TITLE: Relationships between climate and phylogenetic community structure of fossil pollen assemblages are not constant during the last deglaciation

AUTHORS: Kavya Pradhan\textsuperscript{1,}*\textsuperscript{#a}, Diego Nieto-Lugilde\textsuperscript{1,}*\textsuperscript{#b}, Matthew C. Fitzpatrick\textsuperscript{1}

AUTHOR AFFILIATIONS:

\textsuperscript{1}University of Maryland Center for Environmental Science, Appalachian Lab, Frostburg, MD USA

\textsuperscript{#a}Current Address: Department of Biology, University of Washington, Seattle, WA USA

\textsuperscript{#b}Current Address: Departamento de Botánica, Ecología y Fisiología Vegetal, Universidad de Córdoba, Córdoba, España

*CORRESPONDING AUTHOR:

E-mail: mfitzpatrick@umces.edu
Abstract

Disentangling the influence of environmental drivers on community assembly is important to understand how multiple processes influence biodiversity patterns and can inform understanding of ecological responses to climate change. Phylogenetic Community Structure (PCS) is increasingly used in community assembly studies to incorporate evolutionary perspectives and as a proxy for trait (dis)similarity within communities. Studies often assume a stationary relationship between PCS and climate, though few if any studies have tested this assumption over long time periods with concurrent community data. We estimated Nearest Taxon Index (NTI) and Net Relatedness index (NRI), two PCS metrics, of fossil pollen assemblages of angiosperms in eastern North America data over the last 21 ka BP at 1ka intervals. We analyzed spatiotemporal relationships between PCS and seven climate variables, evaluated the potential impact of deglaciation on PCS, and tested for the stability of climate-PCS relationships. The broad scale patterns of PCS, with overdispersion increasing towards the southern and eastern parts of the study area, remained largely stable across time. Most importantly, we found that significant relationships between climate variables and PCS (slope) were unstable as climate changed during the last deglaciation and new ice-free regions were colonized. We also found weak, but significant, relationships between both PCS metrics (i.e., NTI and NRI) and climate and time-since-deglaciation, which were stable even though the baselines (intercepts) changed through time. Overall, our results suggest that (1) PCS of fossil Angiosperm assemblages during the last 21ka BP have had predictable spatial patterns, but (2) the instability in the relationships between PCS and climate brings into question their usefulness in predictive modeling of community assembly.
Introduction

Determining how abiotic and biotic factors influence biological communities is important for understanding community assembly processes and predicting responses to global change. Studies of community dynamics underscore the prominent role of abiotic factors such as climate in mediating assembly processes and in determining community composition via physiological controls on species occurrence (known as environmental filtering; [1–4]. The paleoecological record provides a window into community dynamics across extended periods of climatic change during which groups of co-occurring taxa disassembled and reassembled, sometimes forming no-analog communities [5]. For vegetation, fossil pollen records from sediment cores suggest an influence of climate on plant taxon associations [6] and show clear linkages between climate and community dynamics, such as synchrony between climatic events and compositional changes in plant communities [7,8].

Although climate plays a primary role in shaping communities, biotic processes such as interactions between species and dispersal also are important. For example, dispersal lags associated with post-glacial migration have been shown to influence the geographic ranges of some European plant species and therefore plant community composition [9,10]. Biotic interactions such as competition and facilitation can work in concert with environmental filtering and dispersal constraints to influence community structure [11]. Although the effects of biotic interactions are considered to be most prominent at local scales, their influence has been inferred at the scales of species geographic ranges and other macroecological patterns [12–15].

Static observations provide limited information for differentiating the processes
that generate patterns of community composition, but Phylogenetic Community Structure (PCS) often has been used for this purpose. PCS quantifies the extent to which co-occurring species are phylogenetically related and can serve as a proxy for trait (dis)similarity to infer the relative importance of abiotic and biotic processes in determining community structure [16]. PCS patterns can be compared to random assemblages to test whether the degree of relatedness of co-occurring species differs from that expected by chance [17,18], with nonrandom patterns typically being interpreted as reflecting the outcome of either environmental filtering or competition. Environmental filtering generally is hypothesized to lead to clustered communities (co-occurring species are more related / have greater trait similarity than expected by chance; [19,20], while competition should tend to lead to overdispersed communities (co-occurring species are less related / have more dissimilar traits than expected by chance; [18,19]. It is important to note however that these general predictions for relationships between nonrandom patterns of PCS and the processes responsible for community composition do not always hold given (1) the complex interplay between climate, phylogeny, and biotic processes (mediated through functional traits) and (2) that a particular set of processes can generate multiple patterns, and vice versa [21–24]. In addition, PCS can be influenced by processes other than competition and contemporary environmental filtering, notably the legacy of past climatic conditions due to dispersal lags [25] or intercontinental migrations and in-situ speciation [26].

In addition to complexities that can obscure links between processes and patterns using PCS, studies typically assume that PCS-climate relationships are stable through time, and although community assembly is a dynamic process [8], most studies have analyzed PCS using static community data [27–29]. When studies have examined PCS through time,
most have either used (1) a proxy for time rather than a time series of observations (e.g., chronosequences; [30–33] or (2) relatively short time sequences [34–36]. Studies that have analyzed temporal patterns in PCS most often have found an increase in phylogenetic overdispersion (increasing role of competition) through time [30,31,33,36], while others have found evidence for increased phylogenetic clustering (increasing role of environmental filtering) through time [35].

Extensive paleoecological time series provide an opportunity to examine assemblage dynamics across large spatial extents and extended periods of time as species responded to climatic events, which, when combined with analyses of PCS, may provide inferences regarding the processes shaping community composition that otherwise may be difficult to obtain from analyses of snapshots. In this study, we examine spatiotemporal patterns of fossil pollen assemblages of Angiosperms across eastern North America from the Last Glacial Maximum (LGM; 21 ka BP -- thousands of years before present) to present (0 BP) to assess how PCS evolves as previously glaciated areas are colonized. Specifically, we address three questions: (1) Does PCS of Angiosperm assemblages exhibit nonrandom patterns across space and through time since the LGM?; (2) What is the relationship between PCS and climate, and does this relationship remain stable through time?; and (3) How does deglaciation impact PCS through its influence on subsequent colonization and succession processes?

We expected to find a significant effect of climate on PCS from pollen assemblages, with phylogenetically clustered communities in places and at times where environmental filtering should dominate (e.g., high latitudes and harsh LGM climates), and phylogenetically overdispersed communities in places and at times where the role of
environmental filtering should be reduced (e.g., low latitudes and benign present-day climates). Furthermore, if the PCS-climate relationship is one of the main assembly rules for communities, we predicted that it should remain stable through time. Finally, given the importance of glaciations and the potential impacts of postglacial migrations on plant species and communities, we expected to find a strong signal of time-since-deglaciation on PCS. Our analyses revealed a persistent spatial trend of phylogenetic clustering in the northeast and overdispersion towards the southwestern portion of the study region. Most importantly, relationships between PCS and both climate and time-since-deglaciation were not stable through time.

**Material and Methods**

**Pollen Occurrence Data**

We focus on the last deglaciation (21 ka BP to present) of the eastern and southern half of North America (113°30´ - 53°00´W / 25°00´ - 61°00´N), where time series of fossil pollen records are relatively dense. We used taxon occurrence records from 21 ka BP to 0 ka BP (present) at 1 ka intervals as described by [37] and [38]. Paleoecological records suffer from a host of uncertainties and vary greatly in their degree of temporal and taxonomic resolution. We minimized these issues to the greatest extent possible by selecting fossil-pollen records that (1) have low temporal uncertainty (< 500 years at most sites), (2) use a standardized taxonomy for identification, and (3) estimate occurrence using robust temporal interpolations of relative abundance expressed as the pollen sum for a particular taxon divided by the total sum for all genus-level taxa, rather than divided by the total
upland sum for the site (which includes both genera as well as families and other higher level taxa; [39]. Pollen data quality was calculated as in [37] and [38], and for each 1 ka time step, only sites with a weighted quality value above 0.75 were included. If multiple sites fell within the same 0.5 x 0.5 degrees grid cell (see Climate and Ice Sheet Data section), their pollen abundances were averaged. The majority of grid cells contained pollen assemblages from within a single lake sediment core. We used pollen relative abundances for each climate grid cell and 1 ka time interval. The phylogenetically related taxa pairs *Oxyria-Rumex, Juniperus-Thuja*, and *Ostrya-Carpinus* cannot readily be distinguished from pollen. Hence, those taxa were combined to form a unique branch in the phylogenetic tree (see Phylogenetic Data section). Additionally, *Ambrosia* includes other genera in Asteraceae that have similar pollen morphology and are generally indistinguishable. However, close relatives have similar pollen morphology; as such this aggregation should not bias our results.

All told, the initial dataset included 106 pollen taxa identified to the genus level of both Gymnosperms and Angiosperms. We decided to run analyses only with Angiosperms (n=96) to prevent deep splits in the phylogenies from confounding PCS metrics (see below) and because the low number of Gymnosperm taxa (n=10) prevented estimation of PCS for most of the locations and time slices when Gymnosperms are examined separately. This resulted in 96 Angiosperms taxa at the genus level (S1 Fig). Pollen data are available from the Dryad Digital Repository [38].

**Climate and Ice Sheet Data**

Climate data for North America were obtained from CCSM3 transient simulation [40] and
were subsequently processed and downscaled to yearly, quarterly, and monthly variables at a 0.5 x 0.5 degree grid and 1 ka intervals by [41]. We selected seven uncorrelated variables (Pearson < 0.75) that capture the interplay of water availability and energy, and therefore should influence the taxonomic composition of vegetation, including: minimum temperature of the coldest month (Tmin), maximum temperature of the warmest month (Tmax), minimum precipitation of the driest month (Pmin), maximum precipitation of the wettest month (Pmax), mean yearly actual evapotranspiration (AET), mean yearly ratio of actual and potential evapotranspiration (ETR), and mean yearly water deficit index (WDI).

Climate data are available from the Dryad Digital Repository [41].

To study the effect of postglacial migration on PCS, maps of ice sheet extent at each 1 ka time interval were obtained from [42]. Because of minor temporal mismatches between the ice sheet maps and our climate and pollen datasets, we assigned each ice sheet layer to the closest 1 ka time interval in our data. Using these ice sheet maps, we classified cells during each of the 1 ka time intervals as either glaciated or deglaciated and then calculated the time since each grid cell was deglaciated (DEGLAC) at each time period. We identified 9 instances in which a grid cell at the same time period both contained a pollen site and also was considered to be glaciated. These cells were always adjacent to the border of the ice sheet layer and therefore most likely arose due to temporal mismatches between the ice sheet layers and the pollen data and/or low spatial precision of the ice sheet layers. Given the higher integrity of the pollen data, for these few instances we reclassified such conflicting cells as deglaciated.
Phylogenetic Data

To obtain a phylogenetic tree for the pollen taxa, we used the phylogeny from the megatree Open Tree of Life (OToL v.12.3; [43]. This megatree provides a topology for the tree of life based on multiple phylogenetic studies, resolving unstudied branches using known and accepted taxonomy (i.e., based on non-phylogenetic data). [44] have recently shown that OToL megatree provides the most similar results to purpose-built phylogenies compared to other available megatrees. We queried OToL for Spermatophyta (seed plants) using the taxonomic names of the pollen taxa at the genus level, retaining the 109 taxa in the initial pollen dataset (106 taxa after combining indistinguishable pollen taxa). Mismatches between pollen taxonomy and OToL taxonomy were resolved manually (e.g. Prunus pollen type is Amygdaleae in OToL; S1 Fig). To add branch lengths, which OToL does not provide, we obtained ages from the datelife project (http://datelife.org/; [45] for as many nodes as possible (n=36 of 98 nodes; S1 Fig), including Gymnosperm taxa. Lastly, we used the BLADJ algorithm in Phylocom [46] to estimate ages for the remaining nodes, thereby providing an estimate of branch lengths for the entire tree. We removed Gymnosperm taxa after calculating branch lengths, because the old split between Gymnosperms and Angiosperms helps to constrain node ages for the entire tree. Phylogenetic data are available from the Open Tree of Life https://blog.opentreeoflife.org [43]. We used the ‘rotl’ [47] and ‘datelife’ packages [48] in R [49].

Analysis

We calculated both Nearest Taxon Index (NTI) and Net Relatedness Index (NRI) to quantify
PCS in each occupied grid cell and at each 1 ka time slice using the ‘Picante’ package [50] in R [49]. NRI and NTI represent different components of evolutionary history [17,18]. NRI is calculated based on mean phylogenetic distance, which takes into account deep phylogenetic divergences, while NTI is calculated on the basis of mean nearest taxon distance, which takes more recent divergences into account. To standardize effect size for both metrics (i.e., NRI and NTI), the observed values were compared to a distribution of values from the “independentswap” null model (with default parameters), which randomizes community data matrix maintaining species occurrence frequency and sample species richness [51]. Sensitivity analysis with all null models available in ‘Picante’ were run. Each null model provided a completely different set of values of NTI and NRI, but the spatiotemporal pattern, the results from all subsequent analysis, and hence also main conclusions of our study remained the same and are not reported for the sake of simplicity.

We mapped the values of NTI and NRI at each fossil pollen site in each 1 ka time step to assess spatiotemporal patterns. We used ordinary least square (OLS) regression to quantify the relationship between NTI and NRI and each of the climatic drivers and time-since-deglaciation (e.g. NRI ~ TMIN). PCS metrics exhibited spatial autocorrelation and for many variables were not homoscedastic. To address these issues, we used Quantile Regression (QR) and error based Spatial Autoregressive (SAR) models. Because OLS and QR models exhibited spatial autocorrelation in the residuals and QR models did not show additional insight, we only report results from OLS, which is easier to interpret, and from SAR models that fix bias in parameters estimation. The SAR models were fit using the ‘spdep’ package [52] in R [49].

Because we were interested in testing whether or not PCS-climate relationships
were stable through time, we fit several OLS and SAR models with three levels of complexity for each climate variable: (1) Stable-Relationship (PCS ~ climate), which assumes that the relationship between PCS and the climate remained constant through time; (2) Stable-Slope (PCS ~ time + climate); which assumes that the slope of the relationship was stable through time but the intercept may have changed; and (3) Changed-Relationship (PCS ~ climate*time), which assumes that both the slope and the intercept may have changed through time. We compared the three versions of the OLS and SAR models using ANOVA. The most supported model in the ANOVA provides inference on whether the intercept and/or the slopes of PCS-climate relationships have changed through time. To avoid biases caused by low sample sizes, time periods that had less than 10 sites with PCS values were excluded (i.e., 17-21 ka BP). SAR models were fit using three different distance-based weight matrices: 120, 360, and 480 km. For each climate variable, we selected the distance that resulted in the lowest residual Moran’s I values when fitting the Stable-Relationship model. The selected distances were then used to fit the Stable-Slope and Changed-Relationship SAR models. To study the effect of postglacial migration on PCS, we additionally plotted the average PCS values across all cells of the same time-since-deglaciation (DEGLAC) for each time period.

**Results**

**Spatiotemporal patterns in PCS**

Across all twenty-two 1 ka time steps, plant assemblages recorded in fossil pollen exhibited NRI values (Fig 1a) ranging from -3.00 (overdispersion) to 2.86 (clustering), and NTI
values (Fig 1b) ranging from -3.42 to 2.34. Both NRI and NTI exhibited variation across space, with positive values (clustering) tending to occur in the east (slope = 0.015, $r^2 = 0.078$, $p < 0.001$ for NRI; slope = 0.015, $r^2 = 0.054$, $p < 0.001$ for NTI) and north (slope = 0.019, $r^2 = 0.037$, $p < 0.001$ for NRI; slope = 0.020, $r^2 = 0.025$, $p < 0.001$ for NTI) of the study area. Although the range of NRI and NTI values varied through time, the broad scale spatial pattern remained largely stable (Fig 1).

**Fig 1. Spatial pattern of phylogenetic community structure through time.** a) Net Relatedness Index (NRI) and b) Nearest Taxon Index (NTI) for all cells with pollen occurrences in each time period (for clarity only a subset of time periods are plotted). Higher values (purple) indicate more clustered communities (organisms are more related) and lower values (yellow) indicate overdispersed communities (organisms are less related). Ice sheet extent at each time period is represented with a white polygon.

**Climate-PCS relationship**

Most OLS models relating NRI or NTI with climate variables were significant (Table 1). Among the OLS models with stable slope and intercept through time (i.e., Stable-Relationship), NRI had a positive relationship with Pmin, Pmax, AET, ETR, and WDI, and a negative relationship with Tmin and Tmax (Table 1, Figs 2 and S2). NTI had a negative relationship with all variables except for Pmin, ETR and WDI (Table 1, Fig S3). For the Stable-Slope models, all relationships were similar to Stable-Relationship models, except for NRI~AET which had a negative relationship. For the Changed-Relationships model, the intercepts and slopes for both NRI and NTI varied through time (Figs 3 and 4) and there
was no consistent trend in the fluctuation between positive and negative slopes for either PCS metric. However, all OLS models, including those for Changed-Relationship, explained only a small amount of variation in PCS (mean adjusted $r^2 = 0.028$ and 0.019 SD for NRI and mean adjusted $r^2 = 0.043$ and 0.024 SD for NTI; Table 1). Variance explained tended to increase from the simplest model (Stable-Relationship models) to the most complex (Changed-Relationship models; Table 1).

Fig 2. Relationship between Net Relatedness Index (NRI) and three climate variables and their change through time according to the three fitted models. Gray shading in the scatter plots represents the count of points falling in each bin (hexagons). The panels on the left (a, c, e) represent the overall relationship according to ordinary least square regression when pooling all time periods (red lines). The panels on the right (c, d, f) show the relationship between NRI and climate variables as estimated with the data for different time periods, with green lines representing Stable-Slope Model, blue lines representing Changed-Relationship Model, and the red line showing the overall relationship from panels on the left for comparison. Shaded areas represent the confidence intervals at 95% for the regression lines.

Fig 3. Model intercepts through time. Intercepts of OLS and SAR models for each variable plotted across time for a) Stable-Slope and b) Changed-Relationship models.
Table 1. Parameters of ordinary least square regression (OLS) and spatial autoregressive (SAR) models relating two metrics of phylogenetic community structure (PCS; net relatedness index – NRI – and nearest taxon index – NTI) with seven climate variables. Each combination of PCS metric and climate variable was modeled three times, allowing distinct levels of variation to study the evolution of the parameters through time: stable-relationship, stable-slope, and changed relationship. SAR models reported here were fit selecting neighbors at distances of 480 km for NRI and 360 km for NTI.

<table>
<thead>
<tr>
<th>PCS metric</th>
<th>Model type</th>
<th>Var.</th>
<th>Stable-Relationship</th>
<th></th>
<th>Stable-Slope</th>
<th></th>
<th>Changed-Relationship</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tmin</td>
<td>OLS</td>
<td></td>
<td></td>
<td>-0.007</td>
<td>0.012</td>
<td>***</td>
<td>-0.009</td>
</tr>
<tr>
<td>Tmax</td>
<td></td>
<td></td>
<td></td>
<td>0.769</td>
<td>-0.024</td>
<td>***</td>
<td>-0.026</td>
</tr>
<tr>
<td>Pmin</td>
<td></td>
<td></td>
<td></td>
<td>-0.004</td>
<td>0.004</td>
<td>***</td>
<td>0.004</td>
</tr>
<tr>
<td>Pmax</td>
<td></td>
<td></td>
<td></td>
<td>0.232</td>
<td>0</td>
<td>ns</td>
<td>0</td>
</tr>
<tr>
<td>AET</td>
<td></td>
<td></td>
<td></td>
<td>0.469</td>
<td>0</td>
<td>0.032</td>
<td>***</td>
</tr>
<tr>
<td>ETR</td>
<td></td>
<td></td>
<td></td>
<td>-0.397</td>
<td>0.657</td>
<td>0.028</td>
<td>***</td>
</tr>
<tr>
<td>WDI</td>
<td></td>
<td></td>
<td></td>
<td>0.077</td>
<td>0</td>
<td>0.057</td>
<td>***</td>
</tr>
<tr>
<td>DEGLAC</td>
<td></td>
<td></td>
<td></td>
<td>0.362</td>
<td>0</td>
<td>0.021</td>
<td>***</td>
</tr>
<tr>
<td>Tmin</td>
<td>SARerr</td>
<td></td>
<td></td>
<td>0.155</td>
<td>0</td>
<td>0.143</td>
<td>ns</td>
</tr>
<tr>
<td>Tmax</td>
<td>(480 km)</td>
<td></td>
<td></td>
<td>-0.252</td>
<td>0.012</td>
<td>0.146</td>
<td>***</td>
</tr>
<tr>
<td>Pmin</td>
<td></td>
<td></td>
<td></td>
<td>-0.064</td>
<td>0.005</td>
<td>0.149</td>
<td>***</td>
</tr>
<tr>
<td>Pmax</td>
<td></td>
<td></td>
<td></td>
<td>0.042</td>
<td>0.001</td>
<td>0.144</td>
<td>ns</td>
</tr>
<tr>
<td>AET</td>
<td></td>
<td></td>
<td></td>
<td>0.182</td>
<td>0</td>
<td>0.143</td>
<td>ns</td>
</tr>
<tr>
<td>ETR</td>
<td></td>
<td></td>
<td></td>
<td>-0.007</td>
<td>0.202</td>
<td>0.144</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>WDI</td>
<td>0.159</td>
<td>0.147</td>
<td>***</td>
<td>0.153</td>
<td>***</td>
<td>0.165</td>
</tr>
<tr>
<td>-------</td>
<td>------</td>
<td>-------</td>
<td>-------</td>
<td>------</td>
<td>-------</td>
<td>------</td>
<td>-------</td>
</tr>
<tr>
<td>DEGLAC</td>
<td>0.204</td>
<td>0.144</td>
<td>ns</td>
<td>0.154</td>
<td>***</td>
<td>0.19</td>
<td>***</td>
</tr>
<tr>
<td></td>
<td>Tmin</td>
<td>0.006</td>
<td>-0.012</td>
<td>0.022</td>
<td>***</td>
<td>-0.012</td>
<td>0.028</td>
</tr>
<tr>
<td></td>
<td>Tmax</td>
<td>0.885</td>
<td>-0.027</td>
<td>0.034</td>
<td>***</td>
<td>-0.03</td>
<td>0.049</td>
</tr>
<tr>
<td></td>
<td>Pmin</td>
<td>-0.14</td>
<td>0.007</td>
<td>0.06</td>
<td>***</td>
<td>0.007</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>Pmax</td>
<td>0.578</td>
<td>-0.003</td>
<td>0.013</td>
<td>***</td>
<td>-0.004</td>
<td>0.027</td>
</tr>
<tr>
<td></td>
<td>AET</td>
<td>0.651</td>
<td>-0.001</td>
<td>0.048</td>
<td>***</td>
<td>-0.001</td>
<td>0.056</td>
</tr>
<tr>
<td></td>
<td>ETR</td>
<td>-0.692</td>
<td>1.029</td>
<td>0.044</td>
<td>***</td>
<td>1.05</td>
<td>0.054</td>
</tr>
<tr>
<td></td>
<td>WDI</td>
<td>0.057</td>
<td>0.001</td>
<td>0.083</td>
<td>***</td>
<td>0.001</td>
<td>0.097</td>
</tr>
<tr>
<td></td>
<td>DEGLAC</td>
<td>0.478</td>
<td>0</td>
<td>0.028</td>
<td>***</td>
<td>0</td>
<td>0.034</td>
</tr>
</tbody>
</table>

|       | Tmin | 0.029 | -0.01 | 0.226 | ***  | -0.003| 0.233 | ns    |
|       | Tmax | 0.324 | -0.006| 0.221 | ns    | 0.001 | 0.233 | ns    |
|       | Pmin | 0.071 | 0.003 | 0.222 | *     | 0.005 | 0.236 | ***  |
|       | Pmax | 0.302 | -0.001| 0.222 | ns    | -0.003| 0.236 | ***  |
|       | AET  | 0.456 | 0     | 0.229 | ***  | 0     | 0.235 | **    |
|       | ETR  | -0.031| 0.263 | 0.222 | ns    | 0.191 | 0.233 | ns    |
|       | WDI  | 0.183 | 0     | 0.225 | ***  | 0     | 0.234 | ns    |
|       | DEGLAC | 0.361 | 0     | 0.232 | ***  | 0     | 0.236 | **    |

*** p<0.001; ** p<0.01; * p<0.05. ns non-significant (p>0.05)

The raw NRI and NTI values exhibited significant spatial structure (S1 Table) at all three neighborhood distance classes (i.e., 120, 360, 480 km), and correspondingly, all of the OLS models showed significant spatial autocorrelation in model residuals (S2 Table). SAR models effectively removed spatial autocorrelation at all three neighborhood distance classes (S3 Table), and for both NRI and NTI, SAR models fitted using a neighborhood of
480 km and 360 km, respectively, were most parsimonious (lowest AIC; S4 Table). Here forward we report results only for these SAR models and indicate explicitly when referring to OLS models.

When spatial autocorrelation was removed, the significance levels of the regression models for both NRI and NTI generally decreased and, in some cases, became non-significant (Tmin, AET and ETR for NRI; and Tmax, Pmax, and ETR for NTI; Table 1). Among the SAR models with stable slope through time (i.e., Stable-Relationship and Stable-Slope), NRI showed a positive correlation with all climate variables other than Tmin, whereas NTI showed a negative correlation with all climate variables and models types except ETR and Pmin in the Stable-Relationship model, and ETR, Pmin, and Tmax in the Stable-Slope model (Table 1). Similar to OLS models, in the Changed-Relationships SAR models, the intercepts and slopes for both NRI and NTI varied through time (Figs 3 and 4) with no clear trend towards more positive or more negative slopes.

**Fig 4. Model slopes through time.** Slopes of OLS and SAR models for each variable plotted across time for the Changed-Relationship models.

The explanatory power of the models increased appreciably in the SAR models compared to the OLS models, and these increases were greater for NTI than NRI models (for NRI: mean adjusted $r^2 = 0.160$ and 0.021 SD across all models; and for NTI: mean adjusted $r^2 = 0.236$ and 0.011 SD; Table 1). SAR models also showed an increase in explained deviance for both PCS metrics among models types, with Stable-Relationship having the lower explained deviance (for NRI: mean adjusted $r^2 = 0.145$, 0.002 SD and for
NTI: mean adjusted \( r^2 = 0.225, 0.004 \text{ SD} \), followed by Stable-Slope (for NRI: mean adjusted \( r^2 = 0.151, 0.002 \text{ SD} \) and for NTI: mean adjusted \( r^2 = 0.234, 0.001 \text{ SD} \)), and Changed-Relationship having the higher value of explained deviance (for NRI: mean adjusted \( r^2 = 0.186, 0.0018 \text{ SD} \) and for NTI: mean adjusted \( r^2 = 0.250, 0.007 \text{ SD} \)). Lastly, in an \textit{ad hoc} analysis we re-ran all models and ANOVA analysis equalizing the sample size across time periods. The results from this sensitivity analysis showed that models were not biased by differing sample sizes among time periods (data not shown).

For both NRI and NTI, ANOVA revealed stronger support for Stable-Slope models than Stable-Relationship models for most climate variables (Table 2). The exceptions here include, for NRI, all climate variables with SAR models and, for NTI, for Tmin and AET in both OLS and SAR models, and ETR and WDI in OLS models. Changed-Relationship models also generally had greater support than Stable-Slope models, except for NTI for ETR and WDI in OLS models.
Table 2. ANOVA-based model selection of ordinary least square regression (OLS) and spatial autoregressive (SAR) models relating net relatedness index (NRI) and nearest taxon index (NTI) with seven climate variables. Each ANOVA was run for each combination of PCS metric and climate variable comparing three different models that allow distinct levels of variation to study the evolution of the regression parameters through time: stable-relationship, stable-slope, and changed relationship. SAR models reported here were fit selecting neighbors at distances of 360 km.

<table>
<thead>
<tr>
<th>PCS metric</th>
<th>Model type</th>
<th>Stable-Relationship</th>
<th>Stable-Slope</th>
<th>Changed-Relationship</th>
<th>Selected Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Var. (OLS)</td>
<td>F (OLS)</td>
<td>Df (SAR) L ratio (SAR)</td>
<td>Var. (OLS)</td>
</tr>
<tr>
<td>Tmin</td>
<td>OLS</td>
<td>2961 NA</td>
<td>2943 3.23</td>
<td>2925 10.832</td>
<td>2*** 3***</td>
</tr>
<tr>
<td>Tmax</td>
<td>OLS</td>
<td>2961 NA</td>
<td>2943 3.343</td>
<td>2925 8.024</td>
<td>2*** 3***</td>
</tr>
<tr>
<td>Pmin</td>
<td>OLS</td>
<td>2961 NA</td>
<td>2943 2.216</td>
<td>2925 4.305</td>
<td>2** 3***</td>
</tr>
<tr>
<td>Pmax</td>
<td>OLS</td>
<td>2961 NA</td>
<td>2943 1.954</td>
<td>2925 6.663</td>
<td>2** 3***</td>
</tr>
<tr>
<td>AET</td>
<td>OLS</td>
<td>2961 NA</td>
<td>2943 4.17</td>
<td>2925 9.361</td>
<td>2*** 3***</td>
</tr>
<tr>
<td>ETR</td>
<td>OLS</td>
<td>2961 NA</td>
<td>2943 2.108</td>
<td>2925 2.03</td>
<td>2** 3***</td>
</tr>
<tr>
<td>WDI</td>
<td>OLS</td>
<td>2961 NA</td>
<td>2943 5.605</td>
<td>2925 2.8</td>
<td>2*** 3***</td>
</tr>
<tr>
<td>DEGLAC</td>
<td>OLS</td>
<td>2961 NA</td>
<td>2943 7.189</td>
<td>2925 4.317</td>
<td>2*** 3***</td>
</tr>
<tr>
<td>Tmin</td>
<td>SARerr (480 km)</td>
<td>4 NA</td>
<td>22 18.815</td>
<td>40 236.344</td>
<td>3***</td>
</tr>
<tr>
<td>Tmax</td>
<td>SARerr (480 km)</td>
<td>4 NA</td>
<td>22 20.278</td>
<td>40 120.693</td>
<td>3***</td>
</tr>
<tr>
<td>Pmin</td>
<td>SARerr (480 km)</td>
<td>4 NA</td>
<td>22 12.331</td>
<td>40 103.117</td>
<td>3***</td>
</tr>
<tr>
<td>Pmax</td>
<td>SARerr (480 km)</td>
<td>4 NA</td>
<td>22 14.77</td>
<td>40 119.259</td>
<td>3***</td>
</tr>
<tr>
<td>AET</td>
<td>SARerr (480 km)</td>
<td>4 NA</td>
<td>22 22.084</td>
<td>40 202.106</td>
<td>3***</td>
</tr>
</tbody>
</table>
Impact of deglaciation on PCS

Time-since-deglaciation (DEGLAC) ranged from 0 to 21 ka and both PCS metrics showed a positive relationship with DEGLAC (Table 1). Similar to climate variables, DEGLAC explained a small amount of variation in OLS models but this amount increased in SAR
models. Furthermore, the amount of variation explained also increased with increasing model complexity (Stable-Relationship to Changed-Relationship). Model selection using ANOVA indicated that the Stable-Relationship model was the least supported for both OLS and SAR analyses for both PCS metrics, with the exception of OLS models for NTI. Intercepts for DEGLAC show similar trends to those observed for climate variables in Stable-Slope and Changed-Relationship models (Fig 3). When aggregating (by averaging) PCS values for each combination of DEGLAC and time period (Fig 5), we found phylogenetically clustered communities (positive PCS metrics) in sites that had deglaciated more recently (low DEGLAC) than in sites that had been deglaciated for longer time periods (high DEGLAC). However, we also found that the patterns became less apparent when the effect of climate variables was removed using multiple regression models that included all climate variables and plotting the averaged residuals for each combination of DEGLAC and time period (Fig 5).

**Fig 5. Pattern of time-since-deglaciation on Net Relatedness Index (NRI).** Each cell in the heatmap represents the average NRI value for cells of a particular time-since-deglaciation at a particular time period. Grey squares indicate absence of the particular time-since-deglaciation class for that time period. Blue colors represent higher NRI values (NRI > 0) and, hence, clustered community structure, whereas red colors represent lower NRI (NRI < 0) values and, hence, overdispersed community structure.

**Discussion**

Our analyses of the spatial and temporal dynamics fossil pollen assemblages suggest that
changes in vegetation in eastern North America over the last 21 ka have been accompanied by substantial changes in PCS. Angiosperm assemblages tended to become more phylogenetically clustered in the northeastern parts of the study area and more overdispersed in the southwestern regions. Most importantly, the relationships between PCS and climate were not stable through time and exhibit substantial autocorrelation. Finally, we also found that the relationship between time-since-deglaciation and PCS also has not been constant through time.

Both the range of PCS values and their skewed distribution towards positive values are consistent with results reported by previous studies [26,53]. Although fossil pollen data are biased towards certain plant clades (i.e. wind pollinated tree taxa), they include plant genera that represent some of the most widespread and abundant plant taxa within the study region and therefore should characterize general vegetation patterns. However, the coarse taxonomic resolution of the fossil pollen data (e.g., genus level) used in our analyses may impact PCS metrics due to a lack of terminal branches in the phylogenetic tree and may explain the differences in PCS values when compared with species-level studies [26,53] performed at a similar spatial scale as in this study. Genus-level data would likely tend to influence NTI more than NRI, given that NTI assigns greater emphasis to recent splits in the phylogeny than in NRI [17,18]. Our use of genus-level data also may explain differences in the range of NRI and NTI values.

Null model selection also can affect the range and skewness of PCS metrics. However, our sensitivity analyses incorporating different null models produced similar results and conclusions regarding the relationship between NRI / NTI and climate and time-since-deglaciation even though the range of PCS values varied (data not shown). The
selected taxa pool and the clade under consideration also could influence PCS patterns [26,54]. Lastly, most previous studies analyzed plot-based data at the species level, whereas we are analyzing records of taxa preserved in lake sediments, which may not represent local communities. For example, in a comparison of similarities between community patterns measured using plots versus pollen cores, [55] found that pollen samples best represented plant assemblages at large spatial extents (10 arcminutes and above).

**PCS and spatiotemporal patterns**

Previous studies examining spatial patterns of PCS in the northern hemisphere for angiosperms reported similar results with more clustering in the north and overdispersion towards the south [53,56]. Our results for Angiosperms contradict previous descriptions of PCS patterns for Gymnosperms [25,53]; this, in fact, agrees with the idea that PCS varies with the clade under consideration.

Although a few studies have examined current or past PCS at large geographical extents, none that we are aware of also have addressed changes in spatial patterns of PCS through time. We found that PCS metrics tended to become more positive though time, suggesting that plant communities tend towards phylogenetic clustering as they (re)assemble in response to climate changes. Phylogenetic clustering could be attributed to increasing importance of habitat filtering, assuming a perfect match between phylogenies and functional traits. However, testing this prediction is difficult given that the relative importance of climate, biotic interactions, and other factors such as the dynamics of succession in determining these patterns remains unclear.
Relationship between PCS and climate through time

Our results generally agree with previous studies in the effect of the different climate variables on PCS (i.e., we found similar relationships but with a much lower variance explained). For instance, a large amount of variation in PCS can usually be explained by temperature and precipitation variables [25,26,53,54]. Although climate is known to have a central role in structuring communities when measured on the basis of species richness [10,57] or beta diversity [58], our results suggest that it has a more limited role in explaining PCS (Table 1), at least at the coarse spatial and taxonomic scales of our study. Note that we found strong spatial autocorrelation in PCS and that accounting for spatial autocorrelation increased the explanatory power of certain climate variables, while also revealing non-significant relationships with other variables (i.e. mainly ETR and Pmax). Incorporating spatial effects also inverted the direction of the relationship for several variables (Tmax for NRI or Pmin for NTI; Table 1). While our findings regarding relationships between annual temperature and NRI differ from those reported for pine and podocarp clades based on OLS analysis [25], we intentionally removed gymnosperms from our analysis. Hence differences in the relationships between climate and NRI might be attributable to differential adaptations of angiosperms and gymnosperms to climate [59].

For climate variables with a significant relationship with PCS, we found that climate-PCS relationships are not stable through time, especially when spatial autocorrelation is taken into account. Independently of spatial autocorrelation, PCS-climate relationships varied through time for all of the environmental variables tested (Table 2). Changes in both intercepts and slopes through time suggest a signal of other factors, notably biotic
interactions or dispersal lags, on PCS patterns. In terms of dispersal lags, previous studies have shown that past climate conditions and associated dispersal limitations can influence present community composition and PCS. For instance, analyses of PCS of global conifer, angiosperm, and palm assemblages revealed that both past and current climate are important predictors for NRI [25,54,60]. Our study, however, highlights the changing importance of climate on PCS and calls for caution when interpreting the results from studies based on single-time periods. Taken together, our results suggest that the signal of climate on PCS either is (1) weak and inconsistent in fossil pollen assemblages of Angiosperms in eastern North America since the LGM, and/or (2) obscured by the coarse spatial, temporal, and taxonomic resolution of the records themselves.

**Time-since-deglaciation and PCS**

Our understanding of how disturbance influences PCS is mainly limited to studies analyzing relatively short time series [30–33,35]. These studies suggest that communities that assemble post-disturbance initially tend to be phylogenetically clustered, with succession processes tending to increase overdispersion through time. Our results are consistent with this pattern (Fig 5).

We show that the relationship between PCS and time-since-deglaciation (DEGLAC) is also unstable across time. Temporal variation in model intercepts, but constant slopes would be consistent with temporal lags between deglaciation and changes in PCS over time as well as migration lags. Relative to other regions (e.g., Europe), we might expect postglacial migration to be less constrained by dispersal barriers in North America where mountain ranges are aligned mainly north-south (notwithstanding the role of the
Laurentian Great Lakes). In addition, recolonization of deglaciated regions by some species could have been driven by isolated patches of vegetation near the ice boundary rather than from distant locations [61]. However, our results suggest that slopes also have been unstable.

**Conclusions**

Although relationships between PCS and environment have the potential to inform predictions, our study suggests that such predictions should be approached with caution. We found a consistent pattern of decreasing PCS values from the northeastern to the southwestern parts of our study area across time, but the ecological processes underlying these patterns cannot be discerned. In addition, not only did we find spatial autocorrelation in PCS, proper accounting of spatial autocorrelation altered the stability of PCS-climate relationships through time, suggesting that static PCS-climate relationships may have limited utility for making predictions under changing environments. This remains true for the relationship between PCS and time-since-deglaciation, which suggests that, in general, communities that initially establish post-disturbance tend to be more dispersed than those that proceed them, but this pattern was not consistent.

**Acknowledgements**

TBD.
References


8. Williams JW, Post DM, Cwynar LC, Lotter AF, Levesque AJ. Rapid and widespread


35. Letten AD, Keith DA, Tozer MG. Phylogenetic and functional dissimilarity does not


45. Sánchez-Reyes LL, O'Meara BC. DateLife: Leveraging databases and analytical tools to reveal the dated Tree of Life. 2019. p. 782094. doi:10.1101/782094


Supporting information

S1 Fig. Phylogenetic tree with angiosperm (blue branches) and Gymnosperm taxa (green branches) from the Open Tree of Life for Spermatophytes. The tree is pruned to the taxa in the fossil pollen dataset. Node labels indicate estimated age of the node as in the ‘datelife’ database. Ages for the rest of the nodes were estimated using the bladj algorithm as implemented in the ‘picante’ package in R.

S2 Fig. Scatterplots for NRI against all climatic variables and time-since-deglaciation (DEGLAC) and regression lines for three OLS models: Stable-Relationship (orange line), Stable-Slope (green line) and Changed-Relationship (blue line).

S3 Fig. Scatterplots for NTI against all climatic variables and time-since-deglaciation (DEGLAC) and regression lines for three OLS models: Stable-Relationship (orange line), Stable-Slope (green line), and Changed-Relationship (blue line).

S1 Table: Spatial autocorrelation of raw NRI and NTI values.

S2 Table: Moran’s I values for residuals of OLS models relating NRI and NTI with all variables (geographic, climatic, and time-since-deglaciation).

S3 Table. Moran’s I values for residuals of SAR models relating NRI and NTI with seven climate variables. Moran’s I values are based on models fitted with error-SAR at the specified distances and are all non-significant (ns).
S4 Table. AIC values for OLS and SAR models relating NRI and NTI with seven climate variables. AIC values of SAR are based on models fitted with error-SAR at the specified distances. The lowest AIC values for each climatic variable and PCS metric (i.e., NRI and NTI) are highlighted in bold.
Figure 1
Figure 3
Figure 4
Figure 2