DOI: 10.1111/jbi.13358

RESEARCH PAPER

WILEY Journal of Biogeography

Genomic evidence of a widespread southern distribution during the Last Glacial Maximum for two eastern North American hickory species

Jordan B. Bemmels[1] | Christopher W. Dick[1,2]

[1]Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI

[2]Smithsonian Tropical Research Institute, Ancon, Panama

Correspondence
Jordan B. Bemmels, Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, Michigan 48109.
Email: jbemmels@umich.edu

Funding information
Horace H. Rackham School of Graduate Studies, University of Michigan; Division of Environmental Biology, National Science Foundation, Grant/Award Number: DDIG 1501159; Department of Ecology and Evolutionary Biology, University of Michigan; National Science Foundation, Grant/Award Number: GRFP Fellowship

Editor: Jack Williams

Abstract

Aim: Phylogeographical studies of temperate forest taxa often infer complex histories involving population subdivision into distinct refugia during the Last Glacial Maximum (LGM). However, some temperate deciduous trees may have been broadly distributed in southeastern North America during the LGM. We investigate genome-wide genetic structure in two widespread eastern North American tree species to determine if range expansion from genetically isolated refugia or from a broader, less genetically subdivided region better explains their post-glacial history.

Location: Eastern North America (ENA).

Taxa: Bitternut hickory (*Carya cordiformis* [Wangenh.] K.Koch) and shagbark hickory (*Carya ovata* [Mill.] K.Koch).

Methods: Genetic diversity and differentiation indices were calculated from >1,000 nuclear SNP loci genotyped in ca. 180 individuals per species sampled across ENA. Genetic structure was investigated using principle component analysis and genetic clustering algorithms. As an additional tool for inference, areas of suitable habitat during the LGM were predicted using species distribution models (SDMs).

Results: Populations across all latitudes showed similar levels of genetic diversity. Most genetic variation was weakly differentiated across ENA, with the exception of an outlier population of *Carya ovata* in Texas. Genetic structure in each species exhibited an isolation-by-distance pattern. SDMs predicted high LGM habitat suitability over much of the southeastern United States.

Main conclusions: Both hickory species likely survived the LGM in low-density populations that were broadly distributed across southeastern North America and not highly genetically differentiated, except that the range-edge Texas population of *Carya ovata* may represent a separate glacial refugium. Over most of ENA, genetic structure in both species is best explained by simple latitudinal range shifts and high gene flow among populations, rather than expansions from multiple, genetically isolated refugia as is characteristic of taxa from other Northern Hemisphere temperate regions of the world.

KEYWORDS
eastern North America, glacial refugia, isolation by distance, Last Glacial Maximum, phylogeography, range expansion, temperate trees

1 | INTRODUCTION

Temperate forests have long served as models for understanding how migrational responses to Pleistocene glaciation gave rise to population genetic structure in terrestrial organisms (Hewitt, 1999, 2000; Petit et al., 2003; Qiu, Fu, & Comes, 2011; Shafer, Cullingham, Côté, & Coltman, 2010; Soltis, Morris, McLachlan, Manos, & Soltis, 2006). In Europe, where many classic phylogeographical paradigms were first established (Lumibao, Hoban, & McLachlan, 2017), temperate taxa typically retreated to glacial refugia in Mediterranean regions (Hewitt, 1999, 2000; Petit et al., 2003). Glacial refugia, as defined here, are relatively small, geographically distinct regions, among which genetic connectivity is low (Bennett & Provan, 2008). After expanding out of refugia following glacial retreat, many European species experienced a progressive loss of genetic diversity due to founder effects during northward migration (Hewitt, 1999, 2000). However, mid-latitude areas often exhibit elevated genetic diversity due to admixture of lineages from different refugia (Petit et al., 2003).

In other temperate regions of the world, phylogeographical patterns were structured by very different geographies and glacial histories. In eastern North America (ENA), early studies tended to emphasize genetic breaks between populations separated by rivers and mountain ranges (Jaramillo-Correa, Beaulieu, Khasa, & Bosquet, 2009; Soltis et al., 2006). In western North America, major refugia existed in the Pacific Northwest and Beringia, with smaller refugia on offshore islands and between continental ice sheets (Shafer et al., 2010). In East Asia, responses to glaciation included not only latitudinal migration, but also elevational and longitudinal migration and in situ persistence (Qiu et al., 2011). The complexity of these classic paradigms has recently been expanded in all four Northern Hemisphere regions to include small, low-density cryptic refugia in areas previously thought unsuitable for habitation by temperate species (McLachlan, Clark, & Manos, 2005; Provan & Bennett, 2008; Qiu et al., 2011; Stewart & Lister, 2001; Willis & Van Andel, 2004). However, the molecular and fossil evidence supporting the existence of cryptic refugia is not universally accepted (Tzedakis, Emerson, & Hewitt, 2013).

Compared to the other three Northern Hemisphere temperate forest regions, ENA is phylogeographically unique for at least three reasons. First, its geography is relatively simple, characterized by a contiguous land mass with only a single north-south mountain range of modest height (i.e. the Appalachians), and generally gradual transitions between ecosystem types. Second, latitudinal temperature gradients during the Last Glacial Maximum (LGM, ca. 21.5 ka; Jackson et al., 2000) were particularly steep, with warm areas located in close proximity to glaciers (Tzedakis et al., 2013). Third, despite numerous phylogeographical studies, well-delineated glacial refugia generally shared by most species have not conclusively been identified. Proposed refugial locations include the Gulf Coast, the Atlantic Coast, Florida, Texas, the Ozark Plateau, the Lower Mississippi River Valley, the Appalachians and interior areas near ice sheets (Barnard-Kubow, Debban, & Galloway, 2015; Griffin & Barrett, 2004; Jaramillo-Correa et al., 2009; Magni, Ducousso, Caron, Petit, & Kremer, 2005; McCarthy & Mason-Gamer, 2016; Morris, Graham, Soltis, & Soltis, 2010; Peterson & Graves, 2016; Soltis et al., 2006), which together sum to nearly the entire unglaciated region of ENA. While some species may have survived in one or more of these distinct refugia, other temperate taxa were likely not restricted to distinct LGM refugia, but were widespread over vast areas of the southeastern United States (Bennett, 1985; Lumibao et al., 2017; Magni et al., 2005; McLachlan et al., 2005; Peterson & Graves, 2016).

The fossil record has provided valuable insight into vegetation dynamics in ENA since the LGM, but has not yet resulted in a definitive account of the phylogeographical history of temperate deciduous tree species from this region for several reasons. Macrofossils of temperate deciduous trees are known from the Lower Mississippi River Valley (e.g. Delcourt, Delcourt, Brister, & Lackey, 1980), but there are few other LGM macrofossil sites available from areas with climates likely to have supported these taxa (Jackson et al., 2000). Fossil pollen of temperate deciduous trees is broadly distributed across southern ENA, but typically represents only a minor portion of the total pollen from most LGM assemblages (Jackson et al., 2000). Instead, coniferous species (e.g. *Picea*, *Pinus*) dominate most pollen assemblages (Davis, 1983; Jackson et al., 2000), and plant communities with no modern analogue were likely geographically widespread (Jackson & Williams, 2004; Jackson et al., 2000). The lack of clearly identifiable temperate deciduous forest communities in the LGM fossil record suggests that localized glacial refugia may not have existed in ENA for these communities as a whole, or if they did exist, they may not be represented in the fossil record. Nonetheless, temperate deciduous species were evidently present in some conifer-dominated communities (Jackson et al., 2000), raising questions about their geographical ranges, population sizes, genetic connectivity among populations, and which populations contributed most to postglacial recolonization.

Given the diversity of phylogeographical hypotheses that have been evoked to explain genetic patterns in ENA taxa, studies assessing genome-wide patterns of genetic variation in widely distributed model species would provide valuable insight into the history of temperate deciduous trees from this region. Surprisingly, we are aware of no such studies, although Eckert et al. (2010) and Nadeau et al. (2015) have conducted genome-wide studies of more narrowly distributed conifers, and many non–genome-wide studies of temperate deciduous trees exist (e.g. McLachlan et al., 2005). Here, we use genome-wide genetic variation to examine the phylogeographical history of two widespread, ENA tree species: bitternut hickory (*Carya cordiformis* [Wagenh.] K.Koch) and shagbark hickory (*Carya ovata* [Mill.] K.Koch). We construct and analyse single-nucleotide polymorphism (SNP) datasets from nearly range-wide collections of each species to characterize geographical patterns of genetic diversity and differentiation across ENA, and build palaeodistribution models to infer areas of high habitat suitability during the LGM. In particular, we aim to determine if genetic structure is best explained by recolonization from genetically isolated and geographically distinct refugia (and if so, where these refugia were located), or by expansion

from a much larger region that was not strongly genetically subdivided.

2 | MATERIALS AND METHODS

2.1 | Study species

Carya cordiformis and *Carya ovata* are wind-pollinated, animal-dispersed trees co-distributed from southern Quebec to eastern Texas (Figure 1). Their ranges roughly correspond to the overall geographical distribution of temperate deciduous forests in ENA. *Carya ovata* additionally occurs in several small, disjunct populations in the Sierra Madre Oriental of northern Mexico (Little, 1971). *C. cordiformis* occupies many habitats but occurs most frequently on mesic soils and bottomlands (Smith, 1990). *C. ovata* is common on a wider variety of sites, but is most frequent on drier uplands (Graney, 1990).

Phylogeographical knowledge is completely lacking in *C. cordiformis*. In *C. ovata*, analysis of cpDNA haplotypes has revealed no clear pattern, as some haplotypes are widespread throughout the entire range and others are more spatially restricted, including in formerly glaciated areas (Lumibao et al., 2017). *Carya* pollen is not typically distinguished to the species level, but LGM-age pollen of *Carya* has been found at low density over large areas of southern ENA (Jackson et al., 2000; Prentice, Bartlein, & Webb, 1991). LGM-age *Carya* macrofossils have been found as far north as western Tennessee (35°N; Jackson et al., 2000), and trace amounts of pollen have been found even farther north in the central portion of the state (36°N; Liu, Andersen, Williams, & Jackson, 2013).

Many of the ca. 13 North American *Carya* species readily hybridize with one another (Fralish & Franklin, 2002). Geographically structured hybridization may impact phylogeographical inferences in tree species (Saeki, Dick, Barnes, & Murakami, 2011; Thomson, Dick, & Dayanandan, 2015), and one limitation of our analyses is that we are unable to assess patterns of hybridization with other *Carya*. However, no stable hybrid zones exist in our species and we consider it unlikely that occasional hybridization would systematically bias genetic structure in a similar way across thousands of loci.

2.2 | DNA sampling and SNP genotyping

Silica-dried leaf tissue was collected from 182 individuals of each species from populations across ENA (Figure 1; Tables S1.1–S1.2). Sampled individuals within each population were separated by a minimum of 50 m (but sometimes up to dozens of km) to minimize the chance of sampling siblings and other close relatives. Sample size varied greatly among populations depending on the number of individuals meeting these requirements that could be located (mean $N = 7.8$; Tables S1.1–S1.2). A representative voucher specimen from each population was deposited in the University of Michigan Herbarium (MICH; collector numbers JBB 79-164).

DNA samples were extracted using Nucleospin Plant II extraction kits (*Macherey-Nagel*; Düren, Germany), and libraries were prepared using a modified double digest Restriction Associated DNA (ddRAD) sequencing protocol following Peterson, Weber, Kay, Fisher, and Hoekstra (2012), with restriction enzymes *EcoRI* and *MseI*. Full details of extraction methods and library preparation are provided in Appendix S1 in Supporting Information. Seven libraries of 72 samples each were sequenced at The Hospital for Sick Children (Toronto, ON) on an *Illumina HiSeq* (Illumina; San Diego, CA) using single-end 50-bp sequencing. To ensure adequate depth of coverage, at least one million raw reads per sample were required to process a sample, and individuals not meeting this target were resequenced in subsequent libraries.

Loci were identified and single nucleotide polymorphisms (SNPs) were genotyped using STACKS 1.44-1.46 (Catchen, Amores, Hohenlohe, Cresko, & Postlethwait, 2011; Catchen, Hohenlohe, Bassham, Amores, & Cresko, 2013). Full details of SNP discovery are provided in Appendix S1. After SNPs were successfully identified, one SNP

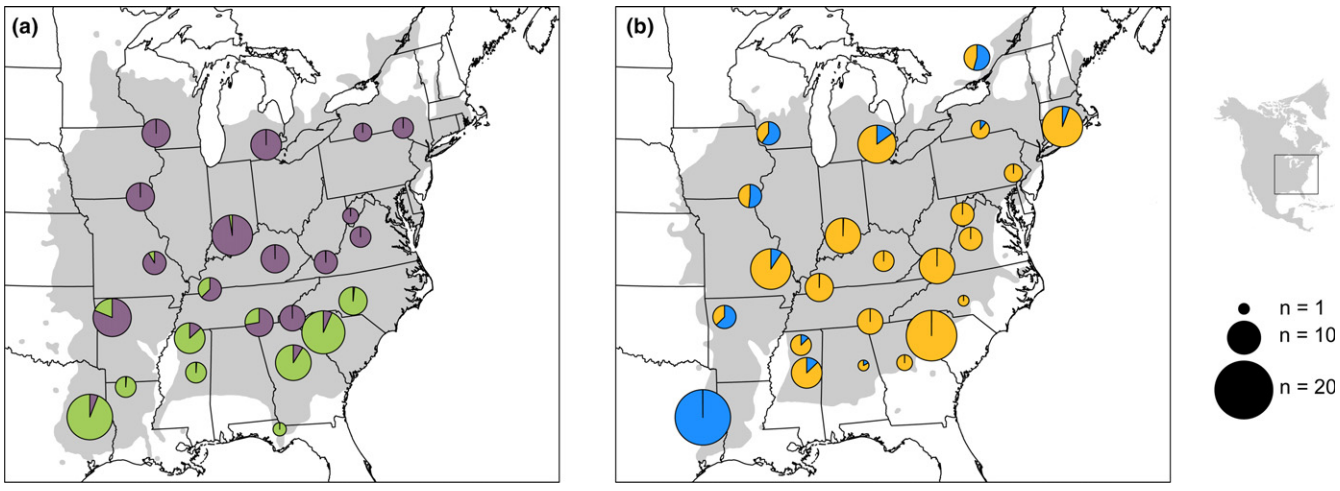

FIGURE 1 Membership of (a) *Carya cordiformis* and (b) *Carya ovata* populations in genetic clusters (different colours on pie charts; $K = 2$) identified using FASTSTRUCTURE (Raj et al., 2014). The geographical distribution of each species is shown in grey (Little, 1971), and the sample size (n) of each population is proportional to the size of the pie chart. Note that optimal $K = 1$–2, but $K = 1$ is not shown as all individuals of each species would belong to the same genetic cluster

genotype per locus was exported from STACKS using the *populations* tool, retaining only SNPs with a minimum genotyping rate of 75% (−r 0.75) and a minimum minor allele frequency (MAF) of 3.3% (−min_maf 0.033), the lowest detectable MAF in at least one population of each species, following Massatti and Knowles (2014). Minimum MAF is an important parameter to consider because it can impact inference of genetic structure (De la Cruz & Raska, 2014). We therefore explored preliminary principal component (PC) analyses (see below) with minimum MAF = 1% and 5%, but found that using the higher minimum MAF (5%) made little qualitative difference in preliminary results. With the lower minimum MAF (1%), small clusters of a few individuals that were outliers along PC axes appeared in *C. cordiformis*. This pattern may arise if the frequency of rare alleles is very similar among closely related individuals that share recent ancestry reflecting local-scale processes (De la Cruz & Raska, 2014). In contrast, we are interested in longer term processes reflecting differences among populations, which were likely better captured with the moderate MAF of 3.3%.

To retain only putatively nuclear SNPs, we removed any SNPs from loci that aligned with a maximum of two mismatches (−v 2) to the *Juglans regia* (Juglandaceae) chloroplast genome (Genbank accession NC_028617.1) or the *Cucurbita pepo* (Cucurbitaceae) mitochondrion genome (NC_014050.1) using BOWTIE 1.2 (Langmead, Trapnell, Pop, & Salzberg, 2009). Extremely variable loci were also excluded as these may represent locus assembly errors; we defined these loci as those with values of θ (Watterson, 1975) above the 95th percentile, with θ calculated for each locus individually using the *R* package "pegas" 0.10 (Paradis, 2010). Individual samples with unusually high levels of missing data across all loci (based on visual inspection) were also excluded.

2.3 | Genetic diversity and divergence

Three genetic diversity parameters were calculated overall and for each population: observed and expected heterozygosity (H_o and H_e respectively), and nucleotide diversity (π). Genetic differentiation (F_{ST}; Nei, 1987) was calculated overall and pairwise between each pair of populations. H_o, H_e and F_{ST} were calculated in the *R* package "hierfstat" 0.04-22 (Goudet, 2005), whereas π was calculated using *populations* in STACKS 1.46 (Catchen et al., 2011, 2013). Genetic diversity and differentiation measures are not reported for populations represented by a single individual.

2.4 | Population genetic structure

To test for isolation by distance (IBD), Mantel tests (Mantel, 1967) were performed to assess the relationship between population pairwise F_{ST} values and geographical distances. Principal component analysis (PCA) was used to investigate genetic relationships among individuals and populations using the *dudi.pca* function in the *R* package "adegenet" 2.0.1 (Jombart, 2008; Jombart & Ahmed, 2011). The NC_e population (Table S1.1) was excluded for *C. cordiformis* because some, but not all individuals from this population formed a distinct genetic cluster, which suggests that several closely related individuals were unintentionally sampled and the high genetic similarity between these individuals could have biased initial PCA results.

Genetic clusters were characterized using FASTSTRUCTURE 1.0 (Raj, Stephens, & Pritchard, 2014), with all populations and individuals included, using the recommended procedure for detecting subtle genetic structure. Initially, the simple prior model was used and the number of clusters (*K*) was varied from 1 to 6 for each species, and *K* was selected using the *chooseK* tool in FASTSTRUCTURE. Then, FASTSTRUCTURE was rerun 100 times using the logistic prior model for the optimal value(s) of *K*, and final estimates of genetic membership of individuals in each genetic cluster were obtained as the average membership from the five runs with the highest likelihood, following Raj et al. (2014). After investigating the broadest level of structure within each dataset, we reran FASTSTRUCTURE on individual genetic clusters to test for substructure within clusters.

2.5 | Palaeodistribution modelling

Species distribution models (SDMs) were constructed to predict the potential distribution of each species during the Last Glacial Maximum (LGM; 21.5 ka). Complete details of SDM construction and data sources are given in Appendix S1. Briefly, occurrence records were obtained from the US Forest Service Forest Inventory Analysis Database (O'Connell et al., 2012), whereas environmental variables were obtained at 2.5-arcminute resolution from the WORLDCLIM 1.4. (Hijmans, Cameron, Parra, Jones, & Jarvis, 2005) and ENVIREM (Title & Bemmels, 2018) databases. SDMs were constructed using MAXENT 3.4.1 (Phillips, Anderson, Dudík, Schapire, & Blair, 2017; Phillips, Anderson, & Schapire, 2006; Phillips, Dudík, & Schapire, 2004) in the *R* package "dismo" (Hijmans, Phillips, Leathwick, & Elith, 2015), with models optimized according to best practices, following Title and Bemmels (2018). Habitat suitability was projected for the LGM according to each of the CCSM4, MIROC-ESM and MPI-ESM-P general circulation models (GCMs), but since projections were similar for all three GCMs, results were averaged into a single map.

3 | RESULTS

3.1 | Genetic diversity and differentiation

The final genetic datasets for *C. cordiformis* and *C. ovata* contained 177 individuals genotyped at 1,046 SNPs, and 180 individuals genotyped at 1,018 SNPs respectively. The overall genotyping rate for both species was 89%.

While some populations were represented by very few individuals (Tables S1.1–S1.2), very small sample sizes are typically sufficient to obtain accurate population genomic measures of genetic diversity and differentiation if calculated across thousands of SNPs (Nazareno, Bemmels, Dick, & Lohmann, 2017; Willing, Dreyer, & van Oosterhout, 2012). For this reason, we did not perform rarefaction in our analyses to match the lowest population sample size. As further

empirical justification for this decision, we found that genetic diversity estimates were uncorrelated with population sample size (Figure S1.1), except for a negative relationship between H_o and sample size in *C. ovata* (R^2 = .21, p = 0.035). However, as H_o is computed on a per-individual basis, there is no theoretical reason to explain how sample size could affect estimates of H_o and we suspect that this correlation is spurious.

Genetic diversity showed little variation among populations for both species (Figure 2). A significant decline in genetic diversity with increasing latitude was not observed for any genetic diversity measure (H_o, H_e, π) for either species. Instead, a significant increase in H_o with increasing latitude was observed in *C. cordiformis* (R^2 = .23, p = 0.021), as was a marginally nonsignificant increase in H_o with increasing latitude in *C. ovata* (R^2 = .19, p = 0.051). Although genetic variation was fairly uniform across latitudes, far northern and far southern populations sometimes showed slightly lower values of H_e and π than typical of mid-latitude populations (Figure 2), as expected for range-edge populations (Jaramillo-Correa et al., 2009). Among-population genetic differentiation (F_{ST}) is low in both species overall (*C. cordiformis*: 0.047; *C. ovata*: 0.038), and among most pairs of populations (Tables S1.3–S1.4).

3.2 | Spatial genetic structure

Spatial genetic structure in both species is weak and dominated by a pattern of IBD. Mantel tests of IBD were statistically significant in both species (*C. cordiformis*: r = .36, p = 0.0017; *C. ovata*: r = .47, $p = 8.1 \times 10^{-5}$; Figure 3).

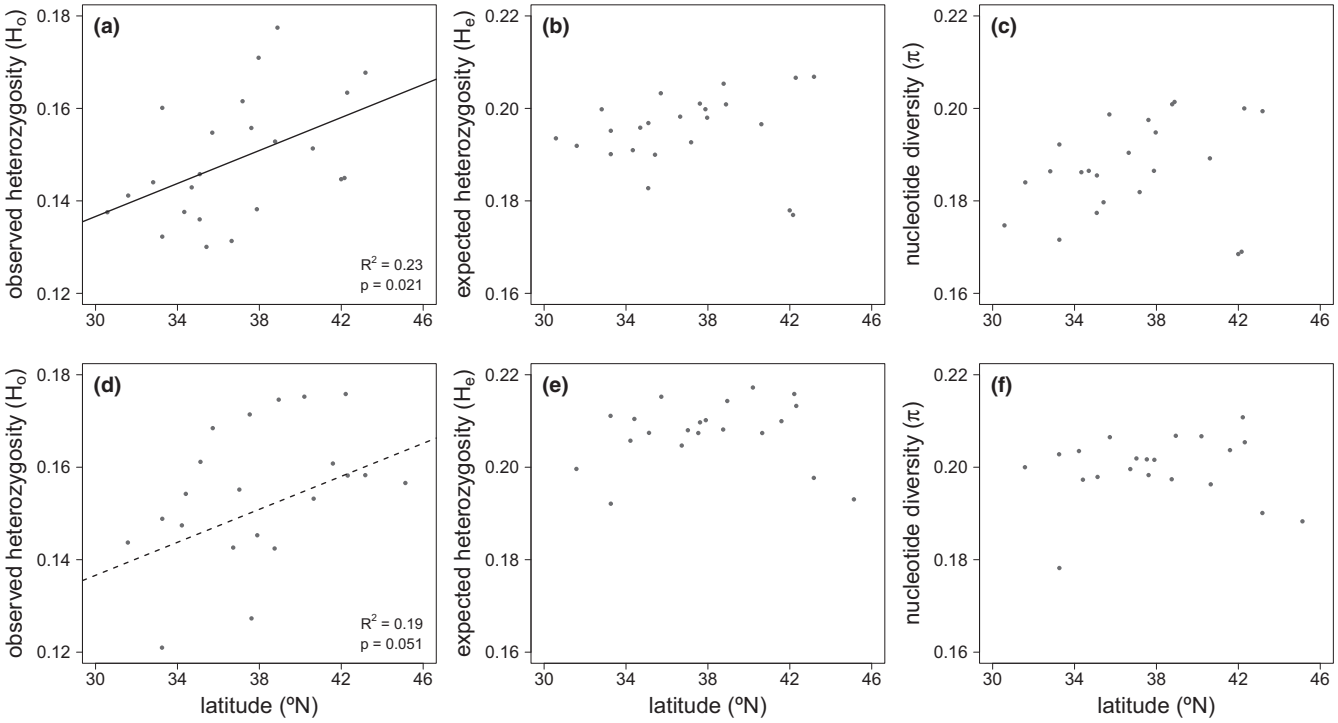

FIGURE 2 Genetic diversity vs. latitude for populations of *Carya cordiformis* (a–c) and *Carya ovata* (d–f). Statistically significant relationships are portrayed as solid lines, and marginally significant relationships as dashed lines

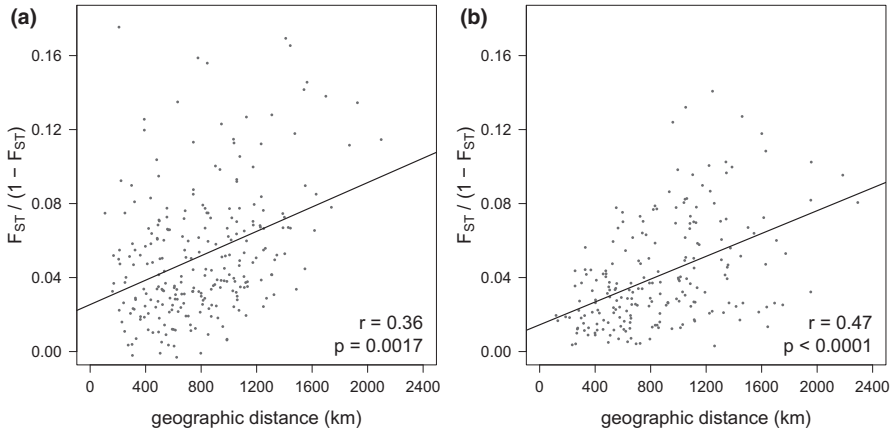

FIGURE 3 Mantel test showing isolation by distance in (a) *Carya cordiformis* and (b) *Carya ovata*. Each dot represents a pair of populations and the *y*-axis is a measure of genetic differentiation between populations

Principal component analysis of genetic variation also revealed an IBD-like pattern, with gradual transitions and substantial overlap among geographical regions, and without clearly defined, distinct genetic clusters over most of the species range (Figure 4). One clear exception to this pattern is that in *C. ovata*, the Texas population (TX; Figure 4b) forms a separate cluster that does not overlap with any other populations. In *C. cordiformis*, PC1 represents a north-south geographical transition, whereas PC2 does not appear strongly related to geography. In *C. ovata*, PC1 distinguishes TX from all other populations, whereas PC2 represent a north-south transition.

Lack of strong genetic structure was also suggested by FASTSTRUCTURE results. Under the model with simple priors, the optimal number of genetic clusters was $K = 1$ for both species. Under the model with logistic priors, which is more useful for detecting subtle structure (Raj et al., 2014), optimal K ranged from 1 to 6. However, the logistic priors model is prone to overfitting (Raj et al., 2014) and $K > 2$ did not produce results that were biologically interpretable. We therefore note that $K = 1$ or 2 is likely the optimal model complexity to explain genetic structure. In both species with $K = 2$, genetic variation was geographically structured but with transition zones between clusters (Figures 5, S1.2). In *C. cordiformis*, the transition was from north to south, whereas in *C. ovata*, the transition was primarily from east to west. No substructure was evident within any genetic cluster for any species, except that within the western cluster for *C. ovata*, optimal $K = 2$ and the Texas population (TX) forms a distinct subcluster relative to the other four populations (AR, IA, ON, WI; data not shown). However, the grouping of these four populations into a distinct subcluster may be only a statistical artefact reflecting the substantial additional membership of each these four western populations in the main eastern cluster.

3.3 | Palaeodistribution modelling

The same four climatic variables were coincidentally retained in the SDMs for both species: maximum temperature of the coldest month, potential evapotranspiration of the warmest quarter, mean annual precipitation and climatic moisture index (Table S1.5). For both species, models were able to predict the current species distribution (Figure 1) very well along the northern and western range edges, but performed more poorly at delineating the southern range edge (Figure 5). This poorer performance may reflect the fact that both species are relatively rare in the southern portion of their ranges, where presence and absence may be determined by local soil type and topography (Graney, 1990; Smith, 1990). For both species, a large, continuous area of high LGM habitat suitability is predicted to have extended over much of the southeastern US, from central Texas to the coast of North Carolina (Figure 5).

4 | DISCUSSION

Carya cordiformis and *Carya ovata* likely survived the LGM over a broad geographical area covering much of southern ENA. This scenario is supported by our genetic results showing weak genetic structure and an IBD pattern, and is compatible with predictions of our palaeodistribution models and the fossil record. Genetic differentiation is weakly geographically structured, without sharp phylogeographical breaks over most of ENA. However, a Texas population (TX) of *C. ovata* is genetically distinct compared to other populations and may be derived from a separate, genetically isolated glacial refugium. We find no evidence of any further subdivision into separate

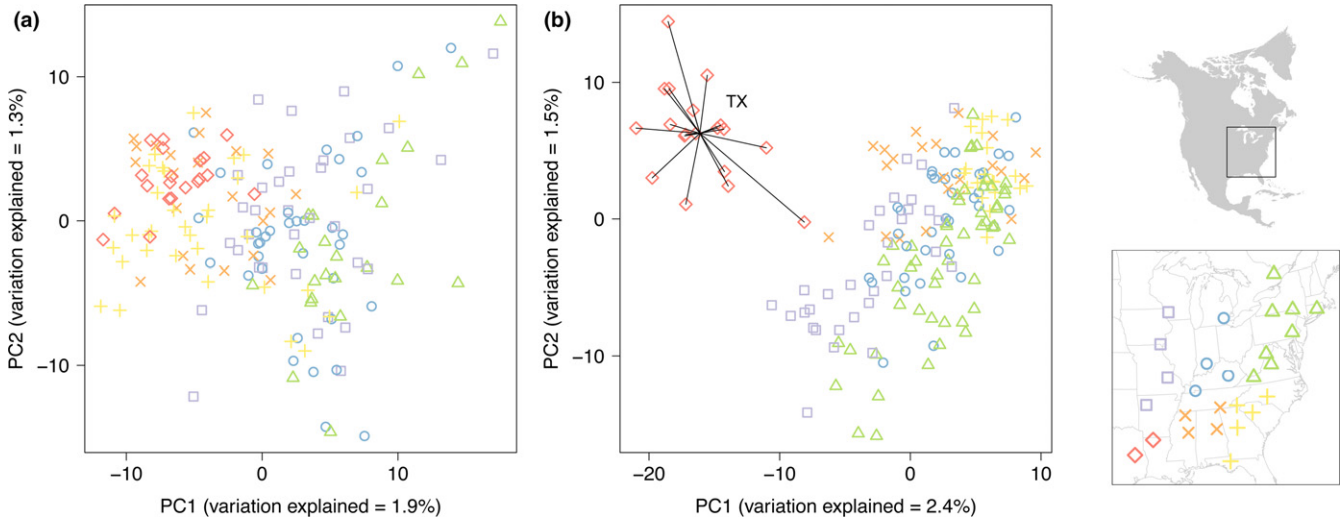

FIGURE 4 Clustering of individuals along the first and second principal component (PC) axes of genetic variation in (a) *Carya cordiformis* and (b) *Carya ovata*. Each symbol represents a single individual, and symbols and colours correspond to the geographical location of the individual (shown in the inset map of eastern North America; geographical regions are delimited here for visualization purposes only). Individuals from the Texas population (TX) of *C. ovata* discussed in the manuscript are connected by black lines in (b). Note that sample size varies greatly for each population, and some populations were not sampled for both species (see Figure 1)

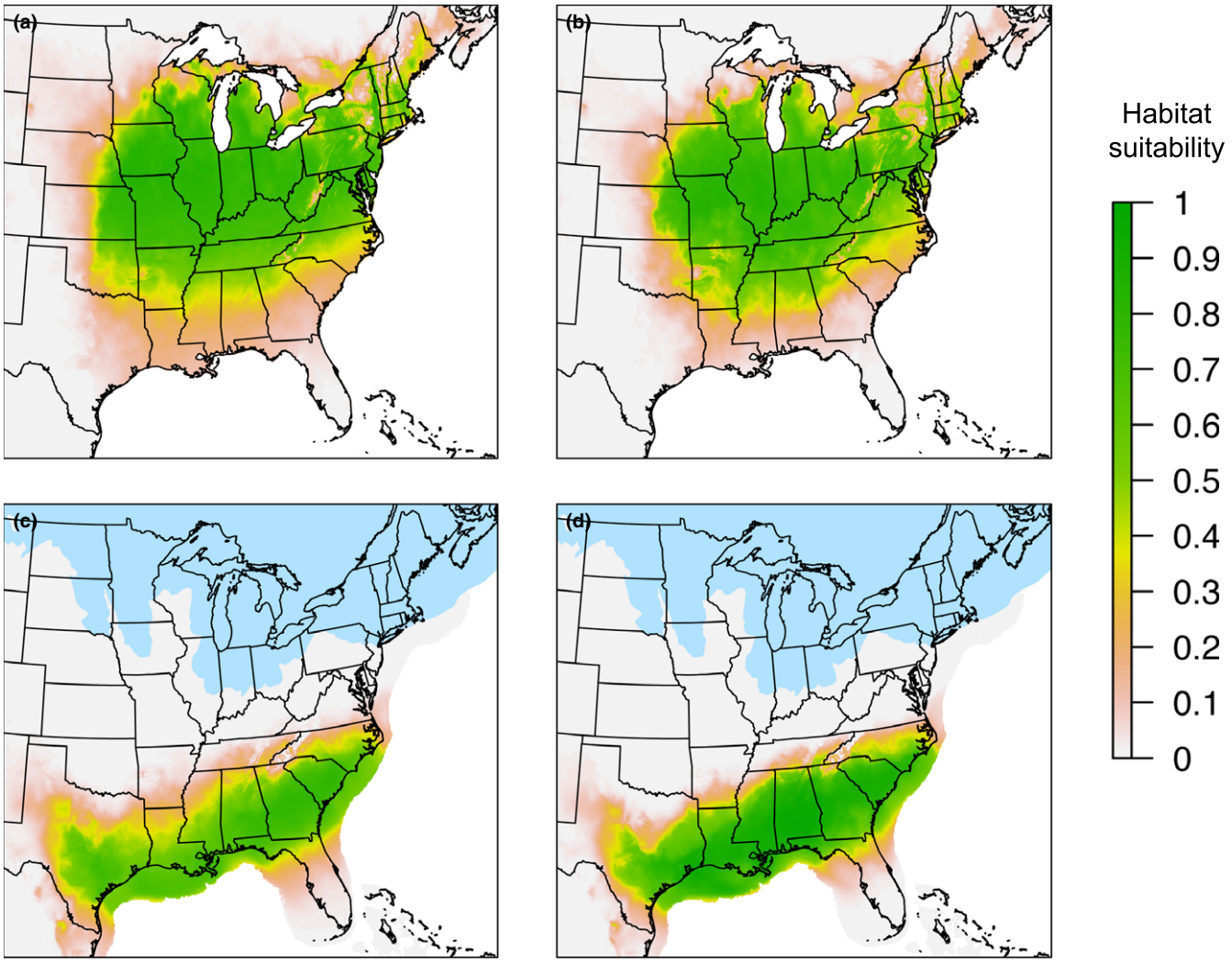

FIGURE 5 Species distribution models showing predicted habitat suitability (colour scale: grey, habitat not suitable; green, maximum habitat suitability) for (a, c) *Carya cordiformis* and (b, d) *Carya ovata*, for climates of both (a, b) the current time period and (c, d) the Last Glacial Maximum (ca. 21.5 ka). Glaciated areas are shown in blue

refugia in either species. Although the geographical distribution of both species was likely quite broad during the LGM, the fossil record indicates that most temperate deciduous trees occurred at low density in communities with no modern analogues (Jackson & Williams, 2004; Jackson et al., 2000; Prentice et al., 1991). Further research into the phylogeographical history of other widespread temperate deciduous trees from ENA is needed, but our results suggest that both hickory species may not have experienced the complex refugial dynamics that structured genetic variation in other Northern Hemisphere temperate forest regions.

4.1 | A widespread distribution during the LGM

Weak genetic structure, lack of strong phylogeographical breaks across most of ENA, and lack of distinct areas of elevated genetic diversity suggest that both species were fairly widely distributed throughout southern ENA during the LGM. Nonetheless, these patterns might also be observed if high gene flow has homogenized populations and eroded historical demographical signatures (e.g. He, Edwards, & Knowles, 2013). Low population genetic differentiation (*C. cordiformis*, F_{ST} = 0.047; *C. ovata*, F_{ST} = 0.038) suggests that gene flow among populations may be quite high, which is typical of widespread, wind-pollinated forest trees due to their large population sizes and capacity for long-distance pollen-mediated gene flow (Alberto et al., 2013; Hamrick, Godt, & Sherman-Broyles, 1992; Savolainen, Pyhäjärvi, & Knürr, 2007). However, the amount of time that has passed since the LGM may have been insufficient for gene flow to completely erode genetic signatures of expansion from a geographically restricted or fragmented LGM distribution. Both species are slow growing and long lived, with peak reproduction occurring in *C. cordiformis* from ages 50 to 125 (Smith, 1990), and in *C. ovata* from ages 60 to 200 (Graney, 1990). Generation times are more than an order of magnitude shorter in species such as small vertebrates and herbaceous plants, yet these species frequently retain phylogeographical structure interpreted to reflect the effects of glaciation (Soltis et al., 2006).

On the other hand, long generation times and large effective population sizes of nuclear DNA (relative to cpDNA or mtDNA) might mean that if periods of range fragmentation and refugial isolation were brief, strongly genetically differentiated populations might not have had sufficient time to develop. Thus, while we hypothesize that a widespread, genetically connected distribution was maintained throughout southern ENA during the LGM, the existence of separate, briefly isolated refugia that have since merged into a more genetically homogenous distribution is also potentially compatible with our results.

While genetic diversity of European taxa was often lost during northward expansion from southern refugia, this pattern is not observed in most temperate tree species from ENA (Lumibao et al., 2017). The relatively uniform levels of population genetic diversity (H_e, π) across the species range (Figure 2) suggest that historical recolonization occurred over a large, slowly expanding region with little loss of diversity during migration (Jaramillo-Correa et al., 2009). However, high gene flow could also have reduced differences in genetic diversity among populations. Whereas H_e and π are relatively uniform across populations, observed heterozygosity (H_o) increases with increasing latitude in both species (Figure 2). Both species are less common in southern areas than in the north (Figure 5; Graney, 1990; Smith, 1990), and it is possible that southern populations could generally be smaller, more isolated, and more prone to reductions in H_o due to inbreeding than northern populations.

Predictions of palaeodistribution models also suggest that climatic conditions were favourable for survival of both species over a broad geographical area (Figure 5). However, SDMs should be interpreted with caution, especially when projecting models to non-analogue climates in novel time periods or geographical areas (Owens et al., 2013). Model projections for the LGM also depend on simulations from general circulation models, which may not accurately reflect true palaeoclimatic conditions (Varela, Lima-Ribeiro, & Terribile, 2015). Despite these concerns, the fossil pollen record is largely compatible with our palaeodistribution models, as low amounts of *Carya* pollen have been found in LGM assemblages across much of the southeastern United States (Jackson et al., 2000; Prentice et al., 1991). Furthermore, major increases in *Carya* pollen have been observed from ca. 16–13 ka in sites as geographically distant as the Missouri Ozarks (Jones, Williams, & Jackson, 2017), central Tennessee (Liu et al., 2013), and South Carolina (Watts, 1980). Although long-distance migration to these sites is possible, expansion from nearby sources is also plausible and would hint that *Carya* may have been broadly longitudinally distributed prior to postglacial expansion.

4.2 | Glacial refugia and sources of postglacial recolonization

We find no evidence of genetically isolated refugia in *C. cordiformis*, but a separate glacial refugium in Texas is strongly suggested by our PCA results for *C. ovata* (Figure 4b). Glacial refugia have previously been inferred in Texas and northern Mexico for several southern *Pinus* and *Prunus* species (Eckert et al., 2010; Schmidtling, 2003; Schmidtling & Hipkins, 1998; Shaw & Small, 2005). However, a separate refugium is not the only possible explanation for the genetic distinctiveness of TX, as we speculate that the ancestors of this population might have historically experienced gene flow with *C. ovata* populations from the mountains of northern Mexico (Little, 1971). Although we have not included any high-elevation Mexican populations in our genetic analyses or palaeodistribution models, we cannot exclude the possibility that these populations may have migrated to lower elevations and come into contact with TX during the LGM. Alternatively, because TX is a range-edge population (Figure 1b), it is possible that genetic drift due to small population size and limited gene flow with other populations could have caused substantial changes in allele frequency in this population. Denser population sampling across the southwestern portion of the species range and of Mexican populations could help distinguish among these scenarios.

The east-west population structuring in *C. ovata* (Figure 1b) detected in fastSTRUCTURE also supports our inference that TX likely represents a separate glacial refugium. Some northern populations show admixture between the western and eastern clusters, but most populations across the species range show substantial membership in the eastern genetic cluster, suggesting that the contribution of the western (TX) lineage to postglacial recolonization was relatively minor. In contrast, the northern and southern genetic clusters in *C. cordiformis* are broadly longitudinally distributed (Figure 1a) and likely do not represent distinct refugia. This pattern suggests that northern LGM populations may have made the greatest contribution to postglacial recolonization in *C. cordiformis*, as has previously been inferred for *Acer rubrum* and *Fagus grandifolia* (McLachlan et al., 2005). Alternatively, recolonization may have occurred from southern areas, but northern areas might have experienced subtle shifts in allele frequency due to genetic drift during northward migration.

While our fastSTRUCTURE results present intriguing hypotheses, IBD is known to bias tests of hierarchical genetic structure (Frantz, Cellina, Krier, Schley, & Burke, 2009; Meirmans, 2012). In particular, such tests are susceptible to incorrect inference of multiple genetic clusters when populations are geographically subsampled from within a single larger cluster subject to IBD (Frantz et al., 2009; Meirmans, 2012). A range-wide IBD pattern is indeed present in our datasets (Figure 3). Rather than indicating the true presence of biologically meaningful genetic clusters, our fastSTRUCTURE results might be a statistical artefact of underlying IBD, especially in *C. cordiformis* for which no sharp genetic breaks were detected (Figure 4a).

4.3 | Implications and future directions

Despite decades of phylogeographical study (Soltis et al., 2006) and synthesis of the fossil record (Davis, 1983; Jackson et al., 2000), a general phylogeographical history of ENA temperate forest taxa remains elusive. Given that *C. cordiformis* and *C. ovata* are common, widespread tree species with a geographical distribution roughly matching that of modern temperate deciduous forests in ENA,

insights from these species may be relevant to predicting the phylogeographical histories of other geographically widespread temperate deciduous trees. While different species likely responded individually to Pleistocene glaciation, our results suggest that at least some temperate tree species were fairly widespread throughout southern ENA during the LGM. One of the most striking findings of our analyses is that both species lack the strong phylogeographical breaks characteristic of taxa from other temperate regions of the world (Hewitt, 2000; Lumibao et al., 2017; Qiu et al., 2011; Shafer et al., 2010), except for the Texas population of *C. ovata*. That both species show similar phylogeographical patterns over most of ENA is somewhat surprising, given substantial differences in their ecology. *Carya ovata* is a habitat generalist (Graney, 1990), but *C. cordiformis* is primarily a mesic, bottomland species (Smith, 1990). Low genetic structure in *C. cordiformis* provides strong evidence that even temperate tree species with specific habitat requirements were not necessarily highly genetically fragmented during the LGM.

Although both species were likely fairly broadly distributed during the LGM, the fossil pollen record suggests that most LGM forest communities across southern ENA were conifer-dominated with no modern analogues (Jackson & Williams, 2004; Jackson et al., 2000). Despite uncertainty about the LGM extent, or even the existence, of the temperate deciduous forest biome (Prentice et al., 2000), several other widespread temperate deciduous tree and plant species are also believed to have expanded from a large area covering much of southern ENA (Bennett, 1985; Magni et al., 2005; McLachlan et al., 2005; Peterson & Graves, 2016). In general, these results support the emerging understanding (Lumibao et al., 2017) that the genetic consequences of Pleistocene glaciation on widespread ENA temperate tree species were very different from those produced by expansion from distinct refugia in Europe (Hewitt, 1999, 2000; Petit et al., 2003).

However, many other temperate plant species experienced more complex phylogeographical histories (e.g. Barnard-Kubow et al., 2015; Eckert et al., 2010; Gonzales, Hamrick, & Chang, 2008; Griffin & Barrett, 2004; Nadeau et al., 2015; Zinck & Rajora, 2016). Species adapted to a narrower range of climatic conditions than *C. cordiformis* and *C. ovata* could have been more likely to experience geographical fragmentation during the LGM. We hypothesize that taxa with a strictly southern, warm-temperate distribution may have become fragmented into distinct far-southern refugia in Florida, Texas, or along the Gulf or Atlantic Coasts (e.g. Eckert et al., 2010; Gonzales et al., 2008). In contrast, cool-temperate and more climatically widespread species such as *C. cordiformis* and *C. ovata* could have survived in more expansive inland areas where cool-temperate conditions likely extended farther north (e.g. Figure 5; see also McLachlan et al., 2005).

Genetically distinct southern populations of temperate species, such as the Texas population of *C. ovata*, have often been identified as high conservation priority (Hampe & Petit, 2005; Médail & Diadema, 2009; Petit et al., 2003). We suggest that any conservation efforts in *C. ovata* should ensure inclusion of populations from Texas and surrounding regions. However, across most of the range of either species, we see little reason to prioritize conservation of southern populations, because these populations are neither genetically distinct nor do they exhibit elevated genetic diversity. On the other hand, most temperate tree species exhibit geographically structured climatically adaptive genetic variation (Aitken & Bemmels, 2016; Savolainen et al., 2007) unlikely to be captured by the putatively neutral SNP markers we employed. Conserving populations from a variety of climates across the species range would therefore likely maximize conservation of adaptively relevant genetic diversity.

To our knowledge, this study is the first to investigate genome-wide genetic variation in a geographically widespread temperate deciduous tree from ENA. Much future work remains to be done to test whether the patterns we observe are generally applicable to other ENA trees with diverse traits and geographical distributions. In addition, application of demographical and coalescent modelling techniques would greatly enhance our understanding of the phylogeographical histories of ENA taxa. Such techniques would allow statistical tests of hypotheses regarding levels of genetic connectivity among populations, changes in population size over time (Barthe et al., 2017), locations of source populations from which postglacial recolonization occurred (He, Prado, & Knowles, 2017), and which ecological factors have most strongly impacted historical migration patterns (Bemmels, Title, Ortego, & Knowles, 2016).

ACKNOWLEDGEMENTS

The authors thank B.J. Belcher, P. Cousineau, R. D'Andrea, J. Fiske, J. Ronson, D. Saenz and P. Tichenor for assistance with fieldwork, and A.T. Thomaz for assistance with genetic analyses. Graduate student support and research funding was provided to J.B.B. by the National Science Foundation (GRFP fellowship; DDIG 1501159), and the University of Michigan EEB Department and Rackham Graduate School.

DATA ACCESSIBILITY

SNP datasets and species occurrence records are available for download from the Deep Blue Data (DOI: 10.7302/Z2JS9NNG).

ORCID

Jordan B. Bemmels http://orcid.org/0000-0001-9996-6996
Christopher W. Dick http://orcid.org/0000-0001-8745-9137

REFERENCES

Aitken, S. N., & Bemmels, J. B. (2016). Time to get moving: Assisted gene flow of forest trees. *Evolutionary Applications*, *9*, 271–290. https://doi.org/10.1111/eva.12293

Alberto, F. J., Aitken, S. N., Alía, R., González-Martínez, S. C., Hänninen, H., Kremer, A., … Savolainen, O. (2013). Potential for evolutionary

responses to climate change—Evidence from tree populations. *Global Change Biology*, *19*, 1645–1661. https://doi.org/10.1111/gcb.12181

Barnard-Kubow, K. B., Debban, C. L., & Galloway, L. F. (2015). Multiple glacial refugia lead to genetic structuring and the potential for reproductive isolation in a herbaceous plant. *American Journal of Botany*, *102*, 1842–1853. https://doi.org/10.3732/ajb.1500267

Barthe, S., Binelli, G., Hérault, B., Scotti-Saintagne, C., Sabatier, D., & Scotti, I. (2017). Tropical rainforests that persisted: Inferences from the Quaternary demographic history of eight tree species in the Guiana shield. *Molecular Ecology*, *26*, 1161–1174. https://doi.org/10.1111/mec.13949

Bemmels, J. B., Title, P. O., Ortego, J., & Knowles, L. L. (2016). Tests of species-specific models reveal the importance of drought in post-glacial range shifts of a Mediterranean-climate tree: Insights from integrative distributional, demographic and coalescent modelling and ABC model selection. *Molecular Ecology*, *25*, 4889–4906. https://doi.org/10.1111/mec.13804

Bennett, K. D. (1985). The spread of *Fagus grandifolia* across eastern North America during the last 18 000 years. *Journal of Biogeography*, *12*, 147–164. https://doi.org/10.2307/2844838

Bennett, K. D., & Provan, J. (2008). What do we mean by 'refugia'? *Quaternary Science Reviews*, *27*, 2449–2455. https://doi.org/10.1016/j.quascirev.2008.08.019

Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., & Postlethwait, J. H. (2011). STACKS: Building and genotyping loci de novo from short-read sequences. *G3: Genes, Genomes, Genetics*, *1*, 171–182. https://doi.org/10.1534/g3.111.000240

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). STACKS: An analysis tool set for population genomics. *Molecular Ecology*, *22*, 3124–3140. https://doi.org/10.1111/mec.12354

Davis, Margaret B. (1983). Quaternary history of deciduous forests of eastern North America and Europe. *Annals of the Missouri Botanical Garden*, *70*, 550–563. https://doi.org/10.2307/2992086

De la Cruz, O., & Raska, P. (2014). Population structure at different minor allele frequency levels. *BMC proceedings*, *8*, S55. https://doi.org/10.1186/1753-6561-8-S1-S55

Delcourt, P. A., Delcourt, H. R., Brister, R. C., & Lackey, L. E. (1980). Quaternary vegetation history of the Mississippi Embayment. *Quaternary Research*, *13*, 111–132. https://doi.org/10.1016/0033-5894(80)90086-1

Eckert, A. J., Van Heerwaarden, J., Wegrzyn, J. L., Nelson, C. D., Ross-Ibarra, J., González-Martínez, S. C., & Neale, D. B. (2010). Patterns of population structure and environmental associations to aridity across the range of loblolly pine (*Pinus taeda* L., Pinaceae). *Genetics*, *185*, 969–982. https://doi.org/10.1534/genetics.110.115543

Fralish, J. S., & Franklin, S. B. (2002). *Taxonomy and ecology of woody plants in North American forests*. New York: John Wiley & Sons.

Frantz, A. C., Cellina, S., Krier, A., Schley, L., & Burke, T. (2009). Using spatial Bayesian methods to determine the genetic structure of a continuously distributed population: Clusters or isolation by distance? *Journal of Applied Ecology*, *46*, 493–505. https://doi.org/10.1111/j.1365-2664.2008.01606.x

Gonzales, E., Hamrick, J. L., & Chang, S. M. (2008). Identification of glacial refugia in south-eastern North America by phylogeographical analyses of a forest understorey plant, *Trillium cuneatum*. *Journal of Biogeography*, *35*, 844–852. https://doi.org/10.1111/j.1365-2699.2007.01834.x

Goudet, J. (2005). HIERFSTAT, a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, *5*, 184–186. https://doi.org/10.1111/j.1471-8286.2004.00828.x

Graney, D. L. (1990). *Carya ovata* (Mill.) K. Koch. In R. M. Burns & B. H. Honkala (Eds), *Silvics of North America: 2. Hardwoods. Agricultural Handbook 654*. Washington, DC: US Department of Agriculture, Forest Service.

Griffin, S. R., & Barrett, S. C. H. (2004). Post-glacial history of *Trillium grandiflorum* (Melianthiaceae) in eastern North America: Inferences from phylogeography. *American Journal of Botany*, *91*, 465–473. https://doi.org/10.3732/ajb.91.3.465

Hampe, A., & Petit, R. J. (2005). Conserving biodiversity under climate change: The rear edge matters. *Ecology Letters*, *8*, 461–467. https://doi.org/10.1111/j.1461-0248.2005.00739.x

Hamrick, J. L., Godt, J. W., & Sherman-Broyles, S. L. (1992). Factors influencing levels of genetic diversity in woody plant species. *New Forests*, *6*, 95–124. https://doi.org/10.1007/BF00120641

He, Q., Edwards, D. L., & Knowles, L. L. (2013). Integrative testing of how environments from the past to the present shape genetic structure across landscapes. *Evolution*, *67*, 3386–3402. https://doi.org/10.1111/evo.12159

He, Q., Prado, J. R., & Knowles, L. L. (2017). Inferring the geographic origin of a range expansion: Latitudinal and longitudinal coordinates inferred from genomic data in an ABC framework with the program X-ORIGIN. *Molecular Ecology*, *26*, 6908–6920. https://doi.org/10.1111/mec.14380

Hewitt, G. (1999). Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society*, *68*, 87–112. https://doi.org/10.1111/j.1095-8312.1999.tb01160.x

Hewitt, G. (2000). The genetic legacy of the Quaternary ice ages. *Nature*, *405*, 907–913. https://doi.org/10.1038/35016000

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, *25*, 1965–1978. https://doi.org/10.1002/(ISSN)1097-0088

Hijmans, R. J., Phillips, S., Leathwick, J., & Elith, J. (2015). Package "dismo". R package. http://cran.r-project.org/web/packages/dismo/index.html.

Jackson, S. T., Webb, R. S., Anderson, K. H., Overpeck, J. T., Webb, T. III, Williams, J. W., & Hansen, B. C. S. (2000). Vegetation and environment in Eastern North America during the Last Glacial Maximum. *Quaternary Science Reviews*, *19*, 489–508. https://doi.org/10.1016/S0277-3791(99)00093-1

Jackson, S. T., & Williams, J. W. (2004). Modern analogs in Quaternary paleoecology: Here today, gone yesterday, gone tomorrow? *Annual Review of Earth and Planetary Sciences*, *32*, 495–537. https://doi.org/10.1146/annurev.earth.32.101802.120435

Jaramillo-Correa, J. P., Beaulieu, J., Khasa, D. P., & Bosquet, J. (2009). Inferring the past from the present phylogeography structure of North American forest trees: Seeing the forest for the genes. *Canadian Journal of Forest Research*, *39*, 286–307. https://doi.org/10.1139/X08-181

Jombart, T. (2008). ADEGENET: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*, 1403–1405. https://doi.org/10.1093/bioinformatics/btn129

Jombart, T., & Ahmed, I. (2011). ADEGENET 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics*, *27*, 3070–3071. https://doi.org/10.1093/bioinformatics/btr521

Jones, R. A., Williams, J. W., & Jackson, S. T. (2017). Vegetation history since the last glacial maximum in the Ozark highlands (USA): A new record from Cupola Pond, Missouri. *Quaternary Science Reviews*, *170*, 174–187. https://doi.org/10.1016/j.quascirev.2017.06.024

Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, *10*, R25. https://doi.org/10.1186/gb-2009-10-3-r25

Little, E. L. (1971). *Atlas of United States trees, volume 1. Conifers and important hardwoods*. Washington, DC: USDA Miscellaneous Publication 1146.

Liu, Y., Andersen, J. L., Williams, J. W., & Jackson, S. T. (2013). Vegetation history in central Kentucky and Tennessee (USA) during the last glacial and deglacial periods. *Quaternary Research*, *79*, 189–198. https://doi.org/10.1016/j.yqres.2012.12.005

Lumibao, C. Y., Hoban, S. M., & McLachlan, J. (2017). Ice ages leave genetic diversity "hotspots" in Europe but not in Eastern North America. *Ecology Letters*, *20*, 1459–1468. https://doi.org/10.1111/ele.12853

Magni, C. R., Ducousso, A., Caron, H., Petit, R. J., & Kremer, A. (2005). Chloroplast DNA variation of *Quercus rubra* L. in North America and comparison with other Fagaceae. *Molecular Ecology*, *14*, 513–524. https://doi.org/10.1111/j.1365-294X.2005.02400.x

Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research*, *27*, 209–220.

Massatti, R., & Knowles, L. L. (2014). Microhabitat differences impact phylogeographic concordance of codistributed species: Genomic evidence in montane sedges (*Carex* L.) from the Rocky Mountains. *Evolution*, *68*, 2833–2846. https://doi.org/10.1111/evo.12491

McCarthy, D. M., & Mason-Gamer, R. J. (2016). Chloroplast DNA-based phylogeography of *Tilia americana* (Malvaceae). *Systematic Botany*, *41*, 865–880. https://doi.org/10.1600/036364416X693964

McLachlan, J. S., Clark, J. S., & Manos, P. S. (2005). Molecular indicators of tree migration capacity under rapid climate change. *Ecology*, *86*, 2088–2098. https://doi.org/10.1890/04-1036

Médail, F., & Diadema, K. (2009). Glacial refugia influence plant diversity patterns in the Mediterranean Basin. *Journal of Biogeography*, *36*, 1333–1345. https://doi.org/10.1111/j.1365-2699.2008.02051.x

Meirmans, P. G. (2012). The trouble with isolation by distance. *Molecular Ecology*, *21*, 2839–2846. https://doi.org/10.1111/j.1365-294X.2012.05578.x

Morris, A. B., Graham, C. H., Soltis, D. E., & Soltis, P. S. (2010). Reassessment of phylogeographical structure in an eastern North American tree using Monmonier's algorithm and ecological niche modelling. *Journal of Biogeography*, *37*, 1657–1667. https://doi.org/10.1111/j.1365-2699.2010.02315.x

Nadeau, S., Godbout, J., Lamothe, M., Gros-Louis, M.-C., Isabel, N., & Ritland, K. (2015). Contrasting patterns of genetic diversity across the ranges of *Pinus monticola* and *P. strobus*: A comparison between eastern and western North American postglacial colonization histories. *American Journal of Botany*, *102*, 1342–1355. https://doi.org/10.3732/ajb.1500160

Nazareno, A. G., Bemmels, J. B., Dick, C. W., & Lohmann, L. G. (2017) Minimum sample sizes for population genomics: An empirical study from an Amazonian plant species. *Molecular Ecology Resources*, *17*, 1136–1147. https://doi.org/10.1111/1755-0998.12654

Nei, M. (1987). *Molecular evolutionary genetics*. New York: Columbia University Press.

O'Connell, B. M., LaPoint, E. B., Turner, J. A., Ridley, T., Boyer, D., Wilson, A. M., … Conkling, B. L. (2012) The Forest Inventory and Analysis Database: Database description and users manual version 5.1.4 for Phase 2. Gen. Tech. Rep. Washington, DC: US Department of Agriculture, Forest Service.

Owens, H. L., Campbell, L. P., Dornak, L. L., Saupe, E. E., Barve, N., Soberón, J. & Peterson, A. T. (2013) Constraints on interpretation of ecological niche models by limited environmental ranges on calibration areas. *Ecological Modelling*, *263*, 10–18.

Paradis, E. (2010). Pegas: An R package for population genetics with an integrated-modular approach. *Bioinformatics*, *26*, 419–420. https://doi.org/10.1093/bioinformatics/btp696

Peterson, B. J., & Graves, W. R. (2016). Chloroplast phylogeography of *Dirca palustris* L. indicates populations near the glacial boundary at the Last Glacial Maximum in eastern North America. *Journal of Biogeography*, *43*, 314–327. https://doi.org/10.1111/jbi.12621

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE*, *7*, e37135. https://doi.org/10.1371/journal.pone.0037135

Petit, R. J., Aguinagalde, I., de Beaulieu, J.-L., Bittkau, C., Brewer, S., Cheddadi, R., … Vendramin, G. G. (2003). Glacial refugia: Hotspots but not melting pots of genetic diversity. *Science*, *300*, 1563–1565. https://doi.org/10.1126/science.1083264

Phillips, S. J., Anderson, R. P., Dudík, M., Schapire, R. E., & Blair, M. E. (2017). Opening the black box: An open-source release of Maxent. *Ecography*, *40*, 887–893. https://doi.org/10.1111/ecog.03049

Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, *190*, 231–259. https://doi.org/10.1016/j.ecolmodel.2005.03.026

Phillips, S. J., Dudík, M., & Schapire, R. E. (2004). A maximum entropy approach to species distribution modeling. *Proceedings of the Twenty-First International Conference on Machine Learning*, *2004*, 655–662.

Prentice, C., Bartlein, P. J., & Webb, T. III (1991). Vegetation and climate change in eastern North America since the Last Glacial Maximum. *Ecology*, *72*, 2038–2056. https://doi.org/10.2307/1941558

Prentice, I. C., Jolly, D., & BIOME 6000 participants (2000). Mid-Holocene and glacial-maximum vegetation geography of the northern continents and Africa. *Journal of Biogeography*, *27*, 507–519. https://doi.org/10.1046/j.1365-2699.2000.00425.x

Provan, J., & Bennett, K. D. (2008). Phylogeographic insights into cryptic glacial refugia. *Trends in Ecology and Evolution*, *23*, 564–571. https://doi.org/10.1016/j.tree.2008.06.010

Qiu, Y. X., Fu, C. X., & Comes, H. P. (2011). Plant molecular phylogeography in China and adjacent regions: Tracing the genetic imprints of Quaternary climate and environmental change in the world's most diverse temperate flora. *Molecular Phylogenetics and Evolution*, *59*, 225–244. https://doi.org/10.1016/j.ympev.2011.01.012

Raj, A., Stephens, M., & Pritchard, J. K. (2014). fastSTRUCTURE: Variational inference of population structure in large SNP data sets. *Genetics*, *197*, 573–589. https://doi.org/10.1534/genetics.114.164350

Saeki, I., Dick, C. W., Barnes, B. V., & Murakami, N. (2011). Comparative phylogeography of red maple (*Acer rubrum* L.) and silver maple (*Acer saccharinum* L.): Impacts of habitat specialization, hybridization and glacial history. *Journal of Biogeography*, *38*, 992–1005. https://doi.org/10.1111/j.1365-2699.2010.02462.x

Savolainen, O., Pyhäjärvi, T., & Knürr, T. (2007). Gene flow and local adaptation in trees. *Annual Review of Ecology, Evolution, and Systematics*, *38*, 595–619. https://doi.org/10.1146/annurev.ecolsys.38.091206.095646

Schmidtling, R. C. (2003). The southern pines during the Pleistocene. *Acta Horticulturae*, *615*, 203–209. https://doi.org/10.17660/ActaHortic.2003.615.19

Schmidtling, R. C., & Hipkins, V. (1998). Genetic diversity in longleaf pine (*Pinus palustris*): Influence of historical and prehistorical events. *Canadian Journal of Forest Research*, *28*, 1135–1145. https://doi.org/10.1139/x98-102

Shafer, A. B. A., Cullingham, C. I., Côté, S. D., & Coltman, D. W. (2010). Of glaciers and refugia: A decade of study sheds new light on the phylogeography of northwestern North America. *Molecular Ecology*, *19*, 4589–4621. https://doi.org/10.1111/j.1365-294X.2010.04828.x

Shaw, J., & Small, R. L. (2005). Chloroplast DNA phylogeny and phylogeography of the North American plums (*Prunus* subgenus *Prunus* section *Prunocerasus*, Rosaceae). *American Journal of Botany*, *92*, 2011–2030. https://doi.org/10.3732/ajb.92.12.2011

Smith, H. C. (1990). *Carya cordiformis* (Wangenh.) K. Koch. In R. M. Burns & B. H. Honkala (Eds), *Silvics of North America: 2. Hardwoods. Agricultural Handbook 654*. Washington, DC: US Department of Agriculture, Forest Service.

Soltis, D. E., Morris, A. B., McLachlan, J. S., Manos, P. S., & Soltis, P. S. (2006). Comparative phylogeography of unglaciated eastern North America. *Molecular Ecology*, *15*, 4261–4293. https://doi.org/10.1111/j.1365-294X.2006.03061.x

Stewart, J. R., & Lister, A. M. (2001). Cryptic northern refugia and the origins of the modern biota. *Trends in Ecology and Evolution*, *16*, 608–613. https://doi.org/10.1016/S0169-5347(01)02338-2

Thomson, A. M., Dick, C. W., & Dayanandan, S. (2015). A similar phylogeographical structure among sympatric North American birches (*Betula*) is better explained by introgression than by shared biogeographical history. *Journal of Biogeography*, *42*, 339–350. https://doi.org/10.1111/jbi.12394

Title, P. O., & Bemmels, J. B. (2018). ENVIREM: An expanded set of bioclimatic and topographic variables increases flexibility and improves performance of ecological niche modeling. *Ecography*, *41*, 291–307. https://doi.org/10.1111/ecog.02880

Tzedakis, P. C., Emerson, B. C., & Hewitt, G. M. (2013). Cryptic or mystic? Glacial tree refugia in northern Europe. *Trends in Ecology and Evolution*, *28*, 696–704. https://doi.org/10.1016/j.tree.2013.09.001

Varela, S., Lima-Ribeiro, M. S., & Terribile, L. C. (2015). A short guide to the climatic variables of the Last Glacial Maximum for biogeographers. *PLoS ONE*, *10*, e0129037. https://doi.org/10.1371/journal.pone.0129037

Watterson, G. A. (1975). On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology*, *7*, 256–276. https://doi.org/10.1016/0040-5809(75)90020-9

Watts, W. A. (1980). Late-Quaternary vegetation history at White Pond on the inner coastal plain of South Carolina. *Quaternary Research*, *13*, 187–199. https://doi.org/10.1016/0033-5894(80)90028-9

Willing, E. M., Dreyer, C., & van Oosterhout, C. (2012). Estimates of genetic differentiation measured by F_{ST} do not necessarily require large sample sizes when using many SNP markers. *PLoS ONE*, *7*, e42649. https://doi.org/10.1371/journal.pone.0042649

Willis, K. J., & Van Andel, T. H. (2004). Trees or no trees? The environments of central and eastern Europe during the Last Glaciation. *Quaternary Science Reviews*, *23*, 2369–2387. https://doi.org/10.1016/j.quascirev.2004.06.002

Zinck, J. W. R., & Rajora, O. P. (2016). Post-glacial phylogeography and evolution of a wide-ranging highly-exploited keystone forest tree, eastern white pine (*Pinus strobus*) in North America: Single refugium, multiple routes. *BMC Evolutionary Biology*, *16*, 56. https://doi.org/10.1186/s12862-016-0624-1

BIOSKETCHES

Jordan Bemmels studies the biogeography of temperate and tropical trees and this work is part of his PhD thesis. Christopher Dick is his thesis advisor, and is broadly interested in tropical tree biogeography and evolution.

Author contributions: J.B.B. and C.W.D. conceived the project; J.B.B. performed fieldwork and laboratory work, and analysed the data; J.B.B. wrote the manuscript with input from C.W.D.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Bemmels JB, Dick CW. Genomic evidence of a widespread southern distribution during the Last Glacial Maximum for two eastern North American hickory species. *J Biogeogr*. 2018;00:1–12. https://doi.org/10.1111/jbi.13358