Broad-scale environmental response and niche conservatism in lacustrine diatom communities

Joseph R. Bennett[1]*, Brian F. Cumming[2], Brian K. Ginn[2,3] and John P. Smol[2]

[1]*Centre for Applied Conservation Research, Faculty of Forestry, 2424 Main Mall, University of British Columbia, Vancouver, British Columbia, Canada V6T 1Z4,* [2]*Paleoecological Environmental Assessment and Research Laboratory (PEARL), Department of Biology, Queen's University, Kingston, Ontario, Canada K7L 3N6,* [3]*Lake Simcoe Region Conservation Authority, 120 Bayview Parkway, Newmarket, Ontario, Canada L3Y 4X1*

ABSTRACT

Aim (1) To resolve theoretical debates regarding the role of environment versus dispersal limitation, the conservatism of niches across distances and the prevalence of environmental specialists in diatom communities. (2) To provide guidance on the use of diatom communities and other microbial analogues to analyse ecological response to environmental change.

Location Eight hundred and ninety-one lakes in five regional datasets from north-western Europe and four regional datasets from north-eastern North America.

Methods Lacustrine diatom communities were analysed at three scales: inter-continental, intra-continental and regional. Nested partial redundancy analyses (RDAs) were used to determine spatial versus environmental components of community variation. Weighted-averaging (WA) regression and calibration, as well as logistic and quadratic regressions, were used to detect niche conservatism and the prevalence of environmental specialists.

Results Community patterns indicate that dispersal limitation acts predominantly at the inter-continental scale, while at the regional (less than *c*. 1,000,000 km^2) scale, a single environmental variable (pH) explains more than five times the community variation as spatial (dispersal-related) variables. In addition, pH niche components appear to be conserved at the inter-continental scale, and environmental specialization does not impose relative rarity, as specialists apparently readily disperse to suitable environments.

Main conclusions Analysis at multiple scales is clearly important in determining the influences of community variation. For diatom communities, dispersal limitation acts most strongly at the broadest scales, giving way to environment at the scales considered by most analyses. The availability of a wide variety of propagules with consistent niches across regions indicates that diatom communities reflect the succession of taxa according to local environmental conditions, rather than disequilibrium with the environment or adaptation of local populations. While multi-scale analyses must be undertaken for other groups to resolve debates over community drivers and determine appropriate scales for prediction, for diatoms (and probably other microbial communities), responses to environmental change can be inferred using analogue datasets from large geographic areas.

Keywords

Analogue technique, diatoms, dispersal limitation, environmental response, Europe, niche conservatism, North America, pH, redundancy analysis (RDA), variation decomposition.

Correspondence: Joseph R. Bennett, Centre for Applied Conservation Research, Faculty of Forestry, 2424 Main Mall, University of British Columbia, Vancouver, British Columbia, Canada V6T 1Z4.
E-mail: jrb5@interchange.ubc.ca

INTRODUCTION

As anthropogenic environmental change intensifies, so does the need to determine the relationships between environments and ecological communities, both to infer past changes as context for current trends and to predict future responses. To systematically do so, ecologists frequently rely on analogues, matching spatially distant communities inhabiting environments corresponding to those that are predicted (Guisan & Thuiller, 2005; Smol, 2008). Such 'space-for-time' substitutions assume a greater role for the environment than dispersal limitation in determining community composition in the focal time frame. However, the relative importance of these factors is actively debated from theoretical perspectives (e.g. Alonso *et al.*, 2006; McGill *et al.*, 2006; Benedetti-Cecchi, 2007) and in differing interpretations of community structure (e.g. Svenning & Skov, 2004, 2007; Welk & Bruelheide, 2006), with increasing recognition that different factors predominate at different spatial scales (Cottenie, 2005). Unfortunately, large datasets with distinct scales at which community determinants may be compared with minimal covariation are scarce. To make crucial predictions regarding ecological/environmental links, ecologists are thus faced with considerable uncertainty in choosing scales at which analogues may be used and predictions may be drawn.

Microbial communities have received particular attention in this debate. Assumptions of virtual ubiquity of microbial propagules (Finlay, 2002) and consistent niches among distant populations have been implicit in the use of microbial community analogues across large distances to model local environmental change (Smol *et al.*, 2001), and pressing environmental issues such as climate change (Verschuren *et al.*, 2000), acid precipitation (Sullivan *et al.*, 1990) and fisheries collapses (Finney *et al.*, 2000) have been examined using these techniques. However, the assumptions of little dispersal limitation and niche conservatism in microbial communities have been challenged (Telford *et al.*, 2006a; Vyverman *et al.*, 2007; Vanormelingen *et al.*, 2008), and the relative roles of distance versus environment in determining microbial community composition are controversial (Martiny *et al.*, 2006), as is the proportion of taxa that do indeed exhibit a discernible environmental response (Pither & Aarssen, 2005; Telford *et al.*, 2006b).

Diatom flora from groups of lakes across large areas have been enumerated to use as analogues for quantitative environmental predictions (Smol, 2008). Diatoms have been assumed to be particularly responsive to their physical environment, due to the importance of the aquatic matrix for organisms with a large surface area to volume ratio and the putative ubiquity of well-adapted propagules over time spans for which environmental changes are tracked. However, the relative roles of dispersal limitation and niche conservatism across distance have rarely been explicitly included in these analyses, and the causes of variation at different scales have remained virtually unexplored. Thus the scale at which the environmental signal may be overwhelmed by dispersal limitation and niche differences of allopatric populations is unknown, as is the extent to which diatom and other microbial communities are indeed exceptions to biogeographic limitations affecting larger organisms.

To address these issues, we analysed diatom assemblages from surficial sediments of 981 lakes from large areas of eastern North America (493 lakes) and north-western Europe (488 lakes) for the contributions of spatial separation and environment at different scales, and for evidence of broad-scale niche conservatism. Our specific goals were threefold.

1. To explore the influence of dispersal limitation and environment on community variability at different scales, using lakes as discrete points on spatial and environmental continua. At scales for which dispersal limitation is predominant, we predict a larger spatial than environmental signal, while the balance would shift to the environmental signal at scales for which environment is the dominant influence on community composition.

2. To investigate evidence for niche conservatism across spatial scales, to determine whether dispersal limitation and metapopulation niche differences lead to divergent environmental niches of conspecifics from distant locations.

3. To test the prediction that environmental generalists have the greatest distributions among sites, and that specialized taxa whose distributions are strongly connected to the environment are relatively uncommon (Hubbell, 2005; Pither, 2007).

The large geographic expanse of our dataset, the separation of lakes by inhospitable terrestrial or oceanic matrix and the nesting of relatively discrete scales allowed a uniquely detailed and precise exploration of spatial versus environmental correlation and niche conservatism.

METHODS

Diatom data

Our European and North American datasets are themselves combinations of regional surveys of between 37 and 237 lakes (Fig. 1) originally used to examine relationships between diatom assemblages and pH to detect signals of acid precipitation (Battarbee *et al.*, 2001; Ginn *et al.*, 2007). These surveys were intended to encompass the range of pH measurements in a contiguous region, to use the relationships between community composition and pH as modern analogues in palaeoecological analyses. Three distinct spatial scales of variation are thus represented in the synoptic dataset: (1) inter-continental (North America versus Europe), (2) intra-continental (comparison of regional surveys), and (3) regional (polynomial terms of spatial coordinates). Environmental variation in the lakes is represented by pH, the only environmental variable that was measured across the entire dataset, and the presumptive determinant of community variation that motivated the original surveys.

Diatom community surveys from surface sediment samples of 493 eastern North American lakes originate from four regional datasets, and are summarized by Ginn *et al.* (2007). Following Ginn *et al.* we removed Lake Earnest, PA, from the dataset, due to an error in water chemistry measurement. Diatom data from surface sediment samples of 488 European lakes were taken from the non-alpine, non-arctic regional
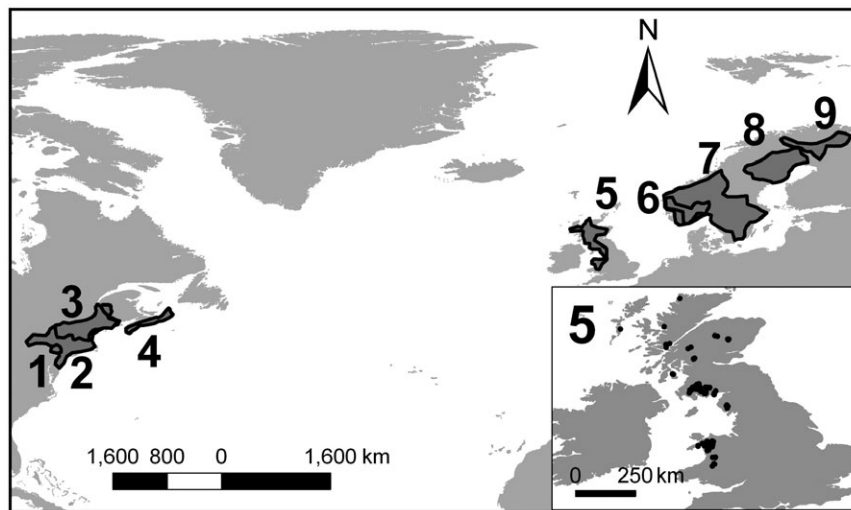
Figure 1 Locations of nine regional diatom surveys incorporated into the synoptic dataset: 1, POCONOS – POCONOS paleolimnological dataset (Sherman & Fairchild, 1994); 2, EMAP – Environmental Monitoring and Assessment Program (Dixit *et al.*, 1999); 3, PIRLA – Paleoecological Investigation of Recent Lake Acidification (Dixit *et al.*, 1993); 4, TEAM – Trends in Eutrophication and Acidification in the Maritimes (Ginn *et al.*, 2007); 5, SWAP-BR – Surface Waters Acidification Programme, British Lakes (EDDI version; Battarbee *et al.*, 2001); 6, NOR – EDDI Norwegian pH dataset; 7, SWAP – Surface Waters Acidification Programme; 8, SWED – EDDI Swedish dataset; 9, FIN – EDDI Finnish pH dataset. Inset: sample locations for region 5, SWAP-BR, showing the scale of inter-regional sampling.

datasets in the European Diatom Database (EDDI) pH dataset, which combined data from several previous studies (Battarbee *et al.*, 2001). Arctic and alpine datasets were not used because their environments are not comparable with those in the North American dataset.

Diatom data for the North American and European datasets were harmonized for taxonomy in order to eliminate replication of taxa under different nomenclatural schemes. North American taxonomy was previously harmonized by Ginn *et al.* (2007) and subsequently updated to current nomenclature (e.g. *Craticula*, *Sellaphora*, etc. as genera in place of the former *Navicula sensu lato*). Taxonomy for the European dataset was originally harmonized according to protocols documented online for the EDDI project (http://craticula.ncl.ac.uk/Eddi/), and then harmonized by us to the North American standards where applicable. While the vast majority of taxa had identical nomenclature schemes, some were classified to a lower taxonomic level (i.e. subspecies, forma or variety) by one dataset. In these cases, the lower taxonomic levels were grouped back to species level in order maintain consistency between the EDDI and North American datasets. The amount of discrete ecological information lost by this process was minimal compared with what would have been lost by eliminating such taxa. A single taxon (*Cymbella* cf. *aequalis* in the North American dataset, *Cymbella aequalis* in the European dataset), was excluded due to taxonomic uncertainty between the European and North American datasets. Table S1 in Supporting Information contains a list of harmonized taxa.

Data analyses

To explore the influence of scale and spatial separation versus environment, community variation was decomposed using nested partial redundancy analyses (RDAs) (Fig. 2) (Borcard *et al.*, 1992; Cushman & McGarigal, 2002) of dominant taxa (> 5% relative abundance in at least three lakes, representing 88% of total abundance in the dataset). Our decomposition of variation was based on the partitioning method outlined by Borcard *et al.* (1992), as adapted by Cushman & McGarigal (2002) for multiple scales. This technique was designed to determine how community variation is influenced by explanatory variables at separate scales. We used it to distinguish the effects of our three spatial scales, and the independent effect of pH as well as its covariation with spatial variables. Identical analyses were also conducted using genus-level data, to determine whether patterns held at higher taxonomic levels.

Diatom data were converted to proportions per sample, as in standard palaeolimnological protocol, and to account for varying numbers of diatom frustules counted per lake. Diatom proportional data were transformed using the Hellinger transformation, as per the recommendations of Legendre & Gallagher (2001) based on analyses of performance of common transformations for community analyses. An initial set of spatial variables comprising polynomial terms ($x, y, xy, x^2, y^2, x^2y, y^2x, x^3, y^3$; Borcard *et al.*, 1992) of centred latitudes and longitudes of lakes was derived, to be used as potential spatial variables in ordinations. This technique was used because the Mantel test may under-represent the spatial component of space/ environment tests (Legendre *et al.*, 2005, 2008), and principal coordinates of neighbouring matrices may reveal spatial structure in the dataset that is not related to dispersal limitation (Tuomisto & Ruokolainen, 2008). The initial set of polynomial terms was subjected to the forward selection procedure of Blanchet *et al.* (2008), which combines selection based on *P*-value ($P < 0.05$) and based on adjusted r^2 (Peres-Neto *et al.*,

2006). In addition, correlation among forward-selected variables was scrutinized, as the polynomial technique can suffer from variance inflation through autocorrelation of spatial terms (Borcard & Legendre, 2002). Variables that were correlated $r > 0.95$ with those that explained more information were removed from the analysis. The following spatial variables were retained: longitude; longitude^2; longitude$^2 \times$ latitude; latitude^2.

In order to attain the decomposition of variance at our three scales, a total of nine partial RDAs were needed (Appendix S1, Tables S2 & S3). This decomposition was modified slightly from that of Cushman & McGarigal (2002) since there is no overlap between continents in either regional identity or spatial variables. RDAs were conducted using CANOCO v.4.5 (ter Braak & Šmilauer, 2002) and variance explained was adjusted for multiple comparisons according to Peres-Neto et al. (2006).

To test for the conservatism of niches over large distances, we conducted weighted-averaging (WA) regressions of taxonomic abundances versus pH for the 70 dominant (> 5% relative abundance in at least three lakes) taxa shared by both Europe and North America (representing 70% of European and 73% of North American dominant taxa). WA with bootstrapping was used to determine pH optima of dominant taxa in separate models calculated for North American and European lakes. WA models were calculated using the program C2 v.1.5 (S. Juggins, University of Newcastle, 2007). The pH optima of taxa shared by both datasets were then examined using least-squares regression. In addition, to determine whether predictive ability of models might diminish across larger spatial scales (due to aggregation of data from allopatric populations with different pH optima), WA calibration with classical deshrinking (Birks et al.,

1990) was used to predict the pH of lakes via diatom assemblages. A synoptic model using all data was run, as were separate models using each regional dataset.

To examine the prediction that the most widely distributed taxa are generalists that are broadly adapted to the most common environments, and that taxa with distinct niches are relatively uncommon (Hubbell, 2005; Pither, 2007), we conducted a least-squares quadratic regression of the number of lakes in which taxa were present versus their inferred optimum pH, using R v.2.8.0. Using presence/absence of the 10 most common diatom taxa from the combined dataset, we also conducted stepwise logistic regressions with pH, region and polynomial terms of spatial coordinates as per Borcard et al. (1992) as explanatory variables, using SAS v.9.2 (SAS Institute, Cary, NC, USA). The significance level for selection was $P < 0.05$. In the stepwise routine, the variable 'continent' was redundant, as all variability at this level would be attributed to 'region' since each region was contained within a single continent.

RESULTS

Scales of community variation

The total variation explained by the RDAs constrained to the spatial and environmental variables (Fig. 2) was 29.5%, corrected for multiple comparisons (Peres-Neto et al., 2006). This result is typical of large ecological datasets (Økland, 1999), but is relatively large given a single environmental variable and broad geographic extent. Of this variation, 10.8% was partitioned at the inter-continental scale. Covariation with pH at this level
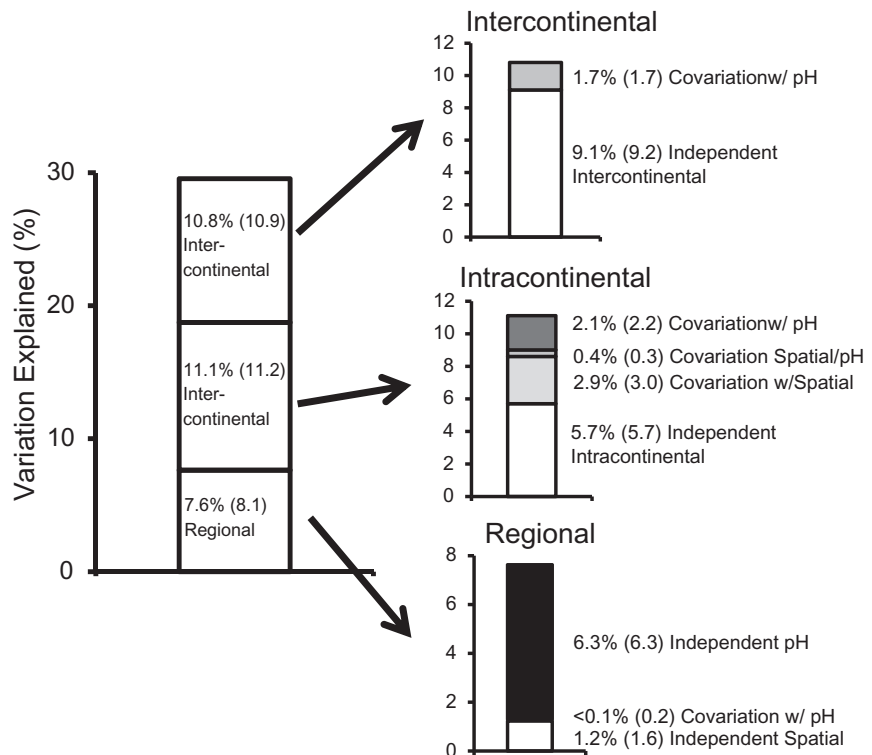

Figure 2 Variation explained by each level of variability explored in nested partial redundancy analyses (RDAs), and the proportion of total variability attributed to factors at each level. Corrected (Peres-Neto et al., 2006) values of explained variation are presented, as are uncorrected values (in brackets), to facilitate comparison with previous studies.

explained only 1.7% of the variation. While inter-continental differences in unmeasured environmental variables may be partly responsible for this variation, dispersal limitation was likely to be a prominent factor at this scale. Indeed, of the 119 dominant taxa in the dataset, 11 were found in all five European regional datasets and none of the North American regional datasets. However, with pH covariation factored out, less than a third of the total variation explained (9.1%) was partitioned at this level, a lower proportion than would be expected if inter-continental dispersal barriers were the predominant determinant of total variability.

The greatest amount (11.1%) of variation was explained at the intra-continental level, indicating systematic differences from region to region. Unsurprisingly, covariation with spatial variables accounted for part of this variation (2.9%), as regions in close proximity had somewhat similar flora (indeed, there was some overlap in the EMAP/PIRLA and SWAP/NOR regional surveys, Fig. 1). Covariation with pH (2.1%) also occurred at this level. After removal of these covariables, about half of the variation at this level (5.7%) was explained only by regional identities. Either inter-regional (e.g. topographic) dispersal barriers or the influence of untested environmental variables may explain this variation.

At the regional scale, representing areas ranging to nearly 1,000,000 km^2, over five times as much variation was explained by pH (6.3%) as by spatial variables (1.2%). Covariation of pH with spatial variables at this level was < 0.1%, providing distinct partitioning of variability. The much stronger influence of pH versus spatial variables at this level indicates that environment is clearly a more powerful driver of community variation than dispersal limitation at the regional scale. At the genus level, the strength of the pH signal at the regional scale was even higher, and was in fact the largest signal of all components at 7.6% variation explained (Fig. S1).

Niche conservatism and environmental response

Optimal pH for conspecifics from Europe and North America were highly correlated ($r = 0.85$, $P < 0.0001$; Fig. 3), indicating that the pH components of niches were largely preserved, even at the inter-continental scale. Generally, higher inferred pH optima for North America (*c.* 6.52 mean optimum) versus Europe (*c.* 6.10 mean optimum) were probably a sampling effect, as mean lake pH was 0.59 units higher in North America than Europe. Even with many lakes in a survey, such unintentional bias in lake choice can lead to different calculated optima (Battarbee *et al.*, 2008). A systematic, directional niche shift in virtually all taxa is extremely unlikely. The strength of prediction of lake pH measurements using diatom assemblages was also comparable for the synoptic and regional data (Table S4), despite the synoptic data deriving from assemblages from across continents, which would dilute the environmental signal if pH components of niches were not preserved. There was no relationship between region size and strength of calibration set ($r = 0.01$, $P = 0.75$).
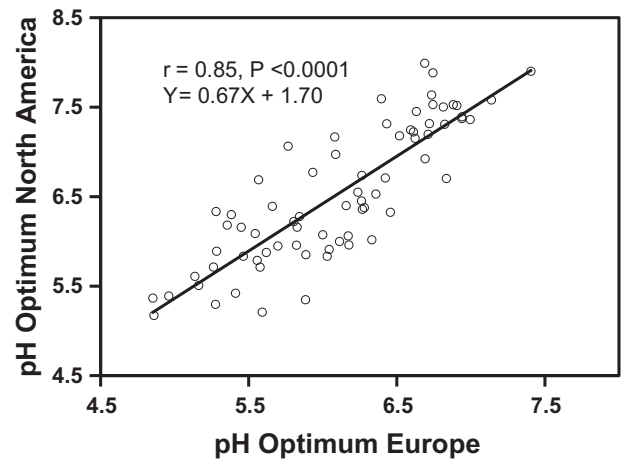

Figure 3 pH optima of taxa found in both North America and Europe ($n = 70$): optima calculated using only North American data versus optima calculated using European data.
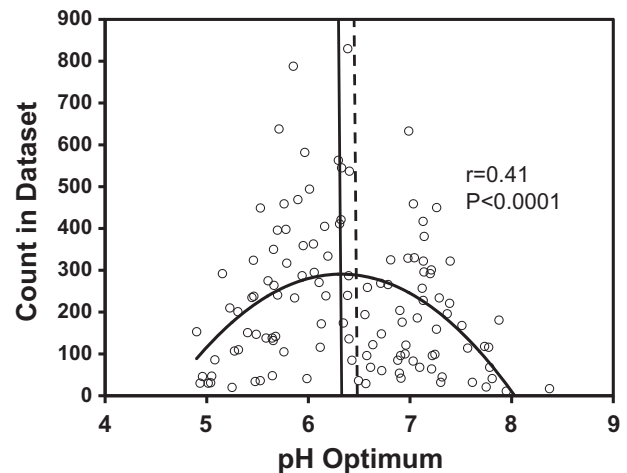

Figure 4 Quadratic regression of the number of lakes where taxa are present versus pH optima of taxa. The vertical solid line represents the optimum pH for count in the dataset (6.32); the hatched line represents the mean pH in the dataset (6.45).

There was a significant but weak quadratic relationship between pH optimum and occurrence in the dataset ($r = 0.41$; $P < 0.0001$ for all terms; Fig. 4). The optimal pH for occurrence was 6.32, which compares favourably with the mean pH in the dataset of 6.45. However, there was considerable scatter in this relationship, and many taxa with high and low pH optima were also common in the dataset. In fact, two of the three most commonly found taxa are apparently low-pH specialists (*Frustulia rhomboides*, 788 observations, pH optimum = 5.85; *Eunotia incisa*, 639 observations, pH optimum = 5.71). Both of these taxa were common due to near-ubiquity in acid lakes: *F. rhomboides* was observed in all regions and 99.4% of all lakes with pH < 6.0 and only 44.3% of lakes with pH > 7.0; *E. insisa* was observed in all regions and 96.4% of lakes with pH < 6.0 and only 24.1% of lakes with pH > 7.0. Because they accessed nearly all of the acidic

Table 1 Forward-selected variables from stepwise logistic regressions of the 10 most commonly-found taxa in the dataset.

Taxon	1st selected variable	2nd selected variable	3rd selected variable	4th selected variable
Tabellaria flocculosa	Region	Long.^{2} × lat.	Lat. × long.	
Frustulia rhomboides	pH	Region		
Eunotia incisa	pH	Region	Lat. × long.	Long.^{2} × lat.
Achnanthidium minutissimum	pH	Region	Lat.^{2} × long.	
Brachysira brebissonii	Region	pH		
Pinnularia biceps	Region	pH		
Fragilariforma exigua	Region	Lat.^{2}	Long.	
Encyonema lunatum	Region			
Chamaepinnularia mediocris	pH	Lat.^{3}	Region	
Psammothidium marginulatum	Region	pH	Long.^{3}	

environments, these taxa were encountered in more lakes than most taxa that were adapted to more common conditions. Of the 10 most widespread taxa, seven exhibited significant relationships with pH (Table 1), a proportion analogous to that found by Telford *et al.* (2006b) for all dominant taxa in a dataset. Specialization therefore did not ensure limited distribution.

DISCUSSION

Each of the three scales in our dataset encompasses large geographic areas, and consequently a pervasive signal of dispersal limitation might have been expected. Under such circumstances a likely scenario would be partitioning of most variation in the RDAs at the inter-continental level, as vastly different floras would be expected in Europe versus North America, given the dispersal barrier of the ocean and the much larger geographic distance between the continents than among and within regions. Diminishing amounts of variation would be explained at the intra-continental and then regional levels, with variation explainable by pH being overwhelmed by that attributed to the large geographic distances at every scale. Under even greater dispersal limitation, floras among regions would be very dissimilar and the regional spatial signal would predominate. Conversely, under a high degree of environmental determinism, pH would explain a large amount of variability, either independently or covarying with scale and/or spatial variables. Under a combination of dispersal limitation and environmental determinism, variability attributed to pH would increase with decreasing spatial scale, assuming either a dominant or subordinate role to spatial variables at the regional (*c*. 10,000 to *c*. 1,000,000 km^2) scale. Our observed pattern obviously reflects the latter combination, from dispersal limitation at the largest (inter-continental) scale, to environment dominating at the still considerable distances represented by the regional scale. If additional environmental variables had been available throughout the dataset, the results of Ginn *et al.* (2007) suggest they would explain additional environmental variation at the intra-continental and regional scales, secondary to that explained by pH.

This finding has significant theoretical and practical implications. Community variation, rather than being a matter of 'dispersal or environment' or 'how much dispersal and how much environment', is driven in this case by 'dispersal then environment' as the spatial focus narrows from the intercontinental to regional scales. While mass effects are conceivable for riverine systems (Soininen & Weckström, 2009), in landscapes of lakes separated by large stretches of terrestrial environment, mass effects precipitating substantial sink populations are extremely unlikely. The community in a lake therefore reflects available propagules from the surrounding region, as well as the local environmental conditions. Over very large distances some community constituents are replaced, but those established in a given environment must be adapted to it. Verleyen *et al.* (2009) found similar patterns in diatom communities, whereby intraregional environmental signals were much stronger than spatial signals and spatial signals increased with distance, though their analysis did not independently partition spatial versus environmental variation (or covariation) across scales, and their interregional and inter-continental comparisons were at a coarser taxonomic scale, potentially obscuring biogeographic differences. Results reported by Van der Gucht *et al.* (2007) suggest that lacustrine bacterial communities respond similarly to the environment. Indeed, we suspect, based on additional evidence provided by Shurin *et al.* (2009), that other small, highly dispersive lacustrine taxa also respond in this way. Less dispersive organisms may, however, respond quite differently: Karst *et al.* (2005) and Girdler & Barrie (2008) found evidence that dispersal limitation for vascular plants may occur at smaller scales, while environment operates as a predominantly larger-scale filter; however, continental-scale analysis of species distributions by Svenning *et al.* (2008) indicate that dispersal limitation may mitigate environmental influence on plant distributions at the largest scales. For such communities, extensive studies combining all pertinent scales are clearly also needed before a comprehensive perspective on community drivers can be achieved.

The apparent conservatism of niches across continents is another key aspect of environmental influence on biogeographic patterns in diatom communities. While niche conservatism is not an uncommon phenomenon in nature (Wiens & Graham, 2005), the fact that it appears to occur at such large geographic distance and among so many diatom taxa indicates that over time-scales relevant to local adaptation, diatom meta-

communities may be geographically very large, and that the conservatism of niches may be a general characteristic of these organisms. Ready availability of a variety of highly fecund propagules with different environmental niches, over shorter time periods than directional change in local environments, could dictate that, in these communities, innate compatibility with the environment may be a more important driver of species distributions than local adaptation. Indeed, in resource-limited environments, such conditions could conceivably suppress local adaptation to changing conditions, if pre-adapted taxa can dominate faster than erstwhile dominants can adapt to change. The fact that there was little evidence that environmental specialization equated with rarity in the dataset is also indicative of a scenario of widely dispersed taxa with a variety of environmental preferences. This finding is contrary to predictions that widely distributed taxa must be generalists, broadly adapted to the most common environments, and that taxa with distinct niches are uncommon (Hubbell, 2005; Pither, 2007), and shows that specialization does not ensure limited distribution provided an organism can disperse to all areas for which it is specialized.

In analyses of spatial versus environmental community patterns, the twin difficulties of covariation of spatial and environmental signals plus the impossibility of knowing whether all important environmental variables have been measured are well recognized (e.g. Cottenie, 2005). In our case, the dominant independent signal of a *single* environmental variable over multiple forward-selected spatial variables at the regional level plus niche conservatism across continents argues strongly for determinism of taxonomic abundances of diatoms by environment at spatial scales of less than *c.* 1 000 000 km^2, and indicates that diatoms, and probably other lacustrine microbial communities, may be used to accurately infer environmental change over large geographic areas. Though replacement of some taxa over very large distances may occur due to dispersal limitation, correlation with predicted environment is maintained as niches of dominant taxa are preserved from region to region, and relatively little dispersal limitation exists at the scale considered by most analyses. Thus, for modelling past environments, a high degree of confidence can exist that if a downcore community closely resembles an extant distant community, then the past environment was similar to that inhabited by the distant community. Considering our results, the very presence of such a community downcore argues for a relatively unimportant role of dispersal limitation over time-scales considered by palaeoecological analyses. Patterns at the genus level indicate even stronger environmental signals at this coarser taxonomic scale (Fig. S1), though analyses on the genus level may miss important biogeographic patterns present at lower levels. Predictions of the exact composition of future floras given pH changes would be more uncertain than relating environment to past assemblages, as many taxa appear to have similar pH optima (Fig. 4); though given a more comprehensive set of environmental and spatial variables, reasonable predictions could probably be made.

The ability to predict community responses to environmental change depends on the interaction of temporal and spatial scales. In a large dataset of diatom communities, we have shown niche conservatism and environmental response over wide geographic areas. Other studies have demonstrated rapid responses of diatom communities to local environmental change, including nearly complete turnover in < 10 years (Laird *et al.*, 1996; Finney *et al.*, 2000). Diatoms, and probably other microbial communities, possess key characteristics that make them excellent indicators of environmental change, compared with other communities for which the environment/community connection is weaker or dispersal limitation occurs at smaller scales. For many such groups, the influences of community composition remain unresolved, while the need for accurate predictions becomes more acute as anthropogenic environmental change escalates. In these cases, extensive, multi-scale analyses similar to ours would help to resolve controversy regarding environment versus dispersal relationships, by locating the scales at which a given factor predominates and at which useful predictions may be made.

REFERENCES

Alonso, D., Etienne, R.S. & McKane, A.J. (2006) The merits of neutral theory. *Trends in Ecology and Evolution*, **21**, 451–457.

Battarbee, R.W., Juggins, S., Gasse, F., Anderson, N.J., Bennion, H., Cameron, N.G., Ryves, D.B., Paives, C., Chalié, F. & Telford, N. (2001) *European Diatom Database (EDDI): an information system for palaeoenvironmental reconstruction*. ECRC Report 81. Available at: http://craticula.ncl.ac.uk/Eddi/jsp/index.jsp (accessed November 2007).

Battarbee, R.W., Monteith, D.T., Juggins, S., Simpson, G.L., Shilland, E.W., Flower, R.J. & Kreiser, A.M. (2008) Assessing the accuracy of diatom-based transfer functions in defining reference pH conditions for acidified lakes in the United Kingdom. *The Holocene*, **18**, 57–67.

Benedetti-Cecchi, L. (2007) Neutral theory and 1/*f* noise make similar predictions of assemblage dynamics. *Trends in Ecology and Evolution*, **22**, 231.

Birks, H.J.B., Line, J.M., Juggins, S., Stevenson, A.C. & Braak, C.J.F. (1990) Diatoms and pH reconstruction. *Proceedings of the Royal Society B: Biological Sciences*, **327**, 263–278.

Blanchet, F.G., Legendre, P. & Borcard, D. (2008) Forward selection of explanatory variables. *Ecology*, **89**, 2623–2632.

Borcard, D. & Legendre, P. (2002) All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecological Modelling*, **153**, 51–68.

Borcard, D., Legendre, P. & Drapeau, P. (1992) Partialling out the spatial component of ecological variation. *Ecology*, **73**, 1045–1055.

ter Braak, C.J.F. & Šmilauer, P. (2002) *CANOCO reference manual and Canodraw for Windows user's guide: software for canonical community ordination, version 4.5.* Microcomputer Power, Ithaca, NY.

Cottenie, K. (2005) Integrating environmental and spatial processes in ecological community dynamics. *Ecology Letters*, **8**, 1175–1182.

Cushman, S.A. & McGarigal, K. (2002) Hierarchical, multi-scale decomposition of species-environment relationships. *Landscape Ecology*, **17**, 637–646.

Dixit, S.S., Cumming, B.F., Birks, H.J.B., Smol, J.P., Kingston, J.C., Uutala, A.J., Charles, D.F. & Camburn, K.E. (1993) Diatom assemblages from Adirondack lakes (New York, USA) and the development of inference models for retrospective environmental assessment. *Journal of Paleolimnology*, **8**, 27–47.

Dixit, S.S., Smol, J.P., Charles, D.F., Hughes, R.M., Paulsen, S.G. & Collins, G.B. (1999) Assessing water quality changes in the lakes of the northeastern United States using sediment diatoms. *Canadian Journal of Fisheries and Aquatic Sciences*, **56**, 131–152.

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

Finney, B.P., Gregory-Eaves, I., Sweetman, J., Dougas, M.S.V. & Smol, J.P. (2000) Impacts of climatic change and fishing on Pacific salmon abundance over the past 300 years. *Science*, **290**, 795–799.

Ginn, B.K., Cumming, B.F. & Smol, J.P. (2007) Diatom-based environmental inferences and model comparisons from 494 northeastern North American lakes. *Journal of Phycology*, **43**, 647–661.

Girdler, E.B. & Barrie, B.T.C. (2008) The scale-dependent importance of habitat factors and dispersal limitation in structuring Great Lakes shoreline plant communities. *Plant Ecology*, **198**, 211–223.

Guisan, A. & Thuiller, W. (2005) Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, **8**, 993–1009.

Hubbell, S.P. (2005) Neutral theory in community ecology and the hypothesis of functional equivalence. *Functional Ecology*, **19**, 166–172.

Karst, J., Gilbert, B. & Lechowicz, M.J. (2005) Fern community assembly: the roles of chance and the environment at local and intermediate scales. *Ecology*, **86**, 2473–2486.

Laird, K.R., Fritz, S.C., Maasch, K.A. & Cumming, B.F. (1996) Greater drought intensity and frequency before AD 1200 in the northern Great Plains, USA. *Nature*, **384**, 552–554.

Legendre, P. & Gallagher, E.D. (2001) Ecologically meaningful transformations for ordination of species data. *Oecologia*, **129**, 271–280.

Legendre, P., Borcard, D. & Peres-Neto, P.R. (2005) Analyzing beta diversity: partitioning the spatial variation of community composition data. *Ecological Monographs*, **75**, 435–450.

Legendre, P., Borcard, D. & Peres-Neto, P.R. (2008) Analyzing or explaining beta diversity? Comment. *Ecology*, **89**, 3238–3244.

McGill, B.J., Maurer, B.A. & Weiser, M.D. (2006) Empirical evaluation of neutral theory. *Ecology*, **87**, 1411–1423.

Martiny, J.B.H., Bohannan, B.J.M., Brown, J.H., Colwell, R.K., Fuhrman, J.A., Green, J.L., Horner-Devine, M.C., Kane, M., Krumins, J.A., Kuske, C.R., Morin, P.J., Naeem, S., Øvreås, L., Reysenbach, A.L., Smith, V.H. & Staley, J.T. (2006) Microbial biogeography: putting microorganisms on the map. *Nature Reviews Microbiology*, **4**, 102–112.

Økland, R.H. (1999) On the variation explained by ordination and constrained ordination axes. *Journal of Vegetation Science*, **10**, 131–136.

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.

Pither, J. (2007) Comment on 'Dispersal limitations matter for microbial morphospecies'. *Science*, **316**, 1124b.

Pither, J. & Aarssen, L.W. (2005) Environmental specialists: their prevalence and their influence on community-similarity analyses. *Ecology Letters*, **8**, 261–271.

Sherman, J.W. & Fairchild, G.W. (1994) The interaction between acidity and nutrient chemistry in softwater lakes of northeastern Pennsylvania. *Journal of the Pennsylvania Academy of Sciences*, **68**, 56–62.

Shurin, J.B., Cottenie, K. & Hillebrand, H. (2009) Spatial autocorrelation and dispersal limitation in freshwater organisms. *Oecologia*, **159**, 151–159.

Smol, J.P. (2008) *Pollution of lakes and rivers: a paleoenvironmental perspective*, 2nd edn. Blackwell Publishing, Malden, MA.

Smol, J., Birks, H.J.B. & Last, W.M. (eds) (2001) *Tracking environmental change using lake sediments. Volume 3, Terrestrial, algal, and siliceous indicators*. Kluwer, Dordrecht.

Soininen, J. & Weckström, J. (2009) Diatom community structure along environmental and spatial gradients in lakes and streams. *Fundamental and Applied Limnology*, **174**, 205–213.

Sullivan, T.J., Charles, D.F., Smol, J.P., Cumming, B.F., Selle, A.R., Thomas, D.R., Bernert, J.A. & Dixit, S.S. (1990) Quantification of changes in lakewater chemistry in response to acidic deposition. *Nature*, **345**, 54–58.

Svenning, J. & Skov, F. (2004) Limited filling of the potential range in European tree species. *Ecology Letters*, **7**, 565–573.

Svenning, J. & Skov, F. (2007) Could the tree diversity pattern in Europe be generated by postglacial dispersal limitation? *Ecology Letters*, **10**, 453–460.

Svenning, J., Normand, S. & Skov, F. (2008) Postglacial dispersal limitation of widespread forest plant species in nemoral Europe. *Ecography*, **31**, 316–326.

Telford, R.J., Vandvik, V. & Birks, H.J.B. (2006a) Dispersal limitations matter for microbial morphospecies. *Science*, **312**, 1015–1015.

Telford, R.J., Vandvik, V. & Birks, H.J.B. (2006b) How many freshwater diatoms are pH specialists? A response to P (2005). *Ecology Letters*, **9**, E1–E5.

Tuomisto, H. & Ruokolainen, K. (2008) Analyzing or explaining beta diversity? Reply. *Ecology*, **89**, 3244–3256.

Van der Gucht, K., Cottenie, K., Muylaert, K., Vloemans, N., Cousin, S., Declerck, S., Jeppesen, E., Conde-Porcuna, J., Schwenk, K., Zwart, G., Degans, H., Vyverman, W. & De Meester, L. (2007) The power of species sorting: local factors drive bacterial community composition over a wide range of spatial scales. *Proceedings of the National Academy of Sciences USA*, **104**, 20404–20409.

Vanormelingen, P., Verleyen, E. & Vyverman, W. (2008) The diversity and distribution of diatoms: from cosmopolitanism to narrow endemism. *Biodiversity and Conservation*, **17**, 393–405.

Verleyen, E., Vyverman, W., Sterken, M., Hodgson, D.A., De Wever, A., Juggins, S., Van de Vijver, B., Jones, V.J., Vanormelingen, P., Roberts, D., Flower, R., Kilroy, C., Souffreau, C. & Sabbe, K. (2009) The importance of dispersal related and local factors in shaping the taxonomic structure of diatom metacommunities. *Oikos*, **118**, 1239–1249.

Verschuren, D., Laird, K.R. & Cumming, B.F. (2000) Rainfall and drought in equatorial East Africa during the past 1100 years. *Nature*, **403**, 410–414.

Vyverman, W., Verleyen, E., Sabbe, K., Vanhoutte, K., Sterken, M., Hodgson, D.A., Mann, D.G., Juggins, S., Van de Vijver, B., Jones, V., Flower, R., Roberts, D., Chepurnov, V.A., Kilroy, C., Vanormelingen, P. & De Wever, A. (2007) Historical processes constrain patterns in global diatom diversity. *Ecology*, **88**, 1924–1931.

Welk, E. & Bruelheide, H. (2006) There may be bias in R/P ratios (realized vs. potential range) calculated for European tree species – an illustrated comment on S (2004). *Journal of Biogeography*, **33**, 2013–2018.

Wiens, J.J. & Graham, C.H. (2005) Niche conservatism: integrating evolution, ecology, and conservation biology. *Annual Review of Ecology, Evolution, and Systematics*, **36**, 519–539.

SUPPORTING INFORMATION

Additional Supporting Information can be found in the online version of this article:

Appendix S1 Nested partial redundancy analyses (RDAs) and prediction of pH for regional datasets.
Figure S1 Variation in genus-level data explained by each level of variability explored in nested partial redundancy analyses (RDAs).
Table S1 Taxonomic harmonization for dominant (5% in three lakes) taxa.
Table S2 Nested partial redundancy analyses (RDAs).
Table S3 Algebra for determining variability in nested partial redundancy analyses (RDAs).
Table S4 Coefficients of determination for subordinate datasets.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

BIOSKETCH

Joseph Bennett is a PhD student in the Centre for Applied Conservation Research at the University of British Columbia. His research concentrates on the interactions of spatial and temporal scales with determinants of community composition on islands and lakes, as well as the development and testing of spatial statistics.

Editor: José Alexandre F. Diniz-Filho