Tree Genetics & Genomes (2019) 15: 74
https://doi.org/10.1007/s11295-019-1380-3

ORIGINAL ARTICLE

Extant population genetic variation and structure of eastern white pine (*Pinus strobus* L.) in the Southern Appalachians

Thomas D. Whitney[1,2,3] · Kamal J. K. Gandhi[1] · J. L. Hamrick[4] · Rima D. Lucardi[2]

Received: 9 June 2019 / Revised: 14 August 2019 / Accepted: 21 August 2019 / Published online: 10 September 2019

Abstract

Eastern white pine (*Pinus strobus* L.) is a widespread conifer in eastern North America. A novel dieback phenomenon, as well as increasing global temperatures contributing to the contraction of suitable habitat, is threatening this species' long-term persistence in the Southern Appalachian Mountains. This southern extent of its current range is where *P. strobus* is hypothesized to have survived in refugial populations during the last glacial maximum. As a result, extant populations located here may have higher levels of ancestral genetic diversity, and by extension, adaptive potential. We genotyped 432 *P. strobus* individuals from 23 sites throughout the Southern Appalachians and another 34 individuals from two reference populations in the northern USA, using 10 established microsatellite markers. Levels of genetic diversity in the southern portion of the range were comparable but not higher than reference northern populations. There was an overall heterozygote deficiency and high inbreeding coefficient ($F_{IS} = 0.173$); however, these values were comparable to published research of *P. strobus* throughout the northern range. There was low overall genetic differentiation ($F_{ST} = 0.055$) among populations in the Southern Appalachians and population structure was best explained by ecoregions. These results show that *P. strobus* in the Southern Appalachians is a fairly heterogenous and admixed species with relatively high genetic diversity mostly partitioned within populations. The Southern Appalachians remains an important area for *P. strobus* conservation, but not necessarily because it is genetically unique.

Keywords Conservation · Genetic differentiation · Genetic diversity · Last glacial maximum · Microsatellites · Refugium

Introduction

The standing genetic variation of a species is essential for adaptation when faced with novel evolutionary pressures, because beneficial alleles that have been filtered through natural selection in the past are readily available at high frequencies (Barrett and Schluter 2008). Diverse gene pools can be especially important for the adaptive potential of long-lived species, such as forest trees, which often endure numerous disturbances over their lifespan (Schaberg et al. 2008). However, the rates of broad, environmental change during the Anthropocene (e.g., climate change, increased colonization by non-native, exotic, and/or invasive pests and pathogens) has become so rapid that some trees, despite their high clinal variation and versatility under a variety of micro-habitat conditions, will likely endure adaptational lags due to their long generation times (Aitken et al., 2008; Alfaro et al. 2014; Kuparinen et al. 2010; Savolainen et al. 2004, 2007). Hence, it is imperative to gather foundational data of the extant genetic variation of forest trees, especially for those species and/or populations that face imminent threats and may have particular conservation value (Potter et al. 2017).

Eastern white pine (*Pinus strobus* L.) is an important ecological and economic tree species in eastern North

Communicated by F. P. Guerra

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11295-019-1380-3) contains supplementary material, which is available to authorized users.

✉ Thomas D. Whitney
thomasdantaswhitney@gmail.com

[1] Present address: Puyallup Research and Extension Center, Washington State University, Puyallup, WA 98371, USA

[2] USDA Forest Service, Southern Research Station, 320 E. Green Street, Athens, GA 30602, USA

[3] WSU Puyallup Research and Extension Center, Washington State University, Puyallup, WA 98371, USA

[4] Department of Plant Biology, 120 Carlton Street, Athens, GA 30602, USA

America, occurring from Minnesota, USA, to Nova Scotia, Canada, and south into northern Georgia, USA, along the Appalachian Mountain range. Its broad distribution is indicative of its versatility, thriving as both an early- and late-successional species, growing at both low (sea-level) and high elevations (up to 1220 m), and outcompeting hardwoods in both riparian valleys and nutrient-poor uplands (Abrams 2001; Wendel and Smith 1990). Due to several favorable ecophysiological traits, *P. strobus* has been the most widely planted tree in eastern North America, often for use in erosion control and reclamation of previously mined or cultivated land (Hepp et al. 2015; Wendel and Smith 1990). Growing in pure or mixed stands as a scattered super-canopy tree, it provides resources and unique habitat qualities for wildlife (Abrams et al. 1995; Rogers and Lindquist 1992). With the continued decline of eastern hemlock (*Tsuga canadensis* (L.) Carriére) due to the invasive hemlock woolly adelgid (*Adelges tsugae* Annand), the ecological and economic importance of *P. strobus*—a frequently co-occurring canopy species—has become even more significant for conservation purposes (Lovett et al. 2006).

In recent centuries, *P. strobus* has endured significant and recurring disturbances due to anthropogenic (e.g., intensive logging and fire exclusion) and biotic factors (e.g., white pine weevil, *Pissodes strobi* Peck, and white pine blister rust, *Cronartium ribicola* J.C. Fisch) (Costanza et al. 2018). Since the beginning of the twenty-first century, a novel and complex dieback phenomenon has also emerged in *P. strobus*, which has renewed concern for the immediate and long-term health of the species. Across its North American range, unique symptoms—including the presence of pathogenic fungal cankers, needle loss, branch dieback, and stem girdling—have arisen in conjunction with native biotic agents: a canker-forming pathogen (*Caliciopsis pinea* Peck) and a scale insect (*Matsucoccus macrocicatrices* Richards) (summarized in Costanza et al. 2018). Furthermore, climate prediction models suggest increasing temperatures are shifting the suitable climatic envelope of *P. strobus* northward at a rapid rate, which would result in significant range contraction at the expense of populations in the lower elevations of the Southern Appalachians (Joyce and Rehfeldt 2013). At the southern extent of its range in the Blue Ridge and Ridge and Valley ecoregions of northwest Georgia and western North Carolina, refugial populations are hypothesized to have survived during the last glacial maximum (LGM), according to fossilized pollen (Davis 1983; Jackson et al. 2000) and molecular phylogeography (Nadeau et al. 2015; Zinck and Rajora 2016). These populations likely served as the source populations for postglacial recolonization northward ~ 22,000 years ago. Thus, the trees currently located in these areas likely harbor the highest levels of extant genetic diversity (Hewitt 1999; but see Petit et al. 2003). South-to-north clines in genetic diversity, consistent with a "southern richness and northern purity" hypothesis (Hewitt 2000), have been reported in other plants and animals (Soltis et al. 1997; Talarico et al. 2019; Wielstra et al. 2013). If *P. strobus* follows a similar pattern, then its most vulnerable populations in the Southern Appalachians may hold exceptional adaptive potential and conservation value.

Prior to 2015, no study had investigated the population genetics of *P. strobus* south of Pennsylvania in the Southern Appalachians region (defined herein as Georgia, Kentucky, North Carolina, South Carolina, Tennessee, Virginia, and West Virginia). As part of two recent phylogeographic studies, Nadeau et al. (2015) and Zinck and Rajora (2016) examined *P. strobus* populations across its entire North American range to reconstruct possible postglacial recolonization routes. However, only 10 of 133 (Nadeau et al. 2015) and 2 of 33 (Zinck and Rajora 2016) sampled populations were located in the Southern Appalachians. The majority of past research on the genetic diversity and structure of *P. strobus* comes from the Canadian and northern USA range (e.g., Beaulieu and Simon 1994; Epperson and Chung 2001; Mehes et al. 2009; Rajora et al. 2000). This geographic imbalance leaves a gap in our knowledge and understanding of the extant genetic diversity and adaptive potential in the area where *P. strobus* survived during and subsequently recolonized after the LGM.

This study aimed to elucidate the genetic diversity, variability, and population structure of *P. strobus* in the Southern Appalachians. Its southern extent remains the most understudied, eminently at-risk, and potentially important part of its range for conservation. We used 10 highly utilized, highly polymorphic microsatellite markers (Echt et al. 1996) to genotype 432 individual *P. strobus* trees from 23 populations within its Southern Appalachian range. Of these populations, 18 were sampled along the main Appalachian Mountain chain where *P. strobus* is fairly contiguous from Virginia to Georgia. The remaining five populations were sampled west of the mountains in Tennessee, Kentucky and West Virginia, where they were found in highly isolated, possibly relict, patches. To understand whether *P. strobus* populations in the Southern Appalachians may be genetic outliers within the range, we made qualitative comparisons to northern populations reported in previous studies, as well as our own sampled and genotyped reference populations (34 individuals) located in New Hampshire and Michigan. Additionally, we modeled the suitable climatic envelope of *P. strobus* at the LGM (~ 22,000 years ago), mid-Holocene (~ 6,000 years ago), and the present day, to contextualize expected and emergent patterns from our genetic analyses. In

elucidating this foundational knowledge gap in *P. strobus* biology, we hoped to determine the similarity, or lack thereof, between the northern North American and Southern Appalachian populations in: (1) the extant genetic diversity within and among populations, (2) levels and patterns of genetic differentiation, and (3) degree of genetic isolation in geographically isolated populations.

Methods

Sample collection

We sampled *P. strobus* needle tissue between 2014 and 2016 from 432 individual trees from 23 sites (populations, hereafter) throughout the Southern Appalachians in the USA, in the states of Georgia, Kentucky, North Carolina, South Carolina, Tennessee, Virginia, and West Virginia, as well as 14 individuals from one reference population in Michigan, and 20 individuals from another reference population in New Hampshire ($N = 466$ individuals from 25 total populations) (Fig. 1, Table 1). The mean size of sampled trees per site ranged from small saplings to poletimber, which ranged from 2 to 15.9 cm diameter at breast height (DBH). Two of the populations were located in state forests, and the remaining populations were located within National Forests (USDA Forest Service). Sites were spaced by at least 10 km, a distance further than *P. strobus* pollen usually disperses (Epperson and Chung 2001). We sampled 14 to 20 individual trees per site, ensuring each tree was $\geq$ 10 m apart to mitigate the sampling of siblings. Within 48 h of collection, needles were preserved at − 80 °C.

Molecular analyses

To assess genetic diversity, variability, and population structure of *P. strobus*, we utilized 12 nuclear microsatellite markers—RPS1b, RPS2, RPS6, RPS12, RPS20, RPS25b, RPS34b, RPS39, RPS50, RPS60, RPS84, and RPS127—previously designed by Echt (1996) (Table S1). These markers have been used extensively in past population genetics studies of *P. strobus* (Chhatre and Rajora 2014; Marquardt et al. 2007; Marquardt and Epperson 2004; Mehes et al. 2009; Mandak et al. 2013; Rajora et al. 2000; Walter and Epperson 2004; Zinck and Rajora 2016).

All DNA extractions utilized the Qiagen DNeasy Plant Tissue Extraction Kit (Qiagen Inc., Chatsworth, California, USA) following the manufacturer's instructions. For each tree, five mature needles, weighed to approximately 0.1 g, were homogenized with a 3.2-mm diameter stainless steel bead within individual 2.0-μL microcentrifuge tubes placed within a Mini-Beadbeater-96 (BioSpec Products, Inc., Bartlesville, OK, USA). Genomic DNA quantity and quality was assessed with a BioPhotometer Plus (Eppendorf, Hamburg, Germany). The majority of polymerase chain reactions (PCRs) to amplify microsatellite regions of interest were conducted in volumes of 12.5 μL, consisting of 2 μL 10× Takara buffer (Takara Bio Inc., Shiga, Japan), 1.5 μL 2.5 mM dNTPs, 0.25 μL both 10 μM forward and reverse primer, 0.1 μL 5 U/mL Takara *Ex Taq*™ polymerase (Takara), and 1–10 ng template DNA. For reactions amplifying RPS12 and RPS50, 0.1 μL 20 mg per mL bovine serum albumin (BSA) was added. The reactions that amplified the loci RPS1b and RPS2 were multiplexed in 25-μL volumes, consisting of 5 μL 10× Takara buffer, 4 μL 2.5 mM dNTPs, 0.15 μL both RPS1b primers and 0.2 μL both RPS2 primers, 0.25 μL 5 U/mL Takara *Ex Taq*™ polymerase, and 1–10 ng template DNA. Each forward primer for every reaction was tagged with a fluorophore: FAM, NED, PET, or VIC (Life Technologies, Carlsbad, CA, USA). Each reverse primer had a GTTT pig-tail attached (Integrated DNA Technologies, Coralville, IA, USA) to reduce stutter during genotyping (Brownstein et al. 1996). Mastercycler Pro S Thermal Cyclers (Eppendorf) were used for PCR amplification under the following touchdown protocol: 2 cycles of 94 °C for 1 min, 60 °C for 1 min, and 70 °C for 35 s; 18 cycles of 93 °C for 1 min, 59–50.5 °C for 45 s decreasing 0.5 °C every cycle, and 70 °C for 35 s; 20 cycles of 92 °C for 30s, 50 °C for 30 s, and 72 °C for 60 s; and a final extension of 70 °C for 5 min. Amplicon sizes were determined on a 3730 capillary sequencer (Applied Biosystems, Foster City, CA, USA) at the Arizona State University DNA Core Lab, with GeneScan LIZ 500 as the size standard (Life Technologies). Allele sizes were visualized and scored using the microsatellite plug-in for GENEIOUS version 10.2.3 (Biomatters, Auckland, New Zealand).

Statistical analyses

Genetic diversity and variability

We tested all pairs of loci for linkage disequilibrium using the probability test, and we tested every locus-population combination for deviations from Hardy-Weinberg equilibrium with exact tests in GENEPOP version 4.2 (Raymond and Rousset 1995). We also calculated null allele frequencies of each locus for each population using the Brookfield (1996) method in GENEPOP and categorized the overall means as either negligible ($r < 0.05$), moderate ($0.05 \leq r < 0.2$), or large ($r \geq 0.2$) according to Chapuis and Estoup (2007).

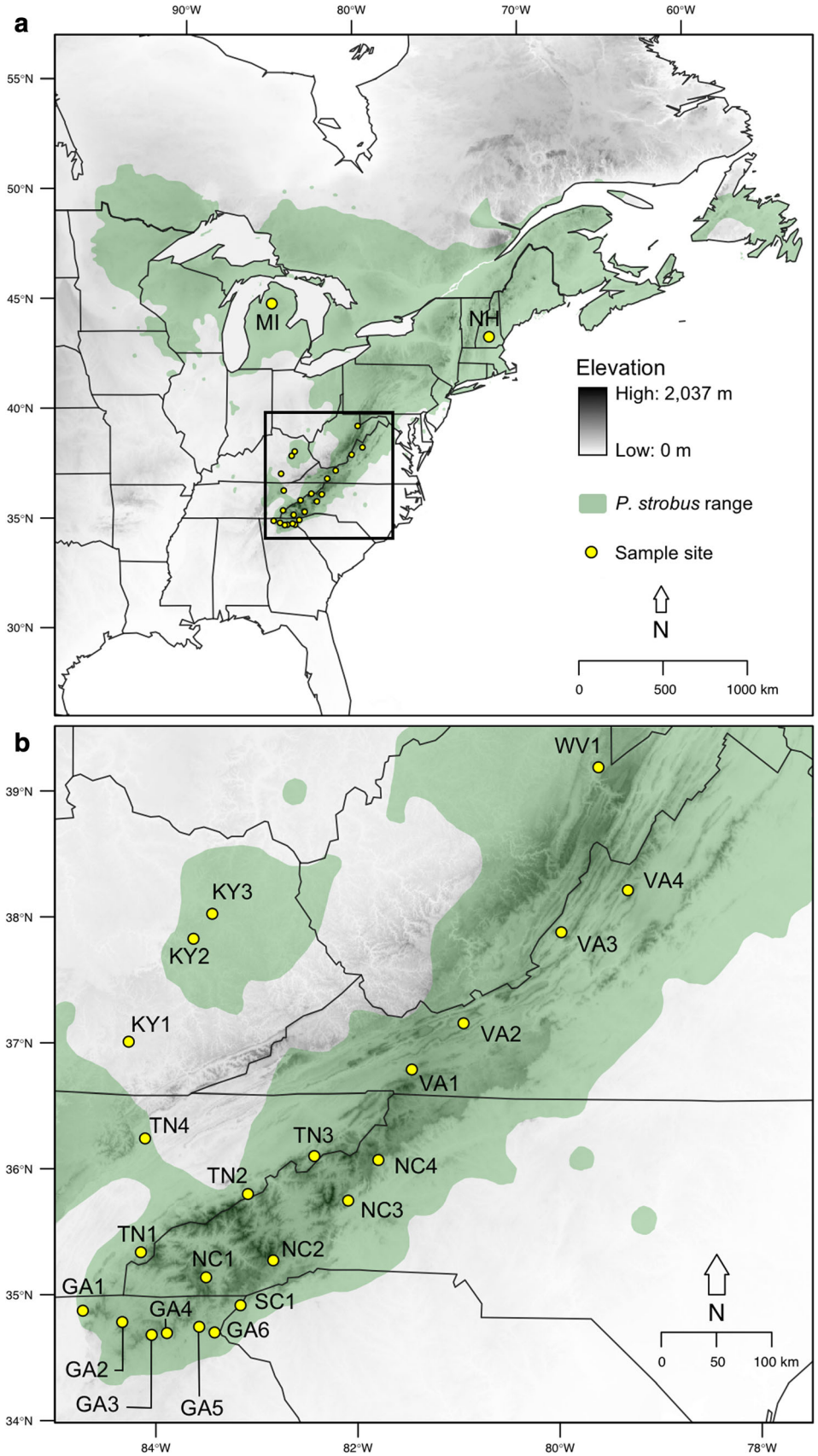

Fig. 1 **a** Sampled population sites of *P. strobus* for molecular analyses. Shaded green area represents the range extent (Little, 1971). Inset (**b**) shows view of sampled populations within the Southern Appalachian region

Genetic diversity was estimated using the effective number of alleles (A_E) and number of private alleles (A_P) in GENALEX version 6.503 (Peakall and Smouse 2006, 2012). Rarefied allelic richness (A_R) was calculated in R version 3.5.1 (R Core Team 2018) using the package "hierfstat" (Goudet 2005). Observed heterozygosity (H_O), expected heterozygosity (H_E), and inbreeding coefficients (F_{IS}) were calculated using the Bayesian Individual Inbreeding Model (IIM) in INEST version 2.2 (Chybicki and Burczyk 2009), which takes into account the presence of null alleles and the upward bias they can cause to F_{IS} estimates (Campagne et al. 2012). The Gibbs sampler in INEST utilized 20,000 burn-in steps, followed by 200,000 cycles, keeping every 1,000th update.

We conducted generalized linear models (GLMs) with identity link functions in R to determine if the surrounding density of conspecifics, a measurement of the degree of isolation, influenced metrics of genetic diversity for our *P. strobus* populations in the Southern Appalachians only. We first used data from the USDA Forest Service Forest Inventory and Analysis (FIA) to create a density raster in R using the "raster" package (Hijmans 2017), where each pixel (250 × 250 m)

Table 1 Population information for microsatellite analysis of *P. strobus*, including location data and estimates of genetic diversity indices

Site ID	Site name	County, state	Ecoregion[a]	Latitude	Longitude	Elev. (m)	Number	A_N	A_E	A_R	A_P	H_O	H_E	F_{IS}
GA1	Mill Creek	Murray, GA	BR	34.8723	−84.7234	291	20	46	2.33	3.99	1	0.504	0.573	0.096
GA2	Rock Creek	Gilmer, GA	BR	34.7821	−84.3334	596	20	46	2.13	3.95	2	0.482	0.543	0.070
GA3	Canada Creek	Union, GA	BR	34.6803	−84.0425	845	15	44	2.10	3.94	2	0.424	0.512	0.117
GA4	Boggs Creek	Lumpkin, GA	BR	34.6940	−83.8914	554	15	47	2.12	4.24	2	0.481	0.591	0.139
GA5	Raper Creek	Habersham, GA	BR	34.7442	−83.5721	510	15	42	2.11	3.89	0	0.457	0.578	0.194
GA6	Panther Creek	Habersham, GA	BR	34.6995	−83.4199	486	18	59	2.57	5.09	4	0.478	0.645	0.226
SC1	Chattooga River	Oconee, SC	BR	34.9154	83.1630	518	20	50	2.53	4.24	1	0.501	0.623	0.208
NC1	Muskrat Valley	Macon, NC	BR	35.1380	− 83.5038	808	19	49	2.70	4.19	0	0.471	0.595	0.176
NC2	Long Branch Trail	Transylvania, NC	BR	35.2729	− 82.8393	926	20	59	2.49	4.65	2	0.436	0.620	0.270
NC3	Little Buck Creek	McDowell, NC	BR	35.7461	− 82.0975	551	20	48	2.33	3.94	0	0.469	0.548	0.099
NC4	Wilson Creek	Avery, NC	BR	36.0696	− 81.7990	890	19	48	2.67	4.23	0	0.477	0.588	0.160
TN1	Tellico Plains	Monroe, TN	BR	35.3375	− 84.1488	625	20	49	2.87	4.40	0	0.532	0.605	0.109
TN2	Green Corner	Cocke, TN	BR	35.7986	− 83.0899	688	19	50	2.47	4.31	0	0.439	0.581	0.204
TN3	Nolichucky River	Unicoi, TN	BR	36.1005	− 82.4328	543	20	52	2.86	4.34	0	0.518	0.613	0.112
TN4	Norris Dam State Park	Campbell, TN	RV	36.2409	− 84.1077	411	20	49	2.71	4.12	0	0.446	0.594	0.199
VA1	Pugh Mountain	Smyth, VA	BR	36.7871	− 81.4679	953	19	56	2.82	4.71	2	0.578	0.668	0.114
VA2	Price Ridge	Bland, VA	RV	37.1551	− 80.9561	792	18	45	2.94	4.12	0	0.479	0.584	0.189
VA3	Falling Springs	Alleghany, VA	RV	37.8771	− 79.9884	451	19	57	3.04	4.76	6	0.421	0.632	0.296
VA4	Deerfield	Augusta, VA	RV	38.2113	− 79.3297	634	20	47	2.66	4.11	1	0.325	0.543	0.353
WV1	Horseshoe Run	Tucker, WV	RV	39.1878	− 79.6211	676	19	38	2.04	3.47	1	0.417	0.496	0.124
KY1	Laurel Road	Laurel, KY	SWA	37.0085	− 84.2697	381	19	49	2.23	4.21	0	0.407	0.586	0.254
KY2	Sheltowee Trace Connector	Menifee, KY	WAP	37.8256	− 83.6284	228	20	46	2.22	3.88	0	0.420	0.544	0.194
KY3	Pretty Ridge	Menifee, KY	WAP	38.0242	− 83.4430	340	18	36	2.31	3.36	0	0.555	0.589	0.079
Mean: Southern Appalachians							18.8	48	2.49	4.18	1.0	0.466	0.585	0.173
NH	Mast Yard State Forest	Merrimack, NH	NECZ	43.2398	− 71.6529	12	20	57	2.86	4.63	3	0.614	0.664	0.046
MI	Poverty Perch	Crawford, MI	NLF	44.7569	− 84.8366	360	14	46	2.67	4.39	1	0.597	0.640	0.073
Mean: all populations							18.6	49	2.51	4.21	1.1	0.477	0.590	0.164

A_N number of alleles, A_E effective number of alleles, A_R rarefied allelic richness, A_P number of private alleles, H_O observed heterozygosity, H_E expected heterozygosity, F_{IS} inbreeding coefficients, *GA* Georgia, *KY* Kentucky, *MI* Michigan, *NC* North Carolina, *NH* New Hampshire, *SC* South Carolina, *TN* Tennessee, *VA* Virginia, *WV* West Virginia

[a] U.S. Environmental Protection Agency, 2013, Level III ecoregions: *BR* Blue Ridge, *NECZ* Northeastern Coastal Zone, *NLF* Northern Lakes and Forests, *RV* Ridge and Valley, *SWA* Southwestern Appalachians, *WAP* Western Allegheny Plateau

holds a value equal to the percentage of *P. strobus* comprising the total composition of trees ≥ 12.7 cm DBH. We then extracted the mean proportional density of *P. strobus* for the pixels surrounding each sample site at four different buffer sizes (radii): 1, 10, 50, and 100 km. In independent GLMs, we assessed the association of proportional conspecific density at these four spatial envelopes with each of the genetic diversity metrics calculated above. We also assessed if mean pairwise F_{ST} was associated with proportional density to determine if more isolated populations were more genetically distant. In a separate set of GLMs, we evaluated the association between latitude and longitude with these metrics of genetic diversity to determine if any geographic clines exist among the 23 sampled populations within the Southern Appalachians ($N = 432$). Latitude and longitude were included in these models simultaneously as covariates.

The program BOTTLENECK (Cornuet and Luikart 1996; Piry et al. 1999) was used to detect recent bottleneck events in our dataset. This program tests for deviations from mutation-drift equilibrium with the assumption that an excess in heterozygotes results from a shrinking population and a deficiency in heterozygotes results from a population expansion. We tested all 25 populations (New Hampshire and Michigan included) separately each with 1,000 permutations. We used the single-step mutation model (SMM) and the two-phase model (TPM) with 95% single-step mutations and 5% multi-step mutations (Piry et al. 1999). Significant excess and deficiency in heterozygosity was determined with one-tailed Wilcoxon signed-rank tests.

Population structure

We used STRUCTURE version 2.3.4 (Pritchard et al. 2000), a genetic clustering algorithm, to assign individuals to genetic groups. There were two sets of analyses, one assessing all individuals, including those from New Hampshire and Michigan ($N = 466$), and one assessing only individuals from the Southern Appalachians focal region ($N = 432$). Each set consisted of two STRUCTURE runs: one with and one without the *LOCPRIOR* option. This parameter, which can improve detection of weak population structure without bias, informs STRUCTURE that individuals sampled at the same location are more likely to share ancestry (Hubisz et al. 2009). We assumed that individuals followed an admixture model and that allele frequencies were correlated among groups (Falush et al. 2003). Each STRUCTURE run varied the number of clusters (K) from 1 to 15 with 20 replicates for each K, using 25,000 burn-in steps, followed by 50,000 Markov Chain Monte Carlo (MCMC) iterations. The most optimal K in each analysis was determined using the LnPr(X|K) and ΔK methods described by Evanno et al. (2005) and implemented in the STRUCTURE HARVESTER pipeline (Earl and Vonholdt 2012).

To visualize pairwise genetic differentiation, we performed principal coordinates analyses (PCoA) using Nei's unbiased genetic distances (Nei 1978) in GENALEX. We also calculated F_{ST} across loci and between all pairs of populations using the *exclusion null alleles* method with 10,000 replicates in FREENA (Chapuis and Estoup 2007), which accounts for the presence of null alleles. The pairwise matrix of linearized F_{ST} values and a pairwise matrix of log-transformed geographic distances (km) of the 23 Southern Appalachian populations ($N = 432$) were used for a Mantel test (Mantel 1967) in GENALEX to detect if an isolation-by-distance (IBD) relationship existed within our focal study area. Despite its known shortcomings (Meirmans 2015), Mantel tests are still useful for determining spatial patterns of genetic variation when results are interpreted with caution (Diniz-Filho et al. 2013).

We conducted analyses of molecular variance (AMOVA) using ARLEQUIN version 3.5 (Excoffier and Lischer 2010) to determine the hierarchical partitioning of genetic variance within individuals, within populations, among populations, and among pre-defined genetic groups in the Southern Appalachian focal region (23 populations, $N = 432$). We conducted three AMOVAs with 10,000 permutations to test for significance. The first test (A) utilized no priors and assumed no genetic structure. The second test (B) informed structure with two pre-defined groups: the five isolated populations west of the main Appalachian Mountain chain (Table 1: TN4, KY1, KY2, KY3, and WV1) and the remaining 18 populations within the main distribution of *P. strobus* along the Appalachian Mountain chain (Fig. 2). The area within the Tennessee River Valley extending northeast along the Kentucky-Virginia border and into central West Virginia lacks *P. strobus* entirely, so this AMOVA sought to determine if this valley might act as a barrier to dispersal, influencing the partitioning of genetic variation. The third test (C) assigned each of the 23 populations based on Level III ecoregion (U.S. Environmental Protection Agency 2013; see Table 1) to inform genetic structure: Blue Ridge (15 populations), Ridge and Valley (5 populations), Southwestern Appalachians ($n = 1$), and West Allegheny Plateau (2 populations) (Fig. 3). This AMOVA examined if the ecoregions of the Southern Appalachians, which vary widely in elevation, aspect, geology, and soils, influenced *P. strobus* genetic differentiation.

Climate suitability modeling

To provide a historical context to the current *P. strobus* distribution and genetic structure, we estimated its suitable climatic envelope at three time points: the LGM (~ 22,000 years ago), the mid-Holocene (~ 6,000 years ago), and the "present day" (1970–2000). We utilized MAXENT version 3.4.0 (Phillips et al. 2006), which uses a maximum entropy algorithm to predict species distribution

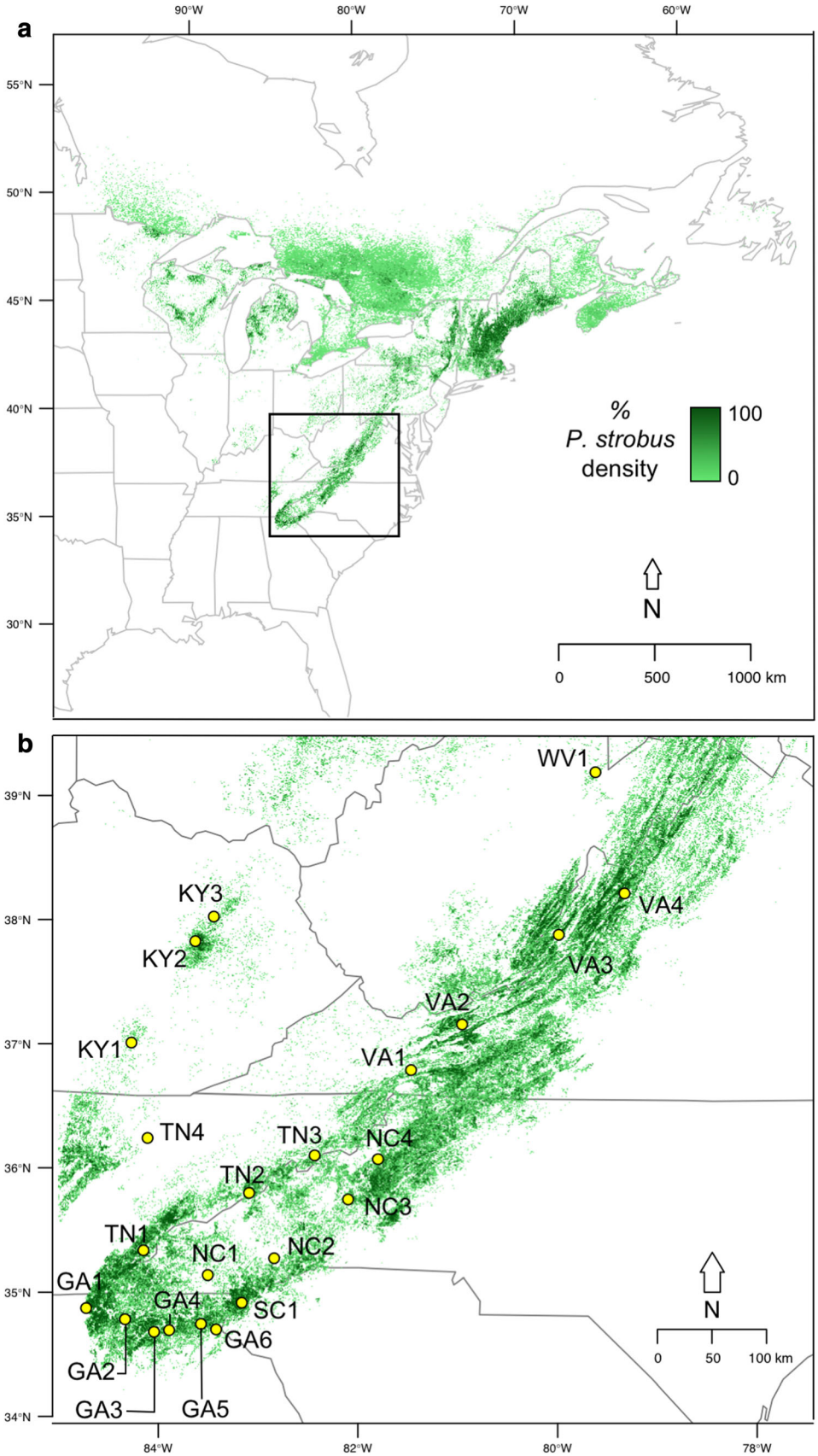

Fig. 2 Relative density of *P. strobus* **a** in North America and **b** within the Southern Appalachian region. The value of each pixel (250 × 250 m) corresponds to the proportion of total tree area (≥ 12.7 cm DBH) that *P. strobus* encompasses

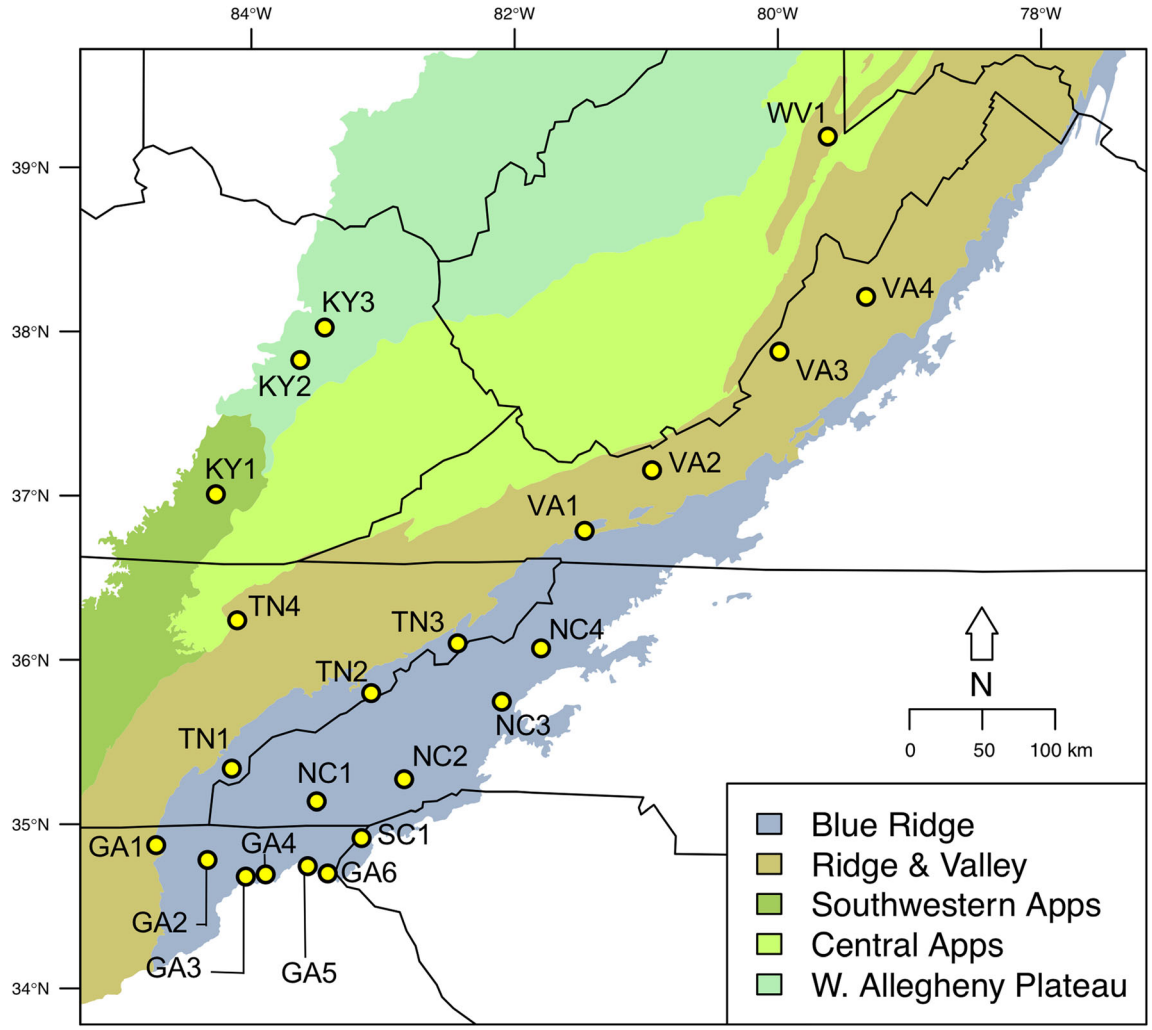

Fig. 3 Sampled populations sites of *P. strobus* from the Southern Appalachians on a Level III ecoregion map (U.S. Environmental Protection Agency, 2013)

based on models of the relationship between occurrence data and in this case, abiotic environmental variables. *P. strobus* occurrence data for MaxEnt was derived from the relative density raster created above (Fig. 2). All cells containing 100% *P. strobus* among trees ≥ 12.7 cm DBH were extracted as points (totaling 55,141) to be used for the model. We obtained 19 climatic variables from the WorldClim database (www.worldclim.org) (Hijmans 2005) as biolayers in our models for the climatic suitability of *P. strobus* (Table S2) at a resolution of 2.5 arc-minutes. The bioclimatic envelope was projected onto LGM and mid-Holocene climate data based on CCSM4. Biolayers BIO2, BIO3, BIO14, and BIO15 were removed from LGM simulations due to their inconsistent estimation among models (Varela et al. 2015). Beyond this, we opted not to use a priori variable selection methods; the regularization procedure in MaxEnt addresses model selection by weighing both informative and uninformative variables appropriately to balance model fit and complexity (Elith et al. 2011). We used a convergence threshold of 10^{-5} and 500 iterations. A randomly chosen set of 3,899 occurrence points was used for training the model, and 1,299 points (25%) were used for testing the model. The model's AUC score (area under the curve of the receiver operating characteristic) was evaluated for predictive accuracy, which varies from 0 (imperfect) to 1 (perfect), with a value of 0.5 indicating random predictive power. An AUC score above 0.7 is considered good model performance (Fielding and Bell 1997).

Results

Microsatellite loci

Two of the microsatellite loci, RPS25b and RPS60, were removed from analyses due to inconsistent amplification and scoring. The 10 remaining microsatellite loci amplified

in this study yielded 128 total alleles across the 466 samples of *P. strobus* (Table 2), ranging from a minimum of 5 alleles (RPS127) to a maximum of 26 (RPS12). There was 100% polymorphism across loci for each population, and each individual had a unique multi-locus genotype. Estimates of null allele frequency were considered either moderate ($0.05 \leq r < 0.2$) or negligible ($r < 0.05$) for all loci. The Bayesian IIM procedure from INEST revealed that both inbreeding and null alleles were important factors; the full model, including inbreeding, null alleles, and genotyping errors (DIC = 19,980.21), fit the data better than either the model assuming no null alleles (DIC = 20,069.64) or the model assuming no inbreeding (DIC = 20,056.59). The total adjusted inbreeding coefficient (F_{IS}) from INEST was 0.187, indicating an overall deficiency of heterozygotes. Observed heterozygosity (mean of 0.492 across all 10 loci) was much lower than expected heterozygosity (mean of 0.620). There were no correlations between mean F_{IS} and mean DBH of sampled trees, indicating no relationship between tree size and level of inbreeding. Of the 250 locus-population combinations, exact tests revealed a significant departure from Hardy-Weinberg Equilibrium in 120 combinations, but with no clear concentration in any particular locus or population. Only one pair of loci, RPS6 and RPS34b, showed significant linkage disequilibrium ($P < 0.001$).

Genetic diversity and variability

Genetic diversity metrics were assessed per population and are summarized in Table 1. There was a mean of 4.86 alleles per locus (A_N) and 4.21 alleles per locus when standardizing for sample size (A_R). Private alleles, which are alleles that only occur in one population, were present in 13 of the 25 populations. Private alleles were absent from all Kentucky populations, all Tennessee populations, three of the four North Carolina populations, a Virginia population, and a Georgia population. The majority of private alleles within the Southern Appalachians were present in the lowest latitude populations in Georgia and South Carolina and also in higher latitude populations in Virginia and West Virginia. No inferential statistics were performed, but the New Hampshire and Michigan reference populations appeared to have similar levels of genetic diversity to the overall values of the Southern Appalachian study region according to most metrics, with the exception that they had comparatively lower inbreeding coefficients ($F_{IS} = 0.046$ and 0.073, respectively).

Table 2 Summary statistics of genetic diversity, *F*-statistics, and null allele frequencies for each of the 10 microsatellite loci used in the study (developed by Echt et al. 1996). The mean F_{IS} value is the null allele-adjusted value from INEST version 2.2 (Chybicki and Burczyk 2009)

Locus	A_N	A_R	H_O	H_E	F_{ST}	F_{IS}	Null
RPS1b	8	3.33	0.296	0.359	0.031	0.112	0.033
RPS2	11	3.82	0.489	0.599	0.030	0.111	0.055
RPS6	12	4.58	0.529	0.678	0.064	0.277	0.114
RPS12	26	9.31	0.686	0.874	0.050	0.399	0.162
RPS20	19	5.94	0.587	0.735	0.060	0.144	0.056
RPS34b	12	3.86	0.464	0.618	0.040	0.429	0.129
RPS39	12	3.02	0.395	0.507	0.067	0.197	0.064
RPS50	16	7.29	0.681	0.815	0.044	0.085	0.046
RPS84	7	2.72	0.331	0.396	0.045	0.116	0.042
RPS127	5	3.02	0.459	0.614	0.160	0.478	0.102
Mean	12.8	4.69	0.492	0.620	0.060	0.187	0.080

A_N total number of alleles, A_R rarefied allelic richness, H_O observed heterozygosity, H_E expected heterozygosity, F_{ST} fixation index, F_{IS} inbreeding coefficient

Generalized linear models (GLMs) provided little support that proportional density of surrounding conspecifics within the Southern Appalachians region influenced genetic diversity, as there was no association between *P. strobus* density and inbreeding coefficient, heterozygosity, or number of rare alleles at any radius. However, in populations where the surrounding proportional density of *P. strobus* increased at radii of 50 and 100 km, respectively, the number of private alleles significantly increased ($F_{1,21} = 5.38$, $t = 2.32$, $P = 0.03$) and allelic richness increased ($F_{1,21} = 4.08$, $t = 2.021$, $P = 0.06$). Additionally, as proportional density at a radius of 100 km increased, mean pairwise F_{ST} values decreased ($F_{1,21} = 3.88$, $t = -1.97$, $P = 0.06$). When investigating if longitudinal or latitudinal patterns in genetic diversity exist, GLMs revealed no association with any of the tested indices.

No sampled population had an excess of heterozygotes from the one-tailed Wilcoxon tests of the SMM and TPM (Table S3), which would be indicative of recent bottleneck event(s). We detected significant heterozygote deficiency, however, in 14 and 16 of the 23 Southern Appalachian populations according to the SMM and TPM, respectively. We also detected a heterozygote deficiency in the New Hampshire but not the Michigan reference population.

Population structure

The maximum ΔK value and the plateau of the LnPr(X|*K*) curve for runs in STRUCTURE indicated the best number of clusters at the highest level of hierarchical analysis ($N = 466$, 25 populations; Fig. 4a) and within the Southern Appalachians only ($N = 432$, 23 populations; Fig. 4b) to be $K = 2$ (Fig. S1). The mean posterior probabilities of each population, including those from New Hampshire and Michigan, were mostly admixed across the landscape and no clear geographic pattern emerged. The STRUCTURE runs conducted with the *LOCPRIOR* option differed

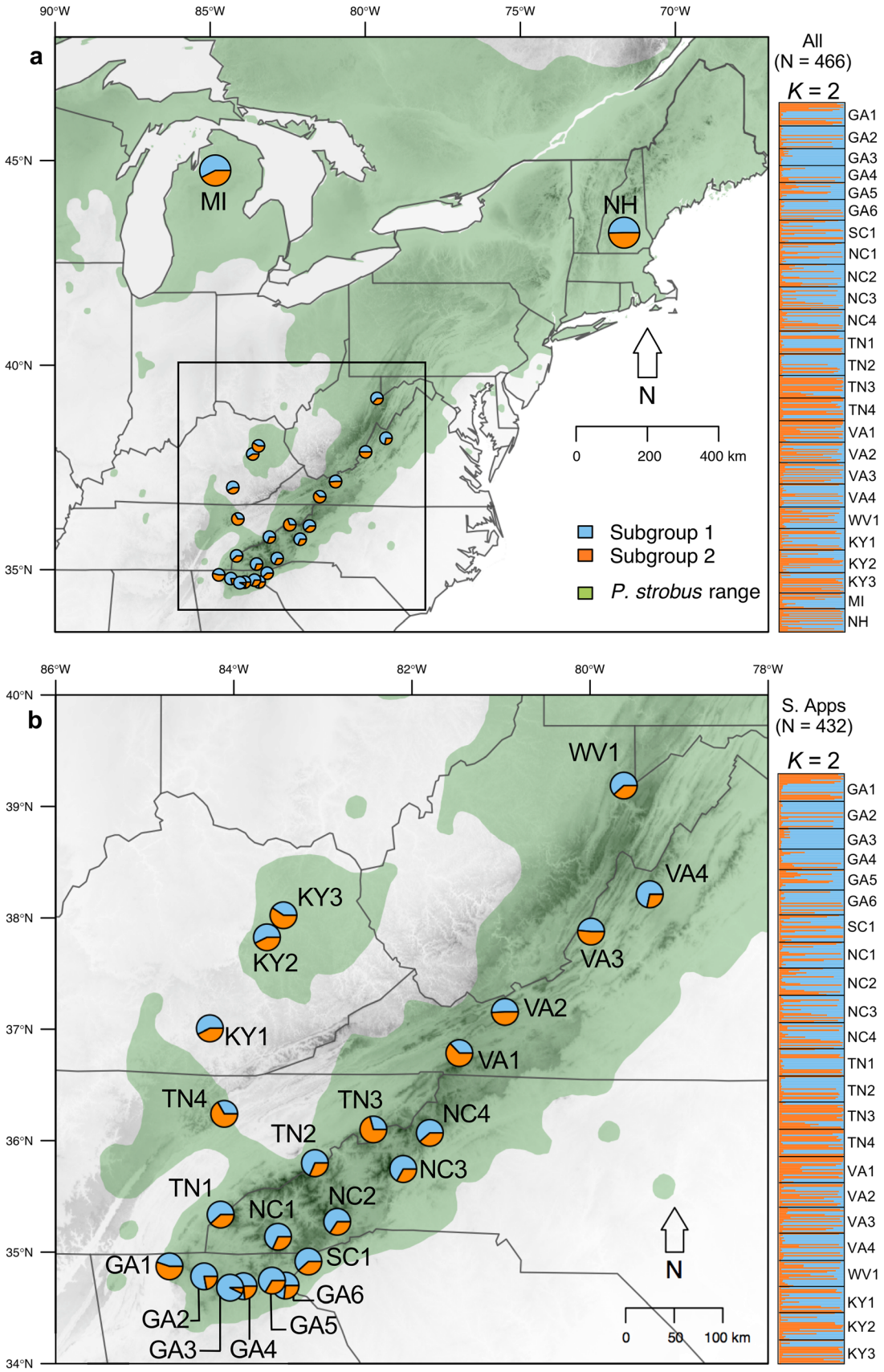

Fig. 4 Population and individual cluster assignments from STRUCTURE for *P. strobus* **a** including all individuals (25 populations, $N = 466$), and **b** including only individuals from the Southern Appalachians (23 populations, $N = 432$)

negligibly, also resulting in $K = 2$, with little population structure emerging at higher K values (Fig. S1).

Nei's unbiased genetic distances, visualized by a Principle Coordinates Analysis (PCoA), accounted for 77.4% of genetic variation explained by the first two axes. Results showed little overall genetic differentiation (Fig. 5a) among the majority of populations, including NH and MI (reference populations). The only populations with substantial differentiation on one or both of the principal coordinates were from Georgia (population GA2), West Virginia (population WV1), and Kentucky (population KY3). These three populations also had the highest mean pairwise F_{ST}, with values from 0.105 to 0.149, whereas all other populations had F_{ST} values less than 0.071 (Table 3, Fig. S2). A second PCoA excluded populations GA2, WV1, and KY3, and resulted in 56% of variation explained by the first two axes (Fig. 5b). The Mantel test revealed a significant association between genetic distance and geographic distance ($R = 0.31$, $P = 0.01$) among the 23 Southern Appalachian populations, suggesting *P. strobus* may have spatial genetic structure consistent with isolation-by-distance in this region (Fig. S3).

The overall F_{ST} value of our entire dataset, calculated using the *exclusion null alleles* method in FreeNA, was 0.060. Within the Southern Appalachians only, the F_{ST} of 0.055 from the AMOVA without any population structure was significant ($P < 0.001$), but 94.5% of genetic variation was partitioned within populations and only 5.5% of

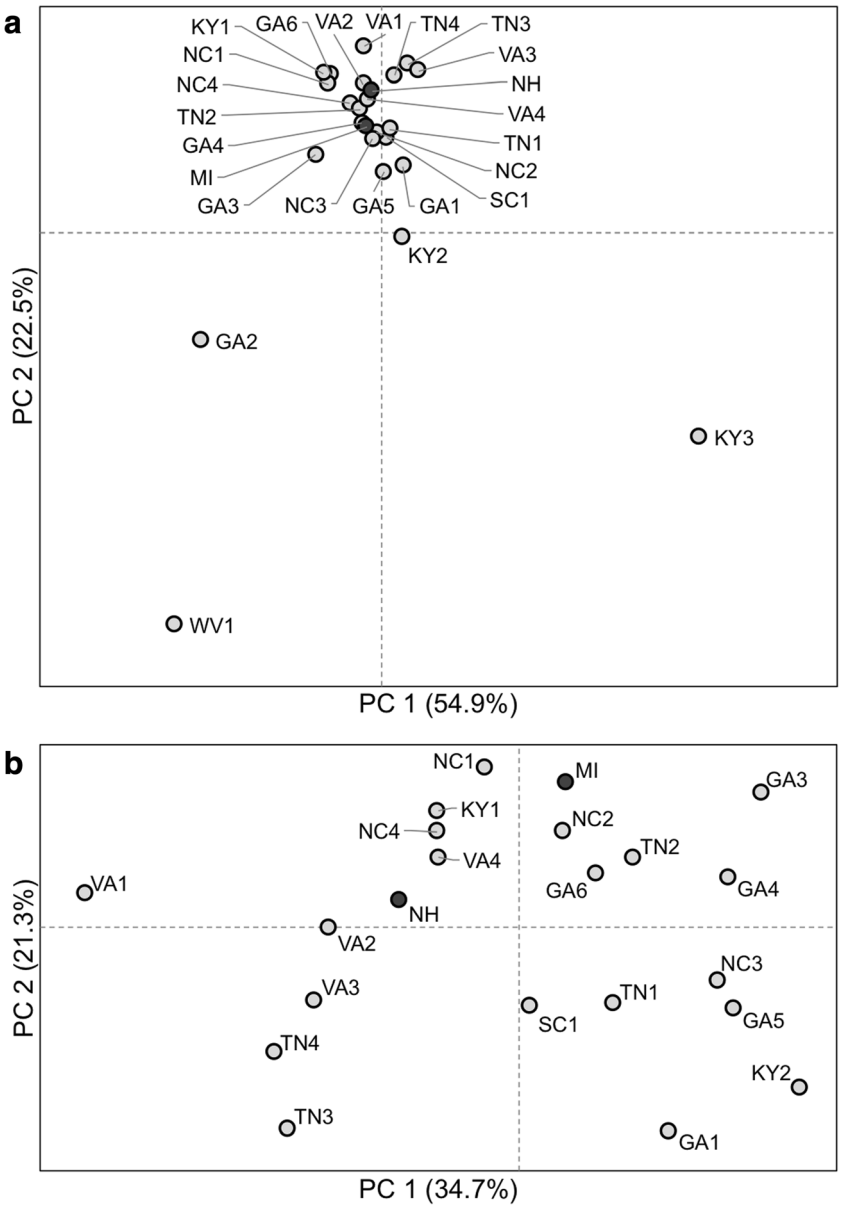

Fig. 5 Principal coordinates analyses (PCoA) based on Nei's unbiased genetic distances of **a** all *P. strobus* populations (25 populations), and **b** all remaining populations after excluding GA2, KY3, and WV1. Dark gray points indicate the two northern populations from Michigan and New Hampshire

Table 3 Population pairwise F_{ST} values (25 populations, 466 individuals) calculated with the "exclusion null alleles" method from FREENA (Chapuis and Estoup 2007). All significant values ($P < 0.05$) are in bold

	GA1	GA2	GA3	GA4	GA5	GA6	SC1	NC1	NC2	NC3	NC4	TN1	TN2	TN3	TN4	VA1	VA2	VA3	VA4	WV1	KY1	KY2	KY3	NH	MI
GA1																									
GA2	**0.113**																								
GA3	**0.077**	0.076																							
GA4	**0.042**	**0.105**	**0.030**																						
GA5	0.004	**0.107**	**0.032**	0.002																					
GA6	0.031	**0.089**	**0.041**	**0.029**	0.017																				
SC1	0.036	**0.113**	**0.059**	**0.060**	0.026	0.015																			
NC1	**0.058**	**0.086**	**0.041**	0.033	0.038	0.023	**0.046**																		
NC2	**0.043**	**0.091**	**0.032**	0.019	0.014	0.011	**0.028**	0.022																	
NC3	**0.027**	**0.111**	**0.044**	0.021	0.010	**0.045**	**0.059**	**0.052**	**0.040**																
NC4	**0.043**	**0.079**	**0.042**	**0.035**	**0.035**	**0.016**	**0.031**	0.001	0.004	**0.039**															
TN1	−0.009	**0.094**	**0.052**	**0.029**	−0.001	0.008	0.023	**0.030**	0.018	**0.018**	**0.017**														
TN2	**0.044**	**0.072**	0.024	0.024	0.007	0.011	**0.030**	**0.030**	0.013	**0.044**	0.021	0.010													
TN3	**0.040**	**0.124**	**0.116**	**0.073**	**0.051**	**0.048**	**0.043**	**0.055**	**0.061**	**0.077**	0.037	**0.040**	**0.061**												
TN4	**0.042**	**0.119**	**0.105**	**0.075**	**0.046**	**0.036**	**0.030**	**0.051**	**0.045**	**0.061**	**0.031**	**0.026**	**0.046**	0.015											
VA1	**0.085**	**0.128**	**0.121**	**0.096**	**0.086**	**0.050**	**0.051**	**0.038**	**0.048**	**0.098**	0.017	**0.058**	**0.062**	**0.026**	0.002										
VA2	**0.044**	**0.106**	**0.087**	**0.059**	**0.053**	**0.017**	**0.030**	**0.042**	0.015	**0.077**	0.014	**0.028**	**0.048**	**0.034**	**0.032**	**0.027**									
VA3	0.041	**0.117**	**0.096**	**0.069**	**0.055**	**0.037**	**0.042**	**0.043**	**0.023**	**0.064**	**0.018**	**0.028**	**0.043**	**0.022**	**0.034**	**0.032**	0.019								
VA4	**0.048**	**0.100**	**0.057**	**0.039**	**0.037**	**0.021**	0.024	**0.027**	0.000	**0.052**	0.004	0.025	**0.031**	**0.048**	**0.037**	**0.034**	0.011	0.012							
WV1	**0.161**	**0.026**	**0.138**	**0.174**	**0.160**	**0.142**	**0.147**	**0.131**	**0.137**	**0.162**	0.114	**0.141**	**0.132**	**0.176**	**0.157**	**0.167**	**0.161**	**0.157**	**0.139**						
KY1	**0.056**	**0.081**	0.035	**0.035**	**0.032**	0.008	**0.035**	0.011	0.011	**0.059**	0.002	**0.023**	0.006	**0.048**	0.032	**0.029**	**0.029**	**0.039**	**0.018**	0.123					
KY2	0.010	**0.129**	**0.066**	**0.045**	0.002	**0.046**	**0.038**	**0.070**	**0.044**	**0.048**	**0.066**	0.018	**0.046**	**0.065**	**0.058**	**0.115**	**0.067**	**0.068**	**0.064**	**0.169**	**0.062**				
KY3	**0.106**	**0.262**	**0.214**	**0.148**	**0.107**	**0.152**	**0.115**	**0.157**	**0.114**	**0.152**	**0.136**	**0.095**	**0.134**	**0.122**	**0.119**	**0.152**	**0.139**	**0.094**	**0.126**	**0.282**	**0.163**	**0.104**			
NH	**0.036**	**0.081**	**0.056**	**0.054**	**0.038**	0.008	0.017	**0.031**	0.010	**0.070**	0.006	0.015	0.012	**0.026**	**0.032**	**0.028**	0.004	0.011	0.018	**0.128**	0.009	**0.052**	**0.117**		
MI	**0.067**	**0.113**	**0.047**	**0.042**	**0.039**	**0.036**	**0.064**	**0.029**	0.018	**0.052**	0.018	**0.039**	0.022	**0.054**	**0.059**	**0.058**	**0.048**	**0.035**	**0.036**	**0.165**	0.030	**0.062**	**0.127**	**0.018**	
Avg	0.052	0.105	0.070	0.056	0.042	0.039	0.048	0.048	0.036	0.062	0.034	0.034	0.040	0.061	0.054	0.067	0.050	0.050	0.042	0.149	0.041	0.063	0.143	0.037	0.053

genetic variation partitioned among populations (Table 4a). The AMOVA with pre-defined structure, separating the five isolated populations west of the Appalachian Mountain chain from the 18 populations located within the mountain range, explained slightly more of the partitioning of genetic variation than the null model ($F_{ST} = 0.056$, $P < 0.001$); however, the amount of variation among groups ($F_{CT} = 0.002$, $P = 0.25$) was not significant (Table 4b). The AMOVA that utilized pre-defined population structure based on respective Level III ecoregion (U.S. Environmental Protection Agency 2013) (Fig. 3), best accounted for the partitioning of genetic variation across the Southern Appalachians ($F_{ST} = 0.062$, $P < 0.001$) and accounted for a significant amount of variation among groups ($F_{CT} = 0.019$, $P = 0.02$) (Table 4c).

Climate suitability modeling

The maximum entropy model performed substantially better than random (AUC = 0.844) (Fig. S4). The maximum temperature in the warmest month (BIO05) and the mean temperature in the warmest quarter (BIO10) were consistently among the highest contributing variables to model performance in all three projections. The suitable climatic envelope for *P. strobus* during the LGM (~ 22,000 years ago) was highest in what is now the North Carolina coastal plain and the Piedmont region of Alabama, Georgia, North Carolina, and South Carolina (Fig. 6a). There is also a narrow, long stretch of suitable habitat directly west of the Appalachian Mountains in northern Georgia and eastern Tennessee, an area of the Ridge and Valley ecoregion where no *P. strobus* exists today (Fig. 2b). By the mid-Holocene warm period (~ 6,000 years ago), the climate suitability for *P. strobus* had shifted northward substantially Fig. 6b), consistent with the steady movement of pollen from the fossil record (Davis 1983). Isolated islands of suitable habitat characterize most of the landscape during this time, including the area within our focal Southern Appalachians study region. The present-day climatic envelope (Fig. 6c) is consistent with the current *P. strobus* range and density (Figs. 1 and 2), with high and contiguous suitability values along the southern half of the Appalachian Mountain chain and moderate habitat suitability west of the mountains where samples were collected from populations TN4, KY1, KY2, KY3, and WV1.

Discussion

Genetic diversity

Levels of genetic diversity in populations of *P. strobus* in the Southern Appalachians were comparably high, but not higher than the two reference populations in the northern latitudes of North America. Additionally, the mean observed heterozygosity across populations of 0.477 falls within the range of values (albeit, on the lower end) reported in recent microsatellite studies from northern populations of 0.432 to 0.740 (Table 5). Allozyme studies (e.g., Beaulieu and Simon 1994; Buchert et al. 1997; Rajora et al. 1998) reported much lower levels of observed

Table 4 Analyses of Molecular Variance (AMOVA) of only individuals within the Southern Appalachians (23 populations, $N = 432$). Tests were conducted either without (A) or with a pre-defined population-group structure (B, C). Significant F-statistics ($P < 0.05$) are bold

Test	Group structure	Source of variation	Degrees of freedom	Sum of squares	Variance components	Percent variation	Fixation indices
		Among populations	22	173	0.14	5.5	
A	None	Within populations	841	2082	2.48	95.5	
		Total	863	2255	2.62		$F_{ST} = \mathbf{0.055}$
	West of vs. within	Between groups	1	10	0.01	0.2	$F_{CT} = 0.002$
B	the Appalachian	Among populations	21	163	0.14	5.4	$F_{SC} = \mathbf{0.054}$
	Mountain chain	Within populations	841	2082	2.48	94.4	
	(two groups[a])	Total	863	2255	2.62		$F_{ST} = \mathbf{0.056}$
	By	Between groups	3	44	0.05	1.9	$F_{CT} = \mathbf{0.019}$
C	ecoregion	Among populations	19	129	0.11	4.3	$F_{SC} = \mathbf{0.044}$
	(four groups[b])	Within populations	841	2082	2.48	93.8	
		Total	863	2255	2.64		$F_{ST} = \mathbf{0.062}$

[a] West group: TN4, KY1, KY2, KY3, and WV1; Within group: the 18 remaining populations

[b] Four U.S. Environmental Protection Agency, 2013, Level III ecoregions (see Table 1 for population groupings): Blue Ridge, Ridge and Valley, Southwestern Appalachians, and Western Allegheny Plateau

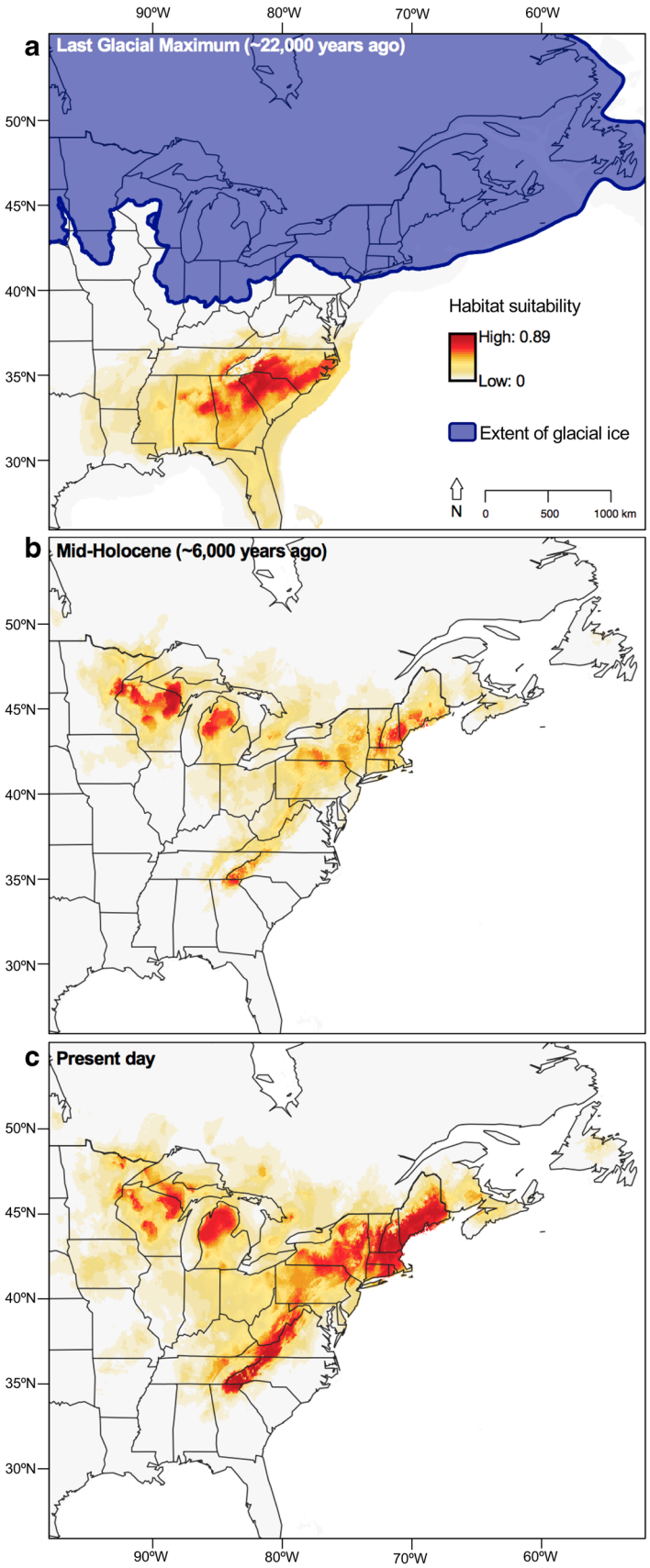

Fig. 6 Modeled climate suitability envelopes of *P. strobus* during **a** the last glacial maxima (~ 22,000 years ago), **b** mid-Holocene (~ 6,000 years ago), and **c** present day (1970–2000)

Table 5 Comparison of the genetic diversity estimates and *F*-statistics reported for *P. strobus* in studies utilizing microsatellite markers developed by Echt et al. (1996). Other smaller-scale studies using these markers, as well as studies using allozymes, were omitted

Publication	Microsatellites[a]	Range	Populations	Number	A_R	H_O	H_E	F_{IS}	F_{ST}
This study	**1b, 2, 6, 12, 34b, 39, 50, 84, 127**	USA: GA, KY, MI, NC, NH, SC, TN, VA, WV	25	466	4.21	0.477	0.590	0.164	0.060
Mehes et al. 2009	**1b**, **2**, **12**, **20**, 25b, **50,** 118b	Canada: NB, NL, NS, ON, PE, QC	10	300–400	*NA*	0.740	0.802	0.072	0.084
Mandak et al. 2013	**1b, 2, 12**, 25b, **34b, 39, 50, 84**, 118b, **127**	USA: CT, MA, ME, MI, NH, NY, PA, VT	30	592	4.23	0.432	0.528	0.215	0.025
Chhatre and Rajora 2014	**1b, 2, 12, 20**, 25b, **34b, 39, 50**, 118b, 119, **127**	Canada: ON	6	614	10.58	0.525	0.608	0.139	0.083
Zinck and Rajora 2016	**1b, 2, 6, 12, 20**, 25b, **34b, 39, 50**, 118b, 119, **127**	Canada: NB, NL, ON, QC USA: ME, MN, NC, NH, NY, PA, VA	33	1650	10.36	0.680	0.740	0.100	0.104

[a] "RPS" microsatellite markers from Echt et al. (1996). Bold denotes a marker also used in this study

GA Georgia, *KY* Kentucky, *ME* Maine, *MI* Michigan, *MN* Minnesota, *NC* North Carolina, *NH* New Hampshire, *NY* New York, *PA* Pennsylvania, *SC* South Carolina, *TN* Tennessee, *VA* Virginia, *WV* West Virginia, *NB* New Brunswick, *NL* Newfoundland and Labrador, *NS* Nova Scotia, *ON* Ontario, *PE* Prince Edward Island, *QC* Quebec

heterozygosity (0.121–0.265), but this is likely due to allozyme loci being less variable than microsatellite loci (Hedrick 1999). From our analysis, the New Hampshire and Michigan reference populations possessed higher allelic richness and heterozygosity than populations in the Southern Appalachians on average. Within the Southern Appalachian region, we found no latitudinal gradients in genetic diversity estimates (from Table 1). The two previous studies to have included any Southern populations yielded both contrasting and affirmatory results. Zinck and Rajora (2016) reported a south-to-north decrease in several metrics, including allelic richness and heterozygosity, but Nadeau et al. (2015) found no latitudinal clines in genetic diversity. In fact, Nadeau et al. (2015) showed there was slightly less genetic diversity in populations at the southern extent of their sampling. They attributed this pattern to genetic drift acting on small and sparse northern Georgia refugial populations during the LGM (Jackson et al. 2000) and admixture during postglacial recolonization with a second refugium in the mid-Atlantic. Similarly, the highest levels of genetic diversity in the congeneric red pine (*Pinus resinosa* Ait.) occur in the northern part of its range, likely due to admixture between refugia in the Southern Appalachians and the mid-Atlantic during postglacial recolonization (Walter and Epperson 2001).

Due to severe inbreeding depression (Fowler 1965; Johnson 1945), *P. strobus* has a higher outcrossing rate than what is typical in other conifers (Beaulieu and Simon 1995) and has been reported with low levels of inbreeding (Epperson and Chung 2001; Marquardt and Epperson 2004; Rajora et al. 2000). More recent studies, genotyping individuals over a broader distribution and with the same microsatellite loci used in this study, report fairly high mean F_{IS} values ranging from 0.072 to 0.215 (Table 5); the mean null allele-adjusted inbreeding coefficient of only our Southern Appalachian populations ($F_{IS} = 0.173$) was comparable within this range. There was also an overall deficit in heterozygosity, similar to northern *P. strobus* populations. Although we found no significant correlation between mean tree size (DBH) and mean inbreeding coefficient in our data, the fact that we sampled young trees in general (< 16 cm DBH) may have also inflated our F_{IS} values, because as tree stands age, selection occurs against inbred progeny (Hufford and Hamrick 2003; Strauss 1986). However, inbreeding may yet be slightly stronger in the Southern Appalachians, as the New Hampshire and Michigan reference populations analyzed in this study had lower F_{IS} values. Zinck and Rajora (2016) reported a similar pattern with higher inbreeding coefficients in a single Virginia ($F_{IS} = 0.20$) and a single North Carolina ($F_{IS} = 0.21$) population relative to the overall mean of mostly northern populations ($F_{IS} = 0.10$). The fact that populations in the Southern Appalachians are generally smaller and more scattered (Abrams 2001) may explain why signals of inbreeding may be stronger than populations located further north.

The degree of isolation for *P. strobus* populations in the Southern Appalachians influenced some but not all metrics of genetic diversity. Heterozygosity and F_{IS} remained mostly unchanged, but number of private alleles decreased as the proportional density of conspecifics also decreased at a large geographic scale (radii of 50 and 100 km). Chhatre and Rajora (2014) similarly found that disjunct populations of *P. strobus* in Ontario, Canada, had reduced number of private alleles (and reduced allelic richness), but had similar heterozygosity and F_{IS} values when compared to continuous populations. Although we found no evidence of a genetic bottleneck in any population, the smaller size of more isolated populations in our dataset may account for the lack of private alleles in general. Due to the longevity (up to 450 years in age) of

P. strobus, it is possible that mechanisms like genetic drift have not yet reduced heterozygosity or increased inbreeding levels in more isolated populations. Only at the largest geographic scales and the highest degrees of isolation were proportional *P. strobus* densities correlated with any differences in genetic diversity indices. This is probably indicative of the great dispersal ability of *P. strobus*. Wind-dispersed pollen and seed usually travel less than 1 km, but in open landscapes they have the potential to travel several kilometers (Munzbergova et al. 2010; Williams 2010; Wilson and McQuilkin 1963). This long-range dispersal ability likely maintains a high level of genetic diversity among distant and/or isolated *P. strobus* populations, likely reducing the impacts or slowing the trajectory of genetic drift.

Population structure

Little overall genetic differentiation is typical of gymnosperms (Hamrick and Godt 1996) and has also been observed among *P. strobus* populations of northern North America. In our Southern Appalachians focal range, we found a low mean F_{ST}, comparable with northern studies (Table 5). Although low, this value ($F_{ST} = 0.055$) was significant with no a priori structure (Table 4a), suggesting there is a non-negligible amount of genetic differentiation. The Mantel test further suggested there is a significant spatial genetic structure consistent with isolation-by-distance, a finding that is to be expected of wind-pollinated plants (Hamrick and Nason 1996) and is consistent with *P. strobus* in northern populations (e.g., Epperson and Chung 2001; Marquardt et al. 2007; Nijensohn et al. 2005; but also see Mandak et al. 2013). Nevertheless, population structure appeared to be weak overall among the analyzed 23 Southern Appalachian populations in this study, where we did not identify a significant pattern of geographic substructure. Although STRUCTURE revealed $K = 2$, the posterior probabilities of cluster assignments suggested nearly universal admixture across populations (Fig. 4). This was likely an erroneous default result, because STRUCTURE cannot assess $K = 1$ as a potential scenario (Janes et al. 2017). The PCoA echoed this, as 22 of the 25 total populations, including those from New Hampshire and Michigan, were clustered closely together (Fig. 5a). Overall, genetic diversity within the Southern Appalachians was overwhelmingly partitioned within populations, and not among them, a result also consistent with northern *P. strobus* (Table 5).

Categorization by Level III ecoregions (U.S. Environmental Protection Agency 2013) best explained genetic differentiation in the Southern Appalachians, accounting for the most variation among *P. strobus* populations and between groups (Table 4c, Fig. 3). Successive uplift events formed the ancient Appalachian Mountain chain starting over 1 billion years ago. This created geologically varied ecoregions in the Southern Appalachians, characterized by their distinct topography, soils, and climatic conditions. Despite weak overall *P. strobus* differentiation in the Southern Appalachians, ecoregional differences may best explain extant population structure and may be indicative of possible local adaptation to specific ecological zones.

This was the first study to explore the genetic diversity and structure of *P. strobus* in Kentucky. These three populations (KY1, KY2, and KY3), as well as the Tennessee population on the western side of the Tennessee River Valley (TN4) and the West Virginia population (WV1), are all isolated from the main distribution of *P. strobus* occurring along the main Appalachian Mountain chain. These five isolated populations are also quite small in terms of surrounding conspecific density (Fig. 2b). When sampling these *P. strobus* patches, we had no indication as to if or when they were insularized; there are no records to suggest they are relict glacial populations or are remnants of a once contiguous *P. strobus* corridor that was present during postglacial recolonization. Isolated populations are often of high conservation interest, because they may harbor unique alleles, potentially available for ongoing or future evolutionary change (Fady et al. 2016), and therefore we found it important to sample from these populations. However, only one (WV1) harbored a private allele. In fact, genetic diversity was fairly comparable to levels in the other 18 populations. Two of the populations (KY3 and WV1) were highly differentiated, but the other three showed high relatedness with the rest of the Southern Appalachian populations (Table 3). Fewer than half of these geographically distinct populations were genetically very distinct, suggesting their current isolation may be a recent phenomenon (Hewitt 1999, Provan and Bennett 2008).

Climate suitability modeling

Fossilized pollen evidence indicates *P. strobus* refugia existed south of the Appalachian Mountains in northwestern Georgia during the LGM (Jackson 2000). As the glaciers receded, *P. strobus* first recolonized cooler, higher elevations and continued range expansion northeastward along the Appalachian Mountain chain, arriving in Virginia's Shenandoah Valley (where populations VA3 and VA4 are located) ~13,000 years ago. According to our climatic suitability envelope during the LGM, suitable habitats existed in the area surrounding the southern extent of the Appalachian Mountain chain, including areas west in what is present-day Tennessee, USA. In addition to populating the cooler, higher elevations of the Appalachian Mountains after glacial retreat, migrants from these populations may also have moved west into higher elevation sites opposite the Ridge and Valley ecoregion along the Cumberland Plateau. This may explain the presence of isolated *P. strobus* patches in this area today (e.g., populations TN4, KY1, KY2, KY3, and WV1). By the mid-Holocene, *P. strobus* had recolonized an area of North America similar in extent to its current distribution (Davis 1983), but our climate suitability envelope during this time period suggests

it may have existed as patchy and sparse populations. This was especially true in the Southern Appalachians and the area west of the mountain chain, where climatic suitability was low to moderate ~6,000 years ago. Since then, climatic suitability has improved within this region, potentially facilitating population expansions. We propose that isolated stands of *P. strobus* currently in Central Tennessee and Kentucky are not relict populations. Their genetic similarity to the other Southern Appalachian populations, as well as the presence of suitable climatic habitat in this area over time, suggests there has been high historic gametic connectivity through what may have been a relatively contiguous corridor at one time. We advise caution in blindly prioritizing isolated, disjunct populations of *P. strobus* in the Southern Appalachians for conservation.

Summary and conclusions

Our snapshot of the standing genetic variability and population structure of *P. strobus* in the Southern Appalachians provides little evidence that this region is a hotspot of exceptionally high extant and/or ancestral genetic diversity. Instead, our genetic analyses suggest that *P. strobus* is a highly admixed species throughout North America as a result of frequent and rangewide genetic exchange since glaciation. Natural and anthropogenic phenomena, as well as our methodological choices, may help to contextualize our findings: (1) the mating system of *P. strobus* and the lack of barriers on the landscape facilitate long-distance gametic exchange. Although the Southern Appalachian Mountains served as a discontinuity driving vicariance in several species during glaciation (Soltis et al. 2006), they have been considered a weak barrier to gene flow for wind-dispersed trees like *P. strobus* (Nadeau et al. 2015). Consistent south-to-north pollen and seed movement over many long-lived generations since the LGM may explain the shallow population structure and consistently high genetic diversity throughout the its range. (2) Anthropogenic activities within the last century may have also affected our estimates of genetic diversity and population structure. As part of the "New Deal" in the 1930s through the 1950s, the Tennessee Valley Authority and the Civilian Conservation Corps cleared significant swaths of land throughout the Southern Appalachians and in many cases replanted millions of seedlings where reforestation was needed for marginal land reclamation and timber (e.g., Vimmerstedt 1962). Some records indicate these replanted seedlings were locally sourced, but it is impossible to know the origin and extent of all planted *P. strobus* stands in this region. Occurring only ~80 years ago (less than a single *P. strobus* generation), there is a possibility that such activities by the federal government may have dampened signal(s) of population structure and/or altered detectable levels of genetic diversity within this region. (3) Although the microsatellite markers we chose for genetic analyses have been used extensively, they may not have allowed for optimal temporal resolution or allelic breadth to detect substantial genetic differentiation among groups. Nadeau et al. (2015) reported the presence of substructure ($K = 4$) within their "southern group" (comprising of eight sample sites), based on *P. strobus* genotypes from ~150 single nucleotide polymorphisms (SNPs). Future investigations into the population structure of *P. strobus* in the Southern Appalachians should consider utilizing SNPs or developing additional microsatellite markers for improved resolution.

Estimating variation in selectively neutral markers is an important barometer for the adaptive potential in trees (Jump et al. 2009). With its great capacity for gene flow, *P. strobus* populations in the Southern Appalachians and across its entire North American range, including the highly isolated populations tested herein, appear to be highly diverse and admixed. Despite the lack of clear population structure found in our study, provenance tests have shown *P. strobus* to have substantial clinal variation in certain phenotypic traits according to latitude and climate (e.g., Housset et al. 2018; Joyce and Rehfeldt 2013). These phenotypic differences, as well as the high genetic variation and high genetic mobility of *P. strobus,* bodes well for its adaptability (Hamrick 2004). Although populations in the Southern Appalachians were not uniquely high in genetic diversity as we hypothesized, they remain an important part of the range deserving of conservation priority.

Acknowledgements We are grateful to those who assisted with *P. strobus* sampling: Chris Asaro (USDA Forest Service), Brittany Barnes (University of Georgia), Megan Bartz (USDA Forest Service), Amanda Block (USDA Forest Service), William Caudill (USDA Forest Service), Lori Chamberlain (Virginia Department of Forestry), Beth Christensen (USDA Forest Service), Kara Costanza (University of New Brunswick), Don Grosman (Arborjet), TJ Hall (Pennsylvania DCNR), Bob Heyd (USDA Forest Service), Erica Roberts (USDA Forest Service), Derek Robertson (USDA Forest Service), Laura Vincent (USDA Forest Service), and Joe White (USDA Forest Service). We also thank Nathan Havill, Joseph Nairn, Kenneth Ross, and two anonymous reviewers for their helpful comments on an earlier version of this manuscript. This research was supported in part by the USDA Forest Service, Southern Research Station (13CA11330129056 and 16CS11330129045), Southern Region-Forest Health Protection, and the University of Georgia (2015) Innovative and Interdisciplinary Research Grant.

Data archiving statement Microsatellite genotypes at all loci for all samples can be found in the supplementary information (Table S4).

References

Abrams MD (2001) Eastern white pine versatility in the presettlement forest. Bioscience 51:967–979

Abrams MD, Orwig DA, Demeo TE (1995) Dendroecological analysis of successional dynamics for a presettlement-origin white-pine mixed-oak forest in the Southern Appalachians, USA. J Ecol 83:123–133

Aitken SN, Yeaman S, Holliday JA, Wang TL, Curtis-McLane S (2008) Adaptation, migration or extirpation: climate change outcomes for tree populations. Evol Appl 1:95–111

Alfaro RI, Fady B, Vendramin GG, Dawson IK, Fleming RA, Saenz-Romero C, Lindig-Cisneros RA, Murdock T, Vinceti B, Navarro CM, Skroppa T, Baldinelli G, El-Kassaby YA, Loo J (2014) The role of forest genetic resources in responding to biotic and abiotic factors in the context of anthropogenic climate change. Forest Ecol Manag 333:76–87

Barrett RDH, Schluter D (2008) Adaptation from standing genetic variation. Trends in Ecol Evol 23:38–44

Beaulieu J, Simon JP (1994) Inheritance and linkage relationships of allozymes in *Pinus strobus* L. Silvae Genet 43:253–261

Beaulieu J, Simon JP (1995) Mating system in natural populations of eastern white pine in Quebec. Can J For Res 25:1697–1703

Brookfield JFY (1996) A simple new method for estimating null allele frequency from heterozygote deficiency. Mol Ecol 5:453–455

Brownstein MJ, Carpten JD, Smith JR (1996) Modulation of non-templated nucleotide addition by taq DNA polymerase: primer modifications that facilitate genotyping. Biotechniques 20:1004–1010

Buchert GP, Rajora OP, Hood JV, Dancik BP (1997) Effects of harvesting on genetic diversity in old growth eastern white pine in Ontario, Canada. Conserv Biol 11:747–758

Campagne P, Smouse PE, Varouchas G, Silvain JF, Leru B (2012) Comparing the van Oosterhout and Chybicki-Burczyk methods of estimating null allele frequencies for inbred populations. Mol Ecol Resour 12:975–982

Chapuis MP, Estoup A (2007) Microsatellite null alleles and estimation of population differentiation. Mol Biol Evol 24:621–631

Chhatre VE, Rajora OP (2014) Genetic divergence and signatures of natural selection in marginal populations of a keystone, long-lived conifer, eastern white pine (*Pinus strobus*) from northern Ontario. Plos One 9

Chybicki IJ, Burczyk J (2009) Simultaneous estimation of null alleles and inbreeding coefficients. J Hered 100:106–113

Cornuet JM, Luikart G (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. Genetics 144:2001–2014

Costanza KKL, Whitney TD, McIntire CD, Livingston WH, Gandhi KJK (2018) A synthesis of emerging health issues of eastern white pine (*Pinus strobus*) in eastern North America. For Ecol Manag 423:3–17

Davis MB (1983) Quaternary history of deciduous forests of eastern North America and Europe. Ann Mo Bot Gard 70:550–563

Diniz-Filho JA, Soares TN, Lima JS, Dobrovolski R, Landeiro VL, Telles MPD, Rangel TF, Bini LM (2013) Mantel test in population genetics. Genet Mol Biol 36:475–485

Earl DA, Vonholdt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour 4:359–361

Echt CS, May-Marquardt P, Hseih M, Zahorchak R (1996) Characterization of microsatellite markers in eastern white pine. Genome 39:1102–1108

Elith J, Phillips SJ, Hastie T, Dudik M, Chee YE, Yates CJ (2011) A statistical explanation of MaxEnt for ecologists. Divers Distrib 17: 43–57

Epperson BK, Chung MG (2001) Spatial genetic structure of allozyme polymorphisms within populations of *Pinus strobu*s (Pinaceae). Am J Bot 88:1006–1010

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14:2611–2620

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour 10:564–567

Fady B, Aravanopoulos FA, Alizoti P, Mátyás C, von Wühlisch G, Westergren M, Belletti P, Cvjetkovic B, Ducci F, Huber G, Kelleher CT, Khaldi A, Kharrat MBD, Kraigher H, Kramer K, Mühlethaler U, Peric S, Perry A, Rousi M, Sbay H, Stojnic S, Tijardovic M, Tsvetkov I, Varela MC, Vendramin GG, Zlatanov T (2016) Evolution-based approach needed for the conservation and silviculture of peripheral forest tree populations. For Ecol Manag 375:66–75

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164:1567–1587

Fielding AH, Bell JF (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. Environ Conserv 24:38–49

Fowler DP (1965) Natural self-fertilization in three jack pines and its implications in seed orchard management. For Sci 11:55–58

Goudet J (2005) HIERFSTAT, a package for R to compute and test hierarchical F-statistics. Mol Ecol Notes 5:184–186

Hamrick JL (2004) Response of forest trees to global environmental changes. For Ecol Manag 197:323–335

Hamrick JL, Godt MJW (1996) Effects of life history traits on genetic diversity in plant species. Philos T Roy Soc B 351:1291–1298

Hamrick JL, Nason JD (1996) Consequences of dispersal in plants. In: Rhodes OE Jr, Chesser RK, Smith MH (eds) Population dynamics in ecological space and time. University of Chicago Press, Chicago, pp 203–236

Hedrick PW (1999) Perspective: Highly variable loci and their interpretation in evolution and conservation. Evolution 53:313–318

Hepp TE, Vimmerstedt JP, Smalley GW, McNab WH (2015) Estimating yields of unthinned eastern white pine plantations from current stocking in the Southern Appalachians. For Sci 61:114–122

Hewitt G (2000) The genetic legacy of the quaternary ice ages. Nature 405:907–913

Hewitt GM (1999) Post-glacial re-colonization of European biota. Biol J Linn Soc 68:87–112

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. Int J Climatol 25:1965–1978

Hijmans RJ, van Etten J, Cheng J, Mattiuzzi M, Sumner M, Greenberg JA, Lamigueiro OP, Bevan A, Racine EB, Shortridge A, Hijmans MRJ (2017) Package 'raster'. R package

Housset JM, Nadeau S, Isabel N, Depardieu C, Duchesne I, Lenz P, Girardin MP (2018) Tree rings provide a new class of phenotypes for genetic associations that foster insights into adaptation of conifers to climate change. New Phytol 218:630–645

Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. Mol Ecol Resour 9:1322–1332

Hufford KM, Hamrick JL (2003) Viability selection at three early life stages of the tropical tree, *Platypodium elegans* (Fabaceae, Papilionoideae). Evolution 57:518–526

Jackson ST, Webb RS, Anderson KH, Overpeck JT, Webb T, Williams JW, Hansen BCS (2000) Vegetation and environment in Eastern North America during the Last Glacial Maximum. Quaternary Sci Rev 19:489–508

Janes JK, Miller JM, Dupuis JR, Malenfant RM, Gorrell JC, Cullingham CI, Andrew RL (2017) The K=2 conundrum. Mol Ecol 26:3594–3602

Johnson LPV (1945) Reduced vigour, chlorophyll deficiency, and other effects of self-fertilization in *Pinus*. Can J For Res 23:145–149

Joyce DG, Rehfeldt GE (2013) Climatic niche, ecological genetics, and impact of climate change on eastern white pine (*Pinus strobus* L.): guidelines for land managers. For Ecol Manag 295:173–192

Jump AS, Marchant R, Penuelas J (2009) Environmental change and the option value of genetic diversity. Trends Plant Sci 14:51–58

Kuparinen A, Savolainen O, Schurr FM (2010) Increased mortality can promote evolutionary adaptation of forest trees to climate change. For Ecol Manag 259:1003–1008

Lovett GM, Canham CD, Arthur MA, Weathers KC, Fitzhugh RD (2006) Forest ecosystem responses to exotic pests and pathogens in eastern North America. Bioscience 56:395–405

Mandak B, Hadincova V, Mahelka V, Wildova R (2013) European invasion of North American *Pinus strobus* at large and fine scales: high genetic diversity and fine-scale genetic clustering over time in the adventive range. PLoS One 8

Mantel N (1967) Detection of diesease clustering and a generalized regression approach. Cancer Res 27:209–220.

Marquardt PE, Echt CS, Epperson BK, Pubanz DM (2007) Genetic structure, diversity, and inbreeding of eastern white pine under different management conditions. Can J For Res 37:2652–2662

Marquardt PE, Epperson BK (2004) Spatial and population genetic structure of microsatellites in white pine. Mol Ecol 13:3305–3315

Mehes M, Nkongolo KK, Michael P (2009) Assessing genetic diversity and structure of fragmented populations of eastern white pine (*Pinus strobus*) and western white pine (*P. monticola*) for conservation management. J Plant Ecol 2:143–151

Meirmans PG (2015) Seven common mistakes in population genetics and how to avoid them. Mol Ecol 24:3223–3231

Munzbergova Z, Hadincova V, Wild J, Herben T, Maresova J (2010) Spatial and temporal variation in dispersal pattern of an invasive pine. Biol Invasions 12:2471–2486

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. Am J Bot 102:1342–1355

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89:583–590

Nijensohn SE, Schaberg PG, Hawley GJ, DeHayes DH (2005) Genetic subpopulation structuring and its implications in a mature eastern white pine stand. Can J For Res 35:1041–1052

Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in excel. Population genetic software for teaching and research. Mol Ecol Notes 6:288–295

Peakall R, Smouse PE (2012) GENALEX 6.5: genetic analysis in Excel. Population genetic software for teaching and research: an update. Bioinformatics 28:2537–2539

Petit RJ, Aguinagalde I, de Beaulieu JL et al (2003) Glacial refugia: hotspots but not melting pots of genetic diversity. Science 300: 1563–1565

Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. Ecol Model 190:231–259

Piry S, Luikart G, Cornuet JM (1999) BOTTLENECK: a computer program for detecting recent reductions in the effective population size using allele frequency data. J Hered 90:502–503

Potter KM, Jetton RM, Bower A, Jacobs DF, Man G, Hipkins VD, Westwood M (2017) Banking on the future: progress, challenges and opportunities for the genetic conservation of forest trees. New Forest 48:153–180

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959

Provan J, Bennett KD (2008) Phylogeographic insights into cryptic glacial refugia. Trends Ecol Evol 23:564–571

R Core Team (2018) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna

Rajora OP, DeVerno L, Mosseler A, Innes DJ (1998) Genetic diversity and population structure of disjunct Newfoundland and Central Ontario populations of eastern white pine (*Pinus strobus*). Can J Bot 76:500–508

Rajora OP, Rahman MH, Buchert GP, Dancik BP (2000) Microsatellite DNA analysis of genetic effects of harvesting in old-growth eastern white pine (*Pinus strobus*) in Ontario, Canada. Mol Ecol 9:339–348

Raymond M, Rousset F (1995) GENEPOP (Version-1.2) - Population genetics software for exact tests and ecumenicism. J Hered 86: 248–249

Rogers LL, Lindquist EL (1992) Supercanopy white pine and wildlife. In: Stine RA (ed) White pine symposium. Duluth, Minnesota, pp 39–43

Savolainen O, Bokma F, Garcia-Gil R, Komulainen P, Repo T (2004) Genetic variation in cessation of growth and frost hardiness and consequences for adaptation of *Pinus sylvestris* to climatic changes. For Ecol Manag 197:79–89

Savolainen O, Pyhajarvi T, Knurr T (2007) Gene flow and local adaptation in trees. Annu Rev Ecol Evol S 38:595–619

Schaberg PG, DeHayes DH, Hawley GJ, Nijensohn SE (2008) Anthropogenic alterations of genetic diversity within tree populations: implications for forest ecosystem resilience. For Ecol Manag 256:855–862

Soltis DE, Gitzendanner MA, Strenge DD, Soltis PS (1997) Chloroplast DNA intraspecific phylogeography of plants from the Pacific Northwest of North America. Plant Syst Evol 206:353–373

Soltis DE, Morris AB, McLachlan JS, Manos PS, Soltis PS (2006) Comparative phylogeography of unglaciated eastern North America. Mol Ecol 15:4261–4293

Strauss SH (1986) Heterosis at allozyme loci under inbreeding and cross-breeding in *Pinus attenuata*. Genetics 113:115–134

Talarico L, Babik W, Marta S, Mattoccia M (2019) Genetic drift shaped MHC IIB diversity of an endangered anuran species within the Italian glacial refugium. J Zool 307:61–70

U.S. Environmental Protection Agency (2013) Level III ecoregions of the continental United States. U.S. EPA - National Health and Environmental Effects Research Laboratory, Corvallis

Varela S, Lima-Ribeiro MS, Terribile LC (2015) A short guide to the climatic variables of the last glacial maximum for biogeographers. PLoS One 10:e0129037

Vimmerstedt JP (1962) Southern Appalachian white pine plantations site, volume, and yield. USDA Forest Service, Southeastern Forest Experimental Station, p. 15

Walter R, Epperson BK (2001) Geographic pattern of genetic variation in *Pinus resinosa*: area of greatest diversity is not the origin of postglacial populations. Mol Ecol 10:103–111

Walter R, Epperson BK (2004) Microsatellite analysis of spatial structure among seedlings in populations of *Pinus strobus* (Pinaceae). Am J Bot 91:549–557

Wendel GW, Smith HC (1990) *Pinus strobus* L. eastern white pine. In: Burns RM, Honkala BH (eds) Silvics of North America. USDA Forest Service, Washington DC, pp 476–488

Wielstra B, Crnobrnja-Isailovic J, Litvinchuk SN, Reijnen BT, Skidmore AK, Sotiropoulos K, Toxopeus AG, Tzankov N, Vukov T, Arntzen JW (2013) Tracing glacial refugia of *Triturus* newts based on mitochondrial DNA phylogeography and species distribution modeling. Front Zool 10

Williams CG (2010) Long-distance pine pollen still germinates after meso-scale dispersal. American J Bot 97:846–855

Wilson RW, McQuilkin WE (1963) Silvical characteristics of eastern white pine (*Pinus strobus*). USDA Forest Service Northeasten Forest Experimental Station Research Paper 13:1–28

Zinck JWR, Rajora OP (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 Evol Biol 16:56