

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Running title: Diatom beta diversity through time and space

Title: Diatom diversity patterns over the past ~ 150 years across the conterminous United States: identifying mechanisms behind beta diversity

Authors: Amanda K. Winegardner^{*1,2}, Pierre Legendre^{2,3}, Beatrix E. Beisner^{2,4}, and Irene Gregory-Eaves^{1,2}

Affiliations:

1. Department of Biology, McGill University, 1205 avenue du Docteur Penfield, Montréal, Québec, Canada, H3A 1B1

2. Québec Centre for Biodiversity Science, 1205 avenue du Docteur Penfield, Montréal, Québec, Canada, H3A 1B1

3. Département de sciences biologiques, Université de Montréal, C.P. 6128, succursale Centre-ville, Montréal, Québec, Canada H3C 3J7

4. Department of Biological Sciences, Université du Québec à Montréal, C.P. 8888, succursale Centre-ville, Montréal, Québec, Canada H3C 3P8

*corresponding author.

Email addresses: amanda.winegardner@gmail.com, pierre.legendre@umontreal.ca, beisner.beatrix@uqam.ca, irene.gregory-eaves@mail.mcgill.ca

Keywords: Beta diversity, community ecology, diatoms, freshwater, paleoecology, paleolimnology

No. of words in Abstract: 300

No. of words in main body (Introduction through Biosketch): 6279

No. of references: 54

Abstract

Aim: Understanding the magnitude and drivers of freshwater diversity over the last 150 years provides essential insights for developing scenarios of future change. Here we quantify and identify drivers of spatial and temporal beta diversity in diatom assemblages between historical and modern times.

Location: United States of America

Methods: Using sedimentary genus-level diatom data from 169 lakes, and species-level data for 52 lakes, we computed spatial beta diversity across all lakes and within ecoregions for 2007 and pre-1850 CE time points. We also computed local contributions to beta diversity (LCBD) and analysed them with respect to environmental variables. Total beta diversity was partitioned into replacement and abundance difference components to identify mechanisms possibly responsible for spatial beta at each time point. Temporal Beta diversity Indices (TBI) were also computed for each lake by comparing the diatom data of all lakes at the time points. TBIs were decomposed into taxon losses and gains to facilitate interpretation. TBIs and their components were related to contemporary land cover.

Results: Temporal beta diversity varied significantly as a function of forest cover with higher temporal beta in lakes from watersheds with contemporary lower forest cover. Spatial beta diversity was similar between the historical and 2007 time points. Lakes with substantial local contributions to beta diversity were differentiated by water quality and land cover variables at a local scale, but showed no systematic regional pattern.

Main conclusions: Spatial beta diversity of diatoms across the US does not appear to have changed between pre-1850 CE and 2007, suggesting that broad-scale land use and hydrological alteration of the landscape has not homogenized these communities. Temporal beta diversity

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

occurred through genus gains and losses and was significantly related to land cover in
watersheds. These analyses, pairing spatial and temporal beta diversity, provide insight into the
mechanisms maintaining diatom diversity.

For Peer Review

Introduction

Modern geological time is commonly referred to as the Anthropocene, a designation recognizing the extent to which humans presently dominate processes and life on the Earth (Crutzen, 2002). Although there is still some debate over its exact onset date (Lewis & Maslin, 2015; responses, Hamilton, 2015; Zalasiewicz *et al.* 2015), the Anthropocene concept has been quickly taken up by the scientific community and the general public. Thus, much recent biodiversity research has sought to understand the distribution and drivers of biodiversity loss concomitant with habitat losses documented over the Anthropocene (Steffen *et al.* 2011), which are expected to compromise ecosystem functioning and services (Cardinale *et al.* 2012). However, several studies, including recent meta-analyses, have shown that temporal species richness trends (local or regional) may be flat or that positive and negative relationships between species richness and time may occur in equal proportions (e.g. Vellend *et al.* 2013; Dornelas *et al.* 2014) or even show only modest losses (Newbold *et al.* 2015) over the last few centuries. Nonetheless, it is more generally accepted that there have been substantial shifts in community composition (Vellend *et al.* 2013; Dornelas *et al.* 2014; Newbold *et al.* 2015) implicating both species losses *and* gains in biodiversity and associated ecosystem shifts (Wardle *et al.* 2011; McGill *et al.* 2015). Advancing our understanding in this area is particularly important in freshwater ecosystems, which hold a disproportionately concentrated diversity relative to other environments (Strayer & Dudgeon, 2010).

Recent literature has highlighted the importance of different metrics of biodiversity and scales of study. Many biodiversity studies have focused on species richness in a site (alpha diversity), or across a region or several sites (gamma diversity). Beta diversity focuses on the differentiation of communities among sites or through time in terms of number of species and

1
2
3 86 composition (Whittaker, 1972). Spatial beta diversity is particularly interesting because it can
4
5 87 identify sites and regions that are exceptional across a landscape owing to degraded or enhanced
6
7
8 88 diversity. Temporal beta diversity can also indicate degrees of change in composition through
9
10 89 time at single locations. Despite the usefulness of considering numerous biodiversity metrics at
11
12 90 multiple scales, very few studies have attempted to consider more than a few trends
13
14 91 simultaneously (McGill *et al.* 2015). The majority of biodiversity change studies have also been
15
16 92 largely restricted to the last 50 years, with only a handful extending beyond this time frame to
17
18 93 encompass the entire Anthropocene. This literature has been recently synthesized to develop
19
20 94 predictions of various biodiversity trends during the Anthropocene (McGill *et al.* 2015, Table 1),
21
22 95 together with a call to test these predictions.
23
24
25
26

27 96 Aquatic systems provide a unique opportunity to quantify several biodiversity trends over
28
29 97 different spatial and temporal scales through the use of the historical archives found in lake
30
31 98 sediments (Gregory-Eaves & Beisner, 2011). Diatoms are especially well suited to this type of
32
33 99 study because of their wide use as indicators of environmental change in the Anthropocene
34
35 100 (Smol & Stoermer 2010); making them useful for evaluating relationships between time and beta
36
37 101 diversity, such as in Vellend *et al.* (2013) and Dornelas *et al.* (2014)). We analysed a large
38
39 102 paleolimnological dataset to quantify diversity trends across both space and time and explore
40
41 103 mechanisms behind diatom assemblage variation across the conterminous US during the
42
43 104 Anthropocene. We used sediment cores collected as part of the 2007 National Lakes Assessment
44
45 105 (NLA), United States Environmental Protection Agency (U.S. EPA;
46
47 106 http://water.epa.gov/type/lakes/lakessurvey_index.cfm) to examine patterns of diatom genera and
48
49 107 species richness, as well as alpha, beta and gamma diversities for 169 lakes. To identify areas of
50
51 108 high beta diversity and relate these to water quality and land cover, we computed spatial beta
52
53
54
55
56
57
58
59
60

1
2
3 109 diversity across all lakes at both the historical (pre-1850 CE) and modern (2007) time points.
4
5
6 110 We also calculated temporal beta diversity (change over the last ~ 150 years) within each lake to
7
8 111 identify compositional variation between the historical and modern time points (Figure 1).
9

10
11 112 The goal of our study was to investigate the relative magnitude of spatial beta diversity
12
13 113 across different ecoregions at each time point, as well as the magnitude of temporal beta
14
15 114 diversity for each lake. We were also interested in mechanisms behind spatial and temporal beta
16
17 115 diversity to better characterize what differences in beta diversity might mean in an ecological
18
19
20 116 context. For example, if lakes in a certain geographic area showed higher change through time
21
22 117 on average than a different region, what environmental drivers or community dynamics might
23
24 118 explain this pattern? To explore mechanisms behind spatial and temporal beta diversity, we
25
26
27 119 employed partitioning techniques described in Legendre (2014) and Legendre & Salvat (2015).
28
29
30 120 We partitioned spatial beta diversity into components explained by the replacement of
31
32 121 genera/species as well as differences in taxa abundances. For temporal beta diversity, we
33
34 122 partitioned this metric into beta diversity explained by either taxa loss or gains.
35

36
37 123 While our study focused on a specific set of lakes, we wanted to ensure that these results
38
39 124 would be broadly applicable, and so couched our hypotheses within the trends identified by
40
41 125 McGill *et al.* (2015). Specifically, McGill *et al.* (2015) developed hypotheses and trends related
42
43 126 to alpha, gamma and beta diversity at three different geographic scales (biogeographic,
44
45
46 127 metacommunity and local) which can be compared to the scales used in this study: continental
47
48 128 US, ecoregion and individual lakes (Table 1). For spatial beta diversity, there could be many
49
50 129 different expectations for how spatial beta diversity might change through time. For example,
51
52
53 130 spatial beta diversity might be less across a large landscape in modern times because of increases
54
55 131 in biotic homogenization (as shown in Table 1). This could happen even in a landscape where
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

alpha diversity is increasing through time (as suggested in Table 1) or where alpha diversity is constant through time as found in studies such as Vellend *et al.* (2013)). Through time, McGill *et al.* (2015) notes that alpha diversity generally remains constant at a local-scale (lake-scale). However, we predicted that we would find significant temporal beta diversity at that scale reflecting changes in community composition through time. By combining these hypotheses with partitioning methods for both spatial and temporal beta diversity, we can test hypotheses as well as provide insight into mechanisms behind observed trends.

Methods

Description of the 2007 National Lakes Assessment (NLA)

The 2007 NLA survey took place from May to October 2007, sampling over 1000 lakes and reservoirs from the conterminous U.S. (U.S. EPA, 2009). Lakes were all >0.04 km² and deeper than 1 m (U.S. EPA, 2009). The lakes were selected using a combination of probabilistic design and specifically targeted ‘reference’ lakes identified by state and tribal partners (U.S. EPA, 2011-2012). The 2007 NLA data includes water quality measurements, land cover metrics, and compositional data for zooplankton and phytoplankton, including diatom assemblages from both water column and lake sediment samples. The data are publicly available from: http://water.epa.gov/type/lakes/NLA_data.cfm. Further details of the sampling are summarized in Beaulieu *et al.* (2013) and Winegardner *et al.* (2015).

Sediment core screening

NLA field teams collected sediment cores from a subset of the sampled lakes, mainly where sampling teams estimated that the sediment bank at the deepest point of the lake would be undisturbed, such that the bottom part of the core would reach sediments representing pre-1850

1
2
3 155 CE conditions. Pre-1850 CE is a date used in many paleolimnological studies as a cut-off for
4
5
6 156 distinguishing between pre- and post-industrial conditions (e.g. see Smol (1992); Rühland *et al.*
7
8 157 (2003); Hyatt *et al.* (2011)). While we recognised that there was the potential for some sites in
9
10 158 this study to have experienced anthropogenic impacts prior to 1850 CE, the majority of
11
12
13 159 disturbance affecting diatoms across the U.S. landscape was expected to have occurred post-
14
15 160 1850 CE, during the Anthropocene. Cores were collected from the deepest point of the lake
16
17 161 using a modified Kajak-Brinkhurst corer (Glew, 1989). Both the top and bottom 1 cm interval of
18
19 162 the sediment was saved for diatom enumeration, with up to 500 diatom valves counted using
20
21 163 standardized methods from the U.S. Geological Survey National Water Quality Assessment
22
23 164 (Charles *et al.* 2003, U.S. EPA, 2011-2012). Thus, for each lake where sediment coring occurred,
24
25 165 there are diatom assemblage data for both modern (2007) conditions as well as bottom of the
26
27 166 core (historical) conditions. To account for different core lengths and variation in sedimentation
28
29 167 across lakes, we used screening criteria to select lakes for our study, ensuring that the bottom
30
31 168 sediment samples used represented historical (pre-1850 CE) conditions. We used a three-fold
32
33 169 approach to estimate age of sediment cores (described in S1) and identified 179 sites for which
34
35 170 we were confident that their bottom samples represented pre-1850 CE conditions. All remaining
36
37 171 lakes in the Coastal Plains ecoregion (all located in Florida) were removed because of the
38
39 172 relative large variability in sedimentation rates at these latitudes. The final set of lakes used in
40
41 173 this study (n = 169; see S1 for information on strata) is shown in Figure 2a, along with ranges of
42
43 174 key limnological variables in Figure 2b-e. Note that there is a geographic bias in the NLA data;
44
45 175 bottom core samples that were not collected from many lakes from the mid- and southern
46
47 176 portions of the country where reservoirs dominate because the U.S. EPA could not verify
48
49 177 whether sediments would be unmixed.
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Diversity analyses

All statistical analyses were completed in R v. 3.1.2 (R Core Team, 2014). Enumeration of diatoms was conducted by four different laboratory groups for the 2007 NLA (Pollard, Pers. Comm.). A sub-section of samples were re-counted by a second taxonomist, different from the original counting group for Quality Assurance/Quality Control (QA/QC). These QA/QC procedures resulted in a Percent Taxonomic Disagreement of ~20% between different taxonomists re-counting the same slide at the genus level, a rate of disagreement considered as acceptable to explore large-scale patterns across the full set of lakes (n = 169). For a smaller set of the lakes (n = 52), we completed all the analyses again using the species level data because for this smaller set, all samples had been enumerated by a single group (The Academy of Natural Sciences, Philadelphia PA). We calculated diatom species or genus richness in three different ways: total richness for all taxa, rarefied richness for all taxa, and rarefied richness for dominant taxa only. “Dominant” species and genera were those having greater than 2% relative abundance in at least one sample from either surface or historical sediments. We calculated rarefied taxa richness, correcting for the total number of valves counted for each sample and setting all samples to a cut-off of 300 valves (150 individuals; and a point where taxa richness had plateaued; *rarefy()* function in *vegan* (Oksanen *et al.* 2016). Thus, the single sample with ~300 valves could be compared to samples with a higher total abundance (the majority had ~500 valves counted). All subsequent analyses were done using relative abundance data and only dominant (>2% relative abundance in at least one sample) diatom taxa. We calculated alpha diversity using a Shannon index and Simpson’s index (evenness) using *diversity()* in *vegan* and further transformed Shannon-Weiner diversity (*H*) to an effective diversity number ($\exp(H)$) as suggested by Ellison (2010). Quantitative (relative abundance) data was used for all genus-level

201 and species-level analyses; however key beta diversity findings were verified using presence-
202 absence data as well.

203 We computed spatial beta diversity for all sites (either $n = 169$ or $n = 52$) across the
204 ecoregions at both the historical and contemporary time points, using non-transformed matrices
205 of diatom relative abundances and the percentage difference dissimilarity described in Legendre
206 & De Cáceres (2013); the percentage difference is the quantitative form of the Sørensen
207 presence-absence index, and was calculated using the R function *beta.div()* and *beta.div.comp()*.
208 As there are many different ways to compute spatial beta diversity, we also verified our spatial
209 beta diversity analyses using the R package *betapart* () (Baselga & Orme, 2012). An
210 overview of these additional analyses are shown in Appendix S2. There were 94 genera in the
211 2007 dataset and 105 genera in the historical dataset (926 species in 2007; 1001 species in the
212 historical set). We also computed spatial beta diversity at the two different time points within
213 five ecoregions, defined by Herlihy *et al.* (2008): Northern Appalachians, Southern Plains,
214 Temperate Plains, Upper Midwest, and Western Mountains (Xeric was excluded from spatial
215 beta diversity computations because $n = 1$) (Figure 2a, $n = 176$). The function *beta.div* () also
216 provided a 'Local Contribution to Beta Diversity' (LCBD index) for each lake as well as a
217 permutational p-value (permutations = 9999), indicating the significance of each LCBD value. P-
218 values were then corrected for multiple testing using Holm's procedure; see Legendre &
219 Legendre (2012, p. 23). LCBD values provide a metric to assess the individual contribution of
220 each lake to the total spatial beta diversity (Legendre & De Cáceres, 2013). We identified the
221 sites with significant LCBD that were the same or different between the historical and 2007
222 sediments and performed a chi-square test of the null hypothesis that processes producing
223 significant LCBD were independent at the two time points, and further verified this result by

1
2
3 224 computing a Pearson correlation between the historical and 2007 LCBD values over all lakes. To
4
5 225 assess individual species' influences on beta diversity, we computed 'Species Contributions to
6
7
8 226 Beta Diversity' (SCBD indices) using the Hellinger transformed abundance data (Legendre & De
9
10 227 Cáceres, 2013). SCBD cannot be computed through a dissimilarity measure. SCBD allowed us to
11
12 228 identify species that were important contributors to spatial beta diversity, both historically and in
13
14 229 modern samples. We used four source functions developed by P. Legendre for use in R for the
15
16 230 beta diversity analyses. We used *beta.div()* (Legendre & De Cáceres, 2013, App. S4),
17
18 231 *beta.div.comp()* (Legendre, 2014, App. S3), *LCBD.comp()* (Legendre, 2014, App. S5) for spatial
19
20 232 beta diversity analyses and *TBI()* (Legendre, 2015) for temporal beta diversity analyses. The first
21
22 233 three functions are available in package *adespatial* in R (Dray *et al.* 2016).
23
24
25
26
27

28 234 We were interested in the mechanisms generating spatial beta diversity at both time
29
30 235 points, namely dynamics referred to as species replacement and abundance or richness difference
31
32 236 Species replacement refers to the "simultaneous gain and loss of species (or individuals
33
34 237 belonging to a particular species, or biomass) along an ecological gradient" (Podani & Schmera,
35
36 238 2011; Baselga & Orme, 2012). This means that as species (measured either in terms of
37
38 239 abundance or biomass) are lost from sites (space, time, etc.), others take their place. On the
39
40 240 contrary, richness difference (or abundance difference when using abundance data) means that
41
42 241 one site (or sampling unit) has more unique species than another (Podani & Schmera, 2011;
43
44 242 Baselga & Orme, 2012). Using *beta.div.comp()* and the percentage difference index (alias Bray-
45
46 243 Curtis index) we partitioned the total beta diversity into replacement and abundance (or richness)
47
48 244 difference (Podani decomposition, *sensu* Podani *et al.* 2013) components such that, in addition to
49
50 245 a mean total beta diversity value for the landscape and each individual ecoregion, we also knew,
51
52 246 for each time point separately, the proportions of the total (spatial) beta diversity that are
53
54
55
56
57
58
59
60

explained by replacement and abundance difference. We then computed LCBD indices from the replacement and abundance difference matrices using function *LCBD.comp()*.

To explain the variation in the extent to which lakes contribute to spatial beta diversity (e.g. the magnitude of LCBD indices), we employed a univariate regression tree (URT) approach. URTs split a response variable (here the vector of LCBD indices) along gradients in explanatory variables, creating groups with similar values of the response (De'ath, 2002). URTs are a good way to identify breaks or thresholds in response variables in relation to explanatory variables without extensive data transformation and produce an intuitive representation of how a response variable varies along one or several environmental gradients. We computed regression trees using two different sets of variables. First, we built a regression tree with each lake's LCBD values of the 2007 data as a response variable, and the following explanatory variables (Supplement S4a for PCA), recorded for each lake in 2007: latitude, longitude, ecoregion, mean Secchi depth (m), chlorophyll a concentration ($\mu\text{g L}^{-1}$), total phosphorus (TP; $\mu\text{g L}^{-1}$), total nitrogen (TN; $\mu\text{g L}^{-1}$), mean temperature ($^{\circ}\text{C}$), specific conductivity ($\mu\text{S cm}^{-1}$ at 25°C) and pH. Second, we built a tree using basin-level land cover variables (summarized by the NLA and based on the 1992 National Land Cover Database; <http://www.mrlc.gov/nlcd1992.php>; Supplement S4b). These variables included percent land cover in each lake basin for the following land cover types: developed (split into low intensity residential, medium intensity residential, high intensity residential, open space), barren, forest (split into deciduous forest, coniferous forest, mixed forest), grasslands, agriculture (split into pasture and row crops), and wetland (split into woody wetlands and emergent herbaceous wetlands) (USGS, 1992). The response variable for these regression trees, LCBD, is a measure of the unique contribution of a specific site to beta diversity in comparison to the rest of the sites in a landscape and as such,

1
2
3 270 each site has a single LCBD value at a given time point. We created the regression trees using
4
5
6 271 the functions *rpart()* and *prune()* from the `rpart` package (Therneau *et al.* 2015). We then
7
8 272 computed these same URTs with transformed Shannon diversity ($\exp(H)$) for the 2007 sediments
9
10 273 as a response variable, to determine whether alpha diversity showed similar patterns in the URTs
11
12 274 as those with LCBD indices. Pruning of the trees was done using the lowest cross-validation
13
14
15 275 relative error (CVRE) for the set of variables, meaning that trees of different sizes were
16
17
18 276 compared and the final tree size was chosen based on the tree with the highest amount of
19
20 277 predictive power. Additionally, the use of CVRE as a method for tree pruning meant that the
21
22 278 smallest tree would be retained over a tree with a lower CVRE as long as the chosen tree was
23
24 279 within one standard error of the smallest CVRE value (De'ath, 2002), resulting in a simple tree
25
26
27 280 with only slightly less predictive power than the tree with the lowest CVRE.

28
29 281 To characterize temporal beta diversity, we used an index that we refer to as Temporal
30
31 282 Beta diversity or TBI which is also described though not named in Legendre & Salvat (2015).
32
33 283 The metric TBI is based on the Podani family of beta diversity indices and is described in
34
35
36 284 Appendix S3. To compute TBI, we used an R function of the same name (*TBI()*) which is also
37
38
39 285 described in S3. We used function *TBI()*, with the percentage difference option, to analyse the
40
41 286 temporal beta diversity between the two time points for each individual lake. The function *TBI()*
42
43 287 computes total temporal beta diversity as well as the components that describe that temporal beta
44
45
46 288 diversity (either taxa loss or taxa gain).

47
48
49 289 Analogous to analysing the relationship between spatial LCBD and land use variables,
50
51 290 we examined the relationship between temporal beta diversity values (and their taxa gain and
52
53 291 loss components) and contemporary land use to test the hypothesis that developed areas (see list
54
55
56 292 of variables used in URT analyses) were associated with higher temporal beta diversity. Using
57
58
59
60

the same approach as with spatial LCBD, we also created a regression tree with either TBI indices or taxa gains in temporal beta diversity as the response variable and land cover types as the explanatory variables.

Results

Results are presented for both the genus-level analyses ($n = 169$) and the species-level analyses ($n = 52$). For simplicity, tables and figures for the genus-level analyses are presented in the main body of the paper, whereas all species-level tables and figures are shown in Appendices S5 and S6 in Supporting Information.

Genus-level results

Spatial diversity results

The different diversity metrics gave considerably distinctive information about ecoregion level patterns in diatom assemblages. For example, the range of rarefied genus richness values differed in some ecoregions between modern and historical sediments. In the Upper Midwest, rarefied richness was substantially larger in modern sediments compared to historical sediments (Table 2).

Total spatial beta diversity across the landscape was similar at both time points (Table 2; results shown with relative abundance data), with taxa replacement consistently representing the dominant mechanism of compositional change among lakes (Table 3). Across all sites, spatial beta diversity using the percentage difference (quantitative data) metric was 0.36 for both historical, and modern sediments. At the ecoregion level, spatial beta diversity varied little between the two time points (Table 2). The genus replacement component dominated in

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

importance for all ecoregions, for both historical and 2007 sediments (Table 3). Spatial beta diversity computed on presence-absence data using the Sørensen metric was 0.23 for both historical and modern sediments, with genus replacement as the dominant mechanism. Because spatial beta diversity was computed from a dissimilarity index, the percentage difference, where distances have an upper bound of 1 and because the sampling effort was the same for each lake, beta has an upper bound of 0.5 in each ecoregion and the values are comparable among ecoregions even if they have different sample sizes (Legendre & De Cáceres, 2013).

We observed considerable variation in the magnitude of LCBD values across the landscape (Figure 3a,b). Overall, we identified only five sites in the historical assemblages and one site in the modern assemblages that had significant LCBD indices (after correction for multiple testing). There was no clear geographic pattern in the distribution of higher relative LCBD values at either time point, or in the sites that had significant LCBD in common between the two time points. In looking at significant LCBD without correcting for multiple testing, we identified 20 sites (12% of total) as having significant LCBD indices in historical assemblages, whereas 17 (10%) had significant LCBDS in the modern assemblages; eight of these sites overlapped in the two time surveys. A McNemar test on a 2 x 2 table (number of sites significant in both time points, number of sites significant historically but not in 2007, number of sites not significant historically but significant in 2007 and number of sites not significant in both), resulted in $P = 0.5$, meaning that we cannot reject the null hypothesis of no effect of the time period on the significance of LCBD values. Additionally, a simple correlation between the historical and 2007 LCBD values showed a significant ($P < 0.05$) correlation of 0.38, showing some relation between historical and modern LCBD values.

The univariate regression tree (URT) using water quality variables as predictors explained 46% ($R^2_{adj} = 0.46$) of the variation in modern LCBD values, and only retained conductivity as an explanatory variable (Figure 4a). Using land cover measures instead of water quality in a URT of the 2007 LCBD values resulted in a slightly weaker model ($R^2_{adj} = 0.4$; Figure 4b), where sites with greater LCBD contributions to beta diversity were generally those with higher proportions of wetland and forest cover in the watersheds, although % of barren land in the watershed also appeared to be an important predictor.

Temporal diversity results

Mean temporal beta diversity computed at the genus level with relative abundance data ranged from 0.4 to 0.5 across the ecoregions (Table 4). The importance of genera loss and gain through time was approximately the same across ecoregions. Similar results were evident with the presence-absence data, where mean temporal beta diversity ranged from 0.3 to 0.4, and again the importance of genera loss and gain was approximately equal across ecoregions. Using temporal beta diversity values as a response variable and % land cover variables as potential predictors, our URT analysis identified % forest as the only significant variable with a threshold level of 36% ($R^2_{adj} = 0.36$; Figure 5). The URT constructed with the genus gain component of temporal beta diversity again retained % forest with a $R^2_{adj} = 0.40$.

Species-level results

Analyses of species-level data allow for additional conclusions to be drawn regarding beta diversity because the autecology of many diatom taxa are relatively well known and finer taxonomic resolution can bear greater insight. Working with a smaller set of lakes located in the Northern Appalachians ($n = 52$) ecoregion, we found landscape patterns similar to our genus-level results, where alpha and spatial beta diversities were approximately equivalent between

1
2
3 361 time points. In terms of LCBD, no sites with significant LCBD was found in this ecoregion in
4
5 362 either historical or modern times. Univariate regression tree of modern LCBD showed longitude,
6
7
8 363 temperature and pH to be key determinants of high LCBD in these lakes (Adj. $R^2 = 0.6$; tree not
9
10 364 shown). We also calculated temporal turnover of diatom species in this reduced set of lakes and
11
12 365 found this to be between 50-60%. With the reduced dataset, Species Contributions to Beta
13
14 366 Diversity (SCBD) could also be examined in relation to spatial beta diversity. The ten species
15
16 367 with the highest contributions to SCBD in historical and 2007 assemblages were mostly
17
18 368 planktonic and five of them were key contributors to SCBD in both data sets (see S6). No
19
20 369 significant correlation was found between these abundances.
21
22
23
24
25 370

26
27 371 **Discussion**
28

29 372 Our analyses identified patterns and drivers of diatom diversity across the conterminous
30
31 373 United States over the last ~150 years and provided support for Anthropocene trends previously
32
33 374 postulated (McGill *et al.* 2015), as well as some unexpected ones (Table 5). Because of the larger
34
35 375 sample size and geographic coverage, our genus-level results are the most robust for drawing
36
37 376 conclusions and are discussed at length here. Across all sites, we found that genus richness
38
39 377 increased in the modern sediments over historical ones at the largest scale; however, this was
40
41 378 mostly driven by the marked changes observed in the Upper Midwest. This pattern may be the
42
43 379 result of introductions of new genera, including dispersal aided by hydrological modification
44
45 380 (Alig *et al.* 2004), a hypothesis consistent with McGill *et al.*'s (2015) conjectures, although other
46
47 381 forms of environmental change are possible. For spatial beta diversity, we found that historical
48
49 382 and modern beta diversity measured at the ecoregion level were similar, although at the site
50
51 383 level, there were some changes in diversity through time. While lakes with the largest TBIs
52
53
54
55
56
57
58
59
60

represent lakes that have changed the most in diatom composition between time points, none had experienced exceptionally large (i.e. significant) changes.

In terms of the mechanism underlying the changes in spatial, we consistently found evidence for replacement-dominated spatial beta diversity within each time point. Additionally, site-specific variables associated with water quality and land cover were significant predictors of the spatial variation in LCBD, but there was no clear regional pattern in the distribution of LCBD indices. Both water quality and land cover varied within regions and so this may be one reason for the lack of geographical pattern.

Beta diversity as an important metric of biodiversity

Alterations to biodiversity are not just via species or genera losses. While the loss of taxa is intuitively and empirically important, taxon gains and compositional shifts can also fundamentally alter ecosystems. In a meta-analysis of plant biodiversity, Vellend *et al.* (2013) found that local-scale species diversity was as likely to increase or decrease through time. More recently, Dornelas *et al.* (2014) analysed time series data from various ecosystems and found distinct differences between alpha and beta diversity, with considerable beta diversity variation, but no net loss in alpha diversity. Taken together with our results, two aspects are highlighted: first, that some critical biodiversity patterns may only be visible through the lens of beta diversity, pointing to the importance of this estimate along with alpha and gamma diversity; and secondly, that beta diversity could be a key component of the biodiversity loss in the Anthropocene. Indeed, we found beta diversity to illuminate key patterns in diatom composition across land use gradients, providing more information than had we focused only on differences in alpha diversity alone. For example, our regression tree analyses showed different driving variables for alpha diversity than components of beta diversity (particularly for water quality)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

(Winegardner, 2016). Similar to Dornelas *et al.* (2014), beta diversity (both spatial and temporal) was significantly related to land cover change, a factor itself related to water quality (Taranu & Gregory-Eaves, 2008).

Sites versus ecoregion-level responses

Contributions to either spatial or temporal diversity varied from lake to lake, even when there was little variation across the ecoregions (based either on historical or 2007 sediments). For example, we found that lakes with very low specific conductivity (less than 18 $\mu\text{S cm}^{-1}$) had the highest LCBD values (spatial), indicating that these sites are particularly unique across the landscape. In our dataset, conductivity was positively correlated ($r = 0.6$) with silica concentration, such that sites with low silica had the most distinctive diatom assemblages, concordant with the known role of ionic strength to influence diatom assemblages (Fritz *et al.* 2010). Spatial LCBD values were greater in basins with lower human (i.e. urban) development, indicating that less disturbed lakes may contribute more diatom heterogeneity to the landscape. Indeed, nutrient enriched lakes (a consequence of most development) can have a homogenizing effect on aquatic assemblages (Olden *et al.* 2004; Donahue *et al.* 2009). We identified a very low threshold for the proportion of development in the watershed that impacted LCBD (Figure 4). The effect on LCBD at a very low level of development might be due to the tendency for humans to modify shorelines even when the rest of the watershed may be intact (e.g., by removing riparian and littoral vegetation; Kaufmann *et al.* 2014), which in turn have substantial consequences for diatoms (Velghe *et al.* 2012) and other organisms (Strayer & Findlay, 2010). The relationship with forest cover indicates that less forest cover may produce downstream environmental conditions that favour specific species or genera, and hence these lakes may contribute more to total beta diversity because they are exceptional in nature.

Land use variables were also informative in describing the observed temporal variation in diatom assemblages, which was higher in lakes with lower forest cover in their basins. While the contemporary forest cover variable does not inform as to whether a change in cover occurred since historical times, it does indicate that lakes with higher basin forest cover currently may have been buffered against diatom assemblage change through time (at least when comparing pre-industrial and contemporary times). Essentially, there was higher temporal beta diversity with reduced forest cover, driven by more gains at the genera-level in low forest, shrubland and wetland lakes.

Dominance of the replacement component in spatial beta diversity

Replacement (gain and loss of species along a gradient) was always the dominant component of spatial beta diversity, regardless of the time point. While a separate analysis, our temporal beta diversity results shed some light onto why replacement is such a dominant component of spatial beta diversity, both historically and in present times. The method used to compute the significance of temporal change for each lake was designed to detect significant changes through time even in the presence of beta diversity, i.e. when community composition is *not* uniform across space. Thus, the temporal changes detected in our study resulted specifically from disproportionate temporal change across the landscape, as opposed to a continent-wide reduction or increase in a specific set of taxa. Leprieur *et al.* (2011) explained that beta diversity dominated by species replacement is more likely to occur in regions with environmental gradients, as opposed to where there has been a spatially uniform environmental change. Broad-scale climatic changes (climate warming) could create a uniform temporal beta diversity pattern across the U.S. landscape. However, no sites had significant temporal beta diversity once corrected for multiple testing, indicating that temporal changes differed from lake to lake,

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

presumably due to local or regional factors. It is also important to note that recent climate change in the U.S. has not manifested itself to the same extent across all regions (Kennedy, 2014). Thus, it would be reasonable to assume that spatial beta diversity is explained by replacement via the mechanism described by Leprieur *et al.* (2011).

We found that gain and loss of abundances on a genus-by-genus basis explained a nearly equivalent proportion of total beta diversity variation between our time points. This could indicate that abundance gain and loss on a genus-by-genus basis were differentially important at various intervals through time. This was shown in Legendre and Salvat (2015), where temporal beta diversity of mollusc communities was partitioned into species gains and losses during five different time intervals (each time interval consisting of 4-29 years), with the importance of the different components alternating between the time periods and from site to site. With only two time points, like our historical and modern points, these finer-scale alterations between genera gains and losses cannot be observed, but could be explored in more detailed sediment core analyses in the future. Environmental variables and mechanisms behind temporal beta diversity may change between time periods.

Beta diversity results within the metacommunity concept

Our study has delineated some specific diversity patterns in the different ecoregions and revealed mechanisms behind these (Table 6) and many of these mechanisms can be related to metacommunity theory via variation partitioning of beta diversity metrics (Giovâni da Silva & Hernández, 2014; Tonkin *et al.* 2016). In our study, the entire landscape, as well as the majority of ecoregions, show approximately equivalent historical and modern beta diversity (though with some small differences discussed below). We can speculate on potential mechanisms driving spatial and temporal beta diversity by considering both spatial and temporal patterns together in

the ecoregions, each of which presumably operates as a metacommunity. For example, in the Northern Appalachians ecoregion, the pattern of heterogeneous diatom communities across the landscape, both historically and currently, could have arisen because taxon losses have matched taxon gains across sites within this metacommunity. On the other hand, diatom assemblages across the modern landscape may be more heterogeneous than historically (e.g. Temperate Plains ecoregion), because dispersal of taxa through time has been sufficiently great enough to allow for taxon (environmental) sorting across the landscape. In the Upper Midwest, we can envision yet another scenario: diatoms may be more homogeneously distributed across space currently because population dynamics have resulted in genera being added to sites in an unsorted way (*sensu* mass effects; Leibold *et al.* 2004) way. Diatoms may also be more homogeneous across an ecoregion metacommunity currently because the taxa eliminated over time are always the same, regardless of local site, as could be the mechanism in the Western Mountains ecoregion. Now that the broad-scale continental patterns have been established, these potential scenarios can now be investigated further by ecoregion through more detailed study within each, using metacommunity theory for guidance.

Conclusions and potential future research

Patterns in spatial and temporal beta diversity varied both across sites and ecoregions, via several possible mechanisms; we have highlighted some of the most important. By considering temporal and spatial beta diversity in the same study, we have been able to provide a holistic view of diatom biodiversity patterns and test hypotheses and predictions for change during the Anthropocene made by McGill *et al.* (2015). Future work considering other trophic levels, as well as data that incorporates multiple historical time points, will be most insightful for generating a broader perspective of biodiversity changes across continents at multiple scales.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

499 Additionally, more work could be done with the temporal beta diversity partitioning to further
500 align the partitioning of temporal beta diversity into taxa losses and gains with the framework
501 outlined in Baselga *et al.* (2013). This would allow for further partitioning of temporal beta
502 diversity into loss and gain components into components explained by taxa replacement and
503 abundance difference, adding further depth to analyses of temporal beta diversity.

504
505 **Acknowledgements**

506 The authors wish to thank Amina Pollard of the NLA programme for help with questions about
507 field sampling, as well as Mihaela Enache and Marina Popatova at the Patrick Center for
508 Environmental Research for help procuring sediment material for sediment dating. Don Charles
509 provided insight on the diatom data collected in the NLA. Julien Gogot and Bassam Ghaleb from
510 GEOTOP-UQÀM completed the radiometric dating. Funding for this project was made possible
511 by an FRQNT Équipe grant awarded to IGE, BEB and PL, as well as an NSERC PGS-D
512 scholarship awarded to AKW. AKW also received additional funding from a McGill
513 International Travel Grant and an Excellence Award from the Québec Centre for Biodiversity
514 Science. This project also received funding from the NARS Campus Challenge award. This
515 study is a contribution to the FRQNT-funded Québec Centre for Biodiversity Science (QCBS-
516 CSBQ).

Supplementary information titles

S1: Description of sediment core screening with respect to length and predicted age of bottom sediment core samples.

S2: Description of additional analyses using the Baselga family of indices

S3: Description of the Temporal Beta diversity Indices (TBI) performed using TBI().

S4: Principal Component Analyses (PCAs) of (a) standardized 2007 water quality variables and (b) land cover variables.

S5: Tables and figures for species-level analyses (n = 52).

S6: Diatom species abundance in historical (Hist_) and 2007 sediments (Surf_) for the species identified as having the highest SCBD values in sediments from both time points (n = 52).

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Biosketch

Amanda Winegardner is a recently-graduated PhD who undertook her studies in the Department of Biology at McGill University, co-supervised by BEB (UQAM) and IGE. PL, BEB and IGE are members of the FRQNT funded Québec Centre for Biodiversity Science (QCBS-CSBQ) and Québec Inter-university Research Group in Limnology and Aquatic Environments (GRIL). **Pierre Legendre** is a specialist in numerical ecology, a Fellow of the Royal Society of Canada (Academy of Science) and a Web of Science Highly Cited Researcher (2014, 2015 and 2016) in Environment/Ecology. He is interested in the ecological and biogeographic processes that organize biodiversity spatially and temporally. **Beatrix Beisner** is a community ecologist working in aquatic ecosystems ranging from experimental mesocosms to landscapes of lakes, linking theory to observation. She is currently the Director of the GRIL. **Irene Gregory-Eaves** is a Tier 2 Canada Research Chair in Freshwater Ecology and Global Change. Her lab uses both paleolimnology and contemporary ecology to study the effect of land use, climate change and pollution on aquatic systems.

References

- Alig, R.J., Kline, J.D., & Lichtenstein, M. (2004) Urbanization on the US landscape: looking ahead in the 21st century. *Landscape and Urban Planning*, **69**, 219-234.
- Baselga A., & Orme C.D.L. (2012) betapart: an R package for the study of beta diversity. *Methods in Ecology and Evolution*, **3**, 808-812.
- Baselga, A. (2013). Separating the two components of abundance-based dissimilarity: balanced changes in abundance vs. abundance gradients. *Methods in Ecology and Evolution*, **4**, 552-557.
- Beaulieu, M., Pick, F., & Gregory-Eaves, I. (2013) Nutrients and water temperature are significant predictors of cyanobacterial biomass in an 1147 lakes data set. *Limnology and Oceanography*, **58**, 1736-1746.
- Cardinale, B.J., Duffy, J.E., Gonzalez, A., Hooper, D.U., Perrings, C., Venail, P., Narwani, A., Mace, G.M., Tilman, D., Wardle, D.A., Kinzig, A.P., Daily, G.C., Loreau, M., Grace, J.B., Larigauderie, A., Srivastava, D.S., and Naeem, S. (2012) Biodiversity loss and its impact on humanity. *Nature*, **486**, 59-67.
- Charles, D.F., Knowles, C., & Davis, R.S. (2003) Protocols for the analysis of algal samples as collected as part of the U.S. Geological Survey National Water-Quality Assessment program. Patrick Center for Environmental Research, The Academy of Natural Sciences. Report No. 02-06.
- Crutzen, P.J. (2002) Geology of mankind. *Nature*, **415**, 23.
- De'ath, G. (2002) Multivariate regression trees: A new technique for modeling species-environment relationships. *Ecology*, **83**, 1105-1117.
- Donahue, I., Jackson, A.L., Pusch, M.T., & Irvine, K. (2009) Nutrient enrichment homogenizes lake benthic assemblages at local and regional scales. *Ecology*, **90**, 3470-3477.
- Dornelas, M., Gotelli, N.J., McGill, B., Shimadzu, H., Moyes, F., Sievers, C., and Magurran, A.E. (2014) Assemblage time series reveal biodiversity change but not systematic loss. *Science* **344**, 296-299.
- Dray, S., Blanchet, G., Borcard, D., Guénard, G., Jombart, T., Larocque, G., Legendre, P., & Wagner, H.H. (2016) adespatial: Multivariate multiscale spatial analysis. R package version 0.0-4. <http://cran.r-project.org/package=adespatial>.
- Ellison, A.M. (2010) Partitioning diversity. *Ecology*, **91**, 1962-1963.
- Fritz, S.C., Cumming, B.F., Gasse, F., & Laird, K.R. (2010) Diatoms as indicators of hydrologic and climatic change in saline lakes in "The Diatoms: Applications for the Environmental and Earth Sciences 2nd edition". Cambridge University Press, Cambridge UK.

1
2
3 604 Giovâni da Silva, P., & Hernández, M.I.M. (2014) Local and regional effects on community
4 605 structure of dung beetles in a mainland-island scenario. *PLoS ONE*, **9**, e111883.
5 606
6 607 Glew, J. (1989) A new trigger mechanism for sediment samples. *Journal of Paleolimnology*, **2**,
7 608 241-24.
8 609
9 610 Hamilton, C. (2015) Getting the Anthropocene so wrong. *The Anthropocene Review*, **2**, 102-107.
10 611
11 612 Herlihy, A.T., Paulsen, S.G., Van Sickle, J., Stoddard, J.L., Hawkins, C.P., & Yuan, L.L. (2008)
12 613 Striving for consistency in a national assessment: the challenges of applying a reference-
13 614 condition approach at a continental scale. *Journal of the North American Benthological Society*,
14 615 **27**, 860-877.
15 616
16 617 Hyatt, C.V., Paterson, A.M., Rühland, K.M., & Smol, J.P. (2011) Examining 20th century water
17 618 quality and ecological changes in the Lake of the Woods, Ontario, Canada: A paleolimnological
18 619 investigation. *Journal of Great Lakes Research*, **37**, 456-469.
19 620
20 621 Kaufmann, P.R., Peck, D.V., Paulsen, S.G., Seeliger, C.W., Hughes, R.M., Whittier, T.R., &
21 622 Kamman, N.C. (2014) Lakeshore and littoral physical habitat structure in a national lakes
22 623 assessment. *Lake and Reservoir Management*, **30**, 192-215.
23 624
24 625 Kennedy, C. (2014) Does “global warming” mean it’s warming everywhere? National Oceanic
25 626 and Atmospheric Administration. Available online: [https://www.climate.gov/news-](https://www.climate.gov/news-features/climate-qa/does-global-warming-mean-it%E2%80%99s-warming-everywhere)
26 627 [features/climate-qa/does-global-warming-mean-it%E2%80%99s-warming-everywhere](https://www.climate.gov/news-features/climate-qa/does-global-warming-mean-it%E2%80%99s-warming-everywhere).
27 628 Accessed 31 August 2015.
28 629
29 630 Legendre, P. (2014) Interpreting the replacement and richness difference components of beta
30 631 diversity. *Global Ecology and Biogeography*, **23**, 1324-1334.
31 632
32 633 Legendre, P. (2015) R function: TBI(). Available online:
33 634 <http://adn.biol.umontreal.ca/~numericaledgeology/FonctionsR/>.
34 635
35 636 Legendre, P., & De Cáceres, M. (2013) Beta diversity as the variance of community data:
36 637 dissimilarity coefficients and partitioning. *Ecology Letters*, **16**, 951-963.
37 638
38 639 Legendre, P., & Legendre, L. (2012) Numerical ecology, 3rd English edn. Elsevier Science BV,
39 640 Amsterdam.
40 641
41 642 Legendre, P., & Salvat, B. (2015) Thirty-year recovery of mollusc communities after nuclear
42 643 experimentations on Fangataufa atoll (Tuamotu, French Polynesia). *Proceedings of the Royal*
43 644 *Society B*, **282**, 20150750.
44 645
45 646 Leibold, M.A., Holyoak, M., Mouquet, N., Amarasekare, P., Chase, J.M., Hoopes, M.F., Holt,
46 647 R.D., Shurin, J.B., Law, R., Tilman, D., Loreau, M., & Gonzalez, A. (2004) The metacommunity
47 648 concept: a framework for multi-scale community ecology. *Ecology Letters*, **7**, 601-613.
48 649
49
50
51
52
53
54
55
56
57
58
59
60

- Leprieur, F., Tedesco, P.A., Hugueny, B., Beauchard, O., Dürr H.H., Brosse, S., & Oberdorff, T. (2011) Partitioning global patterns of freshwater fish beta diversity reveals contrasting signatures of past climate changes. *Ecology Letters*, **14**, 325-334.
- Lewis, S.L., & Maslin, M.A. (2015) Defining the Anthropocene. *Nature* 519: 171-180.
- McGill, B.J., Dornelas, M., Gotelli, N.J., & Magurran, A.E. (2015). Fifteen forms of biodiversity trend in the Anthropocene. *Trends in Ecology and Evolution*, **30**, 104-113.
- Newbold, T., Hudson, L.N., Hill, S.L.L., Contu, S., Lysenko, I., Senior, R.A., Börger, L., Bennett, D.J., Choimes, A., Collen, B., Day, J., De Palma, A., Díaz, S., Echeverria-Londono, Edgar, M.J., Feldman, A., Garon, M., Harrison, M.L.K., Alhusseini, T., Ingram, D.J., Itescu, Y., Kattge, J., Kemp, V., Kirkpatrick, L., Kleyer, M., Correia, D.L.P., Martin, C.D., Meiri, S., Novosolov, M., Pan, Y., Phillips, H.R.P., Purves, D.W., Robinson, A., Simpson, J., Tuck, S.L., Weiher, E., White, H.J., Ewers, R.M., Mace, G.M., Scharlemann, J.P.W., & Purvis, A. (2015) Global effects of land use on local terrestrial biodiversity. *Nature*, **520**, 45-50.
- Oksanen, J., Blanchet, G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Szoecs, E., & Wagner, H.H. (2016) vegan: Community Ecology Package. R package version 2.4-0. <http://cran.r-project.org/package=vegan>.
- Olden, J.D., Poff, N.L., Douglas, M.R., Douglas, M.E., & Fausch, K. (2004) Ecological and evolutionary consequences of biotic homogenization. *Trends in Ecology and Evolution*, **19**, 18-24.
- Podani J., & Schmera D. (2011) A new conceptual and methodological framework for exploring and explaining patterns in presence-absence data. *Oikos*, **120**, 1625-1638.
- Podani, J., Ricotta, C., & Schmera, D. (2013) A general framework for analyzing beta diversity, nestedness and related community-level phenomena based on abundance data. *Ecological Complexity*, **15**, 52-61.
- R Core Team (2014) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available: <http://www.R-project.org/>
- Rühland, K., Priesnitz, A., & Smol, J.P. (2003) Paleolimnological evidence from diatoms for recent environmental changes in 50 lakes across Canadian Arctic Treeline. *Arctic, Antarctic, and Alpine Research*, **35**, 110-123.
- Smol, J.P. (1992) Paleolimnology: an important tool for effective ecosystem management. *Journal of Aquatic Ecosystem Health*, **1**, 49-58.
- Smol, J.P., & Stoermer, E.F. (eds) (2010) The Diatoms: Applications for the Environmental and Earth Sciences 2nd edition. Cambridge University Press, Cambridge UK.




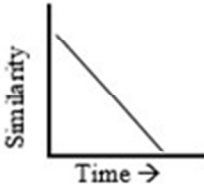
1
2
3 696 Steffen, W., Grinevald, J., Crutzen, P., & McNeill, J. (2011) The Anthropocene: conceptual and
4 697 historical perspectives. *Philosophical Transactions of the Royal Society A*, **369**, 842-867.
5 698
6 699 Strayer, D.L., & Dudgeon, D. (2010) Freshwater biodiversity conservation: recent progress and
7 700 future challenges. *Journal of the North American Benthological Society*, **29**, 344-358.
8 701
9 702 Strayer, D.L., & Findlay, S.E.G. (2010) Ecology of freshwater shore zones. *Aquatic Sciences*, **72**,
10 703 127-163.
11 704
12 705 Taranu, Z.E., & Gregory-Eaves, I. (2008) Quantifying relationships among phosphorus,
13 706 agriculture, and lake depth at an inter-regional scale. *Ecosystems*, **11**, 715-725.
14 707
15 708 Therneau, T., Atkinson, B., & Ripley, B. (2015) rpart: Recursive partitioning and regressions
16 709 trees. R package version 4.1-10. <http://CRAN.R-project.org/package=rpart>
17 710
18 711 U.S. EPA (2009) National Lakes Assessment: A Collaborative Survey of the Nation's Lakes.
19 712 EPA 841-R-09-001. U.S. Environmental Protection Agency, Washington, DC, USA.
20 713
21 714 Tonkin, J.D., Arimoro, F.O., & Haase, P. (2016) Exploring stream communities in a tropical
22 715 biodiversity hotspot: biodiversity, regional occupancy, niche characteristics and environmental
23 716 correlates. *Biodiversity Conservation*, **25**, 975-993.
24 717
25 718 U.S. EPA (2011-2012) National Lakes Assessment. Field Operations Manual. EPA 841-B-11-
26 719 003. U.S. Environmental Protection Agency, Washington, DC, USA.
27 720
28 721 USGS (1992) The USGS Land Cover Institute: NLCD 92 Land Cover Class Definitions.
29 722 Available: <http://landcover.usgs.gov/classes.php>.
30 723
31 724 Velghe K., Vermaire J.C., & Gregory-Eaves I. (2012) Declines in littoral species richness across
32 725 both spatial and temporal nutrient gradients: A palaeolimnological study of two taxonomic
33 726 groups. *Freshwater Biology*, **57**, 2378-2389.
34 727
35 728 Vellend, M., Baeten, L., Myers-Smith, I.H., Elmendorf, S.C., Beauséjour, R., Brown, C.D., De
36 729 Frenne, P., Verheyen, K., & Wipf, S. (2013) Global meta-analysis reveals no net change in local-
37 730 scale plant biodiversity over time. *Proceedings of the National Academy of Sciences*, **110**,
38 731 19456-19459.
39 732
40 733 Wardle, D.A., Bardgett, R.D., Callaway, R.M., & Van der Putten, W.H. (2011) Terrestrial
41 734 ecosystem responses to species gains and losses. *Science*, **332**, 1273-1277.
42 735
43 736 Whittaker, R.H. (1972) Evolution and measurement of species diversity. *Taxon*, **21**, 213-251.
44 737
45 738 Winegardner, A.K. (2016). Aquatic biodiversity patterns along gradients of multiple stressors
46 739 and disturbance histories: Integration of paleoecology and molecular techniques. PhD Thesis,
47 740 McGill University.
48 741
49
50
51
52
53
54
55
56
57
58
59
60

742 Winegardner, A.K., Beisner, B.E., Legendre, P., & Gregory-Eaves, I. (2015) Are the landscape-
743 level drivers of water column and surface sediment diatoms different? *Freshwater Biology*, **60**,
744 267-281.

746 Zalasiewicz, J., Waters, C.N., Barnosky, A.D., Cearreta, A., Edgeworth, M., Ellis, E.C.,
747 Galuszka, A., Gibbard, P.L., Grinevald, J., Hajdas, I., do Sul, J.I., Jeandel, C., Leinfelder, R.,
748 McNeill, J.R., Poirier, C., Revkin, A., Richter, D.B., Steffen, W., Summerhayes, C., Syvitski,
749 J.P.M., Vidas, D., Waple, M., Williams, M., & Wolfe, A.P. (2015) Colonization of the
750 Americas, 'Little Ice Age' climate, and bomb-produced carbon: Their role in defining the
751 Anthropocene. *The Anthropocene Review*, **2**, 117-127.

754 **Tables**

755 **Table 1: Hypotheses and predictions for alpha and beta diversity analyses based on McGill *et al.* (2015, Figure 2) framework.**

Spatial scale	α - or γ -diversity	Temporal β -diversity	Spatial β -diversity
Continental (Biogeographic)	 <p>Richness</p> <p>Time →</p> <p>Higher gamma diversity in modern times</p>	No a priori prediction	 <p>Rate of decay</p> <p>Time →</p> <p>Less spatial beta diversity in modern times (homogenization)</p>
Ecoregion (Metacommunity)	Higher alpha diversity in modern times	No a priori prediction	Less spatial beta diversity in modern times (homogenization- as above)
Lake (Local)	 <p>Richness</p> <p>Time →</p> <p>Alpha diversity stays constant on local scale</p>	 <p>Similarity</p> <p>Time →</p> <p>Beta diversity increases between observations as they are father apart in time</p>	Not applicable

756

Table 2: Spatial beta (β) diversity, mean rarefied genus richness (to 300 valves), mean alpha (α) diversity (Shannon diversity transformed to $\exp(H)$ and Simpson), and gamma (γ) diversity for each ecoregion (Shannon diversity of genus' sums). All calculations used the genus-level data. "Hx" refers to the historical sediments. Richness was rarefied after excluding rare genera (<2% relative abundance). The Xeric ecoregion was excluded in these mean values because there was only one lake in that region.

Ecoregion	β -diversity		Rarefied G		α -diversity (Shannon)		α -diversity (Simpson)		γ -diversity	
	Hx	2007	Hx	2007	Hx	2007	Hx	2007	Hx	2007
All	0.36	0.36	18.0	24.3	2.3	2.2	0.9	0.8	3.3	3.2
Northern Appalachians	0.30	0.31	29.9	27.5	2.5	2.4	0.9	0.8	3.2	3.1
Southern Plains	0.28	0.28	25.0	24.6	2.3	2.4	0.8	0.8	2.8	2.8
Temperate Plains	0.32	0.31	25.7	22.2	2.2	2.1	0.8	0.8	2.8	2.7
Upper Midwest	0.36	0.34	17.0	24.0	2.2	2.1	0.9	0.8	3.2	3.1
Western Mountains	0.37	0.37	25.8	21.8	2.1	1.9	0.9	0.7	3.2	3.0

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Table 3: Explanatory components for historical and 2007 spatial beta diversity, as
computed using *beta.div.comp()*. ‘Repl’ refers to the replacement component; ‘AbDiff’ refers to
the abundance difference component; and ‘Repl/Total’ and ‘AbDiff/Total’ are these two
components with total beta diversity as the denominator.

Ecoregion	Repl	AbDiff	Repl/Total	AbDiff/Total
<i>Historical</i>				
All	0.31	0.04	0.88	0.12
Northern Appalachians	0.29	0.01	0.96	0.04
Southern Plains	0.26	0.02	0.93	0.07
Temperate Plains	0.29	0.03	0.90	0.09
Upper Midwest	0.29	0.07	0.81	0.19
Western Mountains	0.33	0.04	0.91	0.09
<i>2007</i>				
All	0.34	0.01	0.96	0.04
Northern Appalachians	0.31	<0.01	0.10	0.01
Southern Plains	0.28	0	1	0
Temperate Plains	0.28	0.03	0.92	0.08
Upper Midwest	0.33	0.02	0.95	0.05
Western Mountains	0.35	0.02	0.94	0.07

Table 4: Mean (and standard deviation) temporal beta diversity components (abundance loss and gain) for the ecoregions. ‘Total beta’ refers to the mean value of the temporal beta diversity in each region. Temporal beta diversity was computed using the percentage difference index applied to the diatom abundance data; values are in the [0,1] range. Total beta is the sum of ‘Genera loss’ and ‘Genera gain’. Genera loss refers to the component representing loss of abundances on a genus by genus basis between the historical and 2007 time points. Similarly, genera gain refers to the component representing gain of abundances on a genus-by-genus basis. These components were computed on a lake-by-lake basis and then averaged for each ecoregion.

Ecoregion	Loss of genera	Gain of genera	Total beta
Northern Appalachians	0.20 (0.08)	0.18 (0.08)	0.38 (0.16)
Southern Plains	0.21 (0.09)	0.19 (0.06)	0.40 (0.14)
Temperate Plains	0.23 (0.04)	0.22 (0.08)	0.46 (0.09)
Upper Midwest	0.26 (0.10)	0.26 (0.13)	0.51 (0.17)
Western Mountains	0.29 (0.09)	0.25 (0.08)	0.54 (0.17)

803 **Table 5: Summary of our results (in italics) in relation to the McGill *et al.* (2015) framework**

Spatial scale	α - or γ -diversity	Temporal β -diversity	Spatial β -diversity
Continental (Biogeographic)	Gamma diversity across all sites was similar between historical and modern times <i>Does not match McGill et al.(2015) predicted trend</i>	No continental pattern of temporal beta	Spatial beta diversity approximately equivalent between historical and modern times <i>Does not match McGill et al.(2015) predicted trend</i>
Ecoregion (Metacommunity)	Rarefied genus richness higher in modern times in some ecoregions <i>Partial match: true in Upper Midwest only.</i>	No ecoregion pattern of temporal beta	Spatial beta diversity less in modern times than historical times in some ecoregions <i>Does not match McGill et al.(2015) predicted trend</i>
Lake (Local)	Rarefied genus richness higher in modern times for some lakes <i>Does not match McGill et al.(2015) predicted trend</i>	No lakes showed significant temporal beta diversity once corrected for multiple testing <i>Direct comparison to McGill et al. (2015) predicted trend not possible with only two time points</i>	Not applicable

804

805

Table 6: Summary of temporal and spatial beta diversity across ecoregions. For spatial beta diversity, the observation is referring to whether the magnitude of spatial diversity in an ecoregion was higher or lower in historical times. Spatial diversity within each ecoregion is computed between all lakes (pairwise comparisons). Mechanisms column for temporal beta diversity shows when abundance losses or gains (genus by genus) are equivalent or which one dominates.

Ecoregion	Mechanism explaining <i>temporal</i> beta diversity (genus by genus)	Observation from <i>spatial</i> beta diversity
Northern Appalachians	Abundance loss	~ = between contemporary and historical
Southern Plains	Slight abundance loss	= between contemporary and historical
Temperate Plains	Abundance loss	~ = between contemporary and historical
Upper Midwest	=	Lower in contemporary
Western Mountains	Abundance loss	= between contemporary and historical

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Figure Legends

Figure 1: Conceptual diagram of the different forms of beta diversity analysed using surface and bottom sediment core samples from the 2007 National Lakes Assessment (genus- and species-level analyses). Beta diversity *between lakes* (represented by the grey circles) across the landscape was investigated using both bottom (historical) and modern (2007) sediments from cores; resulting in both historical and contemporary estimates of spatial beta diversity. Temporal beta diversity between historical and contemporary conditions was investigated for *each lake*.

Figure 2: Location of lake sites used in this study for genus-level analyses and key limnological variables. (a) Lake sites are classified into six distinct ecoregions: Northern Appalachians (N. Appalachians/NAP, n = 53), Southern Plains (S. Plains/SPL, n = 5), Temperate Plains (TPL, n = 6), Upper Midwest (UMW, n = 69), Western Mountains (W. Mountains/WMT, n = 35), and Xeric (XER, n = 1). Boxplots show the range and median of (b) lake surface area (km²), (c) observed maximum depth (m), (d) pH and (e) TP (µg L⁻¹), where the centre horizontal line is the median, the lower horizontal line the 25th percentile, the upper horizontal line the 75th percentile and points represent outliers.

Figure 3: LCBD values for (a) historical spatial beta diversity and (b) 2007 spatial beta diversity and magnitude of beta diversity (c) for temporal beta diversity (TBI) (genus-level analyses). A lake has a significant LCBD value if $P < 0.05$ after correction for multiple testing, and is coded as “True” (solid circles). LCBD values across all sites (in either historical or 2007) sum to 1. Note that no significance testing is performed for (c).

Figure 4: Univariate regression tree of the 2007 spatial LCBD (n = 169; genus-level) explained by (a) the selected water quality variable Conductivity and (b) land cover variables. Land cover variables are % barren land in a basin; 'PCT_BARREN_BSN', % developed land in a basin; 'PCT_DEVELOPED_BSN', % forest land in a basin; 'PCT_FOREST_BSN', % agricultural land in a basin; 'PCT_AGRIC_BSN', and % wetland in a basin; 'PCT_WETLAND_BSN'. The cumulative R^2_{adj} for the model in (a) is 0.46. The land cover variables are percentages for each lake basin. The cumulative R^2_{adj} for the model in (b) is 0.38.

Figure 5: Univariate regression tree of total temporal beta diversity explained by forest percentage in each basin (n = 169; genus-level). The $R^2_{\text{adj}} = 0.36$.

Figures

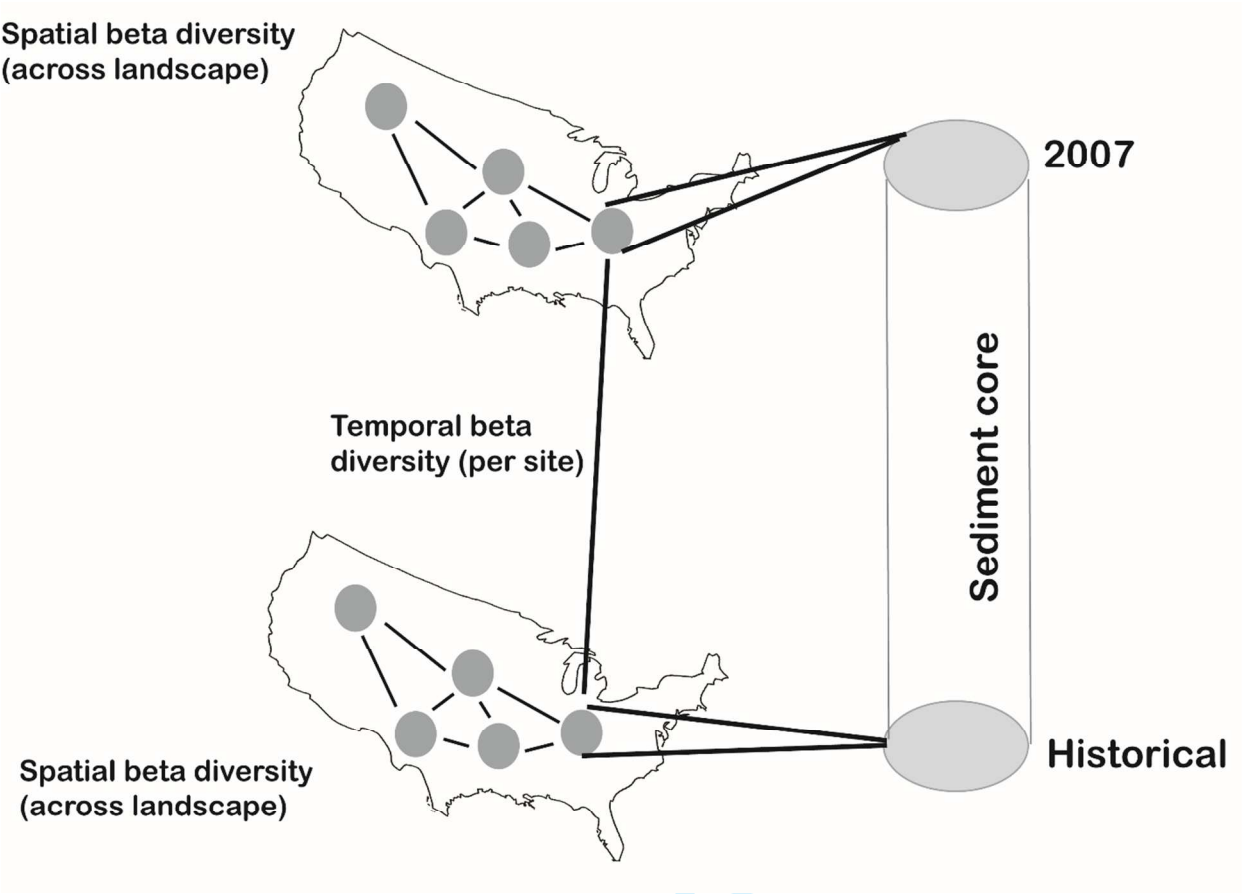


Figure 1

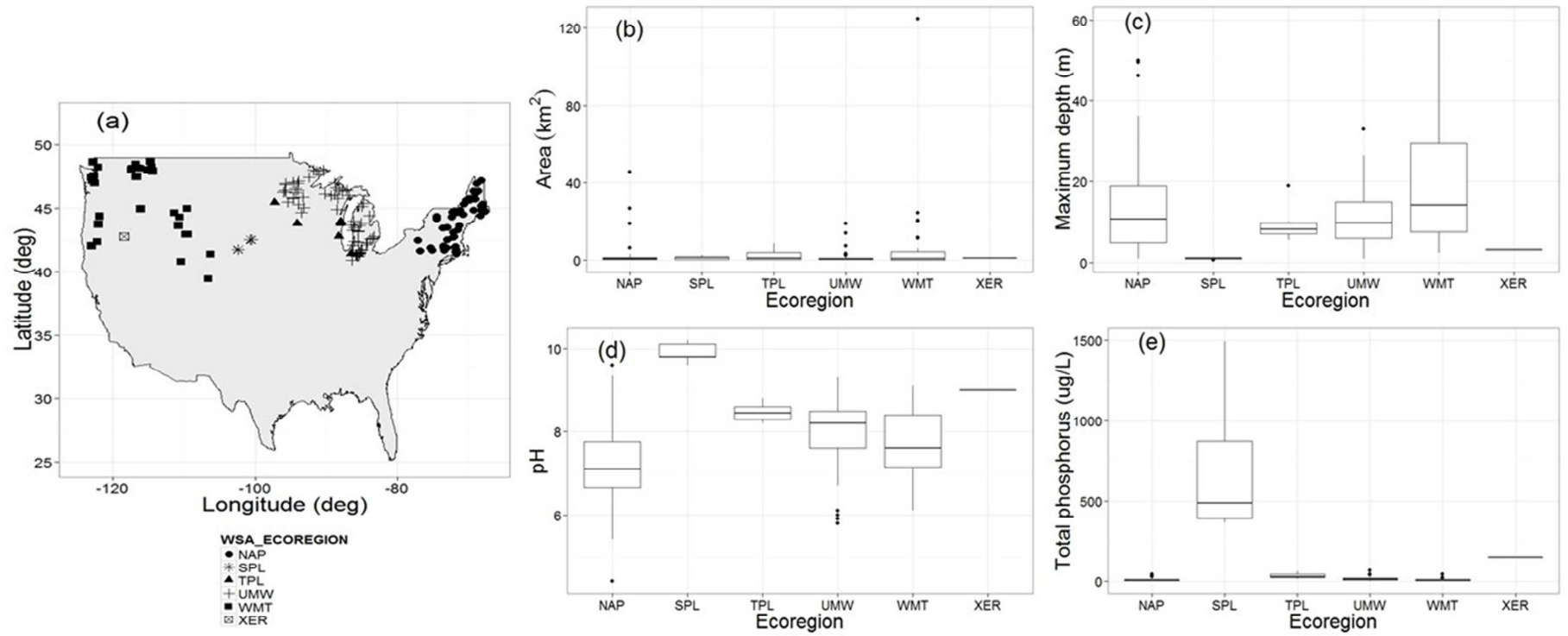


Figure 2

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

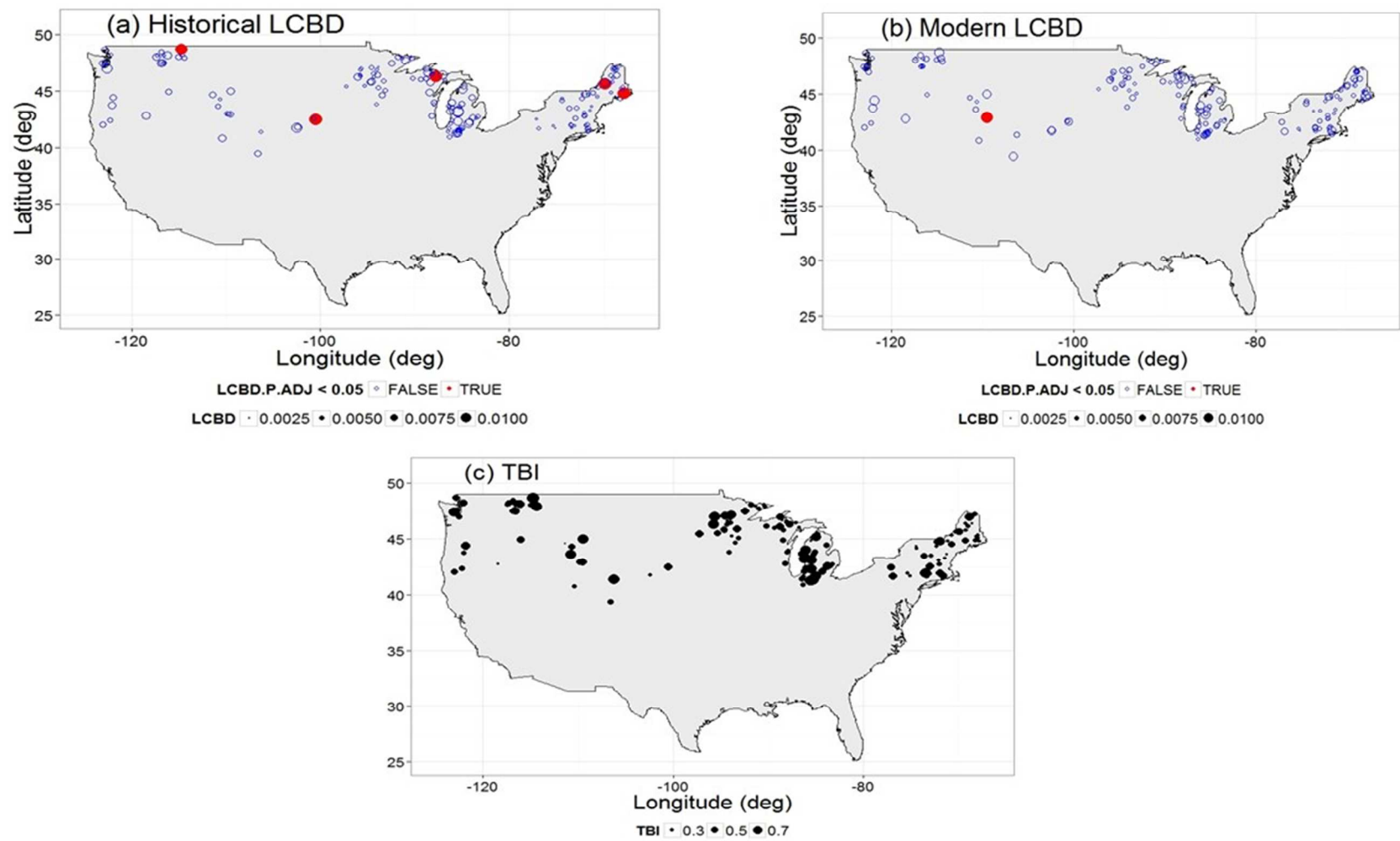


Figure 3

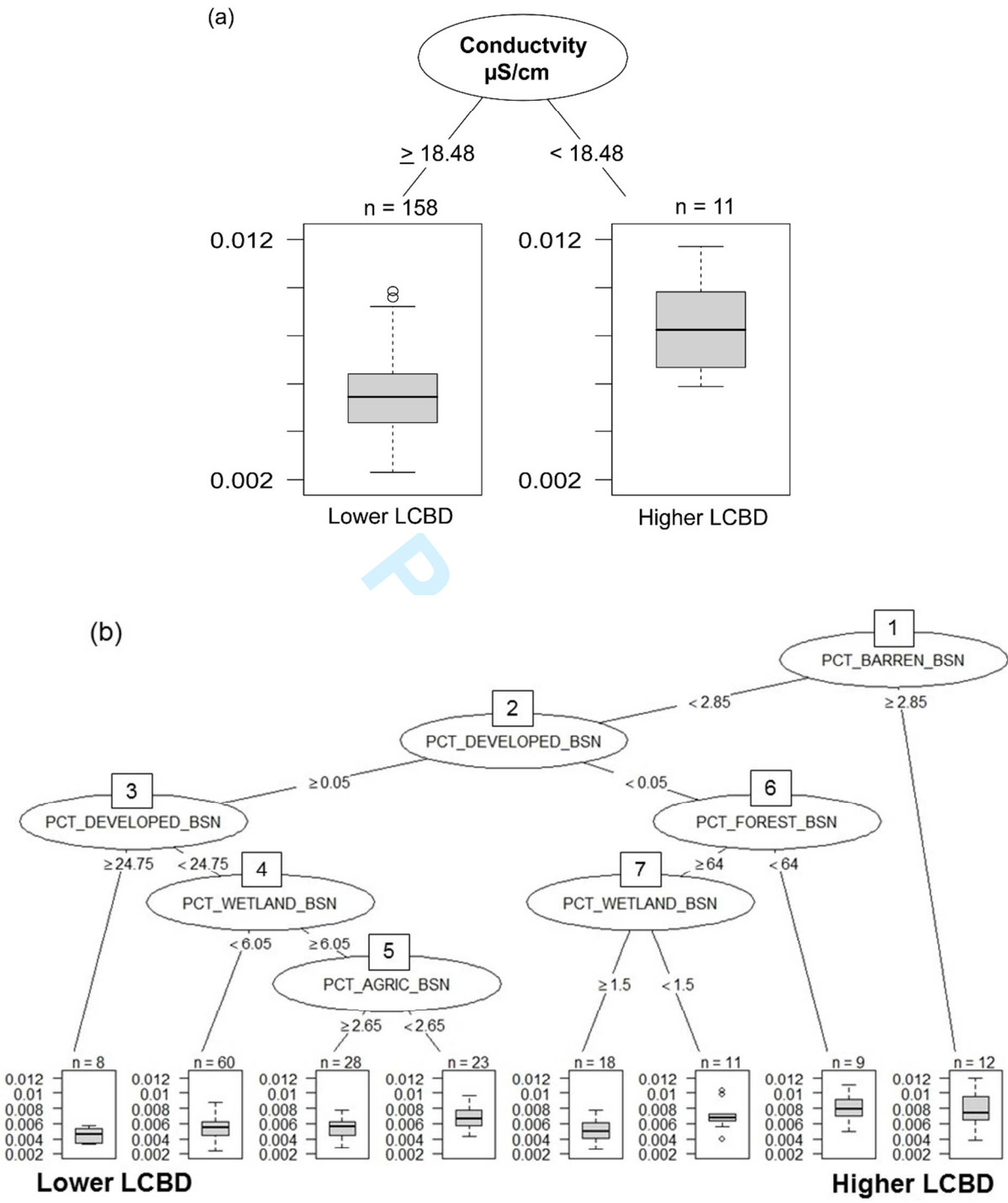


Figure 4

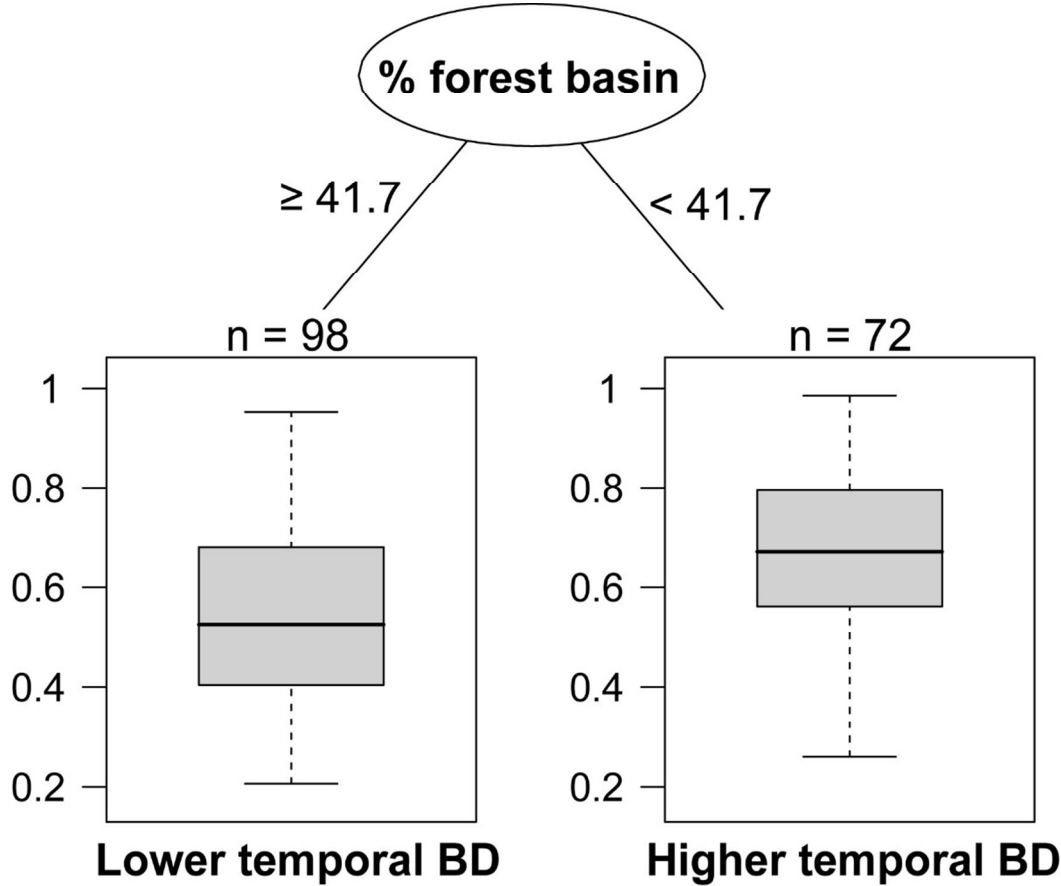


Figure 5

Supporting Information

Appendix to:

Winegardner, A. K., Legendre P., Beisner, B.E., & Gregory-Eaves, I. 2017. Diatom diversity patterns over the past ~ 150 years across the conterminous United States: identifying mechanisms behind beta diversity. *Global Ecology and Biogeography* (in press).

Appendix S1

Description of sediment core screening with respect to length and predicted age of bottom sediment core samples

Lakes sampled in the 2007 NLA were assigned a weight based on their surface area to reflect the portion of all U.S. lakes represented in the survey (see U.S. EPA, 2011-2012 for details). Thus, the 1000+ lakes sampled in 2007 were estimated to represent ~50000 lakes across the U.S. (U.S. EPA, 2009). However, not all of these lakes could be included in this paleolimnological study, firstly because not all lakes were cored and secondly because of differences in sediment age. The need to accurately identify the age of bottom core sediment samples from the 2007 NLA is exemplified in the debate around Bachman *et al.* (2013), which also used surface and bottom core sediment data from the 2007 NLA, but for the purpose of quantifying the extent of eutrophication across the U.S.A. The main criticisms of their work (Smith (2014) and McDonald *et al.* (2014)) revolved around the fact that Bachman *et al.* (2013) had relied on descriptions from the various field teams collecting the sediment cores, as well as (mostly) qualitative criteria from the EPA to classify whether a lake where was one where the bottom of the core was sufficiently deep to have reached sediment from pre-European settlement conditions (hereafter referred to as a “high confidence” (HC) cores, e.g. see U.S. EPA (2010), p.32-33). Indeed, Bachman *et al.* (2013) and later Bachman *et al.* (2014) identified 233 lakes with cores deemed HC by the U.S. EPA based on a number of factors (U.S. EPA, 2010) and used data from bottom samples of these cores in subsequent analyses. However, there is still considerable variation in core length and comments by both Smith (2014) and McDonald *et al.* (2014) suggest that the criteria used to identify these cores preferentially selected samples from relatively short cores (likely not pre-European settlement, or even pre-industrial conditions).

While Bachman *et al.* (2013) and (2014) used the information available on the sediment cores to the greatest extent possible, we utilized a more extensive, three-fold approach to increase the accuracy in identifying cores where the bottom samples likely represent pre-industrial conditions. We also used a pre-1850 CE cut-off to refer to pre-industrial conditions and are not attempting to determine whether the bottoms of cores date back to pre-European settlement conditions. First, by using the length of the cores collected, we estimated the approximate age of the core bottoms using regression equations of latitude and sedimentation rate developed by Brothers *et al.* (2008). We determined that the majority of cores longer than 30 cm in length had bottom sediments estimated to date back to at least 1850 CE. Secondly, we

cross-referenced the list of cores greater than 30 cm in length with the list of designated HC cores. This produced a list of sites with core bottom samples likely older than 1850 CE. Finally, after procuring leftover sediment from a set of bottom core samples (archived at The Academy of Natural Sciences of Drexel University), we randomly selected 35 bottom core samples and further selected an additional 15 bottom core samples from the shortest cores where material was available to undergo radiometric dating. Radiometric dating produced ratios of ^{214}Bi and ^{210}Pb , where activity of ^{210}Pb within two standard errors of ^{214}Bi indicates sediment older than 1850 CE (Dixit *et al.* 1999; Vermaire *et al.* 2012). These radiometric estimates are not as accurate as measuring the decay of unsupported ^{210}Pb throughout a full core, but was the best approximation available since intervals between the top and bottom samples were not kept from the collected cores.

We used chi-square tests to test two hypotheses; first, that age assignment by radiometric dating is independent of core length and second, that age assignment by radiometric dating is independent of the age estimate using the Brothers *et al.* (2008) equation, which accounts for the variation in sedimentation rates in lakes across latitudes. The chi-square test of age assignment and core length ($n = 35$), transformed into binary variables, resulted in a P -value of 0.9, meaning that we cannot reject the null hypothesis that core length is independent of whether a sample is determined to be pre-1850 CE. While we could not use this radiometric dating method on all the candidate bottom samples (due to sample availability and costs), the Chi-square test of age assignment by radiometric dating (for the 35 samples that could be radiometrically dated) and age assignment by regression equations ($n = 35$) resulted in a P -value of 0.6, meaning that we cannot reject the null hypothesis that these two age assignments are independent. As such, both core length and age based on regression estimates appear to be reasonable indicators of core age, especially as the regression estimates factor in the effect of latitude on sedimentation rate. However, age based on regression estimates (i.e. the Brothers *et al.* 2008) provide more robust predictions of ages when compared to radiometric dating.

References

- Bachman, R.W., Hoyer, M.V., & Canfield, Jr., D.E. (2013) The extent that natural lakes in the United States of America have been changed by cultural eutrophication. *Limnology and Oceanography*, **58**, 945-950.
- Bachman, R.W., Hoyer, M.V., & Canfield, Jr., D.E. (2014) Response to comments: Quantification of the extent of cultural eutrophication of natural lakes in the United States. *Limnology and Oceanography*, **59**, 2231-2239.
- Brothers, S., Vermaire, J.C., & Gregory-Eaves, I. (2008) Empirical models for describing recent sedimentation rates in lakes distributed across broad spatial scales. *Journal of Paleolimnology*, **40**, 1003-1019.
- 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*, **42**, 1860-1869.

McDonald, C.P., Lottig, N.R., Stoddard, J.L., Herlihy, A.T., Lehmann, S., Paulsen, S.G., Peck, D.V., Pollard, A.I., & Stevenson, R.J. (2014) Comment on Bachman et al. (2013): A nonrepresentative sample cannot describe the extent of cultural eutrophication of natural lakes in the United States. *Limnology and Oceanography*, **59**, 2226-2230.

Smith, V.H. (2014) Comment: Cultural eutrophication of natural lakes in the United States is real and widespread. *Limnology and Oceanography*, **59**, 2217-2225.

U.S. EPA (2011-2012) National Lakes Assessment. Field Operations Manual. EPA 841-B-11-003. U.S. Environmental Protection Agency, Washington, DC, USA.

Vermaire, J.C., Prairie, Y.T., & Gregory-Eaves, I. (2012) Diatom-inferred decline of macrophyte abundance in lakes of southern Quebec, Canada. *Canadian Journal of Fisheries and Aquatic Sciences*, **69**, 511-524.

Appendix S2

Description of additional analyses using the Baselga family of indices

There is considerable literature focusing on different methodologies relating to the quantification of spatial beta diversity. In a 2011 review of beta diversity methods, Anderson *et al.* (2011) provided clarity on different ways that spatial beta diversity can be interpreted, differentiating between directional spatial beta diversity along a gradient and non-directional spatial beta diversity (a value computed from all pairwise comparisons). The Anderson *et al.* (2011) review also highlighted the fact that there is no shortage of methodological options for the study of spatial beta diversity.

Recent work on beta diversity has drawn attention to the insight gained when beta diversity is partitioned into different explanatory components (e.g., Cardoso *et al.* 2009; Baselga, 2010; Podani & Schmera, 2011; Schmera & Podani, 2011; Baselga, 2012; Baselga & Orme, 2012; Carvalho *et al.* 2013; Podani *et al.*, 2013; and Legendre, 2014). Such analyses allow investigators to quantify the contributions of varying ecological processes to beta diversity, instead of only focussing on the magnitude of a beta diversity index. Early partitioning methods referred to the components of beta diversity as species replacement, richness or abundance difference, and nestedness. Species replacement refers to the “simultaneous gain and loss of species (or individuals belonging to a particular species, or biomass) along an ecological gradient” (Podani & Schmera, 2011; Baselga & Orme, 2012). This means that as species (measured as individuals or biomass) are lost from sites (space, time, etc.), others take their place. Richness difference (or abundance difference when using abundance data) means that one site (or sampling unit) has more unique species than another (Podani & Schmera, 2011; Baselga & Orme, 2012). Nestedness is essentially a special case of richness difference where the species found in one site (or sampling unit) are a subset of the species found in another site having higher richness (Baselga & Orme, 2012). In Baselga (2013) and Podani *et al.* (2013), components are further elaborated for abundance-based beta diversity into balanced variation and abundance gradients.

Legendre (2014) provided a thorough review and critique of the spatial beta diversity partitioning approaches outlined in Baselga & Orme (2012), Baselga (2013), Podani & Schmera (2011) and Schmera & Podani (2011), referring to these slightly different methods as belonging to the ‘Baselga family’ and ‘Podani family’ of analyses respectively. In the present study, we computed the Podani family of indices using the functions *beta.div()* (to obtain total beta diversity and LCBD indices) and *beta.div.comp()* (for partitioning total beta diversity into replacement, richness/abundance difference and nestedness) written by P. Legendre and available in the *adespatial* R package (Dray *et al.* 2017). Another alternative would have been to partition spatial beta diversity using the indices of the Baselga family. These can also be computed with function *beta.div.comp()* of *adespatial*.

Functions *beta.pair()* (for incidence data) and *beta.pair.abund()* (quantitative data) of the *betapart* package (Baselga & Orme, 2012) can be used to compute dissimilarity matrices of the Jaccard and Sørensen indices and their quantitative forms, the Ružička and percentage

difference (aka Bray-Curtis) indices, as well as their decompositions into matrices of replacement and nestedness indices, but these functions do not compute the Baselga-family beta diversity indices or their component indices, *Repl* and *Nes* (following the nomenclature of Legendre 2014). The matrices computed by *beta.pair()* and *beta.pair.abund()* can, however, be used as input into the *LCBD.comp()* function of the *adespatial* package to obtain total beta diversity indices for the Baselga-family indices of replacement and nestedness. Doing the calculations with *beta.div.comp()* or with a combination of *beta.pair()* or *beta.pair.abund()* plus *LCBD.comp()* leads to identical numerical results.

While the focus of this study was on applying beta diversity techniques (Podani family) to the study of diatom variation across the conterminous United States, we also analysed spatial beta diversity using the *betapart* package in order to highlight any similarities or differences between the Podani and Baselga approaches. We computed spatial beta diversity on the genus-level presence-absence and quantitative diatom data for both the modern and historical time points using both the *beta.div.comp()* function of *adespatial* and the *beta.pair()* and *beta.pair.abund()* functions of *betapart* followed by *LCBD.comp()*. For presence-absence data and when using *beta.div.comp()* we ran the analysis using the Sørensen index of the Podani family approach and the Sørensen index of the Baselga family for comparison purposes. We then ran a second analysis (Baselga family) using *beta.pair()*. For quantitative data, we ran the *beta.div.comp()* analysis using the percentage difference (Bray) index of the Podani family approach and the percentage difference index of the Baselga family, again for comparison purposes. We then ran the analysis (Baselga family) using *beta.pair.abund()*.

The analyses performed with the different Podani or Baselga approaches, across the three functions (S2, Table 1), produced identical total beta diversity (BD_{Total}) results. Likewise, results for all components (replacement, richness/abundance difference, nestedness) of the Baselga family, computed using the two computation roads, were also identical.

S2, Table 1: Total spatial beta diversity results for the historical and surface (2007) genus-level diatom data, computed using both the Podani and Baselga approaches. Results are shown for both quantitative and qualitative data with the relevant R package and function and are rounded to four decimal points.

Function	Coefficient	Total spatial beta diversity (BD_{Total})
Qualitative data		
<i>Surface data, 2007</i>		
<i>beta.div.comp()</i>	Podani – Sørensen	0.2301
<i>beta.div.comp()</i>	Baselga - Sørensen	0.2301
<i>beta.pair()</i>	Baselga - Sørensen	0.2301
<i>Historical</i>		
<i>beta.div.comp()</i>	Podani – Sørensen	0.2282
<i>beta.div.comp()</i>	Baselga - Sørensen	0.2282
<i>beta.pair()</i>	Baselga - Sørensen	0.2282
Quantitative data		

<i>Surface data, 2007</i>		
beta.div.comp()	Podani – percentage difference	0.3553
beta.div.comp()	Baselga – percentage difference	0.3553
beta.pair.abund()	Baselga – percentage difference (Bray)	0.3553
<i>Historical</i>		
beta.div.comp()	Podani – percentage difference	0.3553
beta.div.comp()	Baselga – percentage difference	0.3553
beta.pair.abund()	Baselga – percentage difference (Bray)	0.3553

References

- Anderson, M.J., Crist, T.O., Chase, J.M., Vellend, M., Inouye, B.D., Freestone, A.L., Sanders, N.J., Cornell, H.V., Comita, L.S., Davies, K.F., Harrison, S.P., Kraft, N.J.B., Stegen, J.C., & Swenson, N.G. (2011) Navigating the multiple meanings of β diversity: a roadmap for the practicing ecologist. *Ecology Letters*, **14**, 19-28.
- Baselga, A. (2010) Partitioning the turnover and nestedness components of beta diversity. *Global Ecology and Biogeography*, **19**, 134–143.
- Baselga, A. (2012) The relationship between species replacement, dissimilarity derived from nestedness, and nestedness. *Global Ecology and Biogeography*, **21**, 1223–1232.
- Baselga A., & Orme C.D.L. (2012) betapart: an R package for the study of beta diversity. *Methods in Ecology and Evolution*, **3**, 808-812.
- Baselga, A. (2013) Separating the two components of abundance-based dissimilarity: balanced changes in abundance vs. abundance gradients. *Methods in Ecology and Evolution*, **4**, 552-557.
- Cardoso, P., Borges, P.A.V., & Veech, J.A. (2009) Testing the performance of beta diversity measures based on incidence data: the robustness to undersampling. *Diversity and Distributions*, **15**, 1081–1090.
- Carvalho, J. C., Cardoso, P., Borges, P.A.V., Schmera, D., & Podani, J. (2013) Measuring fractions of beta diversity and their relationships to nestedness: a theoretical and empirical comparison of novel approaches. *Oikos*, **122**, 825–834.
- Dray, S., G. Blanchet, D. Borcard, G. Guérande, T. Jombart, G. Larocque, P. Legendre, N. Madi & H. H. Wagner. (2017). adespatial: Multivariate multiscale spatial analysis. R package version 0.0-8. <http://cran.r-project.org/package=adespatial>.

Legendre P. (2014) Interpreting the replacement and richness difference components of beta diversity. *Global Ecology and Biogeography*, **23**, 1324-1334.

Podani J., & Schmera D. (2011) A new conceptual and methodological framework for exploring and explaining patterns in presence-absence data. *Oikos*, **120**, 1625-1638.

Podani J., Ricotta C., & Schmera D. (2013) A general framework for analyzing beta diversity, nestedness and related community-level phenomena based on abundance data. *Ecological Complexity*, **15**, 52-61.

Schmera D., & Podani J. (2011) Comments on separating components of beta diversity. *Community Ecology*, **12**, 153-160.

Appendix S3

Description of the Temporal Beta diversity Indices (TBI) computed using function *TBI()*

Beta diversity is the variation in community composition among sites (spatial beta diversity) or across time (temporal beta diversity). Temporal beta diversity (TBI) was assessed at both the species and genus levels in our study. The results at the genus level are reported in the main paper and those at the species level are in Appendix S4. When comparing surveys carried out at two time periods, the differences in species composition for a site is computed as the sum of two quantities, called *B* and *C* here, which correspond respectively to the losses and gains that occurred between time 1 (T1) and time 2 (T2):

B: abundance loss component, i.e. sum of the losses of abundances on a taxon-by-taxon (species or genera) basis between T1 and T2;

C: abundance gain component, i.e. sum of the gains of abundances on a taxon-by-taxon (species or genera) basis between T1 and T2;

$(B + C)$ is the total change in taxonomic composition (species or genera) between T1 and T2.

A is the sum of the taxonomic abundances (species or genera) common to the observations at T1 and T2.

These quantities are the elements from which the replacement and abundance difference components of total beta diversity can be computed. For example, in the family of indices called the Podani family, replacement is computed as $2 \times \min(B, C)$ and richness difference as $|B - C|$; see Legendre (2014) for equations and details. So, the quantities *B* and *C* were available as a by-product of the decomposition of total beta diversity into replacement and abundance difference components.

The Ružička (1958) dissimilarity index is $D_R = (B + C) / (A + B + C)$ whereas the percentage difference index (Odum, 1950), also known as Bray-Curtis index, is $D_{\%diff} = (B + C) / (2A + B + C)$. The numerator, $(B + C)$, is the same in these two indices. For computations based on taxonomic (species or genera) presence-absence data, the Ružička dissimilarity becomes the Jaccard dissimilarity and the percentage difference becomes the Sørensen dissimilarity; these two dissimilarities are widely used for the analysis of presence-absence data in ecology, genetics and other fields. Hence the study of the *B* and *C* components can be done on presence-absence as well as on taxonomic abundance data. The TBI indices computed in function *TBI()* (below) are the Ružička and percentage difference indices for abundance data, and the Jaccard and Sørensen dissimilarity indices for presence-absence data; they are applied to the comparison of surveys carried out at times T1 and T2. The function also offers an analysis of the *B* and *C* components of these dissimilarities.

Each component, B and C , was divided by a denominator (den), which was $(A+B+C)$ for computations based on the Ružička index and $(2A+B+C)$ for the percentage difference index. After division, each component was in the $[0,1]$ interval and $D = B + C$. We computed temporal beta diversity between the historical diatom assemblage and the 2007 diatom assemblage for each lake, recording total beta diversity (D) as well as the taxonomic loss (B) and gain components (C).

We computed these values for each ecoregion separately and summarized them to examine the relationship between each of these components and latitude and longitude, using OLS regression.

Computations were done using function $TBI()$ in R (Legendre, 2015). Besides computation of the normalized B and C components of D for each site, the function also computes and displays the following summary statistics for all data in the study:

- $mean(B/den)$,
- $mean(C/den)$,
- $mean(D)$,
- $mean(B/den)/mean(D)$ [called $B/(B+C)$ in the table header produced by the function],
- $mean(C/den)/mean(D)$ [called $C/(B+C)$ in the table header produced by the function],
- as well as the sign of the overall change for the whole study (+ for overall gains and – for overall losses) obtained by comparing the overall values of $B/(B+C)$ and $C/(B+C)$.

The loss (B) and gain (C) statistics were first described by Legendre & Salvat (2015) who used them to analyse before-after community composition data.

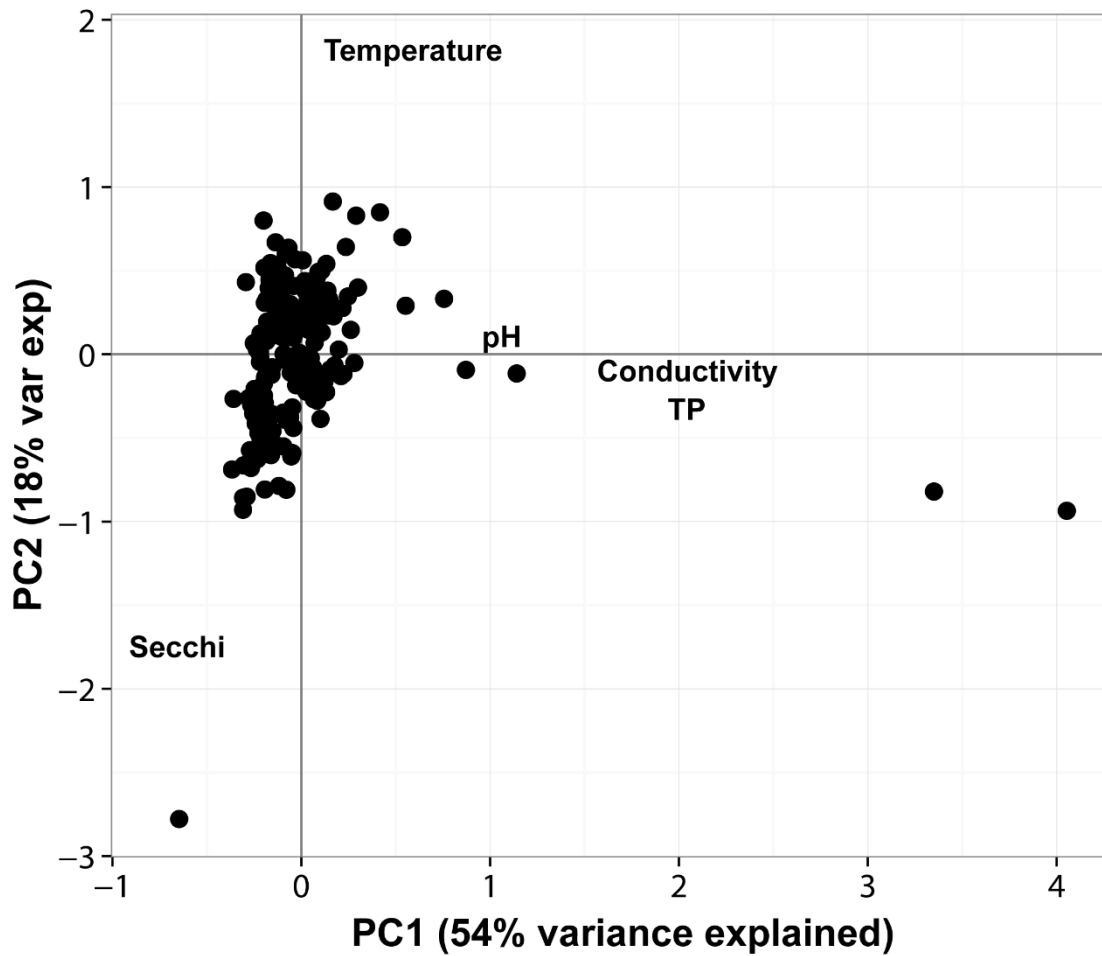
References

- Legendre, P. (2014) Interpreting the replacement and richness difference components of beta diversity. *Global Ecology and Biogeography*, **23**, 1324-1334.
- Legendre, P. (2015) TBI: Difference between multivariate observations at T1 and T2. R function available online at <http://adn.biol.umontreal.ca/~numericalecology/Rcode/>.
- Legendre, P., & Salvat, B. (2015) Thirty-year recovery of mollusc communities after nuclear experimentations on Fangataufa atoll (Tuamotu, French Polynesia). *Proceedings of the Royal Society B*, **282**, 20150750.
- Odum, E. P. (1950) Bird populations of the Highlands (North Carolina) plateau in relation to plant succession and avian invasion. *Ecology*, **31**, 587-605.
- Ružička, M. (1958) Anwendung mathematisch-statistischer Methoden in der Geobotanik (synthetische Bearbeitung von Aufnahmen). *Biologia, Bratislava*, **13**, 647–661.

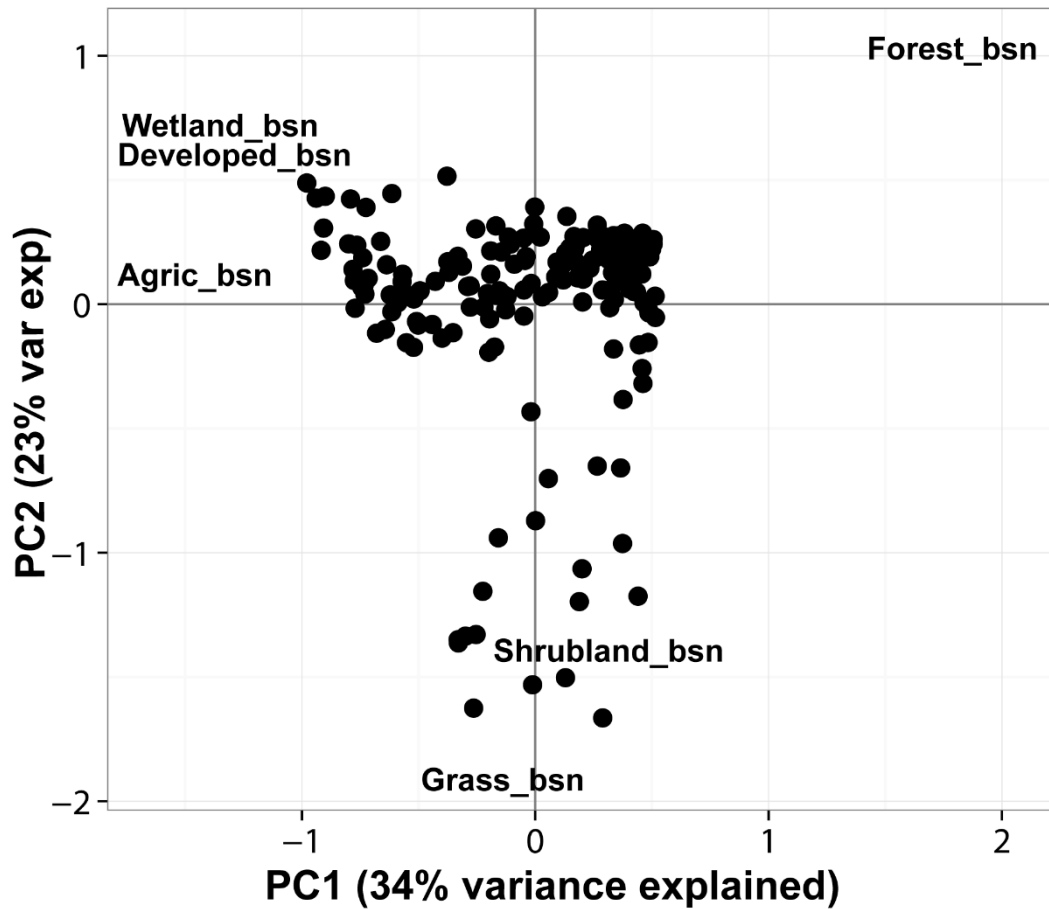
Appendix S4

Principal Component Analyses of (a) standardized 2007 water quality variables and (b) land cover variables

(a) Water quality variables – Chlorophyll *a*, TN and TP were correlated and represented here in the PCA by TP. Conductivity and TP were log transformed.



- (b) Land cover variables (NLCD 1992) – “Forest_bsn” refers to percent forest in basin including all forest types; “Wetland_bsn” refers to percent wetland; “Developed_bsn” refers to residentially developed area in a basin, including low, medium and high values; “Agric_bsn” refers to percent agriculture including all agriculture types (crops, pasture etc.); “Shrubland_bsn” refers to percent shrubland.



Appendix S5

Tables and figures for **species-level** analyses (n=52)

S4, Table 1: Spatial beta (β) diversity, mean rarefied species richness, mean alpha (α) diversity and gamma (γ) diversity for each ecoregion (Shannon diversity of species' sums). Beta diversity was calculated using total variance computed using *beta.div()* based on percentage difference matrices. "Hx" refers to the historical sediments. Species richness was rarefied after rare species (<2% relative abundance) had been excluded. The Xeric, Southern Plains, Temperate Plains, Western Mountains and Upper Midwest ecoregions were excluded in these mean values because there were no sites in those regions for this reduced dataset.

Ecoregion	β -diversity		Rarefied S		α -diversity (Shannon)		α -diversity (Simpson)		γ -diversity	
	Hx	2007	Hx	2007	Hx	2007	Hx	2007	Hx	2007
All	0.40	0.37	56.4	52.8	2.9	2.9	0.9	0.9	4.5	4.4
Northern Appalachians	0.39	0.36	59.8	54.9	3.0	2.9	0.9	0.9	4.5	4.4

S4, Table 2: Explanatory components for historical and 2007 spatial beta diversity, as computed using *beta.div.comp()*. 'Repl' refers to the replacement component; 'AbDiff' refers to the abundance difference component; and 'Repl/Total' and 'AbDiff/Total' are these two components with total beta diversity as the denominator.

Ecoregion	Repl	AbDiff	Repl/Total	AbDiff/Total
<i>Historical</i>				
All	0.40	0	1	0
Northern Appalachians	0.39	0	1	0
<i>2007</i>				
All	0.37	0	1	0
Northern Appalachians	0.36	0	1	0

S4, Table 3: Mean (and standard deviation) temporal beta diversity components for the ecoregions. ‘Total beta’ refers to the mean value of the temporal beta diversity in each region (mean value of the D column in the ‘BCD’ table provided by the function *TBI()*). It was computed using the percentage difference index applied to the diatom abundance data; values are in the [0,1] range. Total beta is the sum of ‘Species loss’ and ‘Species gain’. Species loss refers to the component representing loss of abundances on a species by species basis between the historical and 2007 time points. Species gain refers to the component representing gain of abundances on a species by species basis between the historical and 2007 time points. These components were computed on a lake-by-lake basis and then averaged for each ecoregion.

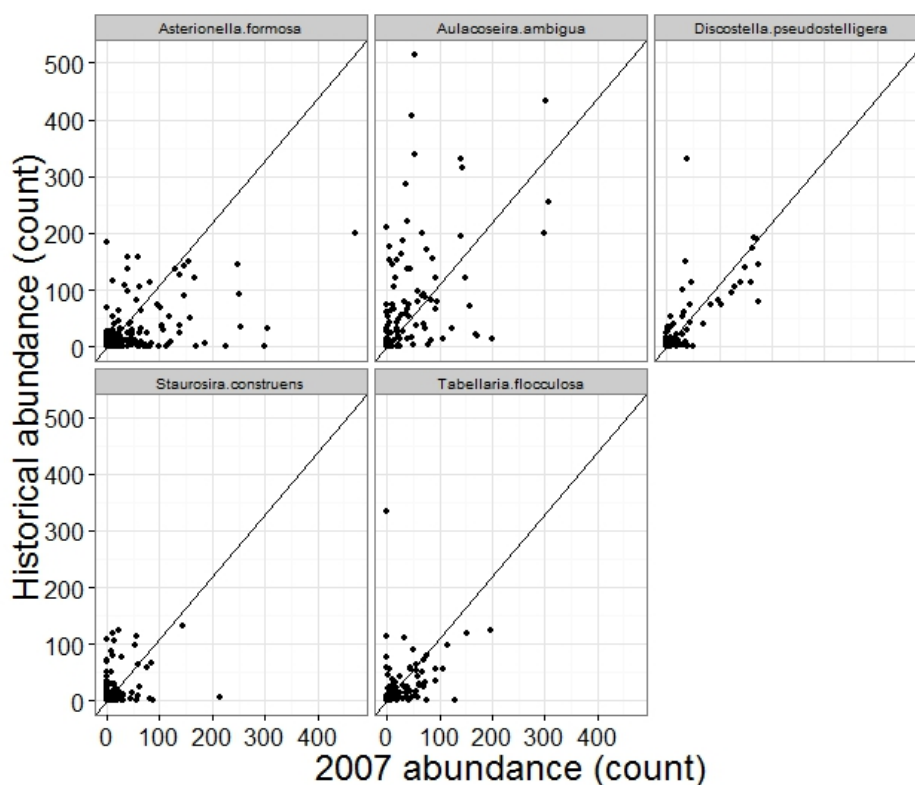
Ecoregion	Species loss	Species gain	Total beta
Northern Appalachians	0.24 (0.08)	0.24 (0.08)	0.48 (0.17)

Appendix S6

Diatom species abundance in historical (Hist_) and 2007 sediments (Surf_) for the species identified as having the highest Species Contribution to Beta Diversity (SCBD) values in sediments from both time points (n=52, species-level analyses)

The ten species with highest contributions in historical assemblages were (in descending order): *Achnanthes biasolettiana*, *Aulacoseira ambigua*, *Discostella pseudostelligera*, *Asterionella formosa*, *Aulacoseira subarctica*, *Staurosira construens*, *Discostella stelligera*, *Cyclotella bodanica*, *Staurosira construens*, and *Tabellaria flocculosa*. The ten highest contributions for the 2007 assemblages were from: *Achnanthes cf. convergens*, *Asterionella formosa*, *Fragilaria crotenensis*, *Aulacoseira ambigua*, *Discostella pseudostelligera*, *Tabellaria flocculosa*, *Cyclotella comensis*, *Cyclotella ocellata*, *Staurosira construens*, and *Pseudostaurosira brevisriata*.

The figure below shows the relationship between diatom abundance in 2007 versus historical sediments for the five species in common between the two time points.



S5, Figure 1: Relationship between diatom abundance in 2007 versus historical sediments for five species in common between the two time points. The 1:1 line is shown to aid in interpretation of lack of correlation.