of New England, editor, *Evolution and ecology of zooplankton communities* (pp. 697–711).
- Colwell, R. K., Mao, C. X., & Chang, J. (2004). Interpolating, extrapolating, and comparing incidence-based species accumulation curves. *Ecology*, *85*, 2717–2727. <https://doi.org/10.1890/03-0557>
- Declerck, S. A., Winter, C., Shurin, J. B., Suttle, C. A., & Matthews, B. (2013). Effects of patch connectivity and heterogeneity on metacommunity structure of planktonic bacteria and viruses. *ISME Journal*, *7*, 533–542. <https://doi.org/10.1038/ismej.2012.138>
- Devercelli, M., Scarabotti, P., Mayora, G., Schneider, B., & Giri, F. (2016). Unravelling the role of determinism and stochasticity in structuring the phytoplanktonic metacommunity of the Paraná River floodplain. *Hydrobiologia*, *764*, 139–156. <https://doi.org/10.1007/s10750-015-2363-5>
- Dias, J. D., Simões, N. R., Meerhoff, M., Lansac-Tôha, F. A., Velho, L. F. M., & Bonecker, C. C. (2016). Hydrological dynamics drives zooplankton metacommunity structure in a Neotropical floodplain. *Hydrobiologia*, *781*, 109–125. <https://doi.org/10.1007/s10750-016-2827-2>
- Dong, X., Li, B., He, F., Gu, Y., Sun, M., Zhang, H., ... Cai, Q. (2016). Flow directionality, mountain barriers and functional traits determine diatom metacommunity structuring of high mountain streams. *Scientific Reports*, *6*, 24711. <https://doi.org/10.1038/srep24711>
- Dray, S., & Dufour, A. (2007). The ade4 package: Implementing the duality diagram for ecologists. *Journal of Statistical Software*, *22*, 1–20.
- Dray, S., Legendre, P., & Peres-Neto, P. R. (2006). Spatial modelling: A comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecological Modelling*, *196*, 483–493. <https://doi.org/10.1016/j.ecolmodel.2006.02.015>
- Dziuban, C., & Shirkey, E. (1974). When is a correlation matrix appropriate for factor analysis? Some decision rules. *Psychological Bulletin*, *81*, 358–361. <https://doi.org/10.1037/h0036316>
- Gonçalves-Souza, T., Romero, G. Q., & Cottenie, K. (2014). Metacommunity versus biogeography: A case study of two groups of Neotropical vegetation-dwelling arthropods. *PLoS ONE*, *9*, e115137. <https://doi.org/10.1371/journal.pone.0115137>
- Guiry, M. D., & Guiry, G. M. (2017). *AlgaeBase. World-wide electronic publication*. Galway, Ireland: National University of Ireland.
- Hájek, M., Roleček, J., Cottenie, K., Kintrová, K., Horsák, M., Pouličková, A., ... Dítě, D. (2011). Environmental and spatial controls of biotic assemblages in a discrete semi-terrestrial habitat: Comparison of organisms with different dispersal abilities sampled in the same plots. *Journal of Biogeography*, *38*, 1683–1693. <https://doi.org/10.1111/j.1365-2699.2011.02503.x>
- Heino, J., Melo, A. S., Siqueira, T., Soininen, J., Valanko, S., & Bini, L. M. (2015). Metacommunity organisation, spatial extent and dispersal in aquatic systems: Patterns, processes and prospects. *Freshwater Biology*, *60*, 845–869. <https://doi.org/10.1111/fwb.12533>

- Heino, J., Soininen, J., Alahuhta, J., Lappalainen, J., & Virtanen, R. (2016). Metacommunity ecology meets biogeography: Effects of geographical region, spatial dynamics and environmental filtering on community structure in aquatic organisms. *Oecologia*, *183*, 121–137.
- Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, *25*, 1965–1978. [https://doi.org/10.1002/\(ISSN\)1097-0088](https://doi.org/10.1002/(ISSN)1097-0088)
- Ibáñez, C., Alcaraz, C., Caiola, N., Rovira, A., Trobajo, R., Alonso, M., ... Prat, N. (2012). Regime shift from phytoplankton to macrophyte dominance in a large river: Top-down versus bottom-up effects. *Science of the Total Environment*, *416*, 314–322.
- Jamoneau, A., Passy, S. I., Soininen, J., Lebourcier, T., & Tison-Rosebery, J. (2018). Beta diversity of diatom species and ecological guilds: Response to environmental and spatial mechanisms along the stream watercourse. *Freshwater Biology*, *63*, 62–73. <https://doi.org/10.1111/fwb.12980>
- Jarvis, A., Reuter, H. I., Nelson, A., & Guevara, E. (2008). Hole-filled SRTM for the globe version 4. Available Database CGIAR-CSI SRTM 90m. Retrieved from <http://www.cgiar-csi.org/data/srtm-90m-digital-elevation-database-v4-1>
- Jenkins, D. G., & Ricklefs, R. E. (2011). Biogeography and ecology: Two views of one world. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *366*, 2331–2335. <https://doi.org/10.1098/rstb.2011.0064>
- Jocque, M., Field, R., Brendonck, L., & Meester, L. D. (2010). Climatic control of dispersal–ecological specialization tradeoffs: A metacommunity process at the heart of the latitudinal diversity gradient? *Global Ecology and Biogeography*, *19*, 244–252. <https://doi.org/10.1111/j.1466-8238.2009.00510.x>
- Lehner, B., & Döll, P. (2004). Development and validation of a global database of lakes, reservoirs and wetlands. *Journal of Hydrology*, *296*, 1–22. <https://doi.org/10.1016/j.jhydrol.2004.03.028>
- Leibold, M. A., Holyoak, M., Mouquet, N., Amarasekare, P., Chase, J. M., Hoopes, M. F., ... Gonzalez, A. (2004). The metacommunity concept: A framework for multi-scale community ecology. *Ecology Letters*, *7*, 601–613. <https://doi.org/10.1111/j.1461-0248.2004.00608.x>
- Lisboa, F. J., Peres-Neto, P. R., Chaer, G. M., Jesus Eda, C., Mitchell, R. J., Chapman, S. J., & Berbara, R. L. (2014). Much beyond mantel: Bringing procrustes association metric to the plant and soil ecologist's toolbox. *PLoS ONE*, *9*, e101238. <https://doi.org/10.1371/journal.pone.0101238>
- Liu, J., Soininen, J., Han, B. P., & Declerck, S. A. J. (2013). Effects of connectivity, dispersal directionality and functional traits on the metacommunity structure of river benthic diatoms. *Journal of Biogeography*, *40*, 2238–2248. <https://doi.org/10.1111/jbi.12160>
- Loewen, C. J. (2017). *Stressor response and spatial dynamics of mountain lake communities*. Doctoral dissertation, University of Alberta, Canada.
- Logue, J. B., Mouquet, N., Peter, H., & Hillebrand, H.; Metacommunity Working Group (2011). Empirical approaches to metacommunities: A review and comparison with theory. *Trends in Ecology & Evolution*, *26*, 482–491. <https://doi.org/10.1016/j.tree.2011.04.009>
- Maechle, M. (2012). *The cluster package: Cluster analysis extended Rousseeuw et al. (2012). Version 1.14*. Retrieved from <https://cran.r-project.org/package=cluster>
- Manguin, E., & Manguin, E. (1964). Contribution à la connaissance des diatomées des Andes du Pérou. Edition du Muséum. Nat., Ser. B, Botanique *12*, 41–98.
- McGlue, M. M., Silva, A., Corradini, F. A., Zani, H., Trees, M. A., Ellis, G. S., ... Assine, M. L. (2011). Limnogeology in Brazil's "forgotten wilderness": A synthesis from the large floodplain lakes of the Pantanal. *Journal of Paleolimnology*, *46*, 273. <https://doi.org/10.1007/s10933-011-9538-5>
- McRae, B. H. (2006). Isolation by resistance. *Evolution*, *60*, 1551–1561. <https://doi.org/10.1111/j.0014-3820.2006.tb00500.x>
- Metzeltin, D., & Lange-Bertalot, H. (2007). Tropical diatoms of South America II. In H. Lange-Bertalot (Ed.), *Iconographia Diatomologica 18*. Königstein, Germany: A.R.G. Gantner Verlag K.G, pp 1–877.
- Moritz, C., Meynard, C. N., Devictor, V., Guizien, K., Labruno, C., Guarini, J.-M., & Mouquet, N. (2013). Disentangling the role of connectivity, environmental filtering, and spatial structure on metacommunity dynamics. *Oikos*, *122*, 1401–1410.
- Nabout, J. C., Siqueira, T., Bini, L. M., & de Nogueira, I. (2009). S. No evidence for environmental and spatial processes in structuring phytoplankton communities. *Acta Oecologica*, *35*, 720–726. <https://doi.org/10.1016/j.actao.2009.07.002>
- Oksanen, J., Blanchet, G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., ... Wagner, H. (2016). *vegan: Community ecology package*. Retrieved from <https://cran.r-project.org/package=vegan>
- Padial, A. A., Ceschin, F., Declerck, S. A., De Meester, L., Bonecker, C. C., Lansac-Tôha, F. A., ... Bini, L. M. (2014). Dispersal ability determines the role of environmental, spatial and temporal drivers of metacommunity structure. *PLoS ONE*, *9*, e111227. <https://doi.org/10.1371/journal.pone.0111227>
- Pandit, S. N., Kolasa, J., & Cottenie, K. (2009). Contrasts between habitat generalists and specialists: An empirical extension to the basic metacommunity framework. *Ecology*, *90*, 2253–2262. <https://doi.org/10.1890/08-0851.1>
- Passy, S. I. (2007). Diatom ecological guilds display distinct and predictable behavior along nutrient and disturbance gradients in running waters. *Aquatic Botany*, *86*, 171–178. <https://doi.org/10.1016/j.aquabot.2006.09.018>
- Passy, S. I. (2017). Framework for community functioning: Synthesis of stress gradient and resource partitioning concepts. *PeerJ*, *5*, e3885. <https://doi.org/10.7717/peerj.3885>
- Peres-Neto, P., & Jackson, D. (2001). How well do multivariate data sets match? The advantages of a Procrustean superimposition approach over the Mantel test. *Oecologia*, *129*, 169–178. <https://doi.org/10.1007/s004420100720>
- Peres-Neto, P. R., Legendre, P., Dray, S., & Borcard, D. (2006). Variation partitioning of species data matrices: Estimation and comparison of fractions. *Ecology*, *87*, 2614–2625. [https://doi.org/10.1890/0012-9658\(2006\)87\[2614:VPOSDM\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2006)87[2614:VPOSDM]2.0.CO;2)
- QGIS Development Team (2013). *QGIS geographic information system. Open source geospatial foundation project*. India: Free Software Foundation.
- R Development Team (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Riera, J. L., Magnuson, J. J., Kratz, T. K., & Webster, K. E. (2000). A geomorphic template for the analysis of lake districts applied to the Northern Highland Lake District, Wisconsin, U.S.A. *Freshwater Biology*, *43*, 301–318. <https://doi.org/10.1046/j.1365-2427.2000.00567.x>
- Riley, S. J. (1999). Index that quantifies topographic heterogeneity. *Intermountain Journal of Sciences: IJS*, *5*, 23–27.
- Rimet, F., & Bouchez, A. (2011). Use of diatom life-forms and ecological guilds to assess pesticide contamination in rivers: Lotic mesocosm approaches. *Ecological Indicators*, *11*, 489–499. <https://doi.org/10.1016/j.ecolind.2010.07.004>
- Rumrich, U., Lange-Bertalot, H., & Rumrich, M. (2000). Diatoms of the Andes, from Venezuela to Patagonia/Tierra del Fuego, and two additional contributions. *Iconographia diatomologica 9*. Königstein, Germany: ARG Gantner Verlag KG, 673 pp.
- Servant-Vildary, S. (1986). Les diatomées actuelles des Andes de Bolivie (Taxonomie, écologie). *Cah Micropaléontologie*, *1*, 99–124.
- Soininen, J., Jamoneau, A., Rosebery, J., & Passy, S. I. (2016). Global patterns and drivers of species and trait composition in diatoms. *Global Ecology and Biogeography*, *25*, 940–950. <https://doi.org/10.1111/geb.12452>

- Steinitz-Kannan, M. (1979). *Comparative limnology of Ecuadorian lakes: A study of species number and composition of plankton communities of the Galapagos Islands and Equatorial Andes*. Ohio, USA: The Ohio State University.
- Steinitz-Kannan, M., Colinvaux, P., & Kannan, R. (1983). Limnological studies in Ecuador: 1. A survey of chemical and physical properties of Ecuadorian lakes. *Archiv für Hydrobiologie*, *65*, 61–105.
- Sylvestre, F., Servant-Vildary, S., & Roux, M. (2001). Diatom-based ionic concentration and salinity models from the south Bolivian Altiplano (15–23°S). *Journal of Paleolimnology*, *25*, 279–295. <https://doi.org/10.1023/A:1011157611619>
- Tapia, P. M., Fritz, S. C., Seltzer, G. O., Rodbell, D. T., & Metiever, S. (2006). Contemporary distribution and late-Quaternary stratigraphy of diatoms in the Junin Plain, central Andes, Peru. *Boletín Sociedad Geológica de Perú*, *101*, 19–42.
- Tonkin, J. D., Death, R. G., Muotka, T., Astorga, A., & Lytle, D. A. (2018). Do latitudinal gradients exist in New Zealand stream invertebrate metacommunities? *PeerJ*, *6*, e4898.
- Valencia, B. G., Matthews-Bird, F., Urrego, D. H., Williams, J. J., Gosling, W. D., & Bush, M. (2016). Andean microrefugia: Testing the Holocene to predict the Anthropocene. *New Phytologist*, *212*, 510–522. <https://doi.org/10.1111/nph.14042>
- Vélez, M. I., Wille, M., Hooghiemstral, H., & Metcalfe, S. (2005). Integrated diatom-pollen based Holocene environmental reconstruction of lake Las Margaritas, eastern savannas of Colombia. *The Holocene*, *15*, 1184–1198. <https://doi.org/10.1191/0959683605hl890rp>
- Verleyen, E., Vyverman, W., Sterken, M., Hodgson, D. A., De Wever, A., Juggins, S., ... Sabbe, K. (2009). The importance of dispersal related and local factors in shaping the taxonomic structure of diatom metacommunities. *Oikos*, *118*, 1239–1249. <https://doi.org/10.1111/j.1600-0706.2009.17575.x>
- Viana, D. S., Santamaría, L., Schwenk, K., Manca, M., Hobæk, A., Mjelde, M., ... Figuerola, J. (2014). Environment and biogeography drive aquatic plant and cladoceran species richness across Europe. *Freshwater Biology*, *59*, 2096–2106. <https://doi.org/10.1111/fwb.12410>
- Vilmi, A., Tolonen, K. T., Karjalainen, S. M., & Heino, J. (2017). Metacommunity structuring in a highly-connected aquatic system: Effects of dispersal, abiotic environment and grazing pressure on microalgal guilds. *Hydrobiologia*, *790*, 125–140. <https://doi.org/10.1007/s10750-016-3024-z>
- Vitousek, P. M., Mooney, H. A., Lubchenco, J., & Melillo, J. M. (1997). Human domination of Earth's ecosystems. *Science*, *277*, 494–499. <https://doi.org/10.1126/science.277.5325.494>
- Vyverman, W., Verleyen, E., Sabbe, K., Vanhoutte, K., Sterken, M., Hodgson, D. A., ... De Wever, A. (2007). Historical processes constrain patterns in global diatom diversity. *Ecology*, *88*, 1924–1931. <https://doi.org/10.1890/06-1564.1>
- Wetzel, C. E., de Bicudo, D. C., Ector, L., Lobo, E. A., Soininen, J., Landeiro, V. L., & Bini, L. M. (2012). Distance decay of similarity in Neotropical diatom communities. *PLoS ONE*, *7*, 10–11.
- Winegardner, A. K., Beisner, B. E., Legendre, P., & Gregory-Eaves, I. (2015). Are the landscape-level drivers of water column and surface sediment diatoms different? *Freshwater Biology*, *60*, 267–281. <https://doi.org/10.1111/fwb.12478>
- Zaharescu, D. G., Hooda, P. S., Burghelca, C. I., & Palanca-Soler, A. (2016). A multiscale framework for deconstructing the ecosystem physical template of high-altitude lakes. *Ecosystems*, *19*, 1064–1079. <https://doi.org/10.1007/s10021-016-9987-9>
- Zorzal-Almeida, S., Soininen, J., Bini, L. M., & Bicudo, D. C. (2017). Local environment and connectivity are the main drivers of diatom species composition and trait variation in a set of tropical reservoirs. *Freshwater Biology*, *62*, 1551–1563. <https://doi.org/10.1111/fwb.12966>

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**How to cite this article:** Benito X, Fritz SC, Steinitz-Kannan M, Velez MI, McGlue MM. Lake regionalization and diatom metacommunity structuring in tropical South America. *Ecol Evol*. 2018;8:7865–7878. <https://doi.org/10.1002/ece3.4305>