



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

Journal:	<i>Global Ecology and Biogeography</i>
Manuscript ID	Draft
Manuscript Type:	Research Papers
Date Submitted by the Author:	n/a
Complete List of Authors:	Winegardner, Amanda; McGill University , Biology Legendre, Pierre; Université de Montréal, Département des Sciences Biologiques Beisner, Beatrix; Université du Québec à Montréal, Département des sciences biologiques Gregory-Eaves, Irene ; McGill University , Biology
Keywords:	Beta diversity, community ecology, diatoms, freshwater, paleoecology, paleolimnology

1  
2  
3 1 **Running title:** Diatom beta diversity through time and space  
4  
5  
6 2  
7  
8 3 **Title:** Diatom diversity patterns over the past ~ 150 years across the conterminous United States:  
9  
10 4 identifying mechanisms behind beta diversity  
11  
12  
13 5  
14  
15 6 **Authors:** Amanda K. Winegardner<sup>\*1,2</sup>, Pierre Legendre<sup>2,3</sup>, Beatrix E. Beisner<sup>2,4</sup>, and Irene  
16  
17 Gregory-Eaves<sup>1,2</sup>  
18  
19  
20 8 **Affiliations:** 1. Department of Biology, McGill University, 1205 avenue du Docteur Penfield,  
21  
22 9 Montréal, Québec, Canada, H3A 1B1  
23  
24 10 2. Québec Centre for Biodiversity Science, 1205 avenue du Docteur Penfield, Montréal, Québec,  
25  
26 11 Canada, H3A 1B1  
27  
28 12 3. Département de sciences biologiques, Université de Montréal, C.P. 6128, succursale Centre-  
29  
30 13 ville, Montréal, Québec, Canada H3C 3J7  
31  
32 14 4. Department of Biological Sciences, University of Québec at Montréal, C.P. 8888, succursale  
33  
34 15 Centre-ville, Montréal, Québec, Canada H3C 3P8  
35  
36  
37 16 \*corresponding author.  
38  
39  
40 17 **Email addresses:** amanda.winegardner@mail.mcgill.ca, pierre.legendre@umontreal.ca,  
41  
42  
43 18 beisner.beatrix@uqam.ca, irene.gregory-eaves@mail.mcgill.ca  
44  
45  
46 19 **Keywords:** Beta diversity, community ecology, diatoms, freshwater, paleoecology,  
47  
48 20 paleolimnology  
49  
50  
51 21 **No. of words in Abstract:** 316  
52  
53 22 **No. of words in main body (Introduction through Biosketch):** 5651  
54  
55 23 **No. of references:** 48  
56  
57  
58  
59  
60

**Abstract**

**Aim:** Understanding the magnitude and drivers of freshwater diversity change 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 176 lakes, and species-level data for 59 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 two time points. TBIs were decomposed into taxon losses and gains to facilitate interpretation. TBIs and their components were related to contemporary land cover variables.

**Results:** Temporal beta diversity varied significantly with forest cover and longitude, with higher values in western regions. Spatial beta diversity was similar between the historical and 2007 time points, with genus replacement explaining almost all variation. There was no systematic pattern in lakes that contributed to spatial beta diversity; however local contributions were explained by a specific structure of water quality and land cover variables.

**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 47 occurred through genus gains and losses and was significantly related to forest and agriculture  
4  
5 48 cover in watersheds, with genus replacement dominating beta diversity at both time points.  
6  
7  
8 49 These analyses, pairing spatial and temporal beta diversity, provide insight into the mechanisms  
9  
10 50 behind diatom diversity.  
11

12 51

13 52

14 53

15 54

16 55

17 56

18 57

19 58

20 59

21 60

22 61

23 62

24 63

25

26

27

28

29

30

31

32

33

34

## Introduction

Modern geological time is commonly referred to as the Anthropocene, a designation recognizing the extent to which humans 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 (e.g. Vellend *et al.* 2013; Dornelas *et al.* 2014) or 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 composition (Whittaker, 1972). Spatial beta diversity is particularly interesting because it can

1  
2  
3 87 identify sites and regions that are exceptional across a landscape owing to degraded or enhanced  
4  
5 88 diversity. Temporal beta diversity can also indicate degrees of change in composition through  
6  
7  
8 89 time at single locations.  
9

10 90 Despite the usefulness of considering numerous biodiversity metrics at multiple scales,  
11  
12 91 very few studies have attempted to consider more than a few trends simultaneously (McGill *et al.*  
13  
14 92 2015). The majority of biodiversity change studies have also been largely restricted to the last  
15  
16  
17 93 50 years, with only a handful extending beyond this time frame to encompass the entire  
18  
19 94 Anthropocene. This literature has been recently synthesized to develop predictions of various  
20  
21 95 metric-scale combination trends during the Anthropocene (McGill *et al.* 2015, Table 1), together  
22  
23 96 with a call to test these predictions.  
24  
25  
26

27 97 Aquatic systems provide a unique opportunity to quantify several biodiversity trends over  
28  
29 98 different spatial and temporal scales through the use of the historical archives found in lake  
30  
31 99 sediments (Gregory-Eaves & Beisner, 2011). We analysed a large paleolimnological dataset to  
32  
33 100 quantify diversity trends across both space and time and explore mechanisms behind diatom  
34  
35 101 assemblage variation across the conterminous US during the Anthropocene. We used sediment  
36  
37 102 cores collected as part of the 2007 National Lakes Assessment (NLA), United States  
38  
39 103 Environmental Protection Agency (U.S. EPA;  
40  
41 104 [http://water.epa.gov/type/lakes/lakessurvey\\_index.cfm](http://water.epa.gov/type/lakes/lakessurvey_index.cfm)) to examine patterns of diatom genera and  
42  
43  
44 105 species richness, as well as alpha, beta and gamma diversities for 176 lakes. To identify beta  
45  
46 106 diversity hot spots and relate these to water quality and land cover, we computed spatial beta  
47  
48 107 diversity across all lakes at both the historical (pre-1850 CE) and modern (2007) time points.  
49  
50  
51 108 We also calculated temporal beta diversity within each lake to identify compositional variation  
52  
53 109 between the historical and modern time points (Figure 1), asking:  
54  
55  
56  
57  
58  
59  
60

- 1) What is the relative magnitude of spatial beta diversity across different ecoregions at each time point?
- 2) What is the magnitude of temporal beta diversity for each lake?
- 3) Which lakes contribute the most to spatial beta diversity? Which lakes show significant change through time?
- 4) Is spatial beta diversity explained mostly by replacement of genera/species or differences in richness and abundances of genera/species across space?
- 5) What are the environmental drivers of spatial beta diversity?
- 6) Is the change through time explained by loss or gain of taxa abundances?
- 7) What are the environmental drivers of temporal beta diversity?

We developed hypotheses based on the trends identified in McGill *et al.* (2015), adapting these predictions to the three scales used in this study: continental US, ecoregion and individual lakes (Table 1).

## 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 >10 acres and deeper than 1 m (U.S. EPA, 2009). The lakes were randomly selected using a combination of probabilistic design and specifically targeted ‘reference’ (pristine/undisturbed) lakes identified by state and tribal partners (U.S. EPA, 2011-2012). After sampling, lakes were re-classified as necessary to three categories representing disturbance levels: least disturbed, intermediate, and highly disturbed relative to other lakes in the same ecoregion (A.T. Herlihy, pers. comm.). The

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

2007 NLA data includes water quality measurements, land use 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](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 CE (pre-industrial) conditions. Cores were collected from the deepest point of the lake using a modified Kajak-Brinkhurst corer (Glew, 1989). Both the top and bottom 1 cm interval of the sediment was saved for diatom enumeration, with up to 500 diatom valves counted using standardized methods from the U.S. Geological Survey National Water Quality Assessment (Charles *et al.* 2003, U.S. EPA, 2011-2012). Thus, for each lake where sediment coring occurred, there are diatom assemblage data for both modern (2007) conditions as well as bottom of the core (historical) conditions. To account for different core lengths and variation in sedimentation across lakes, we used screening criteria to select lakes for our study, ensuring that the bottom sediment samples used represented historical (pre-1850 CE) conditions. We used a three-fold approach to estimate age of sediment cores (described in S1) and identified 179 sites for which we were confident that their bottom samples represented pre-1850 CE conditions. Three Coastal Plains lakes that were very isolated by a few 1000s of kilometres from others in the same ecoregion were removed. The final set of lakes used in this study (n = 176) is shown in Figure 2a, along with ranges of key limnological variables in Figure 2b-e. Note that there was a



geographic bias with bottom core samples were not collected from many lakes from the mid- and southern portions of the country where reservoirs dominate.

### *Diversity analyses*

All statistical analyses were completed in R v. 3.1.2 (R Core Team, 2014). Genus-level was the taxonomic level associated with a higher level of consistency between lab groups enumerating the diatom data for the 2007 NLA (Pollard, Pers. Comm.). As such, we conducted the analyses on the full set of lakes ( $n = 176$ ) with the diatom classified to genus-level. For a smaller set of the lakes ( $n = 59$ ), we completed all the analyses again at the species level data because for this smaller set, all samples were enumerated by a single lab group (The Academy of Natural Sciences). 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 or 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 using the *rarefy()* function in *vegan* (Oksanen *et al.* 2015), correcting for the total number of valves counted for each sample, setting all samples to a cut-off of 300 valves (150 individuals). 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 only dominant (>2% relative abundance) 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).

We used four source functions developed by P. Legendre: *beta.div()* (Legendre & De Cáceres, 2013, App. S4; Legendre, 2013a), *beta.div.comp()* (Legendre, 2014, App. S3),

1  
2  
3 179 *LCBD.comp()* (Legendre, 2013b; Legendre, 2014, App. S5), and *TBI()* (Legendre, 2015) for use  
4  
5  
6 180 in R. We used *beta.div()* to compute spatial beta diversity for all sites (either n = 176 or n = 59)  
7  
8 181 across the landscape at both the historical and contemporary time points, using non-transformed  
9  
10 182 matrices of diatom abundances and the percentage difference method (Legendre & De Cáceres,  
11  
12 183 2013). We also computed spatial beta diversity at the two different time points within six  
13  
14  
15 184 ecoregions, as defined by Herlihy *et al.* (2008): Coastal Plains, Northern Appalachians, Southern  
16  
17 185 Plains, Temperate Plains, Upper Midwest, and Western Mountains (Xeric was excluded from  
18  
19  
20 186 spatial beta diversity computations because n = 1) (Figure 2a, n = 176). The function *beta.div()*  
21  
22 187 also provided a ‘Local Contribution to Beta Diversity’ (LCBD index) for each lake as well as a  
23  
24 188 permutational p-value indicating, the significance of each LCBD value. LCBD values provide a  
25  
26  
27 189 metric to assess the individual importance of each lake on total beta diversity (Legendre & De  
28  
29 190 Cáceres, 2013). We identified the sites with significant LCBD that were the same or different  
30  
31  
32 191 between the historical and 2007 sediments and performed a chi-square test of the null hypothesis  
33  
34 192 that processes producing significant LCBD were independent at the two time points. To assess  
35  
36 193 individual species’ influences on beta diversity, we computed ‘Species Contributions to Beta  
37  
38 194 Diversity’ (SCBD indices) using a Hellinger distance measure (Legendre & De Cáceres, 2013).  
39  
40  
41 195 SCBD allowed us to identify species that were important contributors to spatial beta diversity,  
42  
43  
44 196 both historically and in modern samples.

45  
46  
47 197 We were interested in the mechanisms generating spatial beta diversity at both time  
48  
49 198 points. Using *beta.div.comp()* and the percentage difference index (which is the quantitative form  
50  
51  
52 199 of the Sørensen coefficient), we partitioned total beta diversity into replacement and percentage  
53  
54 200 difference (Podani decomposition) components such that, in addition to a mean total beta  
55  
56 201 diversity value for the landscape and each individual ecoregion, we also knew, for each time  
57  
58  
59  
60

point, the proportions of the total (spatial) beta diversity that are explained by replacement and percentage 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 or partitions with similar values of the response (De'ath, 2002). 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 S3a 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  $25^{\circ}\text{C}$ ) and pH. Second, we built a tree using basin-level land-use variables (summarized by the NLA and based on the 1992 National Land Cover Database; <http://www.mrlc.gov/nlcd1992.php>). 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). We used functions *rpart()* and *prune()* from the *rpart* package (Therneau *et al.* 2015). We then performed these same URTs with transformed Shannon diversity ( $\exp(H)$ ) for the 2007 sediments as a response variable, to determine whether alpha diversity showed similar patterns in the URTs as those with LCBD indices. Pruning of the trees was done using the lowest cross-

validation relative error (CVRE) for the set of variables, meaning that a higher number of variables would be retained than if we had selected a tree with fewer splits whose cross-validation relative error value is within one standard error of the smallest CVRE value.

We then analysed temporal beta diversity between the two time points for each individual lake. We used function *TBI()*, with the percentage difference option, to compare the surveys of all lakes at the two time points. This function computes Temporal Beta diversity Indices (TBI) and tests them for significance through a permutation test (permutations = 99999), identifying the lakes where temporal change is exceptionally large between the two surveys (Legendre, pers. comm.). Multiple testing is adjusted for using Holm's procedure (see Legendre & Legendre (2012), p. 23). The TBI are somewhat similar to LCBD indices in spatial beta analysis, highlighting sites experiencing statistically significant community composition changes through time.

Analogous to analysing the relationship between spatial LCBD and land use variables, we examined the relationship between temporal beta diversity values (and their taxa gain and loss components) and contemporary land use to test the hypothesis that developed areas (see list of variables used in URT analyses) were associated with higher temporal beta diversity. We used logistic regression to test the relationship between significant temporal beta diversity and highly developed land cover. We also regressed the TBIs onto a composite land cover variable (Principal Component 1 scores of a Principal Components Analysis of the land cover variables, S3b), and created a regression tree (using the same approach as with spatial LCBD) with either TBI indices or taxa gain 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 ( $n = 176$ ) and the species-level analyses ( $n = 59$ ). For simplicity, tables and figures for the genus-level analyses are presented in the main body of the study, while all species-level tables and figures are shown in S4 and S5.

### Genus-level results

#### *Spatial diversity results*

We found that the different diversity metrics gave considerably distinctive information about landscape level patterns in diatom assemblages. For example, the range of rarefied genus richness differed in some ecoregions between modern and historical sediments. In the Upper Midwest, rarefied richness was about 50% as large for modern as for historical sediments (Table 2). In contrast, alpha diversity, and gamma diversity were approximately equivalent across all ecoregions (Table 2).

Total spatial beta diversities across the landscape (*beta.div()*) were similar at both time points (Table 2), with taxa replacement consistently the dominant mechanism of compositional change among lakes (Table 3). Across all sites, spatial beta diversity using the percentage difference metric was 0.36 (72% of the maximum percent difference index value of 0.5) for both historical, and modern sediments. At the ecoregion level, spatial beta diversity also varied little at the two time points (Table 2). The genus replacement component dominated in importance for all ecoregions, for both historical and 2007 sediments (Table 3). It is important to note that in this case, spatial beta diversity is comparable amongst ecoregions, even with different sample sizes, because spatial beta diversity was computed from a dissimilarity index that has an upper bound of 1 for the percentage difference index (Legendre & De Cáceres, 2013).

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

We observed considerable variation in LCBD values across the landscape, with some consistency between the two time points (Figure 3), and some of the 2007 variation could be explained by environmental variables. Overall, 22 sites (13% of total) contributed significant LCBD indices in historical assemblages, while 18 sites (10%) contributed significant LCBDs in the modern assemblages; eight of these sites overlapped in the two time surveys. 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. A McNemar test on a two by two 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 a null hypothesis of no effect of time period on significant LCBD values. A simple correlation between the historical and 2007 LCBD values showed a significant ( $P < 0.05$ ) correlation of 0.38. Lakes that were exceptional in historical times (significant LCBD only at the historical time point) were classified as intermediate or highly disturbed based on the U.S. EPA's classification system (i.e. a system that is based on thresholds for TP, TN, chloride,  $\text{SO}_4$ , turbidity and agricultural/developed land cover (A.T. Herlihy, pers. comm.)).

The univariate regression tree (URT) using water quality variables as predictors explained 54% of the variation in modern LCBD values, and only retained conductivity (Figure 4a). When a similar analysis was run on the 2007 alpha diversity (both Shannon diversity ( $H$ ) and  $\exp(H)$ ), only water temperature was retained as a positive predictor in the model, which had  $R^2_{\text{adj}}$  of 0.15 for  $H$  and 0.14 for  $\exp(H)$ . Using land cover measures instead of water quality variables in a URT of the 2007 LCBD values resulted in a slightly weaker model ( $R^2_{\text{adj}} = 0.3$ ;

Figure 4b). With land cover variables, alpha diversity was found to be greater in sites with higher proportions of wetland and forest cover in the watersheds ( $R^2_{\text{adj}} = 0.30$ ).

#### *Temporal diversity results*

Mean temporal beta diversity computed at the genus level ranged from 0.4 to 0.6 across the ecoregions (Table 4). The importance of genera loss and gain through time was approximately the same across ecoregions. Twenty-one of the 176 lakes showed significant temporal beta diversity (Figure 3c; however, after correcting for multiple testing, we failed to identify any significant sites). Only two of the 21 lakes identified as sites of significant TBI were considered “Reference” (i.e. pristine or undisturbed) by US EPA prior to sampling; the others were assumed to have been disturbed and were selected using probabilistic methods. Based on the post-sampling classification of lakes into disturbance categories by the US EPA, most (15 of 21) of the lakes with significant temporal beta diversity were classified as having intermediate disturbance. Significant relationships were found between temporal beta diversity (used as a binary variable; significant or not) and the contribution of the genera loss and genera gain (Figure 5) components ( $P < 0.05$ ) using logistic regression.

The logistic relationship between TBI significance and percent agriculture (2007) was significant ( $P = 0.04$ ), meaning that lakes in regions where agriculture developed in the 20<sup>th</sup> century were more likely to show large differences in diatom community composition between historical times and 2007. The URT constructed with total temporal beta diversity as a response variable and % land cover variables as input retained % agriculture, % developed, % grassland, % forest, % shrubland, and % wetland. After pruning the tree, % forest was the only significant variable retained to explain total temporal beta diversity ( $R^2_{\text{adj}} = 0.36$ ; Figure 6). The URT



1  
2  
3 315 constructed with the genus gain component of temporal beta diversity retained % forest with a  $R^2$   
4  
5  
6 316  $_{adj} = 0.40$ , % shrubland and % wetland had been pruned out.  
7  
8 317

9  
10 318 Species-level results  
11

12  
13 319 We were interested in whether species-level data would allow for additional conclusions  
14  
15 320 to be drawn regarding beta diversity. Working with a smaller set of lakes located in both the  
16  
17 321 Northern Appalachians ( $n = 52$ ) and Coastal Plains ( $n = 7$ ) ecoregions, we found landscape  
18  
19 322 patterns similar to our genus-level results, where alpha and spatial beta diversities were  
20  
21 323 approximately equivalent between time points. Because we were working at the species-level,  
22  
23 324 Species' Contribution to Beta Diversity (SCBD) could also be examined in relation to spatial  
24  
25 325 beta diversity. The ten species with the highest contributions to SCBD in historical and 2007  
26  
27 326 assemblages were mostly planktonic and five of them were key contributors to SCBD in both  
28  
29 327 data sets (see S5). In terms of LCBD, sites with significant LCBD were clustered in the Coastal  
30  
31 328 Plains ecoregion in historical times, but more evenly distributed between the two ecoregions in  
32  
33 329 modern times (S4, Figure 1). Univariate of modern LCBD showed longitude, temperature and  
34  
35 330 pH to be key determinants of high LCBD in these lakes ( $Adj. R^2 = 0.6$ ).  
36  
37  
38  
39  
40

41 331 Similar to the genus level analyses, temporal turnover of diatom species in this reduced  
42  
43 332 set of lakes was between 50-60%. After correcting for multiple testing, no lakes had significant  
44  
45 333 TBI, though this is likely a result of the reduced sample size of this dataset.  
46  
47  
48  
49

50 334  
51 335 **Discussion**  
52

53 336 Our analyses identified patterns and drivers of diatom diversity across the conterminous  
54  
55 337 United States over the last ~150 years and provided support for hypothesized Anthropocene  
56  
57  
58  
59  
60



trends previously identified (McGill *et al.* 2015), as well as some unexpected ones (Table 5). Because of the larger sample size and geographic coverage, our genus-level results are the most robust for drawing conclusions and are discussed at length here. Across all sites, we found that genus richness increased in the modern sediments over historical ones at the continental scale; however this was mostly driven by the marked changes observed in the Upper Midwest. This pattern may be the result of introductions of new genera and dispersal aided by hydrological modification (Alig *et al.* 2004); this is consistent with McGill *et al.* (2015). For spatial beta diversity, we found that historical and modern beta diversity measured at the ecoregion level were similar, although at the site level, there were substantial changes in diversity through time. Sites that contributed significantly to spatial beta diversity were not always those that experienced significant temporal changes in diatom composition. This is not surprising, given that in the spatial beta diversity analyses, the lakes with significant LCBD indices are those that are most exceptional at one time. In the temporal analysis, the lakes with the largest TBIs are those that have changed the most in diatom composition between time points. In this study, sites identified as ‘reference’ lakes prior to sampling were less likely to have experienced significant temporal changes in beta diversity. However, this pattern was less clear when examining the post-sampling U.S. EPA lake classification, for which significant temporal beta diversity was found in all three disturbance categories (least disturbed, intermediate disturbance and highly disturbed).

In terms of the mechanism underlying the changes in spatial and temporal beta diversity, we consistently found evidence for replacement-dominated spatial beta diversity within each time point. Additionally, while there does not appear to be a clear regional pattern in site

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

contributions to modern spatial beta diversity, site-specific variables associated with water quality and land cover were significant predictors of the spatial variation in LCBD.

*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, taxa 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) analyzed 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 ‘crisis’ of 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). 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). Additionally, we saw diatom genera richness increase between historical and modern sediments, which could be related to climatic shifts; this is not unlike studies that demonstrated an increase in species richness with temperature (e.g. Stomp *et al.* 2011).

### 383 *Sites versus ecoregion-level responses*

384 Individual lakes and their contributions to either spatial or temporal beta diversity varied  
385 from site to site, even when variation across the ecoregions did not occur (based either on  
386 historical or 2007 sediments). For example, we found that lakes with very low specific  
387 conductivity (less than  $18 \mu\text{S cm}^{-1}$ ) had the highest LCBD values (spatial), indicating that these  
388 sites are particularly unique across the landscape. In our dataset, conductivity was positively  
389 correlated ( $\text{corr} = 0.6$ ) with silica concentration, such that sites with low silica had the most  
390 distinctive diatom assemblages. The low conductivity and low silica results concord with an  
391 extensive body of literature documenting the responsiveness of diatom assemblages to water  
392 chemistry (Smol & Stoermer, 2010), especially ionic strength (Fritz *et al.* 2010). Spatial LCBD  
393 values were greater in basins with lower human (i.e. urban) development, indicating that less  
394 disturbed lakes may contribute more diatom heterogeneity to the landscape. Indeed, nutrient  
395 enriched lakes (a by-product of most development) can have a homogenizing effect on aquatic  
396 assemblages (Olden *et al.* 2004; Donahue *et al.* 2009). We identified a very low threshold for the  
397 proportion of development in the watershed (Figure 4) and suggest that this is due to the  
398 tendency for humans to modify shorelines even when the rest of the watershed may be intact  
399 (e.g., by removing riparian and littoral vegetation; Kaufmann *et al.* 2014), which in turn have  
400 substantial consequences for diatoms (Velghe *et al.* 2012) and other organisms (Strayer &  
401 Findlay, 2010). The relationship with forest cover indicates that less forest cover may produce  
402 conditions that favour specific species or genera and hence representing more of the total beta  
403 diversity.

404 Land use variables were also informative in describing the observed temporal variation in  
405 diatom assemblage, which was higher in sites with lower forest cover in their basin. While the

1  
2  
3 406 contemporary forest cover variable does not inform as to whether a change in cover occurred  
4  
5 407 since historical times, it does indicate that lakes with higher basin forest cover currently have  
6  
7  
8 408 been buffered to diatom assemblage change through time. Overall, diatom communities were  
9  
10 409 more likely to remain unaltered through time in sites with higher forest cover than where forest  
11  
12 410 cover was less. This result is intuitive, and is reflected in the relationships between land cover  
13  
14 411 variables and the genus gain component of temporal beta diversity; higher genus gain was  
15  
16 412 associated with lower basin percentages of forest, shrubland, and wetland. Essentially, there was  
17  
18 413 higher temporal beta diversity (more genus-level change) with reduced forest cover, driven by  
19  
20 414 more gains at the genera-level in low forest, shrubland and wetland lakes.  
21  
22  
23

24 415 *Dominance of the replacement component in spatial beta diversity*

26  
27 416 Replacement was always the dominant component of spatial beta diversity, regardless of  
28  
29 417 the time point. While a separate analysis, our temporal beta diversity results shed some light onto  
30  
31 418 why replacement is such a dominant component of spatial beta diversity, both historically and in  
32  
33 419 present times. The method used to compute the significance of temporal change for each lake  
34  
35 420 (through permutation) was designed to detect significant changes through time when beta  
36  
37 421 diversity is *not* uniform across space, i.e. if all lakes experienced the same amount of change in  
38  
39 422 community composition through time, one could not identify significant temporal changes. Thus,  
40  
41 423 the temporal changes detected in our study resulted specifically from disproportionate temporal  
42  
43 424 change across the landscape, which could also represent genus replacement in spatial beta  
44  
45 425 diversity. Genus replacement is also more prevalent in regions with variable environments from  
46  
47 426 lake to lake, due to factors like geographic isolation or poor dispersal ability (Leprieur *et al.*  
48  
49 427 2011). Broad-scale climatic changes (climate warming) could create a uniform temporal beta  
50  
51 428 diversity pattern across the U.S. landscape. However, we noted that not all sites had significant  
52  
53  
54  
55  
56  
57  
58  
59  
60

temporal beta diversity, indicating that temporal changes differed from lake to lake, presumably due to local or regional factors. It is also important to note that 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 expect that spatial beta diversity is explained by replacement via the mechanism described by Leprieur *et al.* (2011). Additionally, both alpha and gamma diversities were equivalent across ecoregions, further suggesting replacement as the dominant mechanism of spatial beta diversity.

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 (Figure 5a,b), though this is primarily mathematical evidence. This could mean that both the gain and loss of abundance of specific taxa are important to temporal change, and supported by the mechanisms identified for spatial beta diversity. An alternative explanation is that abundance gain and loss on a genus-by-genus basis was important at different intervals through time. This is shown in Legendre and Salvat (2015), where temporal beta diversity of mollusc communities was partitioned into genus gain and loss for five different time periods (each time period consisting of 5-20 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 genus gain and loss cannot be observed. Environmental variables and mechanisms behind temporal beta diversity may change from time period to time period as well as across temporal scales.

#### *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). The entire landscape, as well as the majority of

1  
2  
3 452 ecoregions, show higher historical beta diversity, with the exception of the Coastal Plains and the  
4  
5 453 Western Mountains, which both had lower modern spatial beta diversity. We can derive potential  
6  
7 454 mechanisms driving beta diversity by considering both spatial and temporal patterns in the  
8  
9 455 ecoregions. For example, in the Northern Appalachians ecoregion, the pattern of heterogeneous  
10  
11 456 diatom communities across the landscape, both historically and currently, could have arisen  
12  
13 457 because taxa losses have matched taxa gains across sites within the ecoregion metacommunity.  
14  
15 458 On the other hand, diatom assemblages across the modern landscape may be more heterogeneous  
16  
17 459 than historically, as they are in the Temperate Plains ecoregion, because dispersal of taxa through  
18  
19 460 time is sufficiently high to allow for genera or species (environmental) sorting across the  
20  
21 461 landscape. In the Upper Midwest, we can envision yet another scenario: diatoms may be more  
22  
23 462 homogeneously distributed across space currently because population dynamics have resulted in  
24  
25 463 genera being added to sites in a non-genera sorting (*sensu* mass effects; Leibold *et al.* 2004) way.  
26  
27 464 Diatoms may also be more homogeneous across space currently because the taxa removed over  
28  
29 465 time are always the same, regardless of site as could be the case in the Western Mountains  
30  
31 466 ecoregion. These potential scenarios can now be investigated further by ecoregion through more  
32  
33 467 detailed study within each, using metacommunity theory for guidance.  
34  
35  
36  
37  
38  
39  
40

41 468 *Conclusions*

42  
43 469 Patterns in both spatial and temporal beta diversity varied both across sites and ecoregions,  
44  
45 470 via several possible mechanisms, of which we have highlighted some of the most important. By  
46  
47 471 considering temporal and spatial beta diversity in the same study, we have been able to provide a  
48  
49 472 holistic view of diatom biodiversity patterns and test hypotheses and predictions for change  
50  
51 473 during the Anthropocene (McGill *et al.* 2015). While the majority of our work was at the genus-  
52  
53 474 level, we have laid out a useful and robust framework for future beta diversity study. Future  
54  
55  
56  
57  
58  
59  
60

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 change at multiple scales.

### Acknowledgements

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

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

**Supplementary information titles**

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

**S2:** Description of the Temporal Beta diversity Indices (TBI) performed using TBI().

**S3:** Principal Component Analyses (PCAs) of (a) standardized 2007 water quality variables and (b) land cover variables.

**S4:** Tables and figures for species-level analyses (n = 59).

**S5:** 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 = 59).

**Biosketch**

**Amanda Winegardner** is a PhD candidate 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 Highly Cited Researcher (2014, 2015) 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.



521 **References**

- 522 Alig, R.J., Kline, J.D., & Lichtenstein, M. (2004). Urbanization on the US landscape: looking  
523 ahead in the 21st century. *Landscape and Urban Planning*, 69, 219-234.
- 524
- 525 Beaulieu, M., Pick, F., & Gregory-Eaves, I. (2013). Nutrients and water temperature are  
526 significant predictors of cyanobacterial biomass in an 1147 lakes data set. *Limnology and*  
527 *Oceanography*, 58, 1736-1746.
- 528
- 529 Cardinale, B.J., Duffy, J.E., Gonzalez, A., Hooper, D.U., Perrings, C., Venail, P., Narwani, A.,  
530 Mace, G.M., Tilman, D., Wardle, D.A., Kinzig, A.P., Daily, G.C., Loreau, M., Grace, J.B.,  
531 Larigauderie, A., Srivastava, D.S., and Naeem, S. (2012). Biodiversity loss and its impact on  
532 humanity. *Nature* 486: 59-67.
- 533
- 534 Charles, D.F., Knowles, C., & Davis, R.S. (2003). Protocols for the analysis of algal samples as  
535 collected as part of the U.S. Geological Survey National Water-Quality Assessment program.  
536 Patrick Center for Environmental Research, The Academy of Natural Sciences. Report No. 02-  
537 06.
- 538
- 539 Crutzen, P.J. (2002). Geology of mankind. *Nature*, 415, 23.
- 540
- 541 De'ath, G. (2002). Multivariate regression trees: A new technique for modeling species-  
542 environment relationships. *Ecology*, 83, 1105-1117.
- 543
- 544 Donahue, I., Jackson, A.L., Pusch, M.T., & Irvine, K. (2009). Nutrient enrichment homogenizes  
545 lake benthic assemblages at local and regional scales. *Ecology*, 90, 3470-3477.
- 546
- 547 Dornelas, M., Gotelli, N.J., McGill, B., Shimadzu, H., Moyes, F., Sievers, C., and Magurran,  
548 A.E. (2014). Assemblage time series reveal biodiversity change but not systematic loss. *Science*  
549 344: 296-299.
- 550
- 551 Ellison, A.M. (2010). Partitioning diversity. *Ecoology* 91: 1962-1963.
- 552
- 553 Fritz, S.C., Cumming, B.F., Gasse, F., Laird, K.R. (2010). Diatoms as indicators of hydrologic  
554 and climatic change in saline lakes in "The Diatoms: Applications for the Environmental and  
555 Earth Sciences 2<sup>nd</sup> edition". Cambridge University Press, Cambridge UK.
- 556
- 557 Glew, J. (1989). A new trigger mechanism for sediment samples. *Journal of Paleolimnology*, 2,  
558 241-243.
- 559
- 560 Hamilton, C. (2015). Getting the Anthropocene so wrong. *The Anthropocene Review*, 2, 102-  
561 107.
- 562
- 563 Herlihy, A.T., Paulsen, S.G., Van Sickle, J., Stoddard, J.L., Hawkins, C.P., & Yuan, L.L. (2008).  
564 Striving for consistency in a national assessment: the challenges of applying a reference-

condition approach at a continental scale. Journal of the North American Benthological Society, 27, 860-877.

Herlihy, A.T. (2015). Personal Communication.

Kaufmann, P.R., Peck, D.V., Paulsen, S.G., Seeliger, C.W., Hughes, R.M., Whittier, T.R., & Kamman, N.C. (2014). Lakeshore and littoral physical habitat structure in a national lakes assessment. Lake and Reservoir Management, 30, 192-215.

Kennedy, C. (2014). Does “global warming” mean it’s warming everywhere? National Oceanic and Atmospheric Administration. Available online: <https://www.climate.gov/news-features/climate-qa/does-global-warming-mean-it%E2%80%99s-warming-everywhere>. Accessed 31 August 2015.

Legendre, P. (2013a). R function: beta.div(). Available online: <http://adn.biol.umontreal.ca/~numericaledocology/Rcode/>.

Legendre, P. (2013b). R function: LCBBD.comp(). Available online: <http://adn.biol.umontreal.ca/~numericaledocology/Rcode/>.

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

Legendre, P. (2015). R function: TBI(). Available online: <http://adn.biol.umontreal.ca/~numericaledocology/FonctionsR/>.

Legendre, P. (2015). Personal Communication.

Legendre, P., & De Cáceres, M. (2013). Beta diversity as the variance of community data: dissimilarity coefficients and partitioning. Ecology Letters 16: 951-963.

Legendre, P., & Legendre, L. (2012). Numerical Ecology, 3<sup>rd</sup> ed. Elsevier Science BV, Amsterdam.

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.

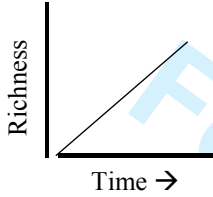
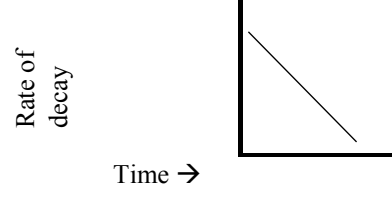
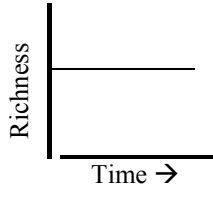
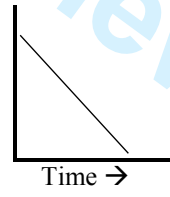
Leibold, M.A., Holyoak, M., Mouquet, N., Amarasekare, P., Chase, J.M., Hoopes, M.F., Holt, R.D., Shurin, J.B., Law, R., Tilman, D., Loreau, M., & Gonzalez, A. (2004). The metacommunity concept: a framework for multi-scale community ecology. Ecology Letters 7: 601-613.

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.
- McGill, B.J. (2015). Land use matters. *Nature* 520: 38-39.
- 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, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., & Wagner, H. (2015). Package 'vegan'. Available: <http://cran.r-project.org/web/packages/vegan/vegan.pdf>.
- 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.
- 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/>
- Smol, J.P., & Stoermer, E.F. (eds) (2010). *The Diatoms: Applications for the Environmental and Earth Sciences* 2<sup>nd</sup> edition. Cambridge University Press, Cambridge UK.
- Steffen, W., Grinevald, J., Crutzen, P., & McNeill, J. (2011). The Anthropocene: conceptual and historical perspectives. *Philosophical Transactions of the Royal Society A*, 369, 842-867.
- Stomp, M., Huisman, J., Mittelbach, G.G., Litchman, E., & Klausmeier, C.A. (2011). Large-scale biodiversity patterns in freshwater phytoplankton. *Ecology*, 92, 2096-2107.
- Strayer, D.L., & Dudgeon, D. (2010). Freshwater biodiversity conservation: recent progress and future challenges. *Journal of the North American Benthological Society* 29, 344-358.
- Strayer, D.L., & Findlay, S.E.G. (2010). Ecology of freshwater shore zones. *Aquatic Sciences*, 72, 127-163.
- Taranu, Z.E., & Gregory-Eaves, I. (2008). Quantifying relationships among phosphorus, agriculture, and lake depth at an inter-regional scale. *Ecosystems* 11: 715-725.
- Therneau, T., Atkinson, B., & Ripley, B. (2015). Package 'rpart'. Available: <http://cran.r-project.org/web/packages/rpart/rpart.pdf>.

1  
2  
3 656 U.S. EPA (2009). National Lakes Assessment: A Collaborative Survey of the Nation's Lakes.  
4 657 EPA 841-R-09-001. U.S. Environmental Protection Agency, Washington, DC, USA.  
5 658  
6  
7 659 U.S. EPA (2011-2012). National Lakes Assessment. Field Operations Manual. EPA 841-B-11-  
8 660 003. U.S. Environmental Protection Agency, Washington, DC, USA.  
9 661  
10 662 USGS (1992). The USGS Land Cover Institute: NLCD 92 Land Cover Class Definitions.  
11 663 Available: <http://landcover.usgs.gov/classes.php>.  
12 664  
13  
14 665 Velghe K., Vermaire J.C., & Gregory-Eaves I. 2012. Declines in littoral species richness across  
15 666 both spatial and temporal nutrient gradients: A palaeolimnological study of two taxonomic  
16 667 groups. *Freshwater Biology*, 57, 2378-2389.  
17 668  
18  
19 669 Vellend, M., Baeten, L., Myers-Smith, I.H., Elmendorf, S.C., Beauséjour, R., Brown, C.D., De  
20 670 Frenne, P., Verheyen, K., & Wipf, S. (2013). Global meta-analysis reveals no net change in  
21 671 local-scale plant biodiversity over time. *Proceedings of the National Academy of Sciences* 110:  
22 672 19456-19459.  
23 673  
24  
25 674 Wardle, D.A., Bardgett, R.D., Callaway, R.M., & Van der Putten, W.H. (2011). Terrestrial  
26 675 ecosystem responses to species gains and losses. *Science* 332: 1273-1277.  
27 676  
28 677 Whittaker, R.H. (1972). Evolution and measurement of species diversity. *Taxon*, 21, 213-251.  
29 678  
30  
31 679 Winegardner, A.K., Beisner, B.E., Legendre, P., & Gregory-Eaves, I. (2015). Are the landscape-  
32 680 level drivers of water column and surface sediment diatoms different? *Freshwater Biology* 60:  
33 681 267-281.  
34 682  
35  
36 683 Zalasiewicz, J., Waters, C.N., Barnosky, A.D., Cearreta, A., Edgeworth, M., Ellis, E.C.,  
37 684 Galuszka, A., Gibbard, P.L., Grinevald, J., Hajdas, I., do Sul, J.I., Jeandel, C., Leinfelder, R.,  
38 685 McNeill, J.R., Poirier, C., Revkin, A., Richter, D.B., Steffen, W., Summerhayes, C., Syvitski,  
39 686 J.P.M., Vidas, D., Waple, M., Williams, M., & Wolfe, A.P. (2015). Colonization of the  
40 687 Americas, 'Little Ice Age' climate, and bomb-produced carbon: Their role in defining the  
41 688 Anthropocene. *The Anthropocene Review*, 2, 117-127.  
42 689  
43  
44 690  
45 691

692 **Tables**693 **Table 1: Hypotheses and predictions for alpha and beta diversity analyses based on McGill *et al.* (2015, Figure 2) framework.**

Spatial scale	$\alpha$ - or $\gamma$ -diversity	Temporal $\beta$ -diversity	Spatial $\beta$ -diversity
Continental (Biogeographic)	 <p>Richness</p> <p>Time →</p> <p>Higher gamma diversity in modern times</p>	No prediction	 <p>Rate of decay</p> <p>Time →</p> <p>Less spatial beta diversity in modern times (homogenization)</p>
Ecoregion (Metacommunity)	<p>Higher alpha diversity in modern times</p>	No prediction	<p>Less spatial beta diversity in modern times (homogenization- as above)</p>
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 farther apart in time</p>	

**Table 2: Spatial beta ( $\beta$ ) diversity, mean rarefied genus richness, mean alpha ( $\alpha$ ) diversity and gamma ( $\gamma$ ) diversity for each ecoregion (Shannon diversity of genus' sums). Beta diversity was calculated using total variance computed using *beta.div()* based on percentage difference matrices. "Hx" refers to the historical sediments. Genus richness was rarefied after rare genera (<2% relative abundance) had been excluded. The Xeric ecoregion was excluded in these mean values because there was only one site in that ecoregion.**

Ecoregion	$\beta$ -diversity		Rarefied G		$\alpha$ -diversity (Shannon)		$\alpha$ -diversity (Simpson)		$\gamma$ -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
Coastal Plains	0.37	0.32	19.2	18.4	1.7	2.0	0.8	0.8	2.5	2.7
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

**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
Coastal Plains	0.36	0.01	0.97	0.03
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
Coastal Plains	0.32	<0.01	0.10	<0.01
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

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 4: 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 ‘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. Genera gain refers to the component representing gain of abundances on a genus by genus 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	Loss of genera	Gain of genera	Total beta
Coastal Plains	0.28 (0.08)	0.27 (0.08)	0.55 (0.16)
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)



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

Spatial scale	$\alpha$ - or $\gamma$ -diversity	Temporal $\beta$ -diversity	Spatial $\beta$ -diversity
Continental (Biogeographic)	Rarefied genus richness higher in modern times  <i>Matches trend</i>	No prediction	Spatial beta diversity approximately equivalent between historical and modern times  <i>Does not match trend</i>
Ecoregion (Metacommunity)	Rarefied genus richness higher in modern times in some ecoregions  <i>Partial match</i>	No prediction	Spatial beta diversity less in modern times than historical times in some ecoregions  <i>Partial match</i>
Lake (Local)	Rarefied genus richness higher in modern times for some lakes  <i>Partial match</i>	12% of lakes showed significant temporal beta diversity  <i>Direct comparison not possible with only two time points</i>	13% (historical) and 10% (2007) of lakes made significant contributions to spatial beta diversity (did not test in lake spatial beta)

742

743

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

744 **Table 6: Summary of temporal and spatial beta diversity across ecoregions.** The  
745 mechanisms column for temporal beta diversity shows when abundance loss (genus by genus) or  
746 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
Coastal Plains	=	Lower in contemporary
Northern Appalachians	Abundance loss	~ = between contemporary and historical
Southern Plains	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

747

## 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 seven distinct ecoregions: Coastal Plains (CPL,  $n = 7$ ), 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 ( $\text{km}^2$ ), (c) observed maximum depth (m), (d) pH and (e) TP ( $\mu\text{g L}^{-1}$ ), where the centre horizontal line is the median, the lower horizontal line the 25<sup>th</sup> percentile, the upper horizontal line the 75<sup>th</sup> percentile and points represent outliers.

**Figure 3: LCBD values for (a) historical spatial beta diversity and (b) 2007 spatial beta diversity and exceptional sites (c) for temporal beta diversity (TBI) (all genus-level).** A lake has a significant LCBD value if  $P < 0.05$ , and is coded as “True” (open circles). LCBD values across all sites (in either historical or 2007) sum to 1. A lake has significant total temporal beta

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 if  $P < 0.05$  before correction for multiple testing and is also coded as “True” (open circles) in (c). Note that no sites had significant TBI after correction for multiple testing.

**Figure 4: Univariate regression tree of 2007 spatial LCBD ( $n = 176$ ; genus-level) explained by (a) 2007 the retained water quality variable and (b) land cover variables.** The units for the water quality variable id: Conductivity:  $\mu\text{S cm}^{-1}$ . The cumulative  $R^2_{\text{adj}}$  for the model is 0.5. The land cover variables are percentage of each lake basin. The cumulative  $R^2_{\text{adj}}$  for the model is 0.3.

**Figure 5: Temporal beta diversity explained by genus gain component ( $n = 176$ ; genus-level).** Panel (a) shows the relationship between genera gain and whether a particular lake site experienced significant ( $P < 0.05$ ) beta diversity (TBI) between historical and 2007 conditions using logistic regression, where “1” means that the temporal beta diversity was significant (not adjusted for multiple testing). Panel (b) shows the eco-regional relationships between mean temporal beta diversity and the proportion of temporal beta diversity explained by variation in genera gain.

**Figure 6: Univariate regression tree of total temporal beta diversity explained by percent forest in a basin ( $n = 176$ ; genus-level).** The  $R^2_{\text{adj}} = 0.36$ .

## Figures

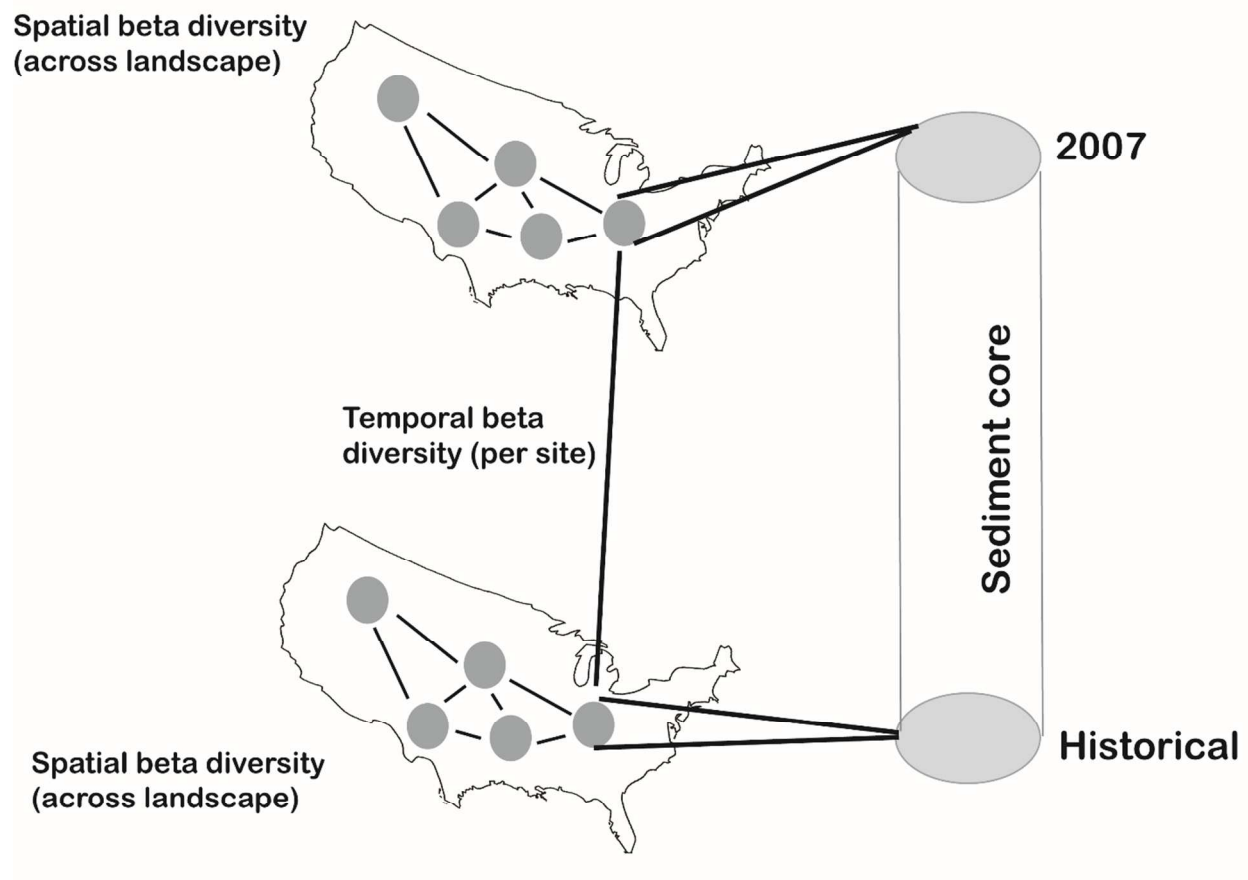


Figure 1

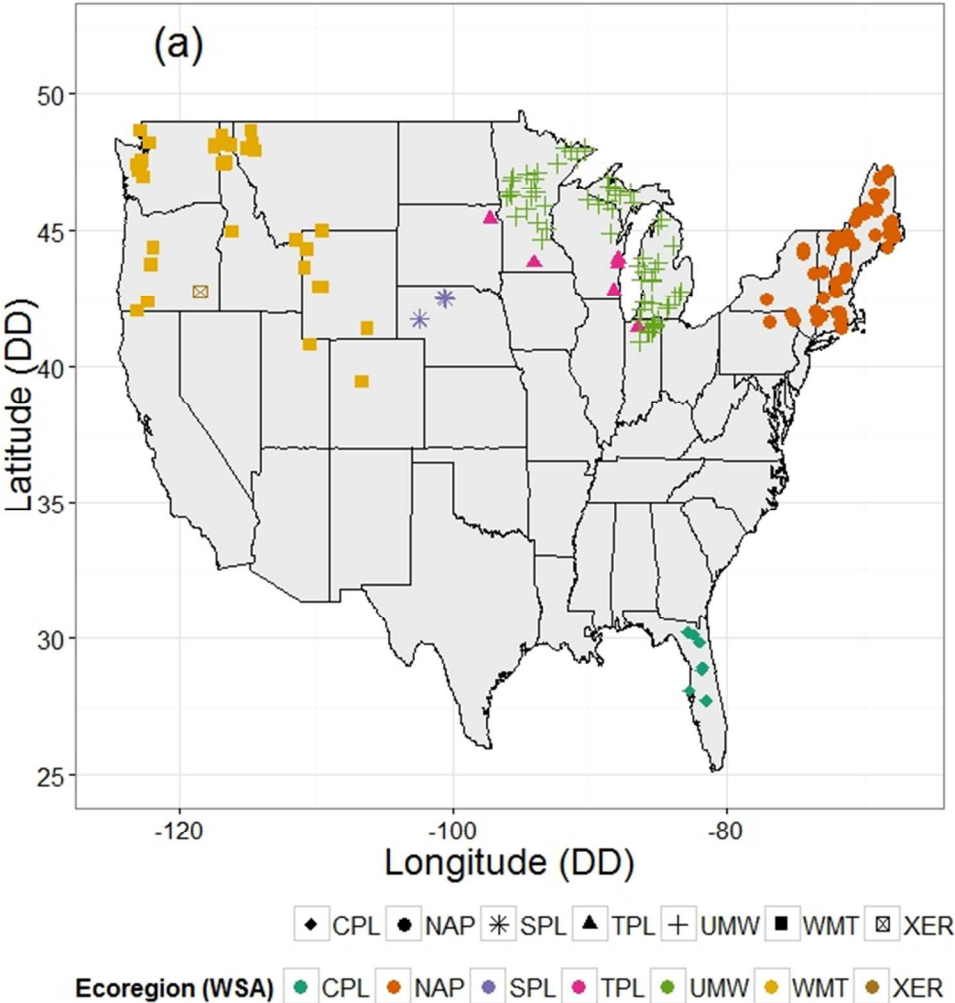


Figure 2(a)

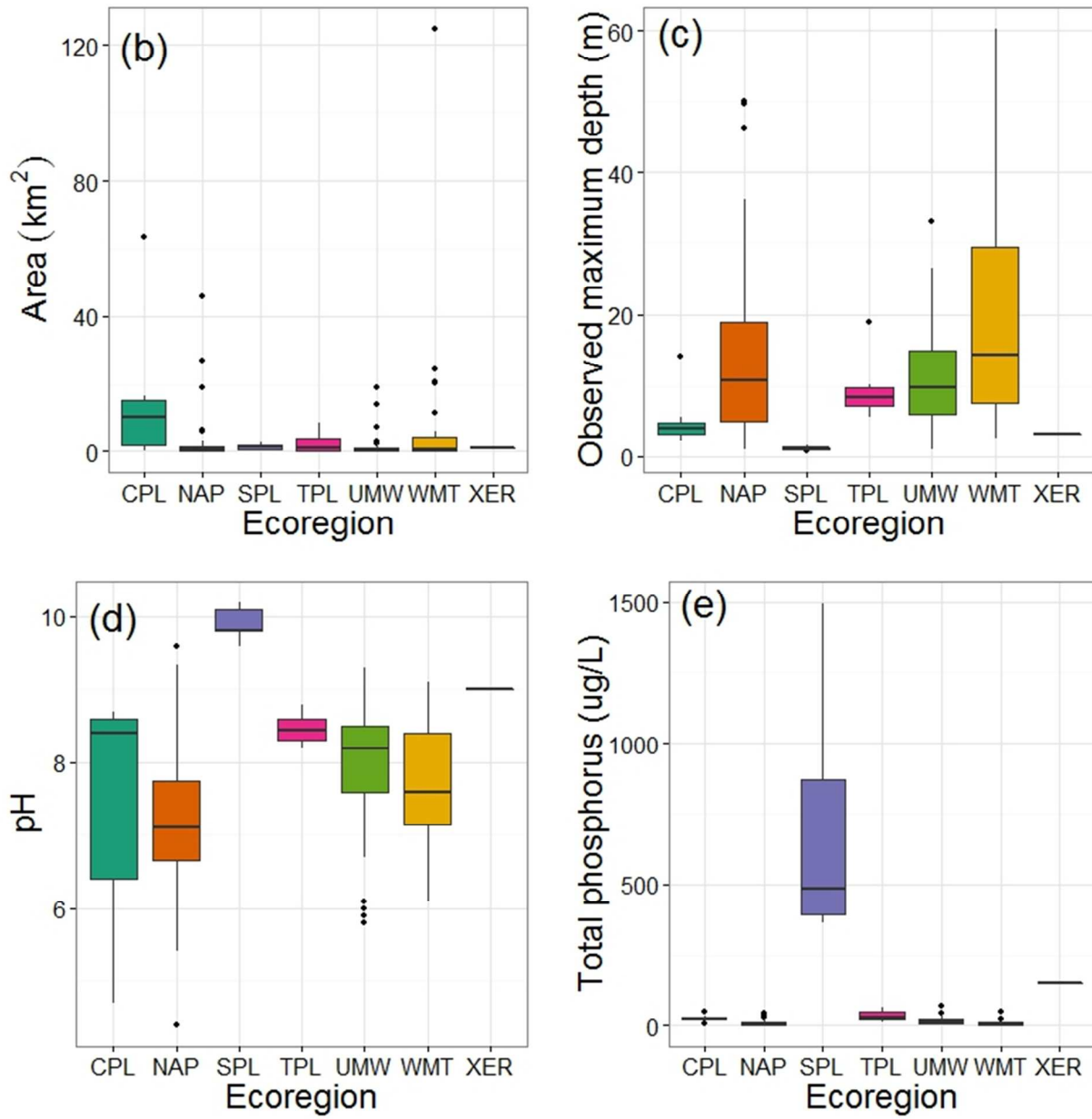


Figure 2(b)

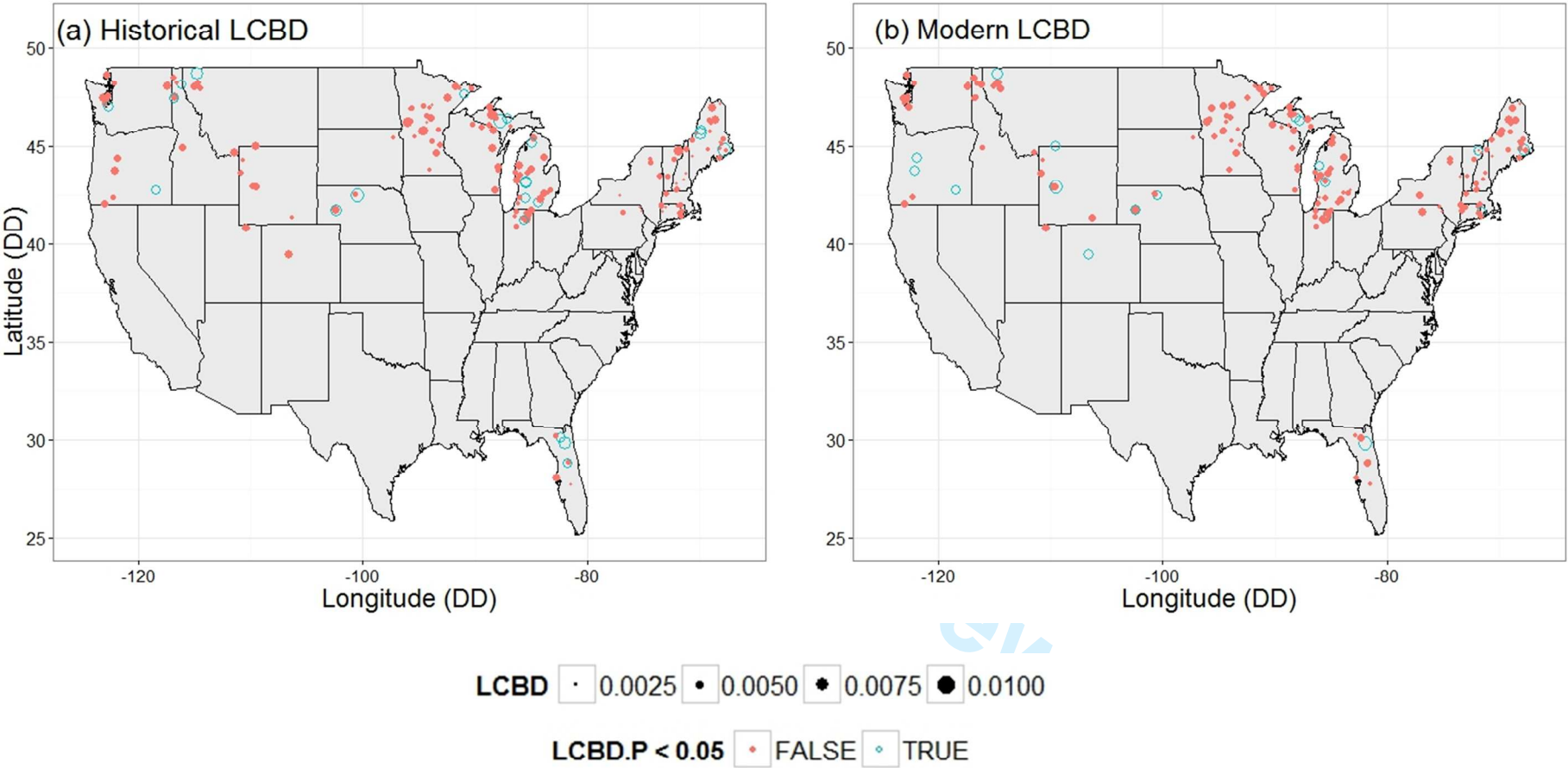


Figure 3(ab)



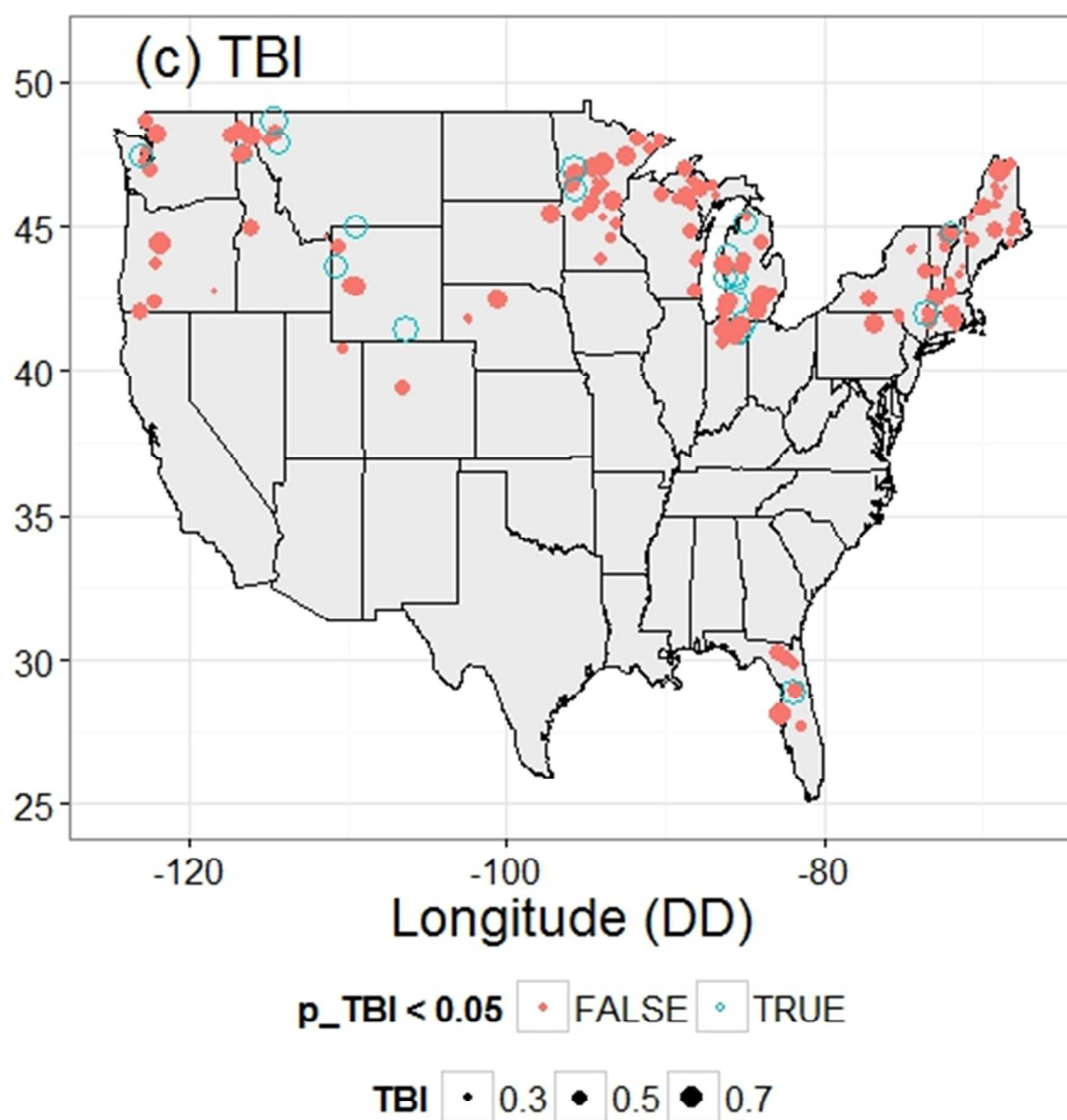


Figure 3(c)

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

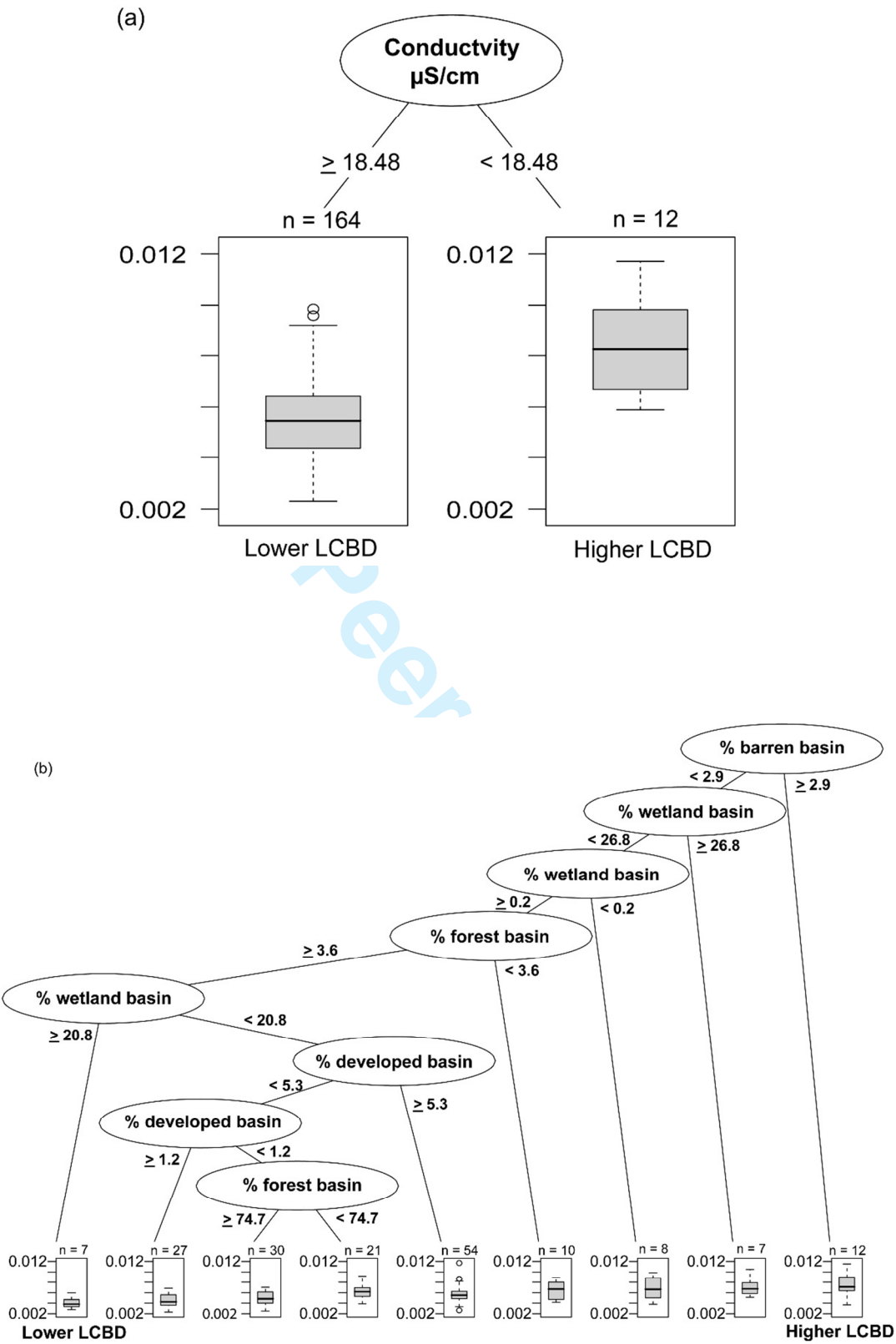


Figure 4

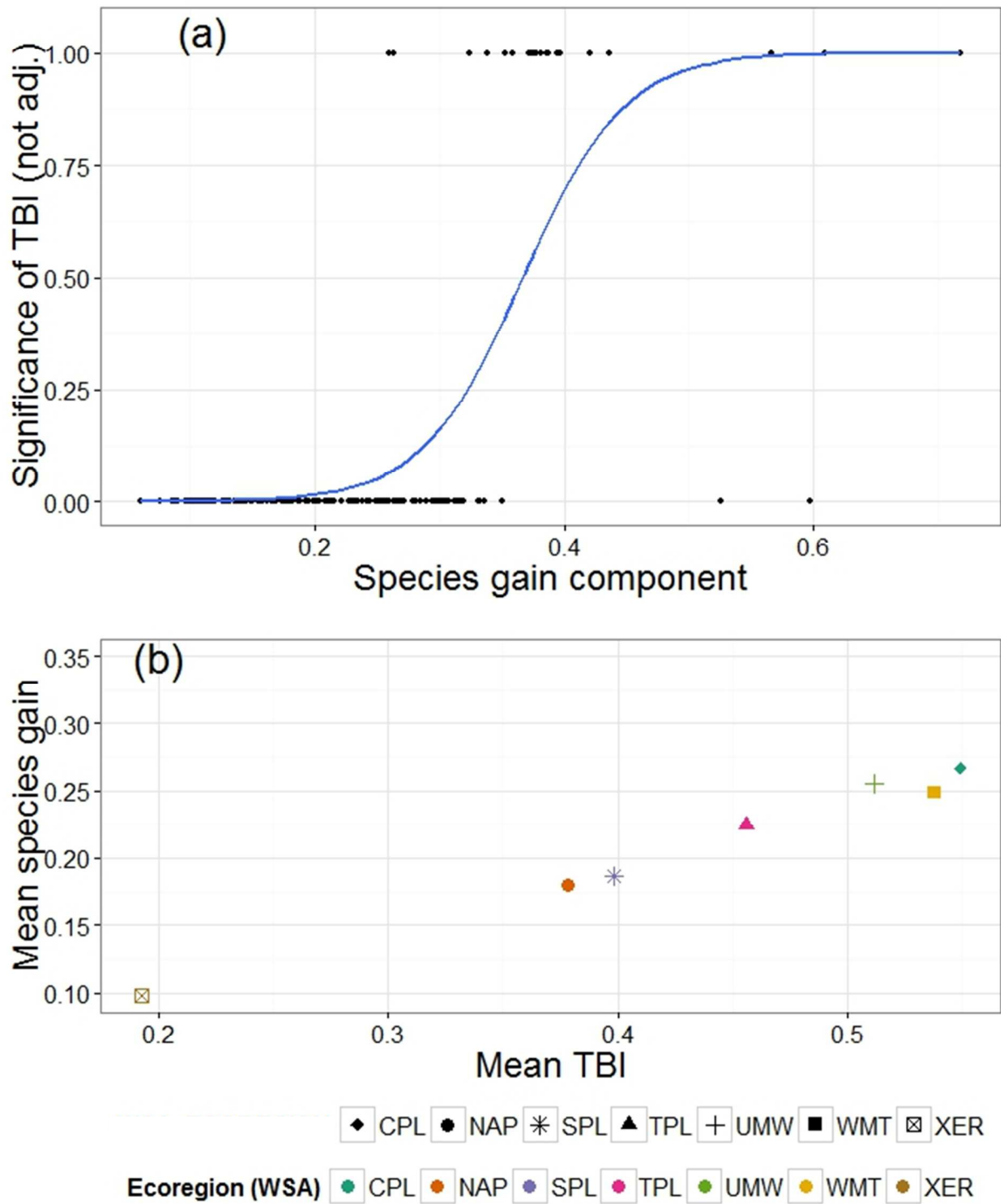


Figure 5

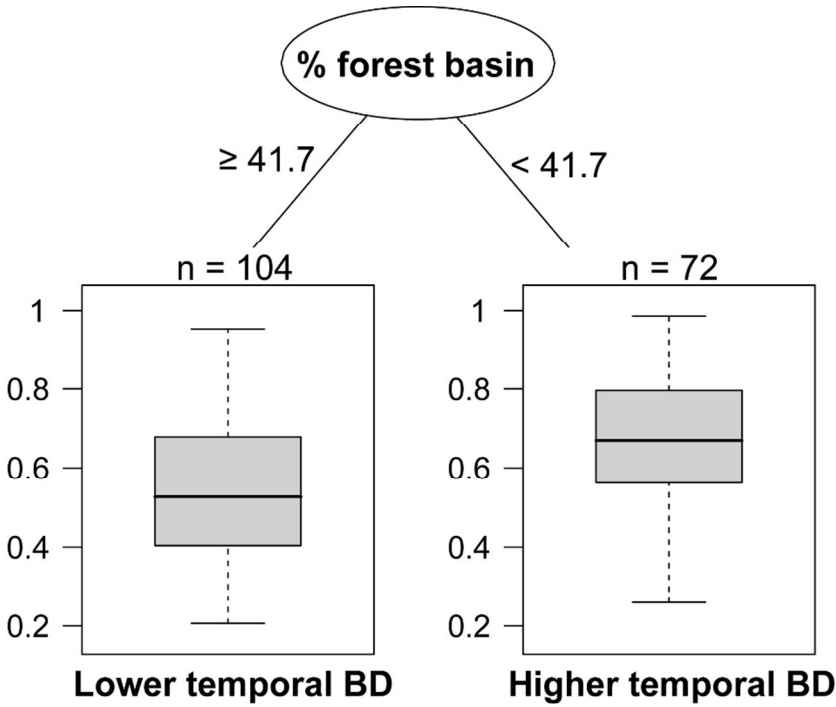


Figure 6

## Supporting Information

### **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  $^{214}\text{Bi}$  and  $^{210}\text{Pb}$ , where activity of  $^{210}\text{Pb}$  within two standard errors of  $^{214}\text{Bi}$  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  $^{210}\text{Pb}$  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.

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.

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

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.

For Peer Review



## S2: Description of the Temporal Beta diversity Indices (TBI) performed using *TBI()*.

The *TBI()* (Legendre, 2015) function outputs four components:

A: Species (or genera that both time points have in common

B: Abundance loss component, i.e. loss of abundances on a species-by-species (or genera-by-genera) basis from the first time point compared to the second time point.

C: Abundance gain component, i.e. gain of abundances on a species-by-species (or genera-by-genera) basis in the second time point compared to the first time point.

D: Total temporal beta diversity ( $B + C$ )

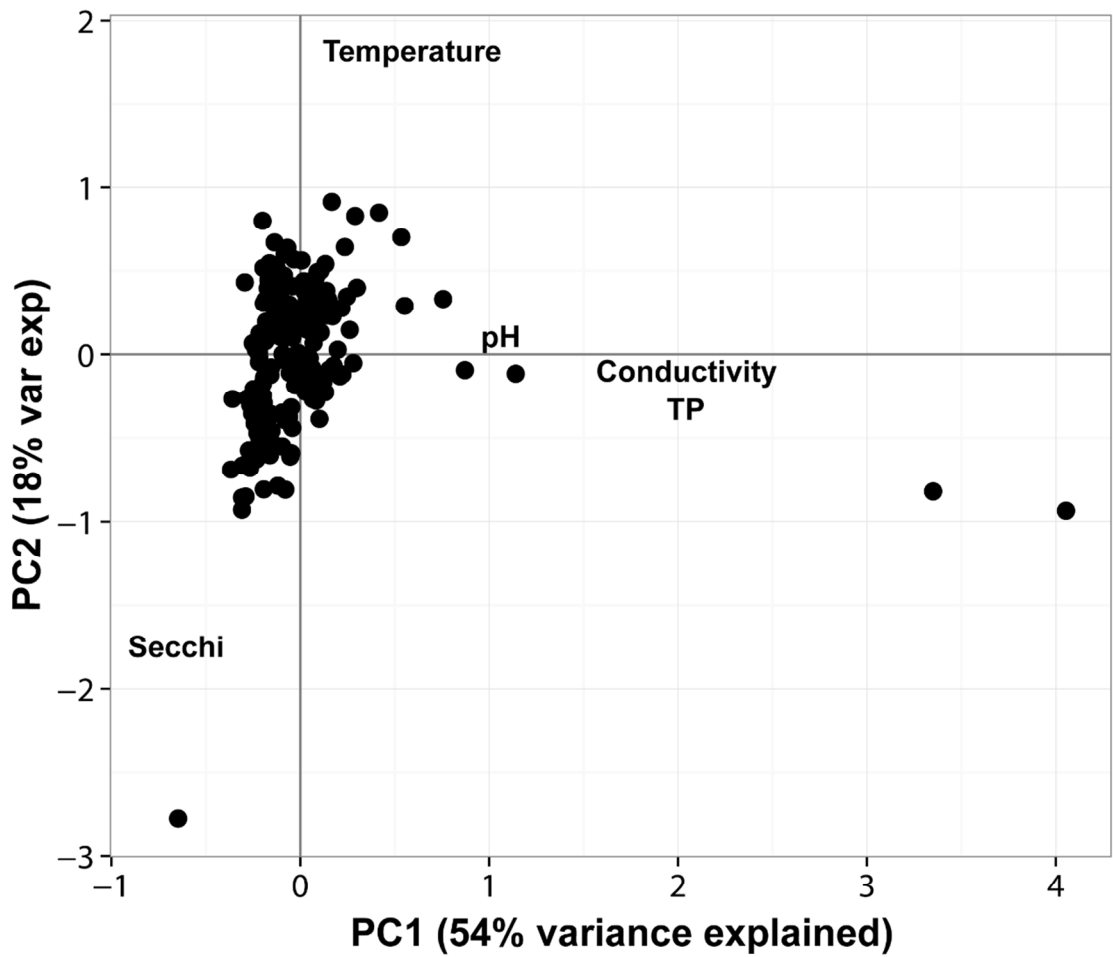
Each component is divided by a denominator, which is  $(2A+B+C)$  for computations based on the percentage difference index. Each component is then in the  $[0,1]$  interval and  $D = B + C$ .

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 species/genus loss ( $B$ ) and gain components ( $C$ ). We summarized these values for each ecoregion and examined the relationship between each of these components and latitude and longitude, using OLS regression. To use the function *decompose.D()*, we needed to combine the 2007 and historical diatom data sets. This “stacking” of datasets results in a number of zeroes being included in the resultant data frame; the changes in gamma diversity between time periods were small (see article Table 2). As such, we assessed the effect of this inclusion of zeros by examining subsets of the data with more or less zeros, but did not detect patterns in the magnitude of beta diversity.

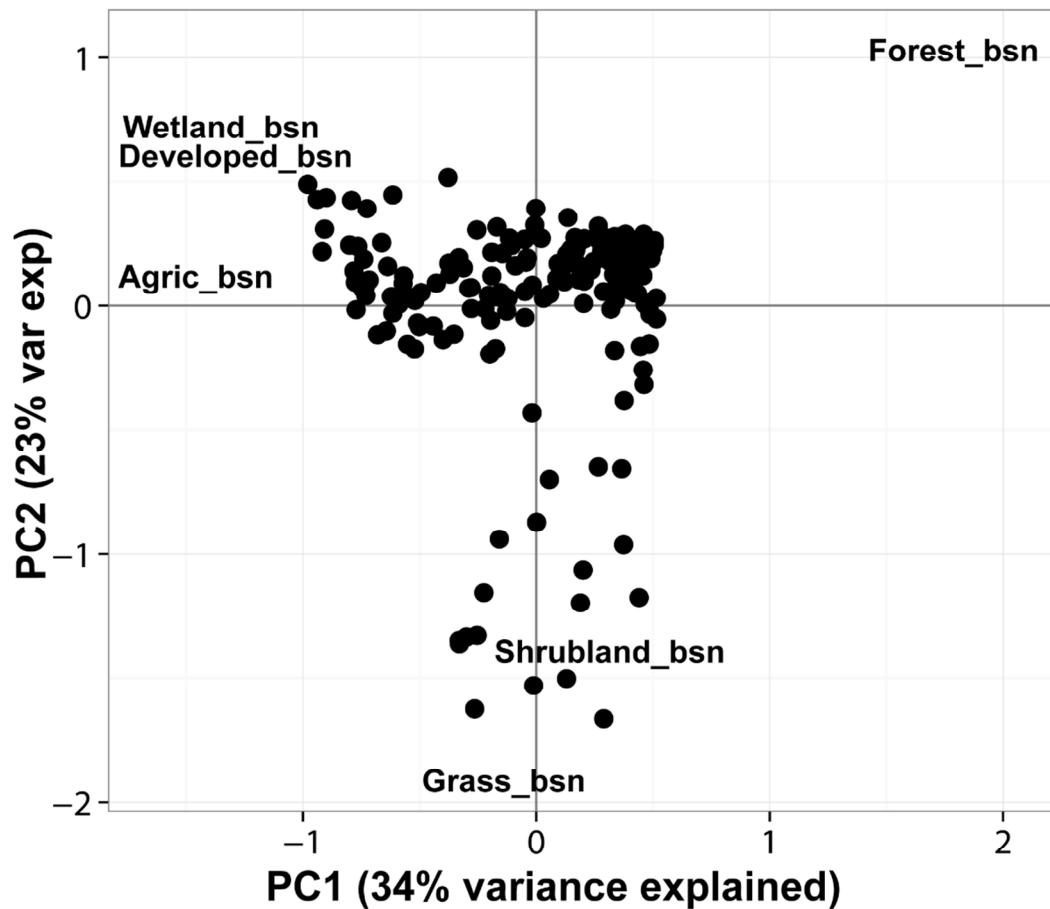
Legendre, P. (2015). R function: *TBI()*. Available online:  
<http://adn.biol.umontreal.ca/~numericecology/FonctionsR/>.

**S3: Principal Component Analyses (PCAs) of (a) standardized 2007 water quality variables and (b) land cover variables.**

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



(b) Land cover variables (NLCD 1992) – “Forest\_bsn” refers to percent forest 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; “Agric\_bsn” refers to percent agriculture including all agriculture types (crops, pasture etc.); “Shrubland\_bsn” refers to percent shrubland.



S4: Tables and figures for species-level analyses (n = 59).

S4, Table 1: Spatial beta ( $\beta$ ) diversity, mean rarefied species richness, mean alpha ( $\alpha$ ) diversity and gamma ( $\gamma$ ) 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. Genus richness was rarefied after rare genera (<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 sample size.

Ecoregion	$\beta$ -diversity		Rarefied S		$\alpha$ -diversity (Shannon)		$\alpha$ -diversity (Simpson)		$\gamma$ -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
Coastal Plains	0.45	0.40	32.7	37.2	2.0	2.5	0.7	0.8	3.4	3.7
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.

<b>Ecoregion</b>	<b>Repl</b>	<b>AbDiff</b>	<b>Repl/Total</b>	<b>AbDiff/Total</b>
<i>Historical</i>				
All	0.40	0	1	0
Coastal Plains	0.45	0	1	0
Northern Appalachians	0.39	0	1	0
<i>2007</i>				
All	0.37	0	1	0
Coastal Plains	0.40	0	1	0
Northern Appalachians	0.36	0	1	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

**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
Coastal Plains	0.3 (0.07)	0.3 (0.07)	0.6 (0.15)
Northern Appalachians	0.2 (0.08)	0.2 (0.08)	0.5 (0.17)



**S4, Figure 1: LCBD values for (a) historical spatial beta diversity and (b) 2007 spatial beta diversity and exceptional sites (c) for temporal beta diversity (TBI). A lake has a significant LCBD value if  $P < 0.05$ , and is coded as “True” (open circles). LCBD values across all sites (in**

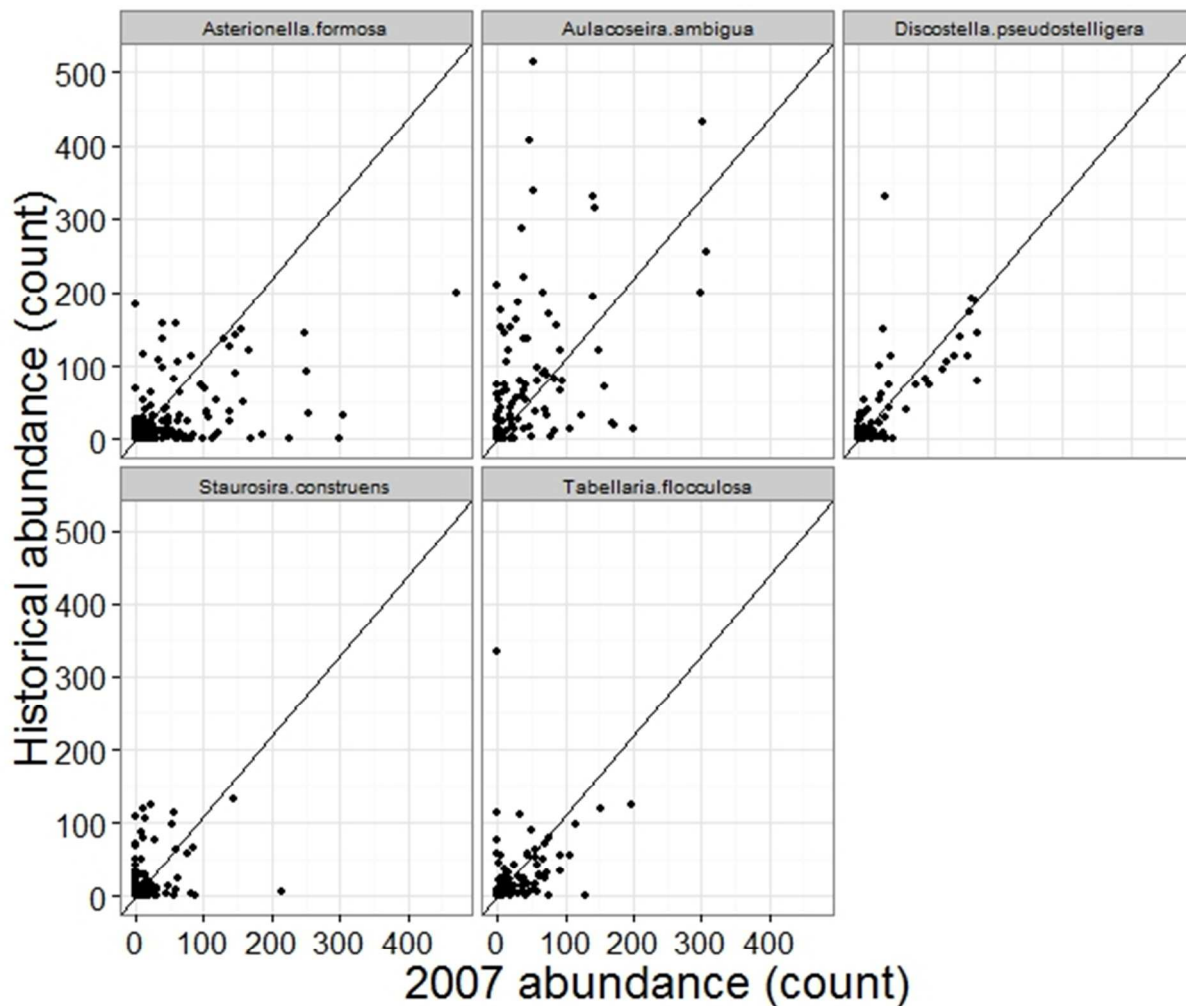
1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

either historical or 2007) sum to 1. Note that none of the TBI values were significant after  
correction for multiple testing in (c)

For Peer Review



**S5: 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 = 59, 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*.

For Peer Review