HISTORICAL PROCESSES CONSTRAIN PATTERNS IN GLOBAL DIATOM DIVERSITY

WIM VYVERMAN,1,9 ELIE VERLEYEN,1 KOEN SABBE,1 KOENRAAD VANHOUTTE,1 MIEKE STERKEN,1 DOMINIC A. HODGSON,2 DAVID G. MANN,3 STEVE JUGGINS,4 BART VAN DE VIVER,5 VIVIENNE JONES,6 ROGER FLOWER,6 DONNA ROBERTS,7 VICTOR A. CHEPURNOV,1 CATHY KILROY,8 PIETER VANORMELINGEN,1 AND AAIKE DE WEVER1

1Ghent University, Biology Department, Protistology and Aquatic Ecology, Krijgslaan 281 S8, 9000 Gent, Belgium
2British Antarctic Survey, High Cross Madingley Road, Cambridge CB30ET UK
3Royal Botanic Garden, Edinburgh, Scotland EH3 5LR UK
4University of Newcastle Upon Tyne, School for Geography Politics and Sociology, Newcastle Upon Tyne NE17RU UK
5National Botanic Garden of Belgium, Department of Cryptogamy (Bryophyta and Thallophyta), Donauwe van Bouchout, 1860 Meise, Belgium
6University College London, Environmental Change Research Centre, Gower Street, London WC1E 6BT UK
7ARC Centre of Excellence in Ore Deposits, Private Bag 79, Hobart, Tasmania 7001 Australia
8National Institute of Water and Atmospheric Research (NIWA), Christchurch 8002 New Zealand

Abstract. There is a long-standing belief that microbial organisms have unlimited dispersal capabilities, are therefore ubiquitous, and show weak or absent latitudinal diversity gradients. In contrast, using a global freshwater diatom data set, we show that latitudinal gradients in local and regional genus richness are present and highly asymmetric between both hemispheres. Patterns in regional richness are explained by the degree of isolation of lake districts, while the number of locally coexisting diatom genera is highly constrained by the size of the regional diatom pool, habitat availability, and the connectivity between habitats within lake districts. At regional to global scales, historical factors explain significantly more of the observed geographic patterns in genus richness than do contemporary environmental conditions. Together, these results stress the importance of dispersal and migration in structuring diatom communities at regional to global scales. Our results are consistent with predictions from the theory of island biogeography and metacommunity concepts and likely underlie the strong provinciality and endemism observed in the relatively isolated diatom floras in the Southern Hemisphere.

Key words: diatoms; dispersal; diversity; latitudinal diversity gradient; metacommunity; microorganisms; protists; theory of island biogeography.

INTRODUCTION

The Baas-Becking (or ubiquity) hypothesis (Baas-Becking 1934, Finlay and Clarke 1999, Finlay 2002) postulates that for microorganisms “everything is everywhere” but “the environment selects.” If this is true, cosmopolitan distribution patterns should prevail and biogeographical patterns such as latitudinal diversity gradients should be absent, once ecological controls are factored out (Hillebrand and Azovsky 2001, Fenchel and Finlay 2004). However, an increasing body of research supports the idea that microorganisms exhibit a biogeography (Martiny et al. 2006) and that regional patterns in microbial diversity (e.g., species-area relationships) are consistent with those observed in macroorganisms (Green et al. 2004, Horner-Devine et al. 2004, Smith et al. 2005). Furthermore, it was recently shown that spatial variation in diatom diversity and species composition cannot be solely driven by local environmental conditions but is also determined by habitat availability for lentic diatom communities (Telford et al. 2006) and spatial attributes in lotic habitats (Soininen 2004, Soininen et al. 2004). To date, it has been difficult however to test the importance of historical (e.g., dispersal limitation, migration) vs. contemporary environmental processes in controlling microbial diversity at the global scale (Martiny et al. 2006). This is largely due to a lack of taxonomically intercalibrated inventories of microbial diversity covering wide geographical regions (Mann 1999, Taton et al. 2003).

Here we assess, for the first time, global patterns of local (in-lake) and regional diatom diversity in 1855 freshwater bodies. We investigate diversity patterns at the genus level because, for most organisms, richness relationships are conserved from species to family level (Hubbell 2001). We thereby alleviate taxonomic bias and avoid the difficulties associated with the taxonomic intercalibration of diatom species data due to (semi-)cryptic variation (Mann 1999). We test the importance of historical processes in regulating global diatom biodiversity and assess the role of environmental conditions as an
alternative explanation for the observed patterns. We predict that if historical factors constrain local and regional diatom communities, (1) local richness should be highest in lake-rich, highly connected regions where local extinction is balanced by high colonization rates, whereas (2) regional diversity should be lowest in isolated lake districts where dispersal limitation is high according to the theory of island biogeography and metacommunity concepts (MacArthur and Wilson 1967, Leibold et al. 2004).

**Materials and Methods**

**Diatom data set development**

We assembled data on the diatom flora and limnology of freshwater bodies in lake districts distributed between 40° and 80° latitude in both hemispheres, together with a more limited set of tropical lakes from Papua New Guinea and Central and East Africa based upon previously published data sets (Fig. 1; see Appendix A for a complete reference list of the data sets used). The primary diatom data consisted of enumerations of species relative abundance in surface sediment samples (mainly sediment-core tops) from 1855 freshwater lakes, each of which typically represents a record of diatom growth in the lake from both pelagic and benthic habitats spanning a few years (Smol and Cumming 2000). Typically between 300 and 600 specimens are counted in routine diatom analysis in the framework of paleolimnological research projects aimed at developing diatom calibration data sets and transfer functions (Battarbee et al. 2001). In order to facilitate taxonomic consistency between the data sets, species were assigned to genera according to the most up-to-date classification (Fourtanier and Kociolek 2003; see Appendix B for a complete list of the genera encountered), which resulted in a data set comprising 89 diatom genera that represents almost the global diversity of freshwater diatom genera known to occur in standing waters. Genus richness is a power function of species richness as is generally found in nature (Enquist et al. 2002, Passy and Legendre 2006). The significant relationship ($R^2 = 0.858$, $P < 0.001$, $n = 1562$ lakes; Appendix C) implies that genus richness is a robust proxy for species richness, which is in agreement with results from various other taxonomic groups (Gaston et al. 1995, Roy et al. 1996, La Ferla et al. 2002) and which will allow us to cautiously extrapolate our observations to the species level (see also Appendix C for additional information).

**Standardization of sample effort**

In order to account for differences in sampling effort in different regions, lakes were grouped in 233 equal-sized grid cells of 10,000 km² and mean local (mLR) and mean regional diatom richness (mRR) were calculated.
for each grid. Preliminary analysis revealed that the number of lakes in each grid can strongly determine the number of diatom genera \( r = 0.530, P < 0.001, n = 233 \) grid cells) as a result of species-accumulation relationships. In each grid cell, mRR was therefore calculated as the number of genera in 10 lakes after 999 permutations of the taxon accumulation curves in the software package Primer 5.1.2 (Primer-E Ltd., Plymouth, UK). This resulted in a data set of 53 grids containing \( \geq 10 \) lakes. On average, 91% (minimum, 71%; maximum, 100%) of the total number of genera in each grid was included in the calculated mRR measure; the total richness of all the lakes in each grid was strongly positively correlated with mRR calculated using our standardization procedure \( r = 0.944, P < 0.001, n = 53 \); mLR was calculated as the average number of genera per lake in each grid cell \( n = 233 \).

Historical factors

We consider dispersal limitation to be a historical event following Martiny et al. (2006), because current taxonomic composition is influenced by past (or recent) dispersal limitation. In order to assess the importance of historical factors influencing diatom richness, we calculated the number of lakes in the equal-sized grids based on the Global Lake Database (Lehner and Döll 2004) using ArcGis 9.0 (ESRI, Redlands, California, USA). Although the Global Lake Database only includes lakes \( \geq 0.1 \) km\(^2\), it can be expected to approximate the global distribution of all freshwater bodies. Lake density (i.e., the number of lakes in the grid divided by the amount of ice-free land area) serves as a measure of the connectivity between lakes in each grid. Habitat availability is approximated by the number of lakes and the surface area of freshwater (as percentage of area) in each grid. The degree of isolation of lake districts is approximated by the area (in percentage of area) covered by sea, ocean, and ice sheets within circles with a radius of 200 km around the center of each grid. This metric is high in isolated oceanic and glaciated areas where the amount of ice-free land (which potentially houses lakes) is minimal. All spatial variables were logarithmically transformed prior to statistical analysis.

Environmental variables

Unfortunately, the database suffers from a relatively high amount of missing values. The environmental variables pH and conductivity were available for a total of 1694 lakes. Conductivity and pH and the concentration of Ca, Na, K, and Mg was available for 1039 lakes; these variables and lake depth and lake surface area were available for a smaller subset of lakes \( n = 681 \) lakes. Soluble reactive phosphate and silicate, ammonium, and nitrate were available for a minority of the lakes.

Regions with a high limnological variability are expected to contain more diatom genera as a result of taxonomic turnover between lakes. We therefore calcu-
lated the variance of each standardized environmental variable which were available for a considerable amount of lakes (i.e., pH, conductivity, and the concentration of Ca, Na, K, and Mg) and approximated total limnological diversity in each grid by calculating the sum of the variances in 10 randomly selected lakes in each grid. Mean annual temperature and the difference between the warmest and coldest month, which approximates seasonality, were obtained using two freely available data sets (Jones and Reid 2001, Hijmans et al. 2005).

Statistical analyses and variation partitioning

Pearson correlations and linear and nonlinear regressions were performed in Statistica 6.0 (StatSoft, Inc., Tulsa, Oklahoma, USA). Principal component analysis (PCA) was performed using CANOCO 4.5 for Windows (ter Braak and Smilauer 2002) in order to compare the limnological variation (in pH, conductivity, and the concentration of Ca, Na, K, and Mg) between the main geographic regions. All environmental variables were logarithmically transformed (except pH), and centered and standardized prior to ordination analysis. PCA with centering and standardization was similarly performed for the larger set of lakes for which only pH and conductivity data were available.

Partial regression analyses were performed using redundancy analyses (RDA) in the program CANOCO 4.5 for Windows (ter Braak and Smilauer 2002) as it implements a user friendly interface by which the effects of particular covariables can be eliminated (Heikkinen et al. 2005). In order to capture nonlinear relationships between the response and the predictor variables we also entered quadratic terms of the predictors (Heikkinen et al. 2005). We first used the forward selection procedure to select only those variables that, combined into one model, contributed significantly ($P < 0.05$) to explaining the variation in the richness measures (Borcard et al. 1992) using the Monte Carlo permutation test (999 permutations). We subsequently partitioned the variation in the response variables local richness and mRR over three groups of predictors, viz. the limnological (including lake morphological) variables, temperature and yearly temperature ranges, and historical factors. The partial analysis resulted in eight fractions if at least one variable of the three classes is significant (see Heikkinen et al. 2005 for detailed information), namely,
unexplained variation and (1) the pure effect of limnological (and lake morphological) variables; (2) the pure effect of temperature related variables; (3) the pure effect of historical variables, and combined variation due to joint effects of (1) and (2), (2) and (3), (1) and (3), and the three groups of variables combined.

RESULTS AND DISCUSSION

Latitudinal gradients in both mLR and mRR were strongly asymmetric between both hemispheres (Fig. 2A, B; see Appendix D for the original genus richness data). At comparable latitudes, lakes in the Southern Hemisphere consistently had lower mLR and mRR than those in the Northern Hemisphere. At temperate latitudes in the Southern Hemisphere, mLR and mRR were comparable to tropical latitudes but declined linearly with increasing latitude between 42° and 70° S. In the Northern Hemisphere, the gradient between mLR and latitude was “hump-shaped” between 55° and 70° N, whereas the gradient between mRR and latitude was virtually flat.

The diversity of the regional diatom pool (mRR) was negatively correlated with the degree of isolation of the individual lake districts (Fig. 3A, Table 1). The isolated sub-Antarctic islands thus consistently had fewer regionally coexisting diatom genera than the highly connected land masses in the Northern Hemisphere at similar latitudes (e.g., Fennoscandia, North America).

**TABLE 1.** Pearson correlation coefficients and significance of the regression analyses between mean regional richness (mRR; n = 53 grid cells) and mean local richness (mLR; n = 233 grid cells) and the historical predictor variables.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Lake density (connectivity)</th>
<th>No. lakes in grid (habitat availability)</th>
<th>Freshwater in grid (%; habitat availability)</th>
<th>Sea and ice sheets (%; isolation)</th>
</tr>
</thead>
<tbody>
<tr>
<td>mRR</td>
<td>0.142 (NS)</td>
<td>0.569***</td>
<td>0.379**</td>
<td>−0.815***</td>
</tr>
<tr>
<td>mLR</td>
<td>0.541***</td>
<td>0.621***</td>
<td>0.260***</td>
<td>−0.305***</td>
</tr>
</tbody>
</table>

*** P < 0.001; ** P < 0.01; NS, not significant.
The effect of isolation was particularly evident when mRR on Amsterdam Island (37°50' S) was compared with less isolated regions at similar latitudes in the Southern and Northern Hemispheres (Fig. 2).

The diversity of local communities (mLR) was positively and significantly correlated with the number of regionally coexisting taxa (mRR; Fig. 3B; \( r = 0.765, P < 0.001, n = 53 \) grid cells), which is observed for most groups of organisms (Hillebrand 2005). The interhemispheric difference in mLR gradients also closely matched habitat availability (Fig. 3C, Table 1) and lake density, which is a proxy for connectivity within lake districts (Fig. 3D, Table 1). Lake density peaks in the northern hemisphere between 55° and 65° N (Fig. 1), on account of the retreat of the last continental ice sheets (Wetzel 2001). In the Southern Hemisphere, the number of lakes is much lower due to the lack of continental land masses between 55° and 65° S, and decreases continuously from the temperate regions toward Antarctica. Local diatom communities in the freshwater lakes under investigation thus appear to be selected from the regional diatom pool; their diversity peaks in regions with numerous, highly connected water bodies where chances of successful colonization increase and local extinction rates are reduced.

Several lines of evidence strongly suggest that, at global to regional scales, historical factors explain significantly more of the observed geographic patterns in genus richness than do the available contemporary environmental conditions. As the patterns in diatom diversity were strongly asymmetrical between both hemispheres (Fig. 2) and diatom richness was highest in the mid latitudes in the North, environmental variables that covary with latitude (e.g., duration of ice cover and stratification regime, light climate, temperature, productivity) were likely of minor importance. Principal component analysis (PCA) of the available water chemistry data (1039 lakes) further revealed that water bodies with contrasting genus diversity from corresponding latitudinal belts in both hemispheres showed a comparable variation in conductivity, major ion composition, and acidity (Fig. 4A). For tropical lakes, only limited data were available on major ion composition. However, pH and conductivity data suggested a greater variation in water chemistry in these lakes than is shown in the subset for which complete data were available (Fig. 4B). Therefore the ranges in the variation of chemical limnology in the tropical lakes most likely span the same gradient lengths as the temperate water bodies. Importantly, our results show that there are no systematic differences in water chemistry between both hemispheres that might underlie the observed interhemispheric differences in diatom diversity. This was confirmed by variation partitioning analyses. Isolation of lake districts accounted for 72.3% of the variation in mRR; the pure effect of this variable was significant and equaled 67.3% (Appendix E). The pure effects of the limnological predictors were smaller (8.5%), whereas none of the temperature related variables was significant. Variation partitioning analysis with historical and temperature related variables alone for a more complete data set (53 grids), revealed that isolation accounts for 68.6% (61.3% unique) and confirmed the small influence of temperature (7.4%,
0.1% unique yet statistically not significant). For local diatom genus richness, historical factors accounted for 35% of the variation (20.2% unique). The pure effects of temperature related and limnological variables were small (5.3% and 3.8%, respectively). In a more restricted data set for which the morphological variables lake depth and lake surface area were additionally available, variation partitioning similarly revealed the importance of historical factors which accounted for 44% of the variation in local diatom richness (25% unique). Temperature related and local environmental variables were less important and explained respectively 18.9% and 26.1% (2.5% and 8.4% unique). We further assessed the relationships between local richness and the available water chemistry and morphological variables using linear and non linear regression analyses. The responses are best described as a constraint spaces; the regressions resulted in generally very weak relations between the local richness and the chemical and morphological variables (Appendix F). For the most important variables regulating diatom community structure (e.g., pH; Telford et al. 2006), maximal and minimal local diatom richness patterns were invariant over a wide range of values; only at low and/or high values was maximal local diversity apparently constrained. We conclude that contemporary ecological conditions alone cannot explain biogeographic trends in local and regional diversity. Unfortunately, for several limnological variables, including nutrients in particular, available data were too limited to be included in our statistical analysis. However, while we anticipate that nutrient levels are related to diversity at the local and regional scale (Chase and Leibold 2002), we do not expect this relationship to override the effect of historical factors in explaining global diatom diversity patterns.

Our data thus strongly suggest that historical processes (i.e., colonization and extinction, dispersion and migration) constrain global patterns in regional and local diatom diversity. In particular, the asymmetry in latitudinal genus richness patterns between 42° S and the pole contradicts the idea that microbial dispersal is virtually unlimited (Baas-Becking 1934, Finlay and Clarke 1999, Finlay 2002). These findings are concordant with the distinct provinciality and high levels of apparent endemism in the isolated diatom floras of Australian (see Plate 1), Antarctic, and Sub-Antarctic freshwaters (Vyverman et al. 1998, Sabbe et al. 2003, Van de Vijver et al. 2005). It will be interesting to see if historical processes operate to a similar extent in other microbial groups, with greatly different life histories and, in particular, survival and reproductive strategies.

Acknowledgments

This research was supported by grants from the Belgian Science Policy Office. E. Verleyen is a senior research fellow with the Fund for Scientific Research–Flanders (Belgium). We thank the EDDI diatomists who provided the data from the Northern Hemisphere: John N. Anderson, Helen Bennion, H. J. B. Birks, John Boyle, Nigel Cameron, Jordi Catalan, Joan Garcia, Françoise Gasse, Atte Korhola, Tom Korsman, André Lotter, Aldo Marchetto, Jane Reed, Patrick Rioual, Peter Rosen, Roland Schmidt, Nadia Solovieva, Jan Weckström, and Sybille Wumsam. We also thank an anonymous reviewer and Sophia Passy for critically reviewing and improving the manuscript. W. Vyverman and E. Verleyen contributed equally to this paper.

Literature Cited


APPENDIX A
References to the source data used to construct the intercalibrated diatom data set (Ecological Archives E088-115-A1).

APPENDIX B
The diatom genera encountered in this study (Ecological Archives E088-115-A2).

APPENDIX C
A figure showing the species–genus richness relationships (Ecological Archives E088-115-A3).

APPENDIX D
A figure showing the interhemispheric asymmetry in local intercalibrated genus richness in 1855 freshwater lakes (Ecological Archives E088-115-A4).

APPENDIX E
Results of the variation partitioning analyses (Ecological Archives E088-115-A5).

APPENDIX F
Figures showing the response of local diatom richness to the available chemical, morphological, and historical variables (Ecological Archives E088-115-A6).