Received: 1 April 2020 | Revised: 24 October 2020 | Accepted: 12 November 2020

DOI: 10.1111/jbi.14054

RESEARCH ARTICLE

Journal of Biogeography WILEY

Ecological niche modelling and phylogeography reveal range shifts of pawpaw, a North American understorey tree

Graham E. Wyatt | J.L. Hamrick | Dorset W. Trapnell

Department of Plant Biology, University of Georgia, Athens, GA, USA

Correspondence
Dorset W. Trapnell, Department of Plant Biology, University of Georgia, Athens, Georgia, USA.
Email: dorset@uga.edu

Funding information
University of Georgia Office of the Vice President for Research awarded to D.W.T., Grant/Award Number: N/A

Handiling Editor: Jack Williams

Abstract

Aim: Predictions of species' responses to accelerating global climate change require an understanding of historical range shifts. However, large-scale phylogeographical studies of Eastern North American understorey plant taxa are relatively scarce. Here we employ ecological niche modelling and genetic analyses for inference of optimal pawpaw habitat in the past and future.

Location: In all, 26 states in the eastern United States.

Taxon: *Asimina triloba* (L.) Dunal (Annonaceae).

Methods: The present-day niche of *Asimina triloba* was modelled in Maxent using seven bioclimatic variables, elevation and location data from field samples and herbarium specimens. To model historically optimal habitats, the present-day model was projected onto rasters of seven bioclimatic variables and elevation representing the last glacial maximum (~22,000 years before present [YBP]) and the mid-Holocene (~6,000 YBP). Predicted habitat suitability for 2070 was also modelled. Additionally, 62 populations were genotyped with nine nuclear microsatellite loci and statistically analysed. Levels and partitioning of genetic variation within and among populations were estimated within a geographical context.

Results: Models indicate that optimal habitat 22,000 YBP was severely restricted to mostly now-submerged Gulf of Mexico and southeastern US coastlines. By 2070, models suggest that optimal habitat will expand substantially northward relative to the present. Species-level genetic diversity (H_E = 0.765) is high and genetic structure among populations is moderate (G_{ST} = 0.202). STRUCTURE indicates that there are two population clusters straddling the Appalachian Mountains.

Main conclusions: Models suggest that 22,000 YBP *A. triloba* was restricted to two major refugia in narrow bands of mostly now-submerged habitat and possibly several small inland refugia. Molecular data are consistent suggesting that the eastern refugia expanded to give rise to the eastern cluster which is characterized by higher genetic diversity. The Gulf of Mexico refugium likely gave rise to populations in the western cluster, which is characterized by lower genetic diversity.

KEYWORDS
Asimina triloba, climate change, ecological niche modelling, glacial refugia, pawpaw, phylogeography, population genetics, range shifts

1 | INTRODUCTION

Among the most important ecological processes for the persistence of species are dispersal and colonization, which enable species to undergo range modifications in response to changing environmental conditions. While individual plants are sedentary, often a species' range can undergo a slow, gradual shift that tracks directional climate change at a pace too subtle for detection in a human lifetime. It is only through the lens of time that natural range shifts generally become evident. Although there is historical evidence of episodes of extremely rapid change that were not linked to continent-wide extinctions, for most of Earth's history environmental change has been gradual allowing incremental shifts in distribution (Hof et al., 2011).

During ice ages, plant taxa occupying non-glaciated habitats displayed varied responses to the cooler, drier conditions, with some species encompassing larger ranges relative to present-day distributions (e.g. *Ceratiola ericoides* [Trapnell et al., 2007] and *Yucca brevifolia* [Cole et al., 2011]). However, most cold-intolerant species in Eastern North America (ENA) retreated to lower latitudes (e.g. Bemmels & Dick, 2018; Davis, 1983; Soltis et al., 2006). After southward retreat, some species had widespread distributions (e.g. *Carya* spp. [Bemmels & Dick, 2018]) while others were confined to one or more smaller refugia (e.g. *Pinus strobus* [Zinck & Rajora, 2016]). Following glacial retreat, many plant taxa migrated northward and enlarged their ranges (Davis & Shaw, 2001; Delcourt & Delcourt, 1987b; McLachlan et al., 2005; Peterson & Graves, 2016). Understanding historical patterns of plant dispersal and movement after the onset of deglaciation 20,000 to 19,000 years before present (YBP) (Clark et al., 2009) provides insight into a species' capacity to alter its range in response to shifting climatic conditions. In recent decades, however, not only has climate change accelerated but land-use change, urbanization, habitat degradation and the modification of ecosystems have been rapid. Understanding how plant species disperse, colonize new habitats and alter their ranges is essential for predicting their response to accelerating global climate change associated with the Anthropocene (Zalasiewicz et al., 2011). Predictions of future responses to a changing environment are possible based on understanding how genetic patterns within and among extant populations have been shaped by historical processes. There is a critical need for such insights and predictive power in the face of direct (e.g. over harvesting) and indirect (e.g. landscape disturbance, habitat loss, climate modification) anthropogenic threats to biodiversity.

Interestingly, some tree species with apparently limited seed dispersal appear to have migrated remarkable distances after the Last Glacial Maximum (LGM). Examples include species of *Quercus*, *Fagus* and *Castanea* (Clark et al., 1998; Johnson & Webb, 1989). This disparity between expected and inferred migration rates also occurs in herbaceous taxa (e.g. *Asarum canadense* [Cain et al., 1998]). Clement Reid (1899) was the first to express perplexity ('Reid's paradox'; Clark et al., 1998; MacDougall, 2003) at the surprisingly rapid rate of northward migration from glacial refugia by large-seeded trees such as oaks. A diffusion model of migration was developed by Skellam (1951) who observed that the inferred migration rate in oaks greatly exceeded expectations. Fossil pollen data have been used to estimate migration rates and document this phenomenon for several tree species (e.g. Clark, 1998). For some tree species, animal vectors have been found to be effective long-distance seed dispersers. For example, small mammals and larger carnivorous mammals dispersed *Prunus mahaleb* seeds 495 and 990 m respectively during a single fruiting event (Jordano et al., 2007), while blue jays dispersed acorns up to 1.9 km (Darley-Hill & Johnson, 1981). Now extinct Pleistocene megafauna may also have played a significant role in dispersing large seeds. Mammoths (*Mammuthus columbi*), tapirs (*Tapirus veroensis*), bison (*Bison bison*), horses (*Equus sp.*) and mastodons (*Mamut americanum*) were present in Southeastern North America as recently as 12,700 YBP (Perrotti, 2018). There is also evidence that horses and mastodons underwent annual migrations that spanned distances in excess of 150 km (Hoppe & Koch, 2007).

Large-scale phylogeographical studies of ENA plant taxa and their responses to post-glacial climate change are relatively scarce for understorey, insect-pollinated species (e.g. Peterson & Graves, 2016). Some studies have used modelling to infer locations of LGM refugia (e.g. Call et al., 2016; Morris et al., 2010; Peterson & Graves, 2016; Ruiz-Sanchez & Ornelas, 2014). Fewer studies have modelled future habitat suitability of ENA plant species (see Call et al., 2016). Our aim is to explore these questions in the widespread ENA tree, *Asimina triloba*, commonly known as pawpaw. Pawpaw occurs in 26 states extending from the eastern US westward to Nebraska, Kansas and Texas, and spanning a latitudinal gradient from Florida to the Great Lakes and southern Ontario (Weakley, 2012). Its northern distribution is well north of the southern limit of the Laurentide Ice Sheet whose uneven southern margin extended as far south as ~39° latitude (Dyke & Prest, 1987). Our specific objectives are to (a) infer the historical distribution of suitable *A. triloba* habitat during the LGM in the Pleistocene epoch (~22,000 YBP) and mid-Holocene epoch (~6,000 YPB) through ecological niche modelling, (b) use modelling to infer the location and extent of suitable *A. triloba* habitat in 2070, (c) investigate the spatial distribution of neutral nuclear genetic variation within and among extant populations of *Asimina triloba* and (d) infer locations of refugia during the LGM and patterns of post-glacial range expansion.

2 | MATERIALS AND METHODS

2.1 | Study species

In the Southeastern US, there are eight native *Asimina* species (Huang et al., 2000), most of which are narrowly distributed across Florida and southern Georgia. A notable exception is *A. triloba* which occurs throughout much of the eastern US (Little, 1977). *Asimina triloba* is an understorey tree that spreads clonally via root suckers (Keener & Kuhns, 1998). It tends to grow at lower elevations (<350 m;

Freeman & Hulbert, 1985) in rich, moist soils in closed-canopy oak-hickory or alluvial forests along creeks or rivers (Murphy, 2001). It occurs less frequently in upland terrain. Flowers emit a yeasty odour (Wyatt, pers. obs.) and are pollinated by weak-flying beetles and flies (Muscidae and Sacrophagidae) (Goodrich et al., 2006; Kral, 1997; Willson & Schemske, 1980). *Asimina triloba* is strongly self-incompatible (Pomper et al., 2010) and produces the largest (5–16 cm long, 3–7 cm wide) edible fruits (berries) of any native North American tree. It is hypothesized that seeds were formerly dispersed by now extinct megafauna while today seeds are typically dispersed by raccoons, coyotes and foxes (Cypher & Cypher, 1999; Willson, 1993; Yeager & Elder, 1945; Wyatt, pers. obs.).

2.2 | Sampling

Leaf samples were collected from a mean of 29.6 individuals (range = 10–62) from each of 62 *A. triloba* populations (Table S1; Figure 1) distributed throughout its US range and spanning 1,835 km. Leaves were obtained from stems separated by ≥10 m to avoid sampling ramets of the same genet. Sizes of the sampling areas were dependent on stem density and varied considerably among populations. Young leaves were collected from mature, flower-bearing stems (≥3 m in height) and snap-frozen in liquid nitrogen for transport to the University of Georgia for genetic analyses. GPS coordinates of each population were recorded.

2.3 | Ecological niche modelling

The realized present-day niche of *A. triloba* was modelled using maximum entropy in Maxent 3.4.0 (Phillips & Dudík, 2008; Phillips et al., 2006). Location data for the 62 populations sampled for genetic analysis (Table S1) as well as 20 additional populations that were not a part of the genetic study presented here were used to compile a presence-only dataset. Location data were supplemented with GPS coordinates from 18 herbarium specimens collected after 1992 (herbaria: BUT, CANM, HBSH, JMUH, LSU, MO, MOR, MWI, NCSC, NCU, OKL, UCAC, USCH and UNC). Prior to 1992, the US government scrambled GPS signals rendering them too inaccurate to be useful. Although this constituted 100 presence records widely distributed across the species' range because Maxent can only use one occurrence record per grid cell (~1 km^2) of climatic layers in BioClim, the number of usable records was reduced to 96. Given the availability of presence-only data for *A. triloba*, we believe that Maxent is the most appropriate and robust statistical model.

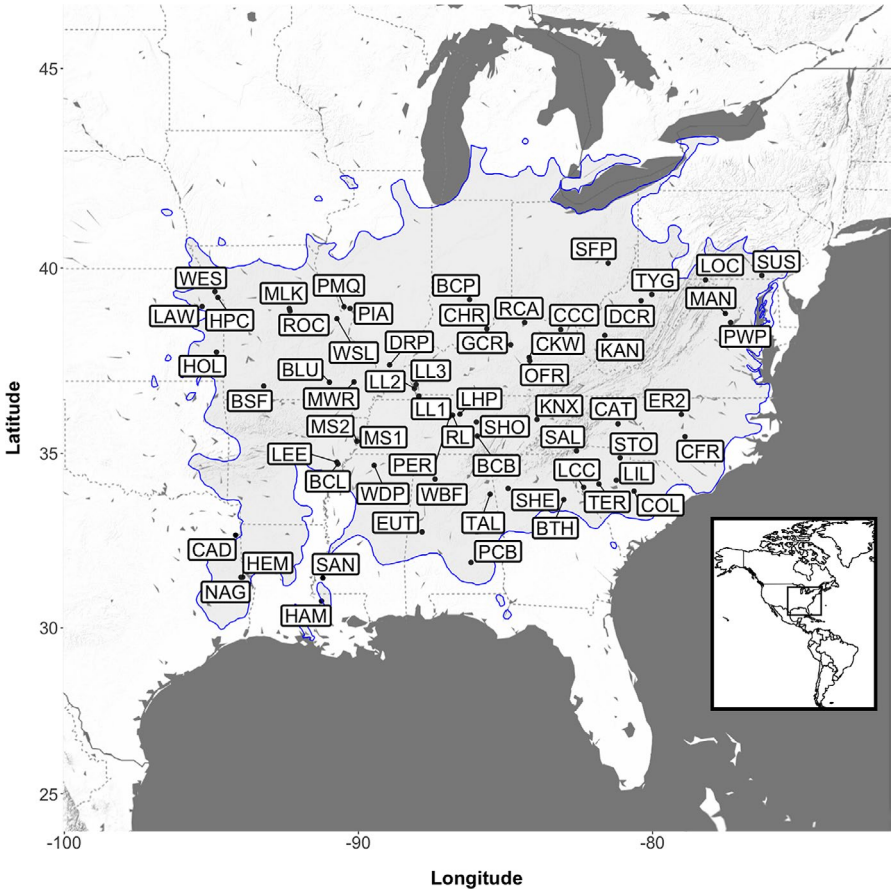

FIGURE 1 Map of 62 populations of *Asimina triloba* in eastern North America using the Mercator projection. The reported range of *A. triloba* (Little, 1977) is shown in grey and outlined in blue

In all, 19 climate variables at 30 arc second resolution were obtained from WorldClim v. 2.0 (Fick & Hijmans, 2017) and elevation data were accessed from the USGS (Archuleta et al., 2017) for present-day models. In all, 12 climate variables were highly correlated (Pearson's correlation coefficient >|0.7|) with the seven most explanatory variables thus were excluded from further analyses. The seven retained variables, as well as elevation data, were utilized for construction of the present-day model (Table S2). Twenty-five percent of the location data were reserved for training. For model construction, we used 10,000 background points, a convergence threshold of 10^{-5}, the maximum number of iterations (500), and prevalence of 0.5. A regularization parameter of 3.0 was selected for calculating the mean area under the receiver operating characteristic curve (AUC) through cross validation. For all models, only those with an AUC score of ≥0.7 were selected (Fielding & Bell, 1997). A threshold of 0.5 was used for calculation of sensitivity and the percent correctly classified (PCC) value. In all, 10 replicates were performed and averaged.

To construct models for inference of suitable *A. triloba* habitat in the past, the present-day model was projected onto rasters of the same seven bioclimatic variables (Table S2) representing the LGM (~22,000 YBP) and the mid-Holocene (~6,000 YBP) as inferred by three global circulation models (GCMs): the Community Climate System Model (CCSM4; Gent et al., 2011), the Model for Interdisciplinary Research on Climate (MIROC; Watanabe et al., 2011) and the Max Planck Institute for Meteorology Earth Systems Model (MPI-ESM; Giorgetta et al., 2013). We also included elevation information from ENVIREM (Title & Bemmels, 2017) for the LGM and the mid-Holocene. Because the three GCMs varied, the inferred probability of suitable habitat at each cell was averaged across models for the LGM and mid-Holocene (Varela et al., 2015). In all, 10 iterations of each GCM were run and the mean standard deviation calculated. To estimate the similarity among GCMs for each time period, the Pearson's correlation coefficient was calculated.

Predicted habitat suitability for 2070 under the Representative Concentration Pathways (RCP) 6.0 and 8.5 trajectories (Vuuren et al., 2011), which are greenhouse gas concentration trajectories, was modelled using the same seven environmental variables as well as elevation data from the USGS, and the same three GCMs (CCSM4, MIROC and MPI-ESM). The mean standard deviation was determined for 10 iterations of each GCM. Pearson's correlation coefficients were calculated to estimate similarity among GCMs under each RCP trajectory.

2.4 | Microsatellite genotyping

Frozen leaf tissue (~0.05 g) was ground to a powder and genomic DNA was extracted using a modified CTAB protocol (Doyle & Doyle, 1987). DNA quality and quantity were evaluated using a ND-1000 Nanodrop® spectrophotometer. Nine nuclear microsatellite loci were amplified with markers developed by Lu et al. (2011; Table S3). A three-primer PCR protocol was used whereby a CAG-tag sequence (Hauswaldt & Glenn, 2003) was added to the 5' end of one primer and a third fluorescently labelled (FAM, HEX, or NED) primer identical to the CAG-tag was included in the reaction. PCR amplification was carried out in 12.5 μL reaction volumes consisting of 0.26 mM molecular grade ddH2O, 1 mM One Taq® standard reaction buffer (New England Biolabs, Ipswich, MA), 0.7 mM MgCl solution (Sigma-Aldrich), 1 mM bovine serum albumin (Thermo-Fisher), 0.2 mM dNTP mix (New England Biolabs, Ipswich, MA), 0.04 mM CAG-tagged primer, 0.4 mM unlabelled primer, 0.36 mM universal fluorescently labelled CAG primer, 0.2 mM One Taq® Hot Start DNA polymerase (New England Biolabs, Ipswich, MA) and 50 ng template DNA.

2.5 | Genetic diversity analyses

A test for null alleles was performed in Genepop version 4.2 using Brookfield's method (Brookfield, 1996; Raymond & Roussett, 1995). Because *A. triloba* is a clonal species, it was necessary to determine the presence of duplicate multi-locus genotypes (MLGs) within populations and to assess whether these represented clones or arose by chance. The likelihood that identical MLGs represent clones can be ascertained based on the number of loci, allele frequencies and the number of stems sampled in the population. The multi-locus probability of identity (PI_m) within populations was calculated as $PI_m = \Pi_s (PI_s)$, where

$$PI_s = \sum_i x_i^4 + \sum_i \sum_{j>i} (2x_i x_j)^2$$

and x_i and x_j are the frequencies of the ith and j^{th} allele at a locus, respectively (Gonzales et al., 2008; Paetkau & Strobeck, 1994). The probability of identity (PI) was adjusted for sample size (N) by $PI = 1-[(1-PI_m)^N]$ (Parker et al., 1998). Duplicate genotypes that represented clones were removed, leaving one individual per MLG for inclusion in subsequent analyses.

Genetic diversity statistics were estimated using GenAlEx version 6.51b2 (Peakall & Smouse, 2006, 2012). Measures of genetic diversity were percent polymorphic loci (P), total number of alleles per population (A_T), effective number of alleles (A_E), private alleles (P_A) (i.e. alleles in a single population), rare alleles (A_{RARE}) (alleles ≤0.1 frequency in a population), observed heterozygosity (H_O), expected heterozygosity (H_E) and Wright's inbreeding coefficient (F_{IS}; Wright, 1922, 1951). Population values were estimated and averaged across all populations while species-level values were calculated by pooling data from all populations.

Observed heterozygosity and Hardy–Weinberg expected heterozygosity were compared for each locus in each population using Wright's inbreeding coefficient (F_{IS}). Deviations from Hardy–Weinberg expectations were tested for significance using χ^2 (Li & Horvitz, 1953). The Bonferroni correction for multiple comparisons was applied using FSTAT (Goudet, 1995).

2.6 | Genetic structure analyses

A Bayesian clustering approach (STRUCTURE version 2.3.4; Pritchard et al., 2000) was used to estimate the number of genetically distinct clusters (*K*) and levels of genetic admixture among populations. In all, 10 independent simulations at each *K*-value from 1 to 20 were run, using a burn-in of 500,000 repetitions and a run length of 1,000,000 Markov chain Monte Carlo (MCMC) iterations. The admixture model was chosen to infer alpha (α). We used a model based on correlated allele frequencies with no a priori assumptions regarding sampling locations. The Evanno et al. (2005) method was used to determine optimal Δ*K* using Structure Harvester (Earl & vonHoldt, 2012). A principal components analysis (PCA) of genotypes was used to check the validity of these findings through R package 'gstudio' (Dyer, 2016; R Core Team, 2013). Genetic structure among populations was also estimated using Nei's (1973) G_{ST} in GenAlEx (Peakall & Smouse, 2006, 2012). G_{ST} was calculated among all 62 populations, among populations within the two genetic clusters identified by STRUCTURE, and for all possible pairs of populations.

2.7 | Geographical patterns

We performed linear and polynomial regressions of the population-level statistics (MLGs, P, A_T, A_E, P_A, A_{RARE}, H_O and H_E) against latitude and longitude and tested for significance to determine whether north-south and/or east-west patterns of genetic diversity exist. To compare levels of genetic diversity in the eastern and western clusters, a two-sample *t*-Test assuming equal variances was performed for all genetic diversity measures.

Isolation-by-distance (IBD; Wright, 1943) was estimated with regression using two approaches: (a) $G_{ST} / (1 - G_{ST})$ (Rousset, 1997) and (b) conditional genetic distance with geographical distance (Dyer et al., 2010) which considers the total genetic covariance among all populations. Each test was evaluated for significance at α = 0.05 level. Additional tests for IBD were performed within clusters identified by STRUCTURE.

3 | RESULTS

3.1 | Ecological niche modelling

For the present-day model, the AUC score is high (0.859), with a sensitivity of 0.90 and a percent correctly classified (PCC) value of 0.77. The most influential bioclimatic variables are minimum temperature of the coldest month (BIO06) and elevation, contributing 54% and 21% respectively to the model. Optimal habitat, defined as a probability of occurrence [*Pr*O] >0.75, encompasses 527,087 km^2 that span a largely contiguous area from Oklahoma to Massachusetts as well as the Southeast and Mid-Atlantic (Figure 2). Of that, 49,316 km^2 occurs at >40° N latitude. Our inferred present-day niche model for *A. triloba* was similar to a model created by Li et al. (2017) and agrees with the contemporary species range.

The LGM projection (Figure 3) infers 70,233 km^2 of optimal habitat, most of which is distributed in narrow bands in now

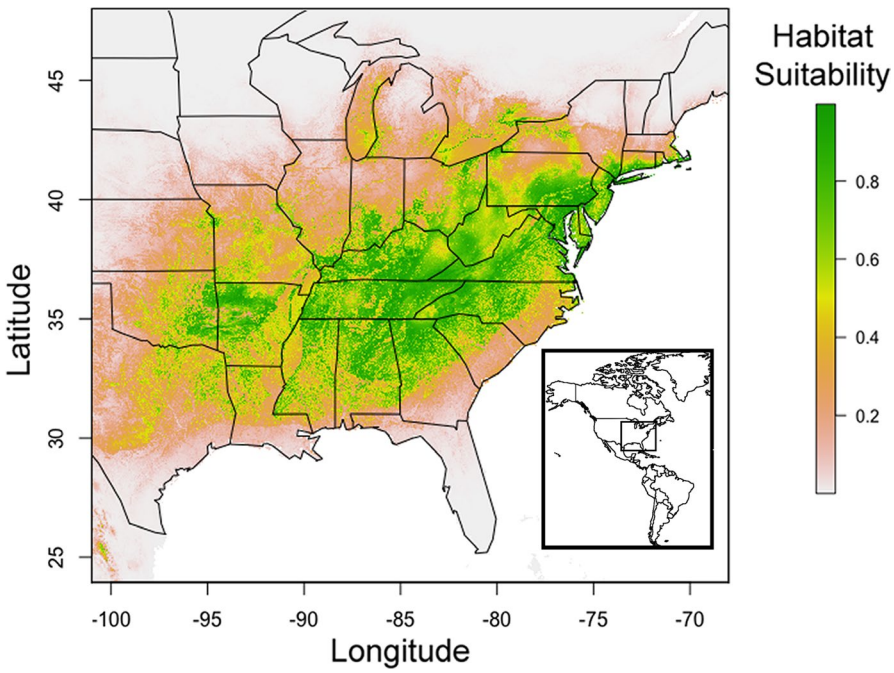

FIGURE 2 Realized present-day niche of *Asimina triloba* in the eastern United States and southeastern Canada modelled through maximum entropy using seven climatic variables (Worldclim) and elevation (USGS). Light pink indicates areas of unsuitable habitat, with yellow and green indicating increasing suitability of habitat for *A. triloba*. The Mercator projection was used

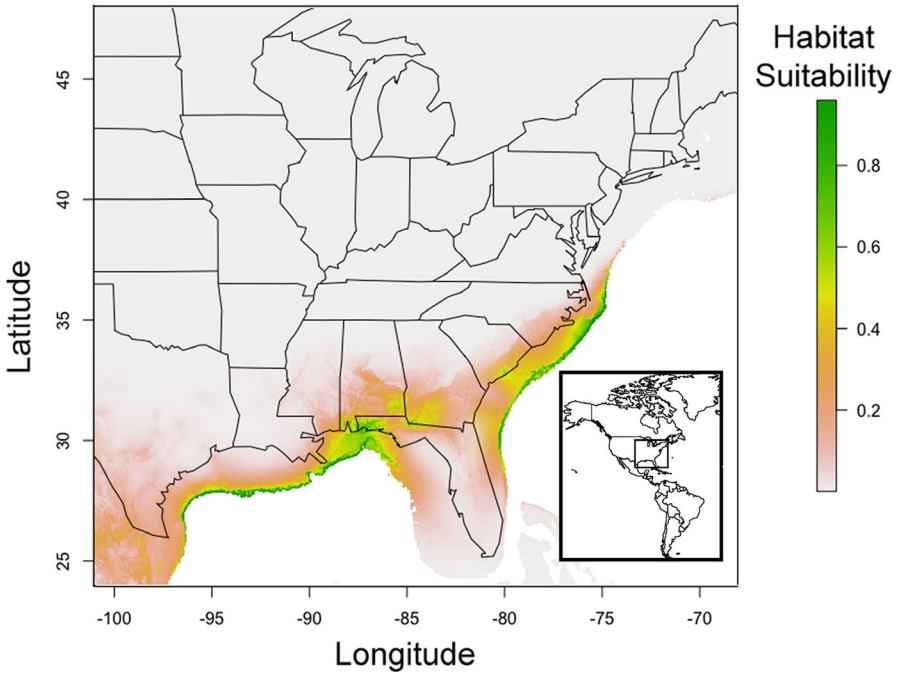

FIGURE 3 Projection of habitat suitability for *Asimina triloba* in the southeastern United States during the Last Glacial Maximum in the Pleistocene epoch (~22,000 YBP), as inferred from the average of three GCMs (CCSM4, MIROC and MPI-ESM) from seven bioclimatic variables as well as elevation (ENVIREM). Light pink indicates areas of unsuitable habitat, with yellow and green indicating increasing suitability of habitat for *A. triloba*. The Mercator projection was used

submerged locations along the Gulf of Mexico (25,923 km^2) and the Southeastern Atlantic coast (40,976 km^2). The model also indicates a limited inland area of suitable habitat in present-day southeastern Alabama and southwestern Georgia (1,265 km^2). Inferred optimal habitat in the LGM was 13.3% of the extent of optimal habitat inferred by the present-day model, a pattern consistent with the

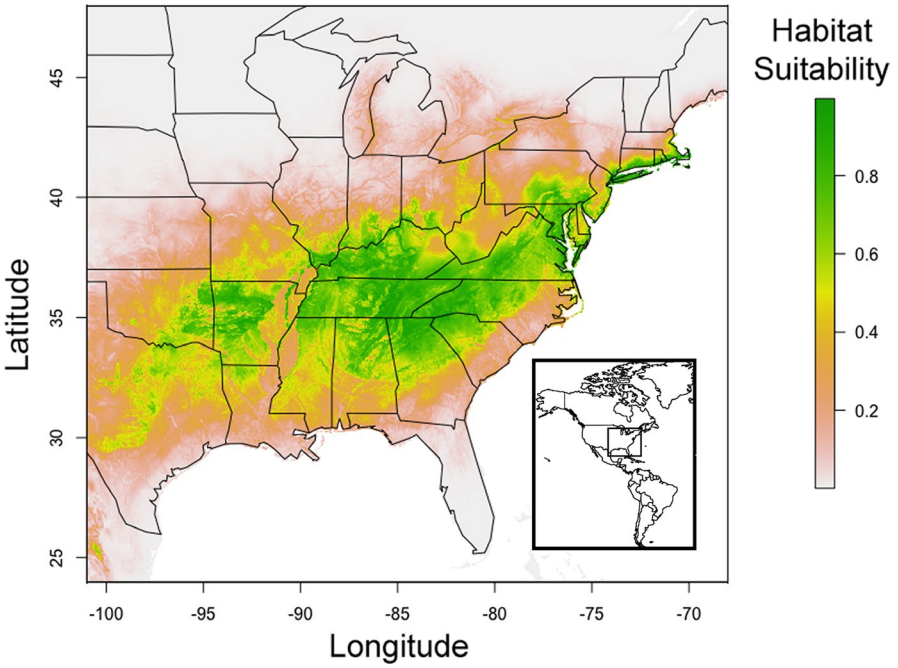

FIGURE 4 Projection of *Asimina triloba* habitat suitability in eastern North America during the mid-Holocene epoch (~6,000 YBP), as inferred from the average of three GCMs (CCSM4, MIROC and MPI-ESM) from seven bioclimatic variables as well as elevation (ENVIREM). Light pink indicates areas of unsuitable habitat, with yellow and green indicating increasing suitability of habitat for *A. triloba*. The Mercator projection was used

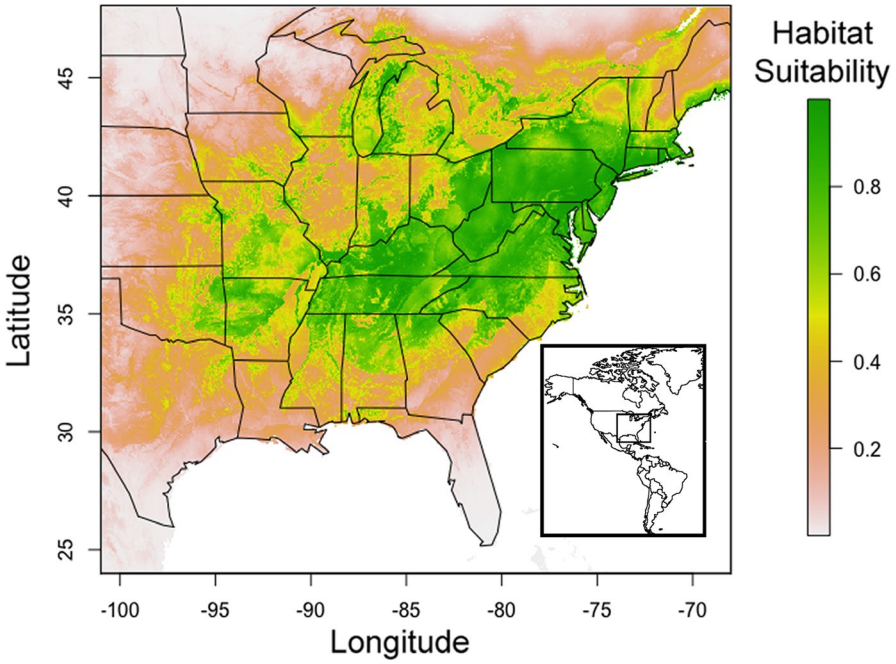

FIGURE 5 Projection of habitat suitability for *Asimina triloba* in the eastern United States and southeastern Canada for the year 2070, as inferred from the average of three GCMs (CCSM4, MIROC and MPI-ESM) under the RCP 6.0 trajectory from seven bioclimatic variables as well as elevation (USGS). Light pink indicates areas of unsuitable habitat, with yellow and green indicating increasing suitability of habitat for *A. triloba*. The Mercator projection was used

cooler, drier Pleistocene climate and with species range contractions reported by Davis et al. (1980). The three GCMs for the LGM have a Pearson's correlation coefficient of 0.650, with highest agreement along the Gulf of Mexico (Figure S1). Mean standard deviation values of iterations for each of the LGM GCMs are illustrated in Figure S2.

The mid-Holocene projection (Figure 4) infers 366,959 km^2 of optimal habitat which is more than five times larger than inferred LGM habitat. Holocene *A. triloba* habitat spanned a range very similar to that of the present-day model, although encompassing a smaller area that was 69.6% of the present-day area.

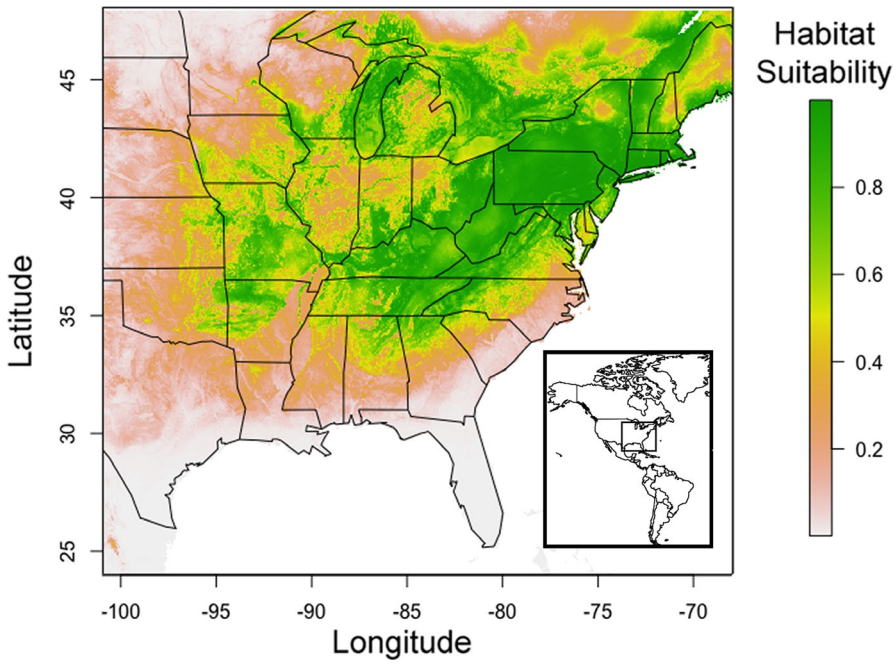

FIGURE 6 Projection of habitat suitability for *Asimina triloba* in the eastern United States and southeastern Canada for the year 2070, as inferred from the average of three GCMs (CCSM4, MIROC and MPI-ESM) under the RCP 8.5 trajectory from seven bioclimatic variables as well as elevation (USGS). Light pink indicates areas of unsuitable habitat, with yellow and green indicating increasing suitability of habitat for *A. triloba*. The Mercator projection was used

TABLE 1 Summary genetic diversity statistics of 62 populations of *A. triloba* arranged in ascending latitudinal order within the two genetic clusters identified by STRUCTURE. *N* = sample size, MLG = number of multi-locus genotypes, *P* (%) = percent of polymorphic loci, A_T = total number of alleles, A_E = effective number of alleles per locus, P_A = number of private alleles, A_{RARE} = number of rare alleles (< 0.1 frequency within a population), H_O = observed heterozygosity, H_E = expected heterozygosity and F_{IS} = inbreeding coefficient. All F_{IS} values are non-significant (p = 0.05)

Population	*N*	MLG	*P* (%)	A_T	A_E	P_A	A_{RARE}	H_O	H_E	F_{Is}
Eastern cluster										
NAG	30	3	78	20	3.1	-	-	0.667	0.586	−0.158
HEM	40	5	89	26	2.8	-	-	0.578	0.544	−0.016
PCB	48	12	89	29	1.9	-	5	0.397	0.391	0.013
EUT	33	30	100	57	3.9	-	9	0.463	0.650	0.231
BTH	37	10	100	54	3.4	2	5	0.620	0.665	0.066
TAL	24	3	89	25	2.2	-	1	0.611	0.480	−0.266
COL	48	12	78	29	1.9	-	1	0.394	0.397	−0.015
SHE	24	5	89	26	2.3	-	1	0.511	0.479	−0.045
LCC	24	18	89	38	2.8	-	4	0.476	0.555	0.147
TER	34	22	89	41	2.9	-	5	0.529	0.579	0.081
LIL	24	21	89	44	2.7	-	4	0.452	0.562	0.194
WBF	43	21	89	53	3.5	-	10	0.578	0.623	0.082
STO	10	5	89	25	2.3	-	2	0.454	0.498	0.129
SAL	30	14	100	43	3.0	-	1	0.563	0.612	0.030
CFR	20	5	89	30	2.5	1	3	0.689	0.550	−0.238
CAT	16	1	67	14	1.7	-	-	0.667	0.333	−1.000
SHO	24	16	89	49	3.4	-	5	0.596	0.566	−0.062
KNX	25	3	100	31	2.8	2	3	0.537	0.600	0.133
PER	48	5	100	37	3.4	-	-	0.583	0.642	0.044
ER2	10	5	67	17	1.7	-	-	0.500	0.326	−0.527
CKW	24	13	100	40	2.5	4	5	0.523	0.526	−0.024
GCR	48	6	78	26	2.4	1	2	0.356	0.471	0.316
CCC	26	13	89	43	3.2	-	4	0.477	0.614	0.222
PWP	48	27	100	56	3.6	-	4	0.540	0.670	0.169
MAN	48	1	33	10	1.1	-	-	0.333	0.167	−1.000
PIA	14	1	44	15	1.4	1	1	0.444	0.222	−1.000
Eastern mean	30.8	10.7	85.1	33.8	2.6	0.42	2.9	0.521	0.511	−0.096
Eastern pooled	800	277	100	129	6.1	11	99	-	0.767	0.297
Western cluster										
HAM	47	2	100	28	2.2	-	-	0.444	0.528	0.170
SAN	21	1	78	16	1.8	1	2	0.778	0.389	−1.000
CAD	39	18	100	31	3.3	1	1	0.633	0.657	0.085
WDP	27	3	89	26	2.5	1	2	0.444	0.531	0.192
BCL	48	27	100	69	2.7	1	8	0.503	0.573	0.125
LEE	24	12	100	56	4.1	-	6	0.643	0.703	0.096
MS1	24	4	78	18	3.5	-	-	0.611	0.674	0.072
MS2	23	1	67	15	0.8	1	1	0.111	0.056	−1.000
BCB	25	8	78	21	1.9	-	-	0.597	0.404	−0.367
RL	27	12	100	44	3.6	2	5	0.681	0.639	−0.064
LHP	20	4	89	22	1.9	-	2	0.611	0.406	−0.442

(Continues)

TABLE 1 (Continued)

Population	N	MLG	P (%)	A_T	A_E	P_A	A_{RARE}	H_O	H_E	F_{Is}
LL1	25	1	56	14	1.6	-	-	0.556	0.278	−1.000
LL2	25	7	78	38	3.4	-	1	0.613	0.580	−0.055
BSF	23	21	89	43	3.1	-	2	0.546	0.543	−0.010
LL3	24	24	78	51	3.7	1	7	0.495	0.577	0.141
BLU	25	9	78	28	2.5	1	1	0.460	0.506	0.062
RCA	25	16	100	49	3.8	1	4	0.490	0.645	0.273
MWR	25	14	78	38	2.3	1	1	0.524	0.442	−0.092
DRP	23	13	100	41	3.3	2	7	0.562	0.599	0.070
OFR	33	6	89	36	2.9	-	2	0.550	0.512	−0.082
HOL	10	8	89	34	2.9	-	1	0.545	0.495	−0.098
KAN	43	10	89	43	3.4	-	-	0.559	0.599	0.090
CHR	30	18	89	45	3.2	-	1	0.580	0.598	0.030
WSL	11	9	100	40	3.5	-	5	0.600	0.647	0.078
ROC	25	1	56	14	1.6	-	-	0.556	0.278	−1.000
MLK	24	2	67	18	1.8	1	1	0.611	0.375	−0.622
PMQ	60	4	89	26	2.2	-	1	0.583	0.441	−0.313
LAW	30	12	89	29	1.9	-	1	0.492	0.392	−0.185
DCR	15	1	44	13	1.4	-	1	0.444	0.222	−1.000
BCP	48	2	67	17	1.7	1	1	0.444	0.292	−0.547
HPC	25	14	89	31	2.1	-	-	0.467	0.469	−0.020
TYG	48	2	89	20	2.0	-	-	0.722	0.444	−0.600
WES	22	9	100	45	3.1	-	3	0.537	0.590	0.078
LOC	38	3	56	16	1.6	-	-	0.278	0.264	−0.134
SUS	30	8	100	37	3.0	-	1	0.602	0.603	0.088
SFP	25	7	100	27	2.1	1	1	0.456	0.482	0.014
Western mean	28.8	8.7	84.4	31.6	2.6	0.44	1.9	0.537	0.484	−0.194
Western pooled	1,037	313	100	126	5.9	16	94	-	0.732	0.226
Species mean	29.6	9.5	84.7	32.5	2.6	0.43	2.3	0.530	0.496	−0.153
Species pooled	1837	590	100	153	6.5	27	144	-	0.765	0.273

The Pearson's correlation coefficient among GCMs is 0.927. The three GCMs and the mean standard deviation values of iterations for each of the Holocene GCMs are shown in Figures S1 and S3, respectively.

The projection for year 2070 under the RCP 6.0 trajectory indicates that optimal habitat will expand northward and increase by 49.6% relative to the present-day model, to 788,548 km^2 (Figure 5) with much habitat currently supporting *A. triloba* remaining suitable. The projection for year 2070 under the RCP 8.5 trajectory indicates that optimal habitat will more than double in size relative to the present-day model, to 1,111,072 km^2 (Figure 6). The three GCMs for the 2070 projection under RCP 6.0 and 8.5 have Pearson's correlation coefficients of 0.866 and 0.902, respectively. The three GCMs for 2070 under RCP 6.0 and 8.5 are shown in Figure S1, and the mean standard deviation values of iterations for each of the 2070 GCMs under the two RCP trajectories are shown in Figures S4 and S5, respectively.

Although other statistical models may have yielded different projections (Araújo et al., 2019), we believe that Maxent is the most appropriate and robust statistical model given the availability of presence-only data for *A. triloba*.

3.2 | Genetic diversity

Null allele frequencies are non-significant for all loci and populations ($p = 0.05$). Mean probability of identity (*PI*) across populations is 1.1×10^{-3} (range = 6.0×10^{-9} to 3.5×10^{-2}) (Table S4), indicating a very high likelihood that identical MLGs within a population are clones. Despite efforts to avoid sampling multiple stems from the same clone, the number of MLGs within populations is consistently lower than the sample size. After removal of duplicate genotypes, the mean number of MLGs per population is 9.5 (range = 1

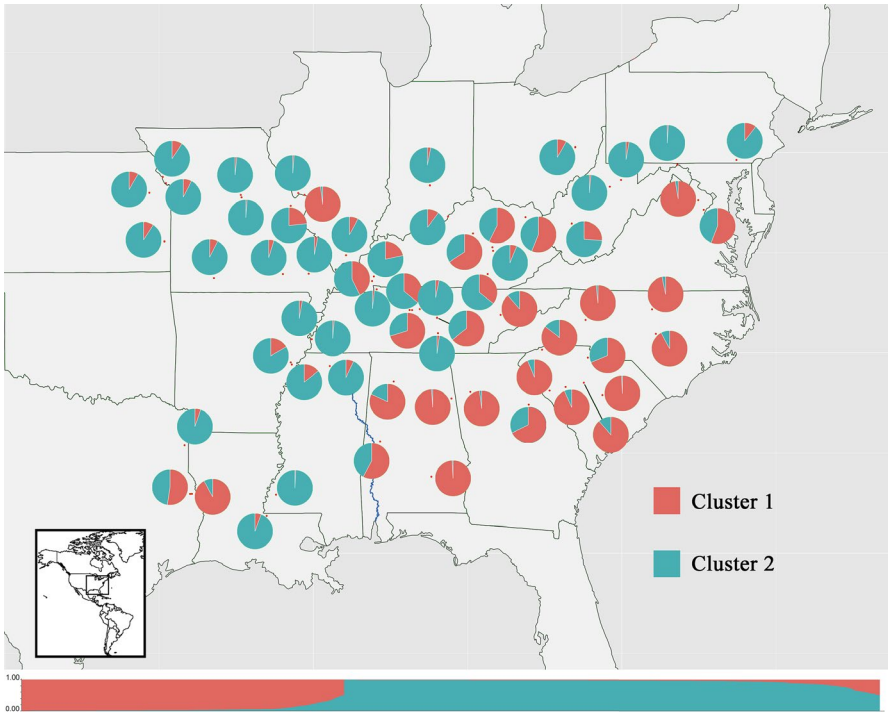

FIGURE 7 Map of 62 *Asimina triloba* populations in the eastern United States with pie charts indicating the proportion of individuals in each population assigned to the eastern cluster (red) and western cluster (blue) identified by STRUCTURE. The Tombigbee River in Alabama is indicated in blue. The Mercator projection was used. Beneath the map is the STRUCTURE bar plot where $K = 2$

to 30) (Table 1). Only one population (LL3) has no duplicate MLGs. The number of MLGs within populations is not significantly correlated with sample size ($r = 0.145$; $p = 0.261$). There was no evidence of MLGs shared by individuals in different populations.

Mean within-population genetic diversity is high as is the pooled species-wide diversity, with a total of 153 alleles (Table 1). Mean allele number per population is 32.5 (range = 10 [MAN] to 69 [BCL]). In all, 27 private alleles were detected in 20 populations, with a maximum of four private alleles in one population (CKW). Both A_T and A_E are significantly correlated with the number of MLGs within populations ($p < 0.00001$ for both). Mean population level expected heterozygosity (H_E) is 0.496 (range = 0.056 [MS2] to 0.703 [LEE]), while species-level $H_E = 0.765$.

Mean population $F_{IS} = -0.153$ (range = −1.00 to 0.316), with negative values in 32 populations, indicating an excess of heterozygosity (Table 1). The number of MLGs within populations and F_{IS} values are significantly correlated ($r = 0.604$; $p = 2.1 \times 10^{-7}$), indicating that excess heterozygosity characterizes populations with a low number of MLGs.

3.3 | Genetic structure

Simulations from STRUCTURE yielded an optimal $K = 2$ (Figure 7). Population assignment to a cluster was based on >50% posterior probability assignment (Table S5). The two clusters fell primarily on either side of the Appalachian Mountains. There is also an east-west discontinuity along the Tombigbee River in Alabama. Some populations had Q-matrix assignments indicating lower genetic affinity to the cluster in which they are geographically situated: examples include populations HEM and NAG (Texas), PIA (Illinois), SUS (Pennsylvania) and PWP (Virginia). PCA results supported the designation of two clusters (Figure S6), although there was some overlap.

Genetic structure (G_{ST}) among all 62 populations is 0.202. Mean pairwise $G_{ST} = 0.227$, with individual pairs ranging from 0.040 (CAD and BCL) to 0.813 (MAN and MS2). Mean pairwise G_{ST} for each population with the remaining 61 populations ranges from 0.147 (WSL) to 0.422 (MAN). Global G_{ST} values among populations within the eastern and western STRUCTURE clusters are nearly identical (0.195 and 0.190, respectively). Mean pairwise G_{ST} values within the eastern and western clusters are 0.226 and 0.216, respectively. The mean pairwise G_{ST} value *between* populations in the eastern and western clusters is 0.251, which is significantly higher than for western populations ($p = 0.009$) but did not differ significantly from pairwise values for eastern populations ($p = 0.083$).

3.4 | Geographical patterns

Two measures of genetic diversity decrease significantly, or nearly so, with increasing latitude when linearly regressed: P ($r = 0.256$; $p = 0.045$) and H_E ($r = 0.246$; $p = 0.054$). Polynomial regression

indicates significantly higher diversity for all measures except P_A and H_O at mid-latitudes than at the southern or northern margins of the species range: MLGs (r = 0.268; p = 0.035), P (r = 0.262; p = 0.039), A_T (r = 0.288; p = 0.023), A_E (r = 0.282; p = 0.026), A_{RARE} (r = 0.328; p = 0.009) and H_E (r = 0.289; p = 0.022). Although private alleles occur throughout the species' range and there is not a significant geographical relationship, 79% of private alleles occur in populations at latitudes below 38° N.

When longitude was linearly regressed against genetic diversity statistics, no significant correlations were found. However, polynomial regression showed that two measures of diversity are significantly correlated with longitude, with higher diversity in the middle of the range: private alleles (r = 0.268; p = 0.035) and rare alleles (r = 0.333; p = 0.008). For nearly all measures, populations within the eastern cluster have higher genetic diversity (Table 1), although the differences are not significant. Only mean population-level and species-level H_O and mean number of private alleles are lower in the eastern cluster.

Significant isolation-by-distance was identified by regression of conditional genetic distance against geographical distance (r = 0.0001; $p = 5.7 \times 10^{-4}$). Given the small r-value for this test, the relationship is unlikely to be biologically meaningful. The regression of G_{ST} / (1 − G_{ST}) and log geographical distance is significant (r = 0.158; p = 0.020). Both analyses revealed significant IBD among populations in the eastern cluster ($r > 0.2$, $p < 0.010$), and an absence of IBD in the western cluster ($p > 0.1$).

4 | DISCUSSION

It has been hypothesized that Southeastern North America served as a Pleistocene refugium for temperate species with much broader current ranges (Hewitt, 2000; Soltis et al., 2006). Based on ecological niche modelling, this was probably true for *A. triloba* as well. While some tree species may have been widespread during the last glacial epoch (e.g. Bemmels & Dick, 2018; Davis & Shaw, 2001; Miller & Parker, 2009), a significant portion of optimal habitat for *A. triloba* (and likely refugia) occurred in narrow coastline bands along the Gulf and Atlantic coasts that currently are mostly below sea level (Figure 3). Small inland sites in what is now southeastern Alabama and southwestern Georgia may also have served as refugia. While the mean standard deviation among iterations of each LGM GCM is high for some areas of inferred habitat (Figure S2) and the Pearson's correlation of coefficient for the three models is only 0.650, this is unsurprising considering the temporal scale of inference. While the precise location of *A. triloba* refugia cannot be known with certainty, we can safely conclude that refugia probably existed at the southeastern edge of terrestrial North America.

While there are no fossil pollen records for *Asimina* because it is zoophilous, there are records for other warm temperate, anemophilous tree species that occupy present-day habitat similar to *A. triloba*. Fossil pollen of *Liquidambar* from ~22,000 YBP has been documented in low abundance in Louisiana and southwestern Mississippi (Jackson & Givens, 1994; Kolb & Freudlund, 1981), southeastern Alabama (Delcourt et al., 1980), the panhandle of Florida (Watts et al., 1992) and St. Catherine's Island, Georgia (Rich et al., 2015). *Nyssa* appears to have been more widespread, with very low abundance pollen records all along the southeastern coastal plain from Texas (Potzger & Tharp, 1947) north to the Chesapeake Bay (Willard et al., 2005), and more inland records from Tennessee (Delcourt, 1979) and Georgia (Watts, 1975). Williams et al. (2004) report that warm mixed forest, suitable for *A. triloba*, occupied coastal areas in the Southeastern US during the LGM, with less-hospitable cool mixed forest to the north. While *Nyssa* does not appear to have retreated to isolated southern refugia, the *Liquidambar* pollen records lend support for the inferred *A. triloba* refugia.

During the post-LGM warming *A. triloba* migrated and expanded its range northward. The mean standard deviation values among iterations of mid-Holocene GCMs are generally low (Figure S3) and the three models are highly correlated (r = 0.927) allowing reasonable confidence in our models. Differences between our LGM and mid-Holocene SDMs are consistent with the finding that the distribution of ENA vegetation underwent substantial change, generally shifting northward approximately 16,000–8,000 YBP, with the most rapid changes occurring 13,000–11,000 YBP (Williams et al., 2004). Due to its high clonality, a phalanx model of expansion could be hypothesized, which entails a slow, steady migration of individuals in a continuous front (Hewitt, 1996). However, since we found no MLGs shared by individuals in different populations, clonal spread was unlikely to be an important mode of range expansion. It is more likely that seeds were dispersed by large- and medium-sized mammals after fruit consumption.

The STRUCTURE and PCA analyses identified two clusters of populations that lie on either side of the Appalachian Mountains and the Tombigbee River drainage. This is consistent with a northward expansion of species during the post-LGM warming with migrants from the Southeastern Atlantic coastal refugium giving rise to the eastern cluster. Simultaneously, new habitat would have become available north of the Gulf of Mexico and along the Mississippi River and Ohio River valleys, which was probably colonized by individuals from the Gulf refugium. Due to its elevation, the southern Appalachian Mountains likely served as a barrier to dispersal between eastern and western genetic clusters. Northward expansion following the LGM has been similarly documented for many tree species (e.g. Davis & Shaw, 2001; Delcourt & Delcourt, 1987a). Similar Gulf and Atlantic coastal refugia have been reported for *Pinus virginiana* (Parker et al., 1997), *Fagus grandifolia* (Morris et al., 2010) and *Carya ovata* (Bemmels & Dick, 2018). Furthermore, the genetic discontinuity in *A. triloba* along the Tombigbee River (Figure 7) is similar to patterns reported for various animal taxa (e.g. Bermingham & Avise, 1986; Soltis et al., 2006).

Further molecular support comes from pairwise G_{ST} values that are lower within the eastern and western clusters (0.226 and 0.216, respectively) than the mean pairwise value between clusters (0.251). The eastern cluster possesses higher (MLGs, P, A_T, A_E, A_{RARE}, H_E and F_{IS}) or equal (A_E) mean population-level genetic diversity values than

the western cluster (Table 1).Only P_A and H_O are lower in the east. A possible explanation is that eastern populations resulted from the northward migration and admixture of propagules from the Atlantic coast and the Alabama/Georgia inland refugia. Periods of glacial maxima are often considerably longer than warmer interglacial periods (Philander, 2008): the most recent glacial period in North America was the Wisconsin Glacial Episode that lasted from 75,000 to 11,000 YBP. During this period of isolation, refugial populations evolved independently, rendering individuals from discrete refugia genetically distinct. Range expansion accompanying post-glacial warming may have led to admixture between separate refugia that produced populations with higher genetic diversity. The Gulf refugium likely gave rise to much of the western cluster with migration following, and perhaps facilitated by, major watersheds such as the Mississippi and Ohio rivers.

Species-level genetic diversity is high (H_E = 0.765) in *A. triloba*, consistent with its life-history traits (i.e. long-lived perennial, outcrossing, animal-dispersed seeds and widespread distribution [Hamrick & Godt, 1996]). At the population level, mean observed heterozygosity (H_O = 0.530) is higher than the mean expected heterozygosity (0.496) (Table 1)—a phenomenon observed in many long-lived woody species (e.g. Frascaria et al., 1993; Huang et al., 1998) and clonal taxa (e.g. Balloux et al., 2003; Pappert et al., 2000). This may result from (a) linkage of neutral markers to regions under selection that favour highly heterozygous individuals (overdominance) and/or (b) disassortative mating between individuals with different alleles (Stoeckel et al., 2006). However, in one population (BTH), samples from three stages of the life cycle (197 seeds, 60 saplings [dbh ≤ 4 cm], and 21 adults) have nearly identical H_O and F_{IS} values (Wyatt, unpublished data), suggesting an absence of selection on linked genes.

Global genetic structure among populations (G_{ST} = 0.202) is high relative to other tree species (Hamrick et al., 1992) and is likely related to low levels of gene flow among populations. Possible causes include (a) limited or short-distance pollen movement between disjunct, alluvial habitats, (b) founder effects and relatively small populations in Pleistocene refugia and extant populations and/or (c) competition between clones after populations have been established. Given that the fly and beetle pollinators of *A. triloba* are weak-flyers (Willson & Schemske, 1980), long-distance pollen movement is presumably rare.

Populations with weak assignment to clusters in STRUCTURE (Table S5; Figure 7) occur primarily in Eastern Tennessee, Kentucky and West Virginia. This region is characterized by the Ridge-and-Valley physiography and would presumably present less of a topographical barrier to dispersal than the higher Blue Ridge Mountains to the southeast. Consequently, admixture from populations originating from eastern and western refugia may have been more likely, a scenario consistent with the finding of higher genetic diversity at intermediate latitudes and longitudes. Lower genetic diversity in populations near refugia may have resulted from reduced gene flow and successive founder effects, a phenomenon observed at the trailing edge of a shifting species range (Comps et al., 2001; Davis & Shaw, 2001; Nei, 1975).

The absence of significant IBD among populations within the western cluster was surprising. A possible reason includes gene flow that likely occurs on a small spatial scale across landscapes. Another possible factor may be historical vicariance. It is worth noting that one of the highest pairwise G_{ST} values (0.572) is between populations LOC and MAN (Figure 1), which are assigned to different clusters although separated by only 114 km. LOC is located at the edge of the Ohio/Mississippi drainage while MAN is in the Chesapeake drainage. Considerable population differentiation owing to watershed assignment among geographically proximate populations was similarly reported in *Veratrum woodii* (Zomlefer et al., 2018).

Significant IBD exists within the eastern cluster, consistent with the hypothesis that these populations resulted from dispersal from two refugia (the Atlantic and Alabama/Georgia refugia) (Figure 3) and sequential founder effects during a northward migration. Another possibility is that these populations resulted from admixture between non-contiguous refugia along the Atlantic coast during northward migration following glacial retreat, as is seen in several European trees and shrubs (Petit et al., 2003).

Often less genetic diversity is observed in northern parts of species' ranges which is thought to result from sequential founder effects during northward expansion (e.g. Conroy & Cook, 2000; Hewitt, &1999, 2000). This was not the case for *A. triloba*. Only percent of polymorphic loci decreased significantly with increased latitude. However, three measures of genetic diversity (H_E, rare alleles and MLGs) are significantly higher at mid-latitudes of *A. triloba's* range. This is consistent with patterns reported by other researchers. Petit et al. (2003) found that populations of 22 European tree species had higher genetic diversity at intermediate latitudes rather than in Pleistocene refugial areas. A similar pattern was seen in *Pinus edulis* in North America (Duran et al., 2012). It is noteworthy that both private alleles and rare alleles are also significantly more diverse at mid-longitudes for *A. triloba* than at the eastern and western margins.

The ecological niche modelling projections of future habitat suitability for *A. triloba* in the year 2070 using the RCP 6.0 and 8.5 trajectories predict an enlargement and northward expansion of suitable habitat and a more continuous distribution of optimal habitat in the eastern mid-latitudes of the range (Figures 5 and 6). The GCMs are well correlated under both RCP 6.0 (r = 0.866) and 8.5 (r = 0.902), with the largest standard deviation values confined to small clusters of cells in the southern Appalachian and Ozark Mountains (Figures S4 and S5). Optimal habitat is expected to increase by ~50% under the RCP 6.0 trajectory versus ~100% under RCP 8.5, which current data suggest is the more realistic trajectory (Schwalm et al., 2020). However, given the low fruit-set observed in populations at the northern edge of the current range (Wyatt, unpublished data) and limited vagility of contemporary seed dispersal vectors (gravity and small mammals), *A. triloba* is unlikely to expand its range rapidly and track environmental changes in lockstep. Newly colonized populations at the edges of the species' range will also likely have reduced genetic diversity due to founder effects. While the projections under RCP 6.0 suggest that *A. triloba* may retain some populations at the southern edge of

the range, these populations are less likely to persist under RCP 8.5. Although our models indicate that optimal habitat will expand significantly, between the possible loss of southern populations and environmental change that outpaces the rate of dispersal and colonization at the northern edge, the realized distribution of *A. triloba* may diminish by 2070.

5 | CONCLUSION

This study infers the locations of Pleistocene refugia and subsequent patterns of post-LGM range expansion based on modelling of optimal habitat and range-wide genetic diversity for the highly clonal and geographically widespread understorey tree, *Asimina triloba*. Ecological niche modelling infers one minor and two major refugia during the LGM located in inland southeastern Alabama/southwestern Georgia and in narrow bands of habitat along the Gulf of Mexico and Southeastern Atlantic coastlines, most of which is submerged today. Contemporary populations sampled across the range of *A. triloba* show high genetic diversity and moderate genetic structuring. Our genetic data strongly suggest the existence of eastern and western clusters divided by the Appalachian Mountains and continuing along the Tombigbee River in Alabama. Populations east of the Appalachian Mountains contain the highest level of genetic diversity and appear to be admixed populations that descended from more than one refugium. As the climate continues its warming trend, suitable habitat for this species is expected to expand northward and shrink somewhat at the southern part of the present-day range, but the rate at which areas of optimal habitat shift is very likely to outpace the rate at which *A. triloba* can colonize and establish new populations, potentially resulting in a shrinkage of the species' range by 2070.

ACKNOWLEDGEMENTS

We thank Earl Kang and Cody Barber for assistance with fieldwork, Patick McLean, Jessica Brown, Dan Bui, Dat Hoang, Kirsten Jackson and Lura Lockard for assistance with laboratory work, Rodney Dyer for analytical suggestions, Patrick Smallwood for feedback on the manuscript, three reviewers for valuable suggestions and the University of Georgia Office of the Vice President for Research for financial support awarded to D.W.T. No permits were required for this research.

DATA AVAILABILITY STATEMENT

Georeferenced occurrence data and genetic data have been deposited with Dryad (https://doi.org/10.5061/dryad.rv15dv46v).

ORCID

Dorset W. Trapnell https://orcid.org/0000-0001-7184-4700

REFERENCES

Araújo, M. B., Anderson, R. P., Barbosa, A. M., Beale, C. M., Dormann, C. F., Early, R., Garcia, R. A., Guisan, A., Maiorano, L., Naimi, B., O'Hara, R. B., Zimmermann, N. E., & Rahbek, C. (2019). Standards for distribution models in biodiversity assessments. *Science Advances*, *5*, eaat4858. https://doi.org/10.1126/sciadv.aat4858

Archuleta, C. M., Constance, E. W., Arundel, S. T., Lowe, A. J., Mantey, K. S., & Phillips, L. A. (2017). The national map seamless digital elevation model specifications. *U.S. Geological Survey Techniques and Methods*, book 11, chapter B9: 1–8.

Balloux, F., Lehmann, L., & de Meeûs, T. (2003). The population genetics of clonal and partially clonal diploids. *Genetics*, *164*, 1635–1644.

Bemmels, J., & Dick, C. W. (2018). Genomic evidence of a widespread southern distribution during the Last Glacial Maximum for two Eastern North American hickory species. *Journal of Biogeography*, *45*, 1739–1750. https://doi.org/10.1111/jbi.13358

Bermingham, E., & Avise, J. C. (1986). Molecular zoogeography of fresh-water fishes in the Southeastern United States. *Genetics*, *113*, 939–965.

Brookfield, J. F. Y. (1996). A simple new method for estimating null allele frequency from heterozygote deficiency. *Molecular Ecology*, *5*, 453–455. https://doi.org/10.1111/j.1365-294X.1996.tb00336.x

Cain, M. L., Damman, H., & Muir, A. (1998). Seed dispersal and the Holocene migration of woodland herbs. *Ecological Monographs*, *68*, 325–347.

Call, A., Sun, Y., Yu, Y., Pearman, P. B., Thomas, D. T., Trigiano, R. N., Carbone, I., & Xiang, Q. (2016). Genetic structure and post-glacial expansion of *Cornus florida* L. (Cornaceae): Integrative evidence from phylogeography, population demographic history, and species distribution modeling. *Journal of Systematics and Evolution*, *54*, 136–151.

Clark, J. S. (1998). Why trees migrate so fast: Confronting theory with dispersal biology and the paleorecord. *American Naturalist*, *152*, 204–224. https://doi.org/10.1086/286162

Clark, J. S., Fastie, C., Hurtt, G., Jackson, S. T., Johnson, C., King, G. A., Lewis, M., Lynch, J., Pacala, S., Prentice, C., Schupp, E. W., Webb, T. III, & Wyckoff, P. (1998). Reid's paradox of rapid plant migration. *BioScience*, *48*, 13–24. https://doi.org/10.2307/1313224

Clark, P. U., Dyke, A. S., Shakun, J. D., Carlson, A. E., Clark, J., Wohlfarth, B., Mitrovica, J. X., Hostetler, S. W., & McCabe, A. M. (2009). The Last Glacial Maximum. *Science*, *325*, 710–714. https://doi.org/10.1126/science.1172873

Cole, K. L., Ironside, K., Eischeid, J., Garfin, G., Duffy, P. B., & Toney, C. (2011). Past and ongoing shifts in Joshua tree distributions support future modeled range contractions. *Ecological Applications*, *21*, 137–149.

Comps, B., Gömöry, D., Letouzey, J., Thiébaut, B., & Petit, R. J. (2001). Diverging trends between heterozygosity and allelic richness during postglacial colonization in the European Beech. *Genetics*, *157*, 389–397.

Conroy, C. J., & Cook, J. A. (2000). Phylogeography of a post-glacial colonizer: *Microtus longicaudus* (Rodentia: Muridae). *Molecular Ecology*, *9*, 165–175. https://doi.org/10.1046/j.1365-294x.2000.00846.x

Cypher, B. L., & Cypher, E. A. (1999). Germination rates of tree seeds ingested by coyotes and raccoons. *American Midland Naturalist*, *142*, 71–76.

Darley-Hill, S., & Johnson, C. (1981). Acorn dispersal by the blue jay (*Cyanocitta cristata*). *Oecologia*, *50*, 231–232. https://doi.org/10.1007/BF00348043

Davis, M. 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

Davis, M. B., & Shaw, R. G. (2001). Range shifts and adaptive responses to Quaternary climate change. *Science*, *292*, 673–679. https://doi.org/10.1126/science.292.5517.673

Davis, M. B., Spear, R. W., & Shane, L. C. K. (1980). Holocene climate of New England. *Quaternary Research*, *14*, 240–250. https://doi.org/10.1016/0033-5894(80)90051-4

Delcourt, H. R. (1979). Late Quaternary vegetation history of the eastern Highland Rim and adjacent Cumberland Plateau of Tennessee. *Ecological Monographs*, *49*, 255–280. https://doi.org/10.2307/1942485

Delcourt, P. A., & Delcourt, H. R. (1987a). Late Quaternary dynamics of temperate forests: Applications of paleoecology to issues of global environmental change. *Quaternary Science Reviews*, *6*, 129–146. https://doi.org/10.1016/0277-3791(87)90030-8

Delcourt, P. A., & Delcourt, H. R. (1987b). Long-term forest dynamics of the temperate zone: A case study of late-Quaternary forests in eastern North America. *Ecological Studies 63* (pp. 107–291). Springer-Verlag.

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

Doyle, J. J., & Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*, *19*, 11–15.

Duran, K. L., Pardo, A., & Mitton, J. B. (2012). From middens to molecules: Phylogeography of the piñon pine, *Pinus edulis*. *Journal of Biogeography*, *39*, 1536–1544. https://doi.org/10.1111/j.1365-2699.2012.02704.x

Dyer, R. J. (2016). gstudio: Tools related to the spatial analysis of genetic marker data. *R Package Version*, *1*(5).

Dyer, R. J., Nason, J. D., & Garrick, R. C. (2010). Landscape modelling of gene flow: Improved power using conditional genetic distance derived from the topology of population networks. *Molecular Ecology*, *19*, 3746–3759. https://doi.org/10.1111/j.1365-294X.2010.04748.x

Dyke, A. S., & Prest, V. K. (1987). Late Wisconsinan and Holocene history of the Laurentide ice sheet. *Géographie Physique Et Quaternaire*, *41*, 237–263. https://doi.org/10.7202/032681ar

Earl, D. A., & vonHoldt, B. M. (2012). Structure harvester: A website and program for visualizing structure output and implementing the Evanno method. *Conservation Genetics Resources*, *2*, 359–361. https://doi.org/10.1007/s12686-011-9548-7

Evanno, G. S., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software structure: A simulation study. *Molecular Ecology*, *15*, 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, *37*, 4302–4315. https://doi.org/10.1002/joc.5086

Fielding, A. H., & Bell, J. F. (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, *24*, 38–49. https://doi.org/10.1017/S0376892997000088

Frascaria, N., Santi, F., & Gouyon, P. H. (1993). Genetic differentiation within and among populations of chestnut (*Castanea sativa* Mill) and wild cherry (*Prunus avium* L.). *Heredity*, *70*, 634–641. https://doi.org/10.1038/hdy.1993.91

Freeman, C. C., & Hulbert, L. C. (1985). An annotated list of the vascular flora of Konza Prairie Research Natural Area, Kansas. *Transactions of the Kansas Academy of Science*, *88*, 84–115. https://doi.org/10.2307/3627880

Gent, P. R., Danabasoglu, G., Donner, L. J., Holland, M. M., Hunke, E. C., Jayne, S. R., Lawrence, D. M., Neale, R. B., Rasch, P. J., Vertenstein, M., Worley, P. H., Yang, Z., & Zhang, M. (2011). The community climatic system model version 4. *Journal of Climate*, *24*, 4973–4991.

Giorgetta, M. A., Jungclaus, J., Reick, C. H., Legutke, S., Bader, J., Böttinger, M., Brovkin, V., Crueger, T., Esch, M., Fieg, K., Glushak, K., Gayler, V., Haak, H., Hollweg, H. D., Ilyina, T., Kinne, S., Kornblueh, L., Matei, D., Mauritsen, T., … Stevens, B. (2013). Climate and carbon changes from 1850 to 2100 in MPI-ESM simulations for the coupled model intercomparison project phase 5. *Journal of Advances in Modelling Earth Systems*, *5*, 572–597.

Gonzales, E., Hamrick, J. L., & Smouse, P. E. (2008). Comparison of clonal diversity in mountain and piedmont populations of *Trillium cuneatum* (Melanthiaceae-Trilliaceae), a forest understory species. *American Journal of Botany*, *95*, 1254–1261.

Goodrich, K. R., Zjhra, M., Ley, C. A., & Raguso, R. A. (2006). When flowers smell fermented: The chemistry and ontogeny of yeasty floral scent in pawpaw (*Asimina triloba*: Annonaceae). *International Journal of Plant Sciences*, *167*, 33–46.

Goudet, J. (1995). FSTAT (Version 1.2): A computer program to calculate F-statistics. *Journal of Heredity*, *86*, 485–486. https://doi.org/10.1093/oxfordjournals.jhered.a111627

Hamrick, J. L., & Godt, M. J. W. (1996). Effects of life history traits on genetic diversity in plant species. *Philisophical Transactions of the Royal Society of London: Biological Sciences*, *351*, 1291–1298.

Hamrick, J. L., Godt, M. 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

Hauswaldt, S. J., & Glenn, T. C. (2003). Microsatellite DNA loci from the diamondback terrapin (*Malaclemys terrapin*). *Molecular Ecology Notes*, *3*, 174–176. https://doi.org/10.1046/j.1471-8286.2003.00388.x

Hewitt, G. M. (1996). Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*, *58*, 247–276. https://doi.org/10.1006/bijl.1996.0035

Hewitt, G. M. (1999). Post-glacial recolonization 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. M. (2000). The genetic legacy of the Quaternary ice ages. *Nature*, *405*, 907–913. https://doi.org/10.1038/35016000

Hof, C., Levinksy, I., Araujo, M. B., & Rahbek, C. (2011). Rethinking species' ability to cope with rapid climate change. *Global Change Biology*, *17*, 2987–2990. https://doi.org/10.1111/j.1365-2486.2011.02418.x

Hoppe, K., & Koch, P. L. (2007). Reconstructing the migration patterns of late Pleistocene mammals from northern Florida, USA. *Quaternary Research*, *68*, 347–352. https://doi.org/10.1016/j.yqres.2007.08.001

Huang, H., Layne, D. R., & Kubisiak, T. L. (2000). RAPD inheritance and diversity in Pawpaw (*Asimina triloba*). *Journal of the American Society of Horticultural Science*, *125*, 454–459. https://doi.org/10.21273/JASHS.125.4.454

Huang, H., Layne, D. R., & Riemenschneider, D. E. (1998). Genetic diversity and geographic differentiations in pawpaw [*Asimina triloba* (L.) Dunal] populations from nine states as revealed by allozyme analysis. *Journal of the American Society Horticultural Science*, *123*, 635–641.

Jackson, S. T., & Givens, C. R. (1994). Late Wisconsinan vegetation and environment of the Tunica Hills Region, Louisiana/Mississippi. *Quaternary Research*, *41*, 316–325. https://doi.org/10.1006/qres.1994.1036

Johnson, W. C., & Webb, T. (1989). The role of blue jays (*Cyanocitta cristata* L.) in the postglacial dispersal of fagaceous trees in eastern North America. *Journal of Biogeography*, *16*, 561–571.

Jordano, P., García, C., Godoy, J. A., & García-Castaño, J. L. (2007). Differential contribution of frugivores to complex seed dispersal patterns. *Proceedings of the National Academy of Sciences*, *104*, 3278–3282. https://doi.org/10.1073/pnas.0606793104

Keener, C., & Kuhns, E. (1998). The impact of Iroquoian populations on the northern distribution of pawpaws in the Northeast. *North American Archaeologist*, *18*, 327–342. https://doi.org/10.2190/AT0W-VEDT-0E0P-W21V

Kolb, C. R., & Fredlund, G. G. (1981). Palynological studies: Vacherie and Rayburn's Domes, north Louisiana salt dome basin. Top Report E530–02200-T-2. *Institute for Environmental Studies*. Louisiana State University.

Kral, R. (1997). Annonaceae Jussieu - Custard-apple Family. In E. Committee (Ed.), *Flora of North America North of Mexico* (pp. 11–20). Oxford University Press.

Li, C. C., & Horvitz, D. G. (1953). Some methods of estimating the inbreeding coefficient. *American Journal of Human Genetics*, *5*, 107–117.

Li, P. S., Thomas, D. C., & Saunders, M. K. (2017). Historical biogeography and ecological niche modelling of the *Asimina – Disepalum* clade

(Annonaceae): Role of ecological differentiation in Neotropical-Asian disjunctions and diversification in Asia. *Evolutionary Biology*, *17*, 188–206.

Little, E. L. Jr. (1977). *Atlas of the United States trees*, *Volume 4: minor Eastern hardwoods*. Pub. 1342. : US Department of Agriculture. 17 p 230 maps.

Lu, L., Pomper, K. W., Lowe, J. D., & Crabtree, S. B. (2011). Genetic variation in pawpaw cultivars using microsatellite analysis. *Journal of the American Society of Horticultural Science*, *136*, 415–521. https://doi.org/10.21273/JASHS.136.6.415

MacDougall, A. (2003). Did Native Americans influence the northward migration of plants during the Holocene? *Journal of Biogeography*, *30*, 633–647. https://doi.org/10.1046/j.1365-2699.2003.00842.x

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

Miller, D. M., & Parker, K. C. (2009). The impacts of the Appalachian Mountains on the post-glacial migration pathways and gene flow of sugar maple (*Acer saccharum*). *Physical Geography*, *30*, 89–104.

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

Murphy, J. L. (2001). Pawpaws, persimmons, and 'possums: On the natural distribution of pawpaws in the northeast. *North American Archaeologist*, *22*, 93–115. https://doi.org/10.2190/TT5Y-YAXJ-0KAU-GUFK

Nei, M. (1973). Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences*, *70*, 3321–3323. https://doi.org/10.1073/pnas.70.12.3321

Nei, M., Maruyama, T., & Chakraborty, R. (1975). The bottleneck effect and genetic variability in populations. *Evolution*, *29*, 1–10. https://doi.org/10.1111/j.1558-5646.1975.tb00807.x

Paetkau, D., & Strobeck, C. (1994). Microsatellite analysis of genetic variation in black bear populations. *Molecular Ecology*, *3*, 489–495. https://doi.org/10.1111/j.1365-294X.1994.tb00127.x

Pappert, R. A., Hamrick, J. L., & Donovan, L. A. (2000). Genetic variation in *Pueraria lobata* (Fabaceae), an introduced, clonal, invasive plant of the southeastern United States. *American Journal of Botany*, *87*, 1240–1245.

Parker, K. C., Hamrick, J. L., Parker, A. J., & Stacy, E. A. (1997). Allozyme diversity in *Pinus virginiana* (Pinaceae): Intraspecific and interspecific comparisons. *American Journal of Botany*, *84*, 1372–1382.

Parker, P. G., Snow, A. A., Schug, M. D., Booton, G. C., & Fuerst, P. A. (1998). What molecules can tell us about populations: Choosing and using a molecular marker. *Ecology*, *79*, 361–382. https://doi.org/10.2307/176939

Peakall, R., & Smouse, P. E. (2006). GenAlEx 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, *6*, 288–295. https://doi.org/10.1111/j.1471-8286.2005.01155.x

Peakall, R., & Smouse, P. E. (2012). GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research – an update. *Bioinformatics*, *28*, 2537–2539. https://doi.org/10.1093/bioinformatics/bts460

Perrotti, A. G. (2018). Pollen and *Sporormiella* evidence for terminal Pleistocene vegetation change and megafaunal extinction at Page-Ladson, Florida. *Quaternary International*, *466*, 256–268.

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.

Petit, R. J., Aguinagalde, I., de Beaulieu, J. L., Bittkau, C., Brewer, S., Cheddadi, R., Ennos, R., Fineschi, S., Grivet, D., Lascoux, M., Mohanty, A., Müller-Starck, G., Demesure-Musch, B., Palmé, A., Martín, J. P., Rendell, S., & 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

Philander, S. G. (2008). *Encyclopedia of Global Warming and Climate Change* (pp. 519–522). SAGE.

Phillips, S. J., & Dudík, M. (2008). Modeling of species distributions with Maxent: New extensions and a comprehensive evaluation. *Ecography*, *31*, 161–175. https://doi.org/10.1111/j.0906-7590.2008.5203.x

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

Pomper, K. W., Lowe, J. D., Lu, L., Crabtree, S. B., Dutta, S., Schneider, K., & Tidwell, J. (2010). Characterization and identification of pawpaw cultivars and advanced selections by simple sequence repeat markers. *Journal of the American Society for Horticulture Science*, *135*, 143–149. https://doi.org/10.21273/JASHS.135.2.143

Potzger, J. E., & Tharp, B. C. (1947). Pollen profile from a Texas Bog. *Ecology*, *28*, 274–280. https://doi.org/10.2307/1930513

Pritchard, J. K., Stephens, J. K., & Donnelly, M. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*, 945–959.

R Core Team (2013). R: A Language and Environment for Statistical Computing.

Raymond, M., & Rousset, F. (1995). GENEPOP (version 1.2): Population genetics software for exact tests and ecumenicism. *Journal of Heredity*, *86*, 248–249. https://doi.org/10.1093/oxfordjournals.jhered.a111573

Reid, C. (1899). *The Origin of The British Flora*, (23–28). Dulau.

Rich, F. J., Vance, R. K., & Rucker, C. R. (2015). The palynology of Upper Pleistocene and Holocene sediments from the eastern shoreline and Central Depression of St. Catherines Island, Georgia, USA. *Palynology*, *39*, 234–247. https://doi.org/10.1080/01916122.2014.945666

Rousset, F. (1997). Genetic differentiation and estimation of gene flow from *F*-statistics under isolation by distance. *Genetics*, *145*, 1219–1228.

Ruiz-Sanchez, E., & Ornelas, J. F. (2014). Phylogeography of *Liquidambar styraciflua* (Altingiaceae) in Mesoamerica: Survivors of a Neogene widespread temperate forest (or cloud forest) in North America? *Ecology and Evolution*, *4*, 311–328.

Schwalm, C. R., Glendon, S., & Duffy, P. B. (2020). RCP8.5 tracks cumulative CO_2 emissions. *Proceedings of the National Academy of Sciences*, *117*, 19656–19657.

Skellam, J. G. (1951). Random dispersal in theoretical populations. *Biometrika*, *38*, 196–218. https://doi.org/10.1093/biomet/38.1-2.196

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

Stoeckel, S., Grange, J., Fernández-Manjarres, J. F., Bilger, I., Frascaria-Lacoste, N., & Mariette, S. (2006). Heterozygote excess in a self-incompatible and partially clonal forest tree species – *Prunus avium* L. *Molecular Ecology*, *15*, 2109–2118.

Title, P. O., & Bemmels, J. B. (2017). 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

Trapnell, D. W., Schmidt, J. P., Quintana-Ascencio, P. F., & Hamrick, J. L. (2007). Genetic insights into the biogeography of the southeastern endemic, Ceratiola ericoides (Empetraceae). *Journal of Heredity*, *98*, 587–593.

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.

Vuuren, D. P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., Hurtt, G. C., Kram, T., Krey, V., Lamarque, J., Masui, T., Meinshausen, M., Nakicenovic, N., Smith, S. J., & Rose, S. K. (2011). The representative concentration pathways: An overview. *Climatic Change*, *109*, 5–31. https://doi.org/10.1007/s10584-011-0148-z

Watanabe, S., Hajima, T., Sudo, K., Nagashima, T., Takemura, T., Okajima, H. Nozawa, T., Kawase, H., Abe, M., Yokohata, T., Ise, T., Sato, H., Kato, E., Takata, K., Emori, S., & Kawamiya, M. (2011). MIROC-SEM 2010: Model description and basic results of CMIP5-20c3m experiments. *Geoscientific Model Development*, *4*, 845–872.

Watts, W. A. (1975). Vegetation record for the last 20,000 years from a small marsh on Lookout Mountain, northwestern Georgia. *Geological Society of America Bulletin*, *86*, 287–291. https://doi.org/10.1130/0016-7606(1975)86<287:VRFTLY>2.0.CO;2

Watts, W. A., Hansen, B. C. S., & Grimm, E. C. (1992). Camel Lake: A 40,000-yr record of vegetational and forest history from Northwest Florida. *Ecology*, *73*, 1056–1066. https://doi.org/10.2307/1940180

Weakley, A. S. Flora of The Southern and Mid-Atlantic States. The University of North Carolina Herbarium, Chapel Hill. Working draft of 30. November 2012.

Willard, D. A., Bernhardt, C. E., Korejwo, D. A., & Meyers, S. R. (2005). Impact of millennial-scale Holocene climate variability on eastern North American terrestrial ecosystems: Pollen-based climatic reconstruction. *Global and Planetary Change*, *47*, 17–35. https://doi.org/10.1016/j.gloplacha.2004.11.017

Williams, J. W., Shuman, B. N., Webb, T., Bartlein, P. J., & Leduc, P. L. (2004). Late-Quaternary vegetation dynamics in North America: Scaling from taxa to biomes. *Ecological Monographs*, *74*, 309–334. https://doi.org/10.1890/02-4045

Willson, M. F. (1993). Mammals as seed-dispersal mutualists in North America. *Oikos*, *67*, 159–176. https://doi.org/10.2307/3545106

Willson, M. F., & Schemske, D. W. (1980). Pollinator limitation, fruit production, and floral display in Pawpaw (*Asimina triloba*). *Bulletin of the Torrey Botanical Club*, *107*, 401–408. https://doi.org/10.2307/2484160

Wright, S. (1922). Coefficients of inbreeding and relationship. *The American Naturalist*, *56*, 330–338. https://doi.org/10.1086/279872

Wright, S. (1943). Isolation by distance. *Genetics*, *28*, 114–138.

Wright, S. (1951). The genetical structure of populations. *Annals of Eugenics*, *15*, 323–354. https://doi.org/10.1111/j.1469-1809.1949.tb02451.x

Yeager, L. E., & Elder, W. H. (1945). Pre- and post-hunting season foods of raccoons on an Illinois goose refuge. *Journal of Wildlife Management*, *9*, 48–53. https://doi.org/10.2307/3795946

Zalasiewicz, J., Williams, M., Haywood, A., & Ellis, M. (2011). The Anthropocene: A new epoch of geologic time? *Philosophical Transactions of the Royal Society*, *369*, 835–841.

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

Zomlefer, W. B., Comer, J. R., Lucardi, R. D., Hamrick, J. L., & Allison, J. R. (2018). Distribution and genetic diversity of the rare plant *Veratrum woodii* (Liliales: Melanthiaceae) in Georgia: A preliminary study with AFLP fingerprint data. *Systematic Botany*, *43*, 858–869.

BIOSKETCH

Graham Wyatt recently earned his Ph.D. under the mentorship of D.W.T. and is currently a lecturer at the University of Georgia. His research interests include pollination ecology, historical processes responsible for maintaining genetic diversity among populations and conservation genetics of rare/endangered plant species.

J.L. Hamrick is an Emeritus Professor of Plant Biology. He is interested in plant evolution, plant population genetics and phylogeography. He is especially interested in how the dispersal of plant pollen and seeds interact with landscape environmental features to effect patterns of genetic diversity.

Dorset Trapnell is an Associate Professor interested in using the genetic signatures contained within contemporary populations to develop insights into historical processes that have shaped the evolutionary trajectory of plant species. Her laboratory employs population genetics to study the evolutionary biology of natural plant populations within spatial and temporal contexts.

Author contributions: This work constituted a part of G.E.W.'s dissertation research. G.E.W. and D.W.T. jointly developed the questions and designed the study. G.E.W. conducted the field work, collected genetic data and with D.W.T. performed analyses. All three authors contributed to the interpretation of the results, and the writing and editing of the manuscript.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Wyatt GE, Hamrick JL, Trapnell DW. Ecological niche modelling and phylogeography reveal range shifts of pawpaw, a North American understorey tree. *J Biogeogr*. 2021;48:974–989. https://doi.org/10.1111/jbi.14054