scientific reports

Check for updates

OPEN

Genetic diversity in North American *Cercis Canadensis* reveals an ancient population bottleneck that originated after the last glacial maximum

Meher Ony[1], William E. Klingeman[2], John Zobel[3], Robert N. Trigiano[1], Matthew Ginzel[4], Marcin Nowicki[1], Sarah L. Boggess[1], Sydney Everhart[5] & Denita Hadziabdic[1]✉

Understanding of the present-day genetic diversity, population structure, and evolutionary history of tree species can inform resource management and conservation activities, including response to pressures presented by a changing climate. *Cercis canadensis* (Eastern Redbud) is an economically valuable understory tree species native to the United States (U.S.) that is also important for forest ecosystem and wildlife health. Here, we document and explain the population genetics and evolutionary history of this deciduous tree species across its distributed range. In this study, we used twelve microsatellite markers to investigate 691 wild-type trees sampled at 74 collection sites from 23 Eastern U.S. states. High genetic diversity and limited gene flow were revealed in wild, natural stands of *C. canadensis* with populations that are explained by two major genetic clusters. These findings indicate that an ancient population bottleneck occurred coinciding with the last glacial maximum (LGM) in North America. The structure in current populations likely originated from an ancient population in the eastern U.S. that survived LGM and then later diverged into two contemporary clusters. Data suggests that populations have expanded since the last glaciation event from one into several post-glacial refugia that now occupy this species' current geographic range. Our enhanced understanding benchmarks the genetic variation preserved within this species and can direct future efforts in conservation, and resource utilization of adaptively resilient populations that present the greatest genetic and structural diversity.

The genetic structure and demographics of many North American plant species have been greatly influenced by climate fluctuations that occurred during the Pleistocene epoch[1,2]. During the last glacial maximum (LGM), which occurred approximately 18,000–21,000 years ago[3–5], the Laurentide Ice Sheet extended from the northernmost portions of North America to 39°N[3]. These events reduced the range of many temperate tree species, forcing them into glacial refugia, which included unglaciated southern regions and suitable micro-environments that were present in northern glaciated regions[1,6]. Many historically contiguous or closely occurring refugia have been identified in the eastern United States (U.S.), but location delineations and number of refugia continue to be debated[6,7]. These refugia are poorly represented within the fossil record, yet the spatial genetic structure and evolutionary histories of many species have been used as evidence of historical refugia[6,8–10].

According to the "range shift following last glacial maximum" hypothesis, many temperate species recolonized and spread into their current distributions after the LGM[11,12]. The result of this recolonization process can be inferred from the genetic structure within extant plant populations, manifested as reduced genetic diversity along colonization routes, and distinct spatial genetic clusters across the newly expanded range of a species[2,13,14]. Patterns of reduced genetic diversity across a range in expansion are common among European temperate plant species[15]. This trend is not as evident among North American tree species, such as *Carya cordiformis* [Wagenh.]

[1]Department of Entomology and Plant Pathology, University of Tennessee, Knoxville, TN, USA. [2]Department of Plant Sciences, University of Tennessee, Knoxville, TN, USA. [3]Department of Forest Resources, University of Minnesota, St. Paul, MN, USA. [4]Department of Entomology, Purdue University, West Lafayette, IN, USA. [5]Department of Plant Pathology, University of Nebraska, Lincoln, NE, USA. ✉email: dhadziab@utk.edu

K. Koch (bitternut hickory) and *C. ovata* [Mill.] K.Koch (shagbark hickory), which present relatively uniform genetic variation in their distributions[15]. Genetic homogeneity in North American plant species can be explained by slow post-glacial expansion into new areas, presence of many refugia occurring in close proximity, and high gene flow across time[15,16]. Nevertheless, phylogeographic studies provide evidence that range dynamics of the post-glacial species populations have contributed more to the current patterns of genetic diversity in temperate tree species than any other ecological force (e.g., central- periphery theory[17], particularly for the populations in the northern edge of a species distribution[11].

The distribution of tree species, their genetic diversity, and population structure are shaped by many factors including climate oscillations, demographic incidents, ecological and environmental variables, and by their distinct biology[11,18,19]. Yet, the role that glaciation has played in the distribution, range, genetic variation, and spatial genetic structure of outcrossing tree species that span a geographically wide range are not well understood, especially across the eastern U.S. To better understand the role of LGM in structuring current species distributions and population structure of temperate tree species in the eastern U.S., we evaluated the spatial population structure of widely distributed forest understory tree *Cercis canadensis* L. (*C. canadensis* var. *canadensis* L.; Fabaceae; eastern redbud). *Cercis canadensis* is a great system to study the role of LGM in eastern U.S. tree species, due to its wide and continuous geographic distribution without any major geographical barriers across eastern U.S.

Cercis canadensis is a self-incompatible[20], deciduous tree native to the midwestern and eastern U.S., as well as northeastern Mexico[21,22]. This species grows well in partial shade, is well adapted to a wide-range of climate conditions and elevations, and can be found in the USDA hardiness zones 4 through 9[23,24]. This relatively small ornamental tree is characterized by its wide, colorful, umbrella-shaped canopy[25], and is a popular landscape tree due to its heart-shaped foliage, compact form, and early spring flowers[23].

When a fine-scale, smaller, and fragmented population of *C. canadensis* was examined with microsatellite loci[26], wild trees maintained high genetic diversity, gene flow, and moderate to high genetic differentiation[27]. Although *C. canadensis* is ecologically important, there is limited knowledge of the contemporary genetic diversity, spatio-temporal genetic structure, gene flow, and past evolutionary history of this species across its native range in the U.S. To address this knowledge gap, we used microsatellite loci to accomplish the following: (1) characterize the genetic diversity of the wild populations of *C. canadensis* within its native range in the U.S.; (2) infer patterns in the spatial genetic structure of *C. canadensis*; and (3) reveal the evolutionary demographics in its native range. We hypothesized that *C. canadensis* wild populations will be genetically diverse and genotypes will be spatially clustered across its native range. We also hypothesized that the genetic structure would be consistent with range expansion that occurred from one of several southern glacial refugia. More specifically, we aimed to explore the following questions: (1) Do genetic diversity and population structures patterns reflect evidence for northern glacial refugium? We expected to detect high genetic diversity with distinct population structure in northern range limit, otherwise a trend of high to low genetic diversity from south to north would indicate recolonization of *C. canadensis* in north from one or more southern glacial refugia; (2) Is there any evidence for micro-refugia in current *C. canadensis* species distribution range? If there were multiple micro-refugia present in close proximity, especially in southern region, we expected to detect genetic homogeneity in the studied populations.

Materials and methods

Sample collection. Leaf samples of *C. canadensis* were collected by authors, collaborators, and citizen scientists (*see acknowledgements*), who sampled specimens across the native range of this species in 23 states in the midwestern and eastern U.S. (Table 1). The use of trees in the present study complied with international, national and/or institutional guidelines. Plants were identified based on collection guide provided to our collectors and confirmed by co-authors (voucher specimen deposited at the University of Tennessee Vascular Herbarium, catalog # TENN-V-0246136). For each collection site, at least 10 non-cultivated *C. canadensis* trees occurring within a one-mile radius were selected and their geographical coordinates were recorded. From each tree, five to seven young and disease-free leaves were collected, held between several pieces of absorbent paper and stored at ambient room temperature in a paper envelope until processing. Leaf samples from 1193 individual trees were collected at 117 collection sites. To avoid the over-representation of trees within a geographical area, we randomly selected a subset of collection sites from geographical areas with more than one collection site sampled. This study used total of 790 trees representing 79 collection sites that span much of the current native geographic range of *C. canadensis*.

DNA extraction. From each tree, 60 to 100 mg of dry leaf tissue was used to isolate DNA. Samples were homogenized four times for 30 s each at 6 m/s using a Beadmill 24 homogenizer (Fisher Scientific, Pittsburgh, Pennsylvania, U.S.) and were kept in liquid nitrogen for 5 min between each pulverization step. The Qiagen DNeasy Plant Mini Kit (Qiagen, Valencia, California, U.S.) was used to isolate genomic DNA (gDNA) from the pulverized samples with the following minor modifications in the manufacturer's provided protocol. Specifically, 2% w/v polyvinylpyrrolidone (PVP) was mixed into the lysis buffer (AP1). Then 8 μl of RNase was added into each sample tube and incubated at 65 °C in a water bath for 45 min. Every two mins each sample tube was inverted gently to mix the sample well. Lastly, samples were incubated at − 20 °C for at least one hour. Ethanol was used to wash the spin columns if there was any visible remaining debris and elution buffer added. Elution buffer was heated to 65 °C before 50 μl was added to the spin columns twice. Concentrations of gDNA were quantified using ND1000 Ultraviolet-Vis Spectrophotometer (NanoDrop Technologies, Wilmington, Delaware, U.S.) and the gDNA was stored at − 20 °C until further use.

State	County name	Site acronym	Latitude	Longitude	N	MLG	H_E	rbard	P value	Pa
Alabama	Jefferson	AL1	33.7197	− 86.7748	10	10	0.51	0.27	0.01	1
Alabama	Jefferson	AL2	33.5670	− 86.6802	10	10	0.45	0.31	0.01	0
Alabama	Tallapoosa	AL3	32.7705	− 85.7854	10	9	0.42	0.35	0.01	0
Arkansas	Pulaski	AR1	34.7601	− 92.3173	10	10	0.48	0.02	0.22	0
Arkansas	Washington	AR2	36.0878	− 94.1679	10	10	0.52	0.03	0.14	0
Arkansas	Washington	AR3	36.1316	− 94.1336	10	10	0.50	0.03	0.13	0
Florida	Gadsden	FL1	30.6264	− 84.8949	9	9	0.33	0.10	0.02	0
Florida	Alachua	FL2	29.7761	− 82.5081	10	10	0.56	0.24	0.01	1
Florida	Dixie	FL3	29.6367	− 82.7913	10	10	0.52	0.24	0.01	0
Georgia	Whitfield	GA1	34.7287	− 85.0314	10	10	0.53	0.26	0.01	0
Georgia	Clarke	GA2	33.8804	− 83.3572	10	10	0.35	0.10	0.02	0
Georgia	Talbot	GA3	32.6655	− 84.5126	10	10	0.34	0.04	0.11	0
Georgia	Clarke	GA4	33.9032	− 83.3869	10	10	0.48	0.06	0.36	0
Iowa	Fremont	IA2	40.6746	− 95.6908	8	8	0.64	0.10	0.02	1
Indiana	Jefferson	IN1	38.7831	− 85.3695	6	6	0.64	0.09	0.37	1
Indiana	Tippecanoe	IN2	40.5306	− 86.9244	10	10	0.60	0.04	0.36	1
Indiana	Montgomery	IN3	40.0494	− 86.9016	9	9	0.57	0.00	0.76	0
Indiana	Parke	IN4	39.8820	− 87.2020	10	10	0.62	0.03	0.52	0
Kansas	Geary	KS1	38.8941	− 96.8544	10	10	0.56	0.05	0.04	0
Kansas	Butler	KS2	38.8681	− 96.8501	10	10	0.52	0.03	0.12	0
Kentucky	Laurel	KY1	37.2145	− 84.1943	7	7	0.51	0.21	0.01	0
Kentucky	Madison	KY2	37.8949	− 84.2743	10	10	0.46	0.01	0.29	2
Kentucky	Bath	KY3	38.1049	− 83.8265	10	10	0.59	0.10	0.01	0
Kentucky	Carter	KY4	38.3641	− 82.8029 z	8	8	0.48	0.19	0.01	0
Maryland	Carroll	MD1	39.3548	− 76.8963	10	10	0.51	0.07	0.02	0
Maryland	Prince George	MD2	38.7473	− 76.9924	6	6	0.56	0.45	0.01	0
Michigan	Washtenaw	MI1	42.3828	− 83.9068	7	7	0.47	0.06	0.05	0
Michigan	Berrien	MI2	41.9052	− 86.3701	9	9	0.63	0.01	0.93	0
Michigan	Berrien	MI3	41.7903	− 86.7627	9	9	0.66	0.07	0.01	0
Missouri	Boone	MO1	38.8401	− 92.2923	10	10	0.59	0.03	0.12	0
Missouri	Genevieve	MO2	38.0137	− 90.2203	10	10	0.54	0.07	0.01	0
Missouri	Boone	MO3	39.0788	− 92.3066	10	10	0.52	0.11	0.01	1
Missouri	Dent	MO4	37.4552	− 91.6869	10	10	0.52	0.06	0.06	0
Mississippi	Pontotoc	MS2	34.1447	− 88.9972	10	10	0.53	0.13	0.01	0
Mississippi	Oktibbeha	MS3	33.4258	− 88.7004	10	10	0.53	0.24	0.01	1
North Carolina	Guilford	NC1	36.0322	− 79.7063	8	8	0.51	0.03	0.43	0
North Carolina	Cabarrus	NC2	35.4578	− 80.5947	9	9	0.47	− 0.01	0.79	1
North Carolina	Mecklenburg	NC3	35.2415	− 80.9845	10	10	0.47	0.18	0.01	0
North Carolina	Wake	NC4	35.9789	− 78.6368	10	10	0.58	0.03	0.06	0
Nebraska	Sarpy	NE1	41.1799	− 95.9181	10	10	0.68	0.02	0.32	0
New York	Monroe	NY1	43.0339	− 77.5503	10	10	0.53	0.35	0.01	0
New York	Tompkins	NY2	42.3988	− 76.5549	9	9	0.61	0.24	0.01	0
Ohio	Hamilton	OH1	39.0857	− 84.3589	9	9	0.67	0.13	0.01	0
Ohio	Hamilton	OH2	39.0390	− 84.3478	10	10	0.63	0.07	0.01	1
Oklahoma	Cherokee	OK2	35.7603	− 94.9070	10	10	0.54	0.05	0.02	0
Pennsylvania	Centre	PA1	40.8038	− 77.8058	10	10	0.58	0.25	0.01	0
Pennsylvania	Huntingdon	PA2	40.5064	− 77.9806	9	9	0.47	0.20	0.01	0
Pennsylvania	Juniata	PA3	40.5658	− 77.2557	6	6	0.52	− 0.02	0.49	0
South carolina	Pickens	SC1	34.6415	− 82.8241	10	10	0.57	0.11	0.01	0
South carolina	Lexington	SC2	34.1716	− 81.3047	8	8	0.56	0.03	0.75	0
Tennessee	Cocke	TN1	35.8211	− 83.1528	8	8	0.46	0.12	0.02	0
Tennessee	Anderson	TN10	36.0054	− 84.2077	10	10	0.32	0.05	0.13	0
Tennessee	Knox	TN11	35.9462	− 83.9178	9	9	0.38	0.00	0.54	0
Tennessee	Hamilton	TN2	35.0758	− 85.1288	10	10	0.50	0.02	0.26	0
Tennessee	Hamilton	TN3	35.0235	− 85.3791	10	10	0.54	0.04	0.11	0
Tennessee	Shelby	TN4	35.0892	− 89.8661	10	10	0.47	0.31	0.01	0
Continued										

State	County name	Site acronym	Latitude	Longitude	N	MLG	H_E	rbard	P value	Pa
Tennessee	Shelby	TN5	35.2723	− 89.6452	10	10	0.36	0.19	0.01	0
Tennessee	Shelby	TN6	35.3255	− 90.0515	10	10	0.55	0.03	0.12	0
Tennessee	Cheatham	TN7	36.1222	− 87.1416	10	10	0.52	− 0.02	0.70	0
Tennessee	Wilson	TN8	36.0564	− 86.4258	10	10	0.55	0.15	0.01	1
Tennessee	Wilson	TN9	36.1683	− 86.5654	10	9	0.44	0.13	0.01	0
Texas	Sabine	TX1	31.4106	− 94.0225	10	10	0.44	0.02	0.25	1
Texas	Smith	TX2	32.4520	− 95.2534	10	10	0.55	− 0.03	0.92	2
Texas	Collin	TX3	33.0794	− 96.5488	9	9	0.51	0.02	0.25	0
Virginia	Prince William	VA1	38.8121	− 77.5524	9	9	0.52	0.10	0.04	0
Virginia	King George	VA2	38.3293	− 77.0908	7	7	0.57	0.06	0.42	1
Virginia	Brunswick	VA3	36.9225	− 77.7582	10	10	0.55	0.06	0.03	0
Virginia	Washington	VA4	36.6563	− 81.9050	10	10	0.55	0.32	0.01	0
Virginia	Radford	VA5	37.1345	− 80.5218	10	10	0.47	0.38	0.01	0
West Virginia	Cabell	WV1	38.3925	− 82.4245	8	8	0.59	0.12	0.01	0
West Virginia	Kanawha	WV2	38.4842	− 81.4354	10	10	0.56	0.20	0.01	1
West Virginia	Braxton	WV3	38.8735	− 80.6293	8	8	0.48	− 0.01	0.70	1
West Virginia	Monongalia	WV4	39.6034	− 79.9922	7	7	0.50	0.11	0.03	0
Average/total					**691**	**689**	**0.67**	**0.05**	**0.01**	**19**

Table 1. Genetic diversity indices across 74 *Cercis canadensis* collection sites from the United States using 12 microsatellite loci. N = total number of samples per collection site, MLG = number of diploid individuals multilocus genotypes after clone correction, H_E = Nei's genotypic diversity corrected for sample size, rbard = linkage disequilibrium, Pa = number of private alleles in each collection site.

Locus	Repeat motif	Allele No	Ar	H_O	H_E	H	Evenness	F_{ST}	F'_{ST}	F_{IS}	Nm	Pa
127spa	(TC)4	12	1.73	0.99	0.75	1.44	0.64	0.05	0.05	− 0.39	0.82	1
168a	(CT)7	13	1.51	0.23	0.70	1.45	0.69	0.24	0.24	0.57	0.58	4
177b	(GA)6	9	1.64	0.23	0.78	1.70	0.80	0.15	0.15	0.66	0.84	0
199a	(GAGA)8	12	1.55	0.23	0.71	1.58	0.65	0.21	0.21	0.59	0.65	1
220a	(TATT)4	11	1.64	0.57	0.80	1.66	0.64	0.19	0.20	0.11	0.70	2
229a	(GAGAG)4	6	1.50	0.15	0.62	1.11	0.77	0.15	0.15	0.71	0.86	0
625a	(GA)4	11	1.57	0.24	0.68	1.32	0.59	0.13	0.13	0.59	0.81	2
658a	(CT)6	9	1.50	0.21	0.60	1.17	0.69	0.13	0.13	0.59	0.89	3
680a	(GT)8	8	1.37	0.01	0.59	1.15	0.66	0.33	0.33	0.98	0.37	0
762a	(TC)7	11	1.52	0.28	0.68	1.41	0.69	0.21	0.21	0.48	0.64	3
780b	(AG)12	16	1.79	0.70	0.86	2.19	0.75	0.07	0.07	0.12	1.70	1
995a	(AG)7	7	1.23	0.01	0.52	0.84	0.82	0.53	0.54	0.99	0.18	2
Average/Total		**10**	**1.55**	**0.32**	**0.69**	**1.42**	**0.70**	**0.19**	**0.19**	**0.43**	**0.75**	**19**

Table 2. Genetic diversity indices of 12 microsatellite loci across 74 collection sites of *Cercis canadensis*. Ar = allelic richness corrected for sample size, H_O = observed heterozygosity, H_E = Nei's genotypic diversity, H = Shannon–Wiener index, F_{ST} = populations fixation index, F'_{ST} = population differentiation, F_{IS} = inbreeding coefficient, Nm = gene flow, Pa = number of private alleles in each collection site.

Microsatellite primers and genotyping conditions. Initially, gDNA was isolated from five *C. canadensis* individuals from the University of Tennessee Gardens (Knoxville, Tennessee, U.S.) and used to evaluate 68 candidate microsatellite loci[26]. Primers for twelve polymorphic microsatellite loci (Table 2) were selected for this study based on the successful amplification and PCR product size agreement with the published data. Microsatellite loci were amplified with polymerase chain reaction (PCR) in a 10 µl reaction mixture containing the following: 1 µl gDNA, 1 µl of 10 µM of each forward and reverse primer, 0.5 µl of dimethyl sulfoxide, 4 µl of GoTaq G2 Hot Start Master Mix (Promega Corp, Madison, Wisconsin U.S.), and 2.5 µl sterile molecular grade water. To assure validity of the data, both a negative control (reaction mixture with water instead of any DNA sample) and a positive control (a DNA sample from the initial screens that amplified across all microsatellite primers) were incorporated for every primer-pair tested. Amplification of DNA with 12 microsatellite loci across all samples was completed in 96 well plates using an Eppendorf Thermocycler (Eppendorf AG, Hamburg, Germany) with the following thermal profile: initial denaturation at 94 °C for 3 min, followed by 35 cycles of denaturation at

94 °C for 30 s, annealing at 55 °C for 30 s, and an extension at 72 °C for 30 s, with a final extension of 72 °C for 4 min.

Amplified PCR products were visualized with QIAxcel Capillary Electrophoresis System (Qiagen) and analyzed with a 15/600 bp internal alignment marker and a 25 to 500 bp DNA ladder. All *C. canadensis* gDNA samples were amplified and visualized against each of the 12 microsatellite loci using the procedure described above. Reactions not producing any amplified products were rerun once before they were considered missing data. Samples with ≥ 40% missing data were discarded. Also, collection sites with more than four samples having ≥ 40% missing data were excluded from the dataset.

Genetic diversity. Using the Excel macro FLEXBIN version[28], raw allele sizes were converted into allelic classes. In this program, alleles were binned into base-pair (bp) size categories by statistical similarities. This binned genetic dataset was used for all of the following statistical analyses, which were completed using R version 3.5.3[29]. Clone-correction of the data was implemented to identify presence of clonal individuals at the collection site level using the R package POPPR version 2.8.2[30,31]. For each collection site, only multi-locus genotypes (MLG) were used to obtain unbiased estimates of allelic frequency from the dataset[32].

R package POPPR was used to calculate the total number of alleles per locus, observed heterozygosity (H_O; number of the heterozygotes present at a locus which is divided by sample size), expected heterozygosity (H_E; calculated as expected heterozygosity per locus[33]), and linkage disequilibrium (rbard); non-random association of alleles between loci). Additionally, the Shannon-Weiner diversity index (H) was calculated for each collection site using POPPR. H considers both allele richness and evenness of the allelic distribution[34]. The number of unique private alleles in collection sites and different loci was estimated in POPPR package. Allelic richness (Ar), a measure of rarefied allelic counts per locus, was estimated using package HIERFSTAT version 0.04–22[35]. Allelic richness is used as an estimate of the long-term evolutionary potential to adapt and persist in a given population[36,37]. The genetic fixation index (F_{ST}), inbreeding coefficient (F_{IS}), and allelic differentiation (F'_{ST})[38,39] were calculated using HIERFSTAT package. Gene flow (Nm) was estimated using GenAlEx 6.5 software (Peakall & Smouse, 2006; Peakall & Smouse, 2012). In the program, Nm was estimated as the effective number of the migrants per locus based on F-statistics.

Population structure. Population structure within the native range of wild *C. canadensis* trees was analyzed using the program STRUCTURE version 2.3.4[40] to which an admixture model was applied. This Bayesian clustering method with Monte Carlo Markov Chain (MCMC) approach was used with the following parameters: 500,000 burn-in period with 1,000,000 MCMC repetitions for 30 independent chains for *K* values from 1 to 18. The resulting output was visualized with STRUCTURE HARVESTER web version 6.94[41]. The optimum *K* value, indicator of population clusters present in the dataset, was calculated utilizing the Evanno method[42]. The estimation of *ΔK* criterion obtained from STRUCTURE HARVESTER were visualized using POPHELPER 2.2.6[43] that merged the 30 independent chains. R packages MAPS version 3.3.0[44] and PLOTRIX version 3.8–1[45] were used to generate pie charts of admixture proportions at K = 2.

Several model-free methods were utilized to investigate the population structure of *C. canadensis* samples. A Neighbor-Joining (NJ) dendrogram was constructed using Nei's genetic distance in POPPR[46,47]. Discriminant Analysis of Principal Components (DAPC) was implemented using package ADEGENET version 2.1.1[48] to visualize the underlying genetic structure of this species in its wide geographical range. This is a two-step multivariate analysis that investigates the genetic variations within populations among the sampled collection sites[49]. At first, a principal component analysis (PCA) was conducted, and then the number of PCA vectors (to explain majority of variance with minimizing over-fit of the DAPC) was selected. Then, a selected number of PCAs were used to reveal differences between groups while minimizing within group variations, as well as ordination of collection sites into distinct groups using discriminant analysis[49,50]. Moreover, missing values were calculated as mean allele frequency and cross-validation analysis was performed to select appropriate PC numbers.

Isolation by distance (IBD) was estimated using the Mantel test[51,52] with 10,000 permutations in package VEGAN version 2.5–6[53] using Euclidean distance. IBD checks for a correlation between genetic distance and geographical distance among the individuals in a dataset. The Mantel test was implemented across the 74 collection sites while considering the whole dataset as one population.

Analysis of Molecular Variance (AMOVA)[54] was carried out using POPPR with 10,000 permutations by sorting the individuals into hierarchical groups to assess the degree of molecular variance partitioned within, between, and among the collection sites. The levels of population hierarchy included: (1) 74 collection sites as one hierarchical group; (2) two groups on the basis of the STRUCTURE analysis; and (3) four major groups on the basis of five major eco-region divisions namely hot continental division (mountain provinces), hot continental division, warm continental division, subtropical division, and prairie division (see Supporting Information Fig. S2) in the midwestern and eastern U.S. (Bailey, 1994). *C. canadensis* collection sites in warm continental division were grouped with hot continental division collection sites. As *C. canadensis* is found in wide range of climate and elevations, we tested if there was any influence of the regional climate patterns in the population structures of *C. canadensis* in AMOVA analysis.

Demographic histories. To investigate and interpret the evolutionary history of *C. canadensis*, we used DIYABC program version 2.1[55,56] that utilized Approximate Bayesian Computation (ABC) statistical methods. For this analysis, collected individuals were pooled into two major groups, based on the STRUCTURE results. To elucidate the evolutionary history of *C. canadensis*, we analyzed competing scenarios in two ABC steps. In the first step, we tested five demographic scenarios using 200,000 simulated pseudo-observed datasets (PODs) wherein: (1) the first two scenarios suggested stepwise divergence of the current two major groups from

an ancient population, (2) a third scenario suggested a single, two-way split of contemporary groups from an ancient unsampled population, and (3) the last two scenarios were based on the hypothesis of divergence of current groups from two separate ancient un-sampled populations. Once the analysis of these scenarios was completed, the two scenarios from the first step that yielded higher logistic regression support were selected as the basis for assessing the second step of ABC. In the second step, seven scenarios were constructed that addressed the possibility of a bottleneck occurrence within the evolutionary history of the species. Over 1,000,000 pseudo-observed datasets (PODs) were simulated under the assumed prior parameter ranges for each scenario. Posterior probabilities of the compared scenarios were estimated to select the best supported scenario[55].

Results

Microsatellite genetic diversity and hierarchical fixation indices.
Twelve microsatellite loci were amplified from 790 *C. canadensis* trees sampled in this study. Due to presence of missing data (missing ≥ 40% SSRs), five of 79 collection sites were excluded and 49 individuals from the remaining 74 sites were discarded, resulting in 691 individuals from 74 collections. Additionally, after deleting two clonal individuals, 689 unique multilocus genotypes from 74 collection sites remained for further data analyses (Table 1). The average null alleles or missing data across the dataset were overall 2.89% (see Supporting Information Fig. S1). Nei's genetic diversity index (H_E) value in the studied 74 collection sites was 0.67, ranging from 0.32 (Anderson county, TN10) to 0.68 (Sarpy county, NE1) (Table 1). Moreover, weak (rbard = 0.05, *P* value = 0.01) but significant linkage disequilibrium value detected in the dataset. Nineteen private alleles were detected within the 74 collection sites (Table 1). In addition, 9 of the 12 microsatellite loci yielded private alleles, and the highest number of private alleles was recovered from locus 168a (Pa = 4, Table 2). The number of alleles per locus ranged from 6 to 13 with a mean of 10 alleles per locus (Table 2). Overall allelic richness (Ar) ranged from 1.23 for locus 995a to 1.79 for locus 780b, with a mean of 1.55, implying a presence of high allelic richness in wild *C. canadensis* individuals. Observed heterozygosity (H_O) across all loci was 0.32, ranging from 0.01 (locus 995a and 680a) to 0.99 (locus 127spa). The overall expected heterozygosity (H_E) across all 12 microsatellite loci was high (H_E = 0.69), ranging from 0.52 (locus 995a) to 0.86 (locus 780b).

The overall Shannon-Weiner diversity index (H) for the 12 loci was 1.42 and ranged from 0.84 (locus 995a) to 2.19 (locus 780b; Table 2). Additionally, high population fixation (F_{ST} = 0.19; ranging from 0.05 to 0.53; Table 2) and population differentiation (F'_{ST} = 0.19; ranging from 0.05 to 0.54) were identified among *C. canadensis* populations. We estimated an inbreeding coefficient (F_{IS}) of 0.43 across all loci, indicating excess homozygotes (Table 2) among the studied *C. canadensis* populations. The average estimated gene flow was 0.75, which indicates that a limited amount of gene flow has occurred among the studied populations (Table 2).

Population structure.
Using Nei's genetic distance, we estimated pairwise F_{ST} values among the 74 collection sites and the values ranged from 0.02 to 0.33 (see Supporting Information Table S1). STRUCTURE results revealed an optimum $\Delta K = 2$, implying that across its wide native range, *C. canadensis* collection sites are divided into two major clusters. Collection sites in the northern-most collection region of the U.S. (Ohio to Nebraska) and mid-south to mid-north (from Texas to Nebraska) were part of the first cluster (designated as north genetic cluster) (Fig. 1). The rest of the collection sites from the northeast (New York) to mid-south (Mississippi) along the Atlantic Ocean coastline belonged to the second cluster (designated as south genetic cluster). Whereas, a constructed NJ dendrogram revealed the presence of two major groups (except KS and TX collection sites that did not group with any major group), which supported the STRUCTURE findings of presence of two genetic clusters (Fig. 2). In addition, the collection site distribution in these two major groups (NJ dendrogram) is similar to the distribution of collection sites in the two STRUCTURE-based clusters. The DAPC biplot further confirmed the presence of genetic structures, primarily along the x-axis with two overlapping clusters (Fig. 3). Therefore, based on additional analyses used in this study, the grouping of *C. canadensis* individuals is best explained with two genetic clusters (Fig. 1–3). These analyses also showed that the majority of the collection sites (except two collection sites from Georgia) grouped in clusters based on their geographical origin.

In the Analysis of Molecular Variance (AMOVA) analysis, the first data arrangement showed that most of the genetic variation was present within 74 collection sites (74.2%, *P* < 0.001) (Table 3). A significant amount of variation was also partitioned among collection sites (25.8%, *P* < 0.001). When the dataset was divided into two genetic clusters based on STRUCTURE results, 69.2% (*P* < 0.001) of the genetic variability was attributed to the location of collection sites (Table 3). There was also a significant amount of variability between the two different genetic clusters (13.9%, *P* < 0.001) and between collection sites within the clusters (17%, *P* < 0.001) (Table 3). When the dataset was partitioned by the major eco-regions groups that are represented across the distribution of *C. canadensis*, only 7.9% (*P* < 0.001) of the variability could be attributed among eco-region groups, versus 18.6% (*P* < 0.001) variability that was attributed among the collection sites within groups (Table 3). The majority of genetic variation was explained among individuals within a collection site, rather than among populations or group levels for all three tested scenarios (Table 3). Nevertheless, the extent of variation that was observed within collection sites and between clusters revealed the presence of genetic structure. AMOVA results were, therefore, congruent with the hierarchical fixation indices and indicated the presence of population structure. However, the lowest amount of variation was found within the groups when the data were divided according to major eco-regions. This finding suggests that partitioning trees within eco-regions cannot be expected to explain the genetic differentiation and population structure observed in *C. canadensis* wild populations. Results from the isolation-by-distance analysis indicated that among *C. canadensis* populations, the geographical distance effect was weak, but linearly correlated (r = 0.08, *P* < 0.001) with genetic distance (see Supporting Information Fig. S2).
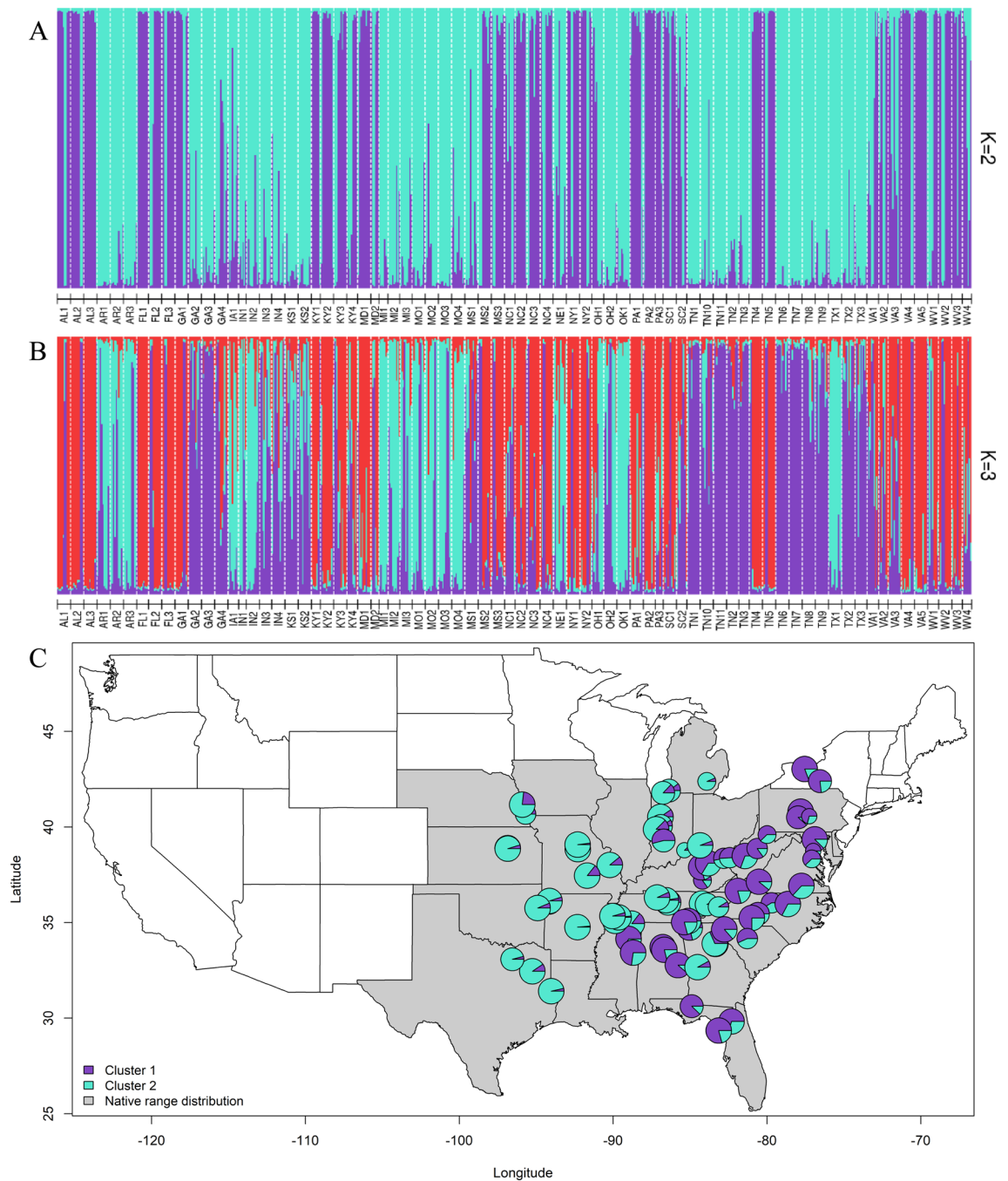
Figure 1. (**A**, **B**, **C**). STRUCTURE bar graphs representing genetic clusters ($\Delta K = 2$–3) among 74 collection sites of *Cercis canadensis* (**A** and **B**). Each vertical bar represents an individual sample and the color of the bar indicates the assignment probability of that individual to belong to one of the identified clusters. Pie charts of admixture coefficients inferred by STRUCTURE ($\Delta K = 2$), plotted across geographical distribution of *C. canadensis* in the eastern United States (C) and 74 collection sites used in this study.

Demographic histories. The DIYABC program with the ABC approach, however, supported the presence of population structure and found evidence for an ancient bottleneck event occurring in *C. canadensis* wild populations. From the first step of the analysis, two probable scenarios were chosen according to their posterior relative support (Scenario 2, posterior probability (P) = 0.39 and Scenario 3, posterior probability (P) = 0.37; Fig. 4A). In these analyses, Scenario 2 provided evidence that the contemporary *C. canadensis* population originated in the southeastern U.S. region from an ancient population and then later, a north population group (first genetic cluster) diverged from the south population group (second genetic cluster). Alternatively, Scenario 3 suggested that both current *C. canadensis* groups (north and south) have split from an ancient, as yet unsampled group (Fig. 4A). In the second step of the ABC analysis (Fig. 4B), the principal component analysis and relative posterior probability tests revealed that Scenario 2a (posterior probability (P) = 0.74, Fig. 4B) was the most
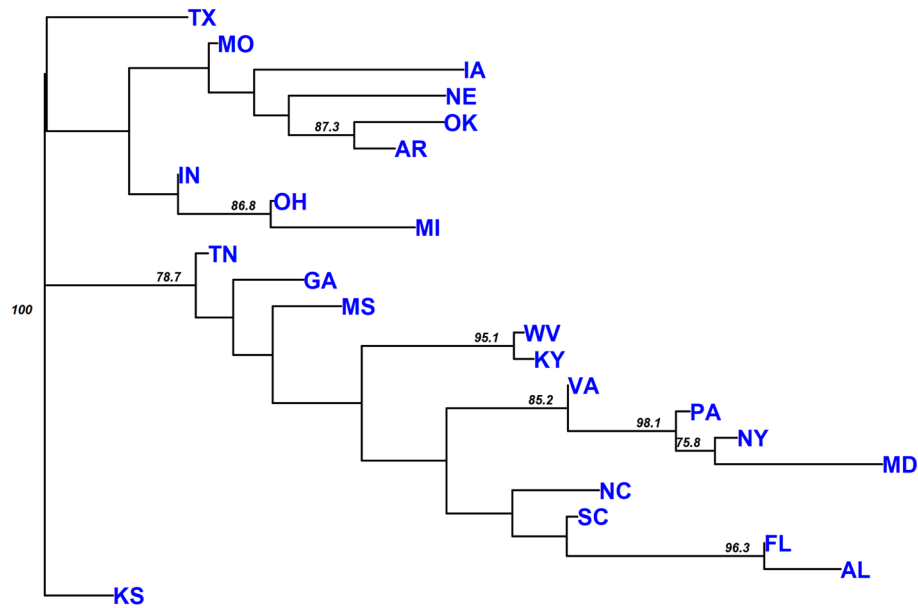
Figure 2. Neighbor-joining tree of 74 collection sites of *Cercis canadensis* using Nei's genetic distance. Numbers indicate the percentage of bootstrap support using 1,000 replications (threshold set at 70%).
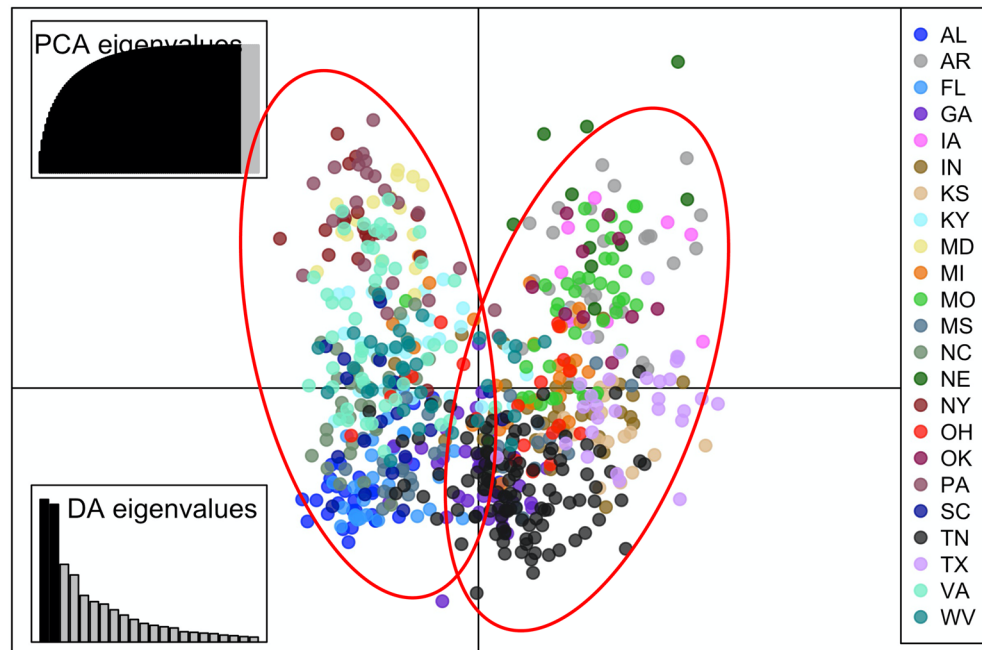
Figure 3. Discriminant Analysis of Principal Components (DAPC) plot of *Cercis canadensis* individuals across 23 states in e U.S. The first 47 principal components explained 94% of the variation in *C. canadensis* individuals in the dataset. Here, allele 154 at locus 199a explained 12% of the variance and allele 102 at locus 220a explained 13% of the variance on the first axis (threshold = 0.06). Data was constructed using 1,000 permutations. Discriminant Analysis eigenvalues are also presented.

supported and therefore had the greatest likelihood of accurately describing the evolutionary processes that are evident within native stands of *C. canadensis*. Thus, we infer from Scenario 2a that from an ancient population, a group of *C. canadensis* in the south first endured a bottleneck, and then later, a group of northern *C. canadensis* diverged from the southern group (Scenario 2a, Fig. 4B).

Estimated posterior parameters of Scenario 2a suggested that a population bottleneck occurred approximately 4,950 generations ago (ranging from 722 to 9,650 generations in simulated datasets), which is approximately

Source of variation	df	Sum Square	Variance Component	% of Variation	P value
(i) 74 collection sites					
Between 74 collection sites	73	1441.17	1.72	25.76	0.001
Within 74 collection sites	561	2785.17	4.96	74.24	0.001
Total	634	4226.35	6.69	100.00	
$F_{ST} = 0.26$					
(ii) Two clusters (STRUCTURE)					
Between two clusters	1	331.78	1.00	13.87	0.001
Between collection sites	72	1109.40	1.22	16.97	0.001
Within collection sites	561	2785.17	4.96	69.15	0.001
Total	634	4226.34	7.18	100.00	
$F_{ST} = 0.31$, $F_{IS} = 0.20$, $F_{CT} = 0.14$					
(iii) Four eco-regions					
Between four eco-regions	3	320.78	0.55	7.88	0.001
Between collection sites within eco-regions	71	1192.76	1.30	18.55	0.001
Within collection sites	600	3087.95	5.15	73.57	0.001
Total	674	4601.49	7.00	100.00	
$F_{ST} = 0.26$, $F_{IS} = 0.20$, $F_{CT} = 0.08$					

Table 3. Analysis of Molecular Variance of *Cercis canadensis* across 12 microsatellite loci for (i) 74 collection sites as one hierarchical cluster, (ii) 74 collection sites divided into two groups (north and south) based on STRUCTURE results (two clusters), and (iii) four eco-regions. F_{ST} = variance among collection sites relative to the total variance. F_{IS} = inbreeding coefficient of individuals relative to population. F_{CT} = variance among groups relative to the total variance.

25,000 years ago, given the average time for a *C. canadensis* tree to reach reproductive maturity is six to seven years[57] (Fig. 4B). Therefore, the bottleneck event most probably occurred during the last glacial period, which ended about 21,000 years ago[3,5]. Later, the northern population diverged from the southern population about 493 generations ago (ranging from 102 to 1,490 generations in simulated datasets) (Fig. 4B). Post-hoc analyses provided goodness-of-fit for this scenario, with the original dataset well embedded in the prior PODs population and nested in the posterior PODs population (see Supporting Information Table S2 for details of this analysis).

Discussion

Wild populations of *C. canadensis* that were sampled across its native range in the U.S. revealed high levels of genetic diversity and population differentiation, the presence of population structure, limited gene flow, and an ancient bottleneck that temporally coincides with the last glacial period in North America. We detected the presence of geographical clusters, longitudinally in the southern region (along U.S. coastal plains), and northern region. Evolutionary history analyses revealed an ancient bottleneck event occurring in the *C. canadensis* population in the south followed by divergence of the northern population from the southern population of *C. canadensis*.

When populations were compared across the ecoregion divisions from which they were collected, ecoregion designations were not associated with population structure and genetic diversity of wild populations of *C. canadensis*. Low genetic variation in *C. canadensis* across ecoregions is not surprising, given that this tree species is well-adapted to a wide range of soil types, environmental conditions and habitats, has not been constrained by any major geographical barriers, and is widespread among the eastern U.S. However, we found evidence of relatively higher genetic diversity among the northernmost collection sites (IA, IN, MI, NE, and OH, and in mid-latitude North America) that are located at the periphery of the contemporary northern range of this species. A plausible explanation for this discrepancy between northernmost samples when compared to the southern collection sites is the probability of one major refugium or admixture between small, but genetically rich, populations in refugial contact zones[58–60]. This effect is most evident in species that experience reproduction via long distance gene flow and local adaptation among sensitive individuals in the distributed population margins. However, unlike European temperate species, eastern north American species maintained high genetic diversity in northern populations[61,62]. This high level of genetic diversity in the *C. canadensis* northern population could be maintained by long-distance seed dispersal events during range expansion from post-glacial northern refugia[62,63].

The ability of *C. canadensis* to maintain high genetic diversity can be influenced by several factors, including wide and continuous geographic distribution, an outcrossing reproductive system, and large effective population size[59,64–66]. Many other temperate tree species sustain high genetic diversity and allelic richness across a wide geographical range even in the presence of environmental stressors[11,14,58], pressure from insect and plant pathogens[67,68], and human disturbances[67,69,70]. A study using microsatellite loci revealed high genetic variation among five Asian *Cercis* spp., averaging 5.7 alleles per locus[71]. A recent study focused on smaller and fragmented *C. canadensis* populations also determined that trees in this species maintain high genetic diversity and allelic richness across the native range[27], congruent with Asian *Cercis* species and several other hardwood tree species[66,72–76].
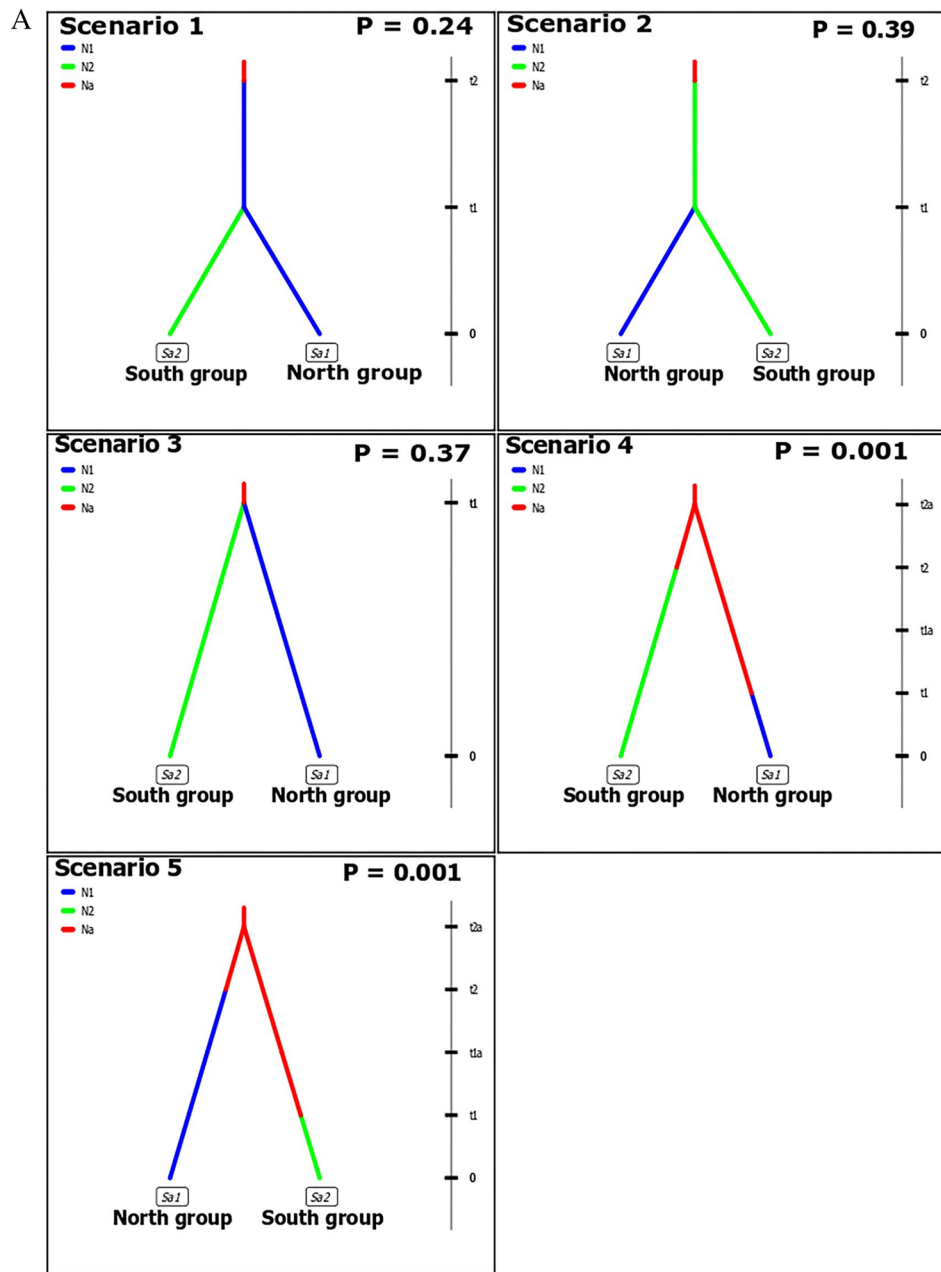
Figure 4. (**A**, **B**). Probable DIYABC evolutionary scenarios for *Cercis canadensis* evolutionary history. Here, current *C. canadensis* populations are north (N1) and south (N2). Also, N1b and N2b represent the populations of N1 and N2 before bottleneck event. On the right side of each scenario, a time scale indicated the timeline of each event (t = 0 is current time, t1-db = bottleneck occurrence, t1/t2 = split of the populations from originating population). Our data was analyzed using two ABC steps resulting in five competing scenarios from the first step (4A), and seven scenarios with possible bottleneck events from the second step analyses (4B). The scenario 2a (B) from step 2 gained the most support in DIYABC analysis and the timeline for the bottleneck (t1-db) and divergence (t1) events of scenario 2a is given in generations. For each scenario, value of relative posterior probability (P) was reported.

In addition to high genetic variability, *C. canadensis* populations also display a wide range of morphological variation across diverse environmental conditions[20,26,77–79]. For example, *Cercis* leaf shape, size, surface pubescence, and other structural features were found to be strongly related to environmental factors, such as temperature and moisture content[79–81]. In *Cercis* spp., these characteristics likely originated through local adaptation to varying climatic pressure[26,82,83]. Morphological variation in *C. canadensis* has led efforts to differentiate the species into the following three varieties: *C. canadensis* var. *canadensis* L., distributed across mesophytic habitats of the eastern U.S., *C. canadensis* var. *mexicana* (Rose) M. Hopkins (Mexican redbud) and *C. canadensis*
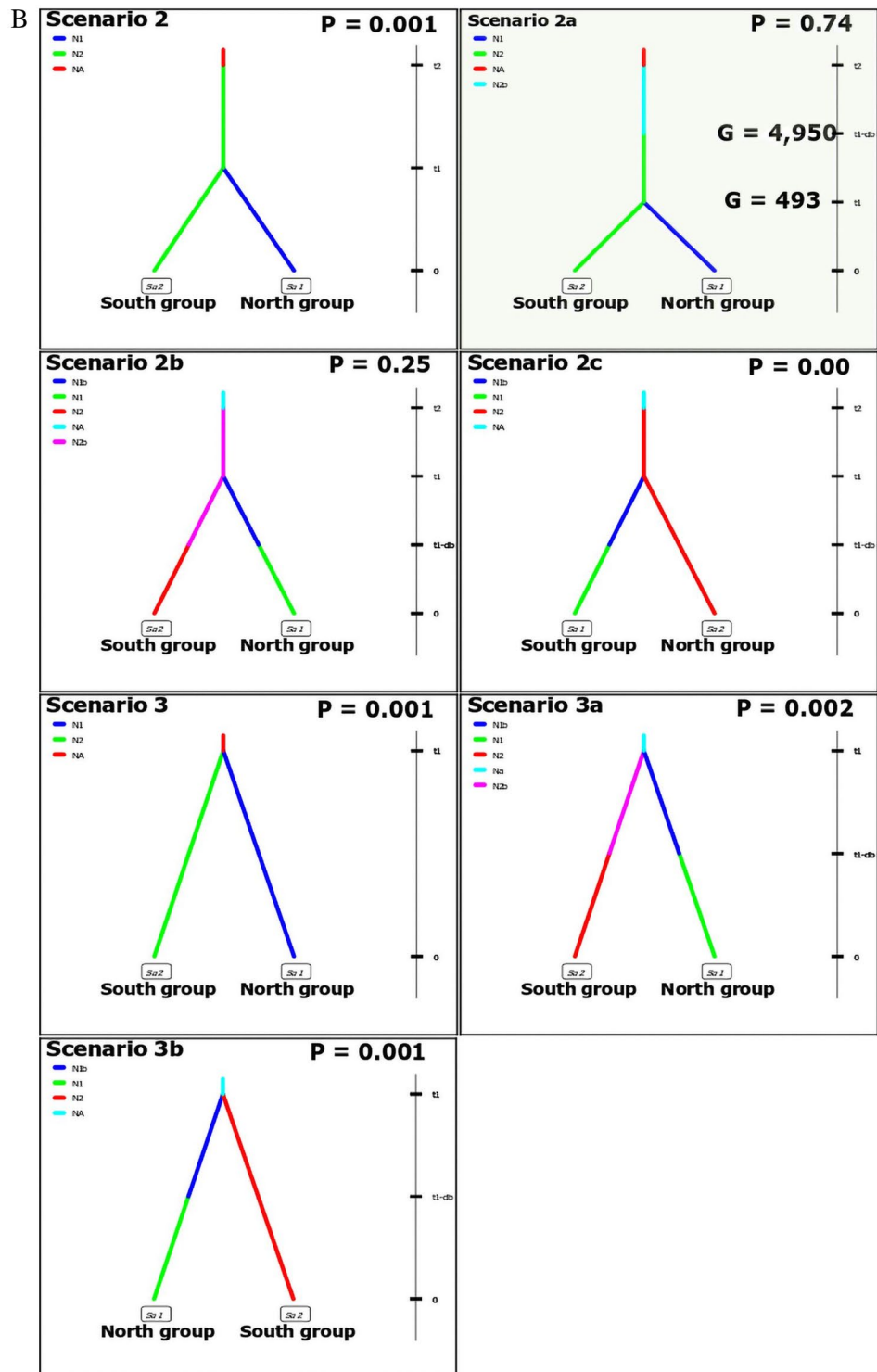
Figure 4. (continued)

var. *texensis* (S. Watson) M. Hopkins (Texas redbud), which are both commonly found in semi-arid regions of central Mexico and southwestern Texas[26,84–86]. However, the validity of such sub-specific classification has been questioned because of the highly continuous pattern of morphological variation in *C. canadensis* populations across its range. Moreover, current phylogenetic studies were unable to provide sufficient support to validate these divisions[84,87]. Because *C. canadensis* var. *mexicana* and *C. canadensis* var. *texensis* were not represented in our study, our data will not assist in resolving this question.

Widely distributed tree species that grow in large populations usually have low genetic differentiation and have limited population structure across their geographic range[15,60,65,66,68,88]. Populations of *Viburnum rufidulum* Raf.[89] and *Cornus florida* L.[68] are temperate tree species that are widely distributed in the southeastern U.S. and

have low genetic differentiation with weak population structures. Across populations of *V. rufidulum* and *C. florida*, high levels of gene flow via pollen and seed dispersal may have reduced genetic variability[75,89]. Contrary to these studies, high genetic differentiation observed among widely distributed populations of *C. canadensis* may be attributed to a limited gene flow, as well as the demographic history of the species.

Similar to many other self-incompatible forest tree species[20], gene flow in *C. canadensis* is dependent upon various pollen and seed dispersal mechanisms. Flight distance of insect pollinators varies from one to several miles[90,91], which would limit long distance gene flow by pollen dispersal among trees. Seedpods and seeds of *C. canadensis* are relatively heavy and typically fall in close proximity to the parental tree. Progeny that survive grow as non-reproductive seedlings during the next several years[57,92] and yield half-sib "neighborhoods" within a localized spatial scale[89,93–96]. Several seed-feeding mammals, such as eastern woodrats (*Neotoma floridana* Ord)[97], and birds including quail[98] contribute to the dispersal of *C. canadensis* seed to some extent. Small rodents and deer may repeatedly eat from the same tree, thus carrying the closely related, half-sibling propagules (if eaten when seeds have matured) for distances restricted to the retention time of fecal scats[57,97,99–101]. To fully understand gene flow patterns and predict changes to *C. canadensis* distribution patterns, it may be helpful to unravel the seed and pollen dispersal methods and efficacy of seed transport by animals that have been associated with this tree species. However, seed transport efficacy will likely be limited to the relatively short distances traveled by these animals during foraging. Fruit consumption rate by animals is also restricted by reliance upon *C. canadensis* fruits as emergency food in late fall or winter, and this behavior would lower efficiency of functional seed dispersal[57,97,99–101]. These events likely limit the gene flow to short distances, create spatial genetic structures, and increase the likelihood of inbreeding at a local level, as revealed in fine scale level assessments of *C. canadensis*[27,69,102]. We also collected *C. canadensis* samples from New York (U.S.) that represent individuals occurring farther north than the reported geographic range of the species. These individuals also could result from open-pollinated escapes subsequent to introduction of *C. canadensis* into managed landscapes.

STRUCTURE analysis of the *C. canadensis* dataset revealed the presence of two geographically distinct clusters, designated as northern and southern clusters, that are divided longitudinally northwest by southeast along a Kentucky-Tennessee-Mississippi transition zone. From this evidence, the southern Appalachian Mountains have not posed a barrier, as populations belonging to the northern STRUCTURE clusters were found on both sides of the Appalachian Mountains. The presence of only two genetic clusters is congruent with the simple postglacial lineage theory presented for eastern North American tree species[59,103]. The most recent glacial event ended approximately 21,000 years ago. By its conclusion, boreal and temperate tree species had shifted closer to mid-latitudes within the eastern U.S., where many species survived within bottlenecked refugial populations[104]. According to our DIYABC supported scenario, a southeastern refugium was also likely to be the major postglacial refugium for *C. canadensis*. This scenario is further supported by several phylogenetic studies that have indicated that southeastern U.S. populations served as one prominent large-scale, post-glacial refugium for many temperate species[14,103,104]. Modern day temperate species including *Fagus grandifolia* Ehrh. (American beech), *Acer rubrum* L. (Red maple), and *C. florida* (Flowering dogwood), for example, likely originated from this southeastern refugium[4,14,105]. *Cercis canadensis* also shares the same geographic distribution as these temperate tree species, and modern-day wild populations of *C. canadensis* are ubiquitous throughout this region.

Our analyses also revealed support for several possible micro-refugia across the eastern U.S., evident in genetic differences among populations and presence of substructures that lack distinct centers. Several studies on different tree species indicated presence of refugia in the eastern U.S. such as southern Appalachian Mountains, southeastern coastal plains, and lower Mississippi River Valley (McLachlan et al., 2005; Potter et al. 2011). Post-glacial *C. canadensis* populations from this geographical range may have spread northward to establish the current species distribution. Post-glacial populations of other tree species from this range are adapted to semi-arid to xeric environments[9,14,103] and present adaptive characteristics that are similarly evident in mid-western *C. canadensis* populations. Moreover, the presence of a number of refugia or fragmented refugia is also can be supported by the high genetic diversity and allelic richness of the modern-day *C. canadensis* populations[60].

Phylogeographical studies of other tree species and animals indicate that they survived as northern cryptic micro-refugia[104,106–108]. Due to insufficient fossil data from the Late Pleistocene in the northern region, it is difficult to conclude the definite presence of northern refugia of the *C. canadensis* populations[5,6]. The best supported DIYABC evolutionary scenario suggests that at the time of the last glacial period *C. canadensis* populations persisted within a southern population group. Therefore, we also find little evidence for the possibility of a northern cryptic refugium for this species. Instead, pre-glacial *C. canadensis* was distributed in midwestern and southeastern U.S. populations, which later survived in one or more post-glacial midwestern and southeastern U.S. refugia. As a consequence of long-term population isolation within the refugial areas, a post-glacial refugial population in the midwestern U.S. may have diverged from the southeastern large refugium population, giving rise to a genetically differentiated northern spatial cluster[70,95]. Also, this post-glacial northern population may have later migrated from the midwestern U.S. to establish the current range distribution.

Ancestors of North American *Cercis* species are thought to have originated under mesic conditions and may have dispersed into North America across the North Atlantic Land Bridge[81,84,109]. According to several studies, ancestral *Cercis* population adapted to the drier environment and then spread into the Northern hemisphere during the mid-Miocene period[84,87]. It is possible, then, that this ancestral, un-sampled, mid-Miocene *Cercis* population gave rise to the southern *C. canadensis* population as suggested by DIYABC Scenario 2a.

This economically and ecologically significant deciduous shade tree species has a number of desirable morphological variations and ornamental characteristics including foliar color and texture, flower color variation, drought tolerance, pathogen resistance, as well as a wide variety of architectural forms[20,26,57]. Fruits and seeds of *C. canadensis* are consumed by several bird species and small mammals[57,97,99,101], and many pollinators depend on this tree for an early season food source[110]. More than three dozen cultivars are available commercially and nursery stock sales of the species contribute to more than $27 M annually in the U.S.[71,111]. When paired with

recent introductions of novel horticultural cultivars with highly desirable characteristics, the value of adaptive traits that are likely to exist across different geographic localities supports the importance in conserving local level diversity of *C. canadensis*. These populations are genetic reservoirs of potential variability that can provide breeding programs with the resources needed to improve selected traits (e.g., limiting seed pod productivity in landscape specimens) and provide additional opportunities for developing high-value cultivars for commercial trade. Future work should also focus on identifying important adaptive traits in wild populations that can be used to help ensure that *C. canadensis* populations will persist and will continue to adapt to a changing climate that is occurring across portions of the current species distribution.

Data Archiving Statement

References

1. Hewitt, G. The genetic legacy of the Quaternary ice ages. *Nature* **405**, 907–913. https://doi.org/10.1038/35016000 (2000).
2. Hewitt, G. Genetic consequences of climatic oscillations in the Quaternary. *Philos. Trans. R. Soc. Lond.* **359**, 183–195. https://doi.org/10.1098/rstb.2003.1388 (2004).
3. Ehlers, J. & Gibbard, P. *Quaternary Glaciations-Extent and Chronology: Part I: Europe* Vol. 2 (Elsevier, New York, 2004).
4. Call, A. *et al.* Genetic structure and post-glacial expansion of *Cornus florida* L. (Cornaceae): Integrative evidence from phylogeography, population demographic history, and species distribution modeling. *J. Syst. Evol.* **54**, 136–151. https://doi.org/10.1111/jse.12171 (2016).
5. Jackson, S. *et al.* Vegetation and environment in eastern North America during the Last Glacial Maximum. *Quatern. Sci. Rev.* **19**, 489–508. https://doi.org/10.1016/S0277-3791(99)00093-1 (2000).
6. Nadeau, S. *et al.* 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. https://doi.org/10.3732/ajb.1500160 (2015).
7. Beaulieu, J. & Simon, J. Genetic structure and variability in *Pinus strobus* in Quebec. *Can. J. For. Res.* **24**, 1726–1733. https://doi.org/10.1139/x94-223 (1994).
8. Provan, J. & Bennett, K. Phylogeographic insights into cryptic glacial refugia. *Trends Ecol. Evol.* **23**, 564–571. https://doi.org/10.1016/j.tree.2008.06.010 (2008).
9. Soltis, D., Morris, A., McLachlan, J., Manos, P. & Soltis, P. Comparative phylogeography of unglaciated eastern North America. *Mol. Ecol.* **15**, 4261–4293. https://doi.org/10.1111/j.1365-294X.2006.03061.x (2006).
10. Mee, J. & Moore, J. The ecological and evolutionary implications of microrefugia. *J. Biogeogr.* **41**, 837–841. https://doi.org/10.1111/jbi.12254 (2014).
11. Hoban, S. *et al.* Range-wide distribution of genetic diversity in the North American tree *Juglans cinerea*: A product of range shifts, not ecological marginality or recent population decline. *Mol. Ecol.* **19**, 4876–4891. https://doi.org/10.1111/j.1365-294X.2010.04834.x (2010).
12. Hampe, A. & Petit, R. Conserving biodiversity under climate change: The rear edge matters. *Ecol. Lett.* **8**, 461–467. https://doi.org/10.1111/j.1461-0248.2005.00739.x (2005).
13. Excoffier, L., Foll, M. & Petit, R. Genetic consequences of range expansions. *Annu. Rev. Ecol. Evol. Syst.* **40**, 481–501. https://doi.org/10.1146/annurev.ecolsys.39.110707.173414 (2009).
14. McLachlan, J., Clark, J. & Manos, P. Molecular indicators of tree migration capacity under rapid climate change. *Ecology* **86**, 2088–2098. https://doi.org/10.1890/04-1036 (2005).
15. Bemmels, J. & Dick, C. Genomic evidence of a widespread southern distribution during the Last Glacial Maximum for two eastern North American hickory species. *J. Biogeogr.* **45**, 1739–1750. https://doi.org/10.1111/jbi.13358 (2018).
16. Jaramillo-Correa, J., Beaulieu, J., Khasa, D. & Bousquet, J. Inferring the past from the present phylogeographic structure of North American forest trees: Seeing the forest for the genes. *Can. J. For. Res.* **39**, 286–307. https://doi.org/10.1139/X08-181 (2009).
17. Eckert, C., Samis, K. & Lougheed, S. Genetic variation across species' geographical ranges: The central–marginal hypothesis and beyond. *Mol. Ecol.* **17**, 1170–1188. https://doi.org/10.1111/j.1365-294X.2007.03659.x (2008).
18. Foll, M. & Gaggiotti, O. Identifying the environmental factors that determine the genetic structure of populations. *Genetics* **174**, 875–891. https://doi.org/10.1534/genetics.106.059451 (2006).
19. Loveless, M. & Hamrick, J. Ecological determinants of genetic structure in plant populations. *Ann. Rev. Ecol. Syst.* **15**, 65–95. https://doi.org/10.1146/annurev.es.15.110184.000433 (1984).
20. Roberts, D., Werner, D., Wadl, P. & Trigiano, R. Inheritance and allelism of morphological traits in eastern redbud (*Cercis canadensis* L.). *Hortic. Res.* **2**, 1–11 (2015).
21. Couvillon, G. *Cercis canadensis* L. seed size influences germination rate, seedling dry matter, and seedling leaf area. *HortScience* **37**, 206–207 (2002).
22. Li, S. *et al.* Methods for breaking the dormancy of eastern redbud (*Cercis canadensis*) seeds. *Seed Sci. Technol.* **41**, 27–35 (2013).
23. Cheong, E. & Pooler, M. Micropropagation of Chinese redbud (*Cercis yunnanensis*) through axillary bud breaking and induction of adventitious shoots from leaf pieces. *In Vitro Cell. Dev. Biol. Plant* **39**, 455–458 (2003).
24. Pooler, M., Jacobs, K. & Kramer, M. Differential resistance to *Botryosphaeria ribis* among *Cercis* taxa. *Plant Dis.* **86**, 880–882. https://doi.org/10.1094/PDIS.2002.86.8.880 (2002).
25. Trigiano, R., Beaty, R. & Graham, E. Somatic embryogenesis from immature embryos of redbud (*Cercis canadensis*). *Plant Cell Rep.* **7**, 148–150. https://doi.org/10.1007/BF00270127 (1988).
26. Wadl, P., Trigiano, R., Werner, D., Pooler, M. & Rinehart, T. Simple sequence repeat markers from *Cercis canadensis* show wide cross-species transfer and use in genetic studies. *J. Am. Soc. Hortic. Sci.* **137**, 189–201. https://doi.org/10.21273/JASHS.137.3.189 (2012).
27. Ony, M. *et al.* Habitat fragmentation influences genetic diversity and differentiation: Fine-scale population structure of *Cercis canadensis* (eastern redbud). *Ecol. Evol.* **10**, 3655–3670. https://doi.org/10.1002/ece3.6141 (2020).
28. Amos, W. *et al.* Automated binning of microsatellite alleles: Problems and solutions. *Mol. Ecol. Resour.* **7**, 10–14. https://doi.org/10.1111/j.1471-8286.2006.01560.x (2007).
29. R: A language and environment for statistical computing (R Foundation for Statistical Computing, Vienna, Austria, 2019).
30. Kamvar, Z., Tabima, J. & Grünwald, N. Poppr: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* **2**, e281. https://doi.org/10.7717/peerj.281 (2014).

31. Kamvar, Z., Brooks, J. & Grünwald, N. Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. *Front. Genet.* **6**, 208. https://doi.org/10.3389/fgene.2015.00208 (2015).

32. Tsui, C. *et al.* Population structure and migration pattern of a conifer pathogen, *Grosmannia clavigera*, as influenced by its symbiont, the mountain pine beetle. *Mol. Ecol.* **21**, 71–86. https://doi.org/10.1111/j.1365-294X.2011.05366.x (2012).

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

34. Shannon, C. E. A mathematical theory of communication. *Bell System Tech. J.* **27**, 379–423 (1948).

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

36. Hurlbert, S. The nonconcept of species diversity: A critique and alternative parameters. *Ecology* **52**, 577–586. https://doi.org/10.2307/1934145 (1971).

37. El Mousadik, A. & Petit, R. High level of genetic differentiation for allelic richness among populations of the Argan tree [*Argania spinosa* (L.) Skeels] endemic to Morocco. *Theor. Appl. Genet.* **92**, 832–839. https://doi.org/10.1007/BF00221895 (1996).

38. Bird, C., Karl, S., Smouse, P. & Toonen, R. In *Phylogeography and Population Genetics in Crustacea* Vol. 19 (eds Held Christoph, Koenemann Stefan, & Schubart Christoph) pp. 31–55 (Boca Raton, FL: CRC Press, 2011).

39. Meirmans, P. & Hedrick, P. Assessing population structure: F_{ST} and related measures. *Mol. Ecol. Resour.* **11**, 5–18. https://doi.org/10.1111/j.1755-0998.2010.02927.x (2011).

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

41. Earl, D. & Bridgett, V. STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.* **4**, 359–361. https://doi.org/10.1007/s12686-011-9548-7 (2012).

42. Evanno, G., Regnaut, S. & Goudet, J. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Mol. Ecol.* **14**, 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x (2005).

43. Francis, R. Pophelper: An R package and web app to analyse and visualize population structure. *Mol. Ecol. Resour.* **17**, 27–32. https://doi.org/10.1111/1755-0998.12509 (2017).

44. Becker, R. & Wilks, A. *MAPS: An R Package to Drae Geographical Maps* (Version package 3.3.0, 2018).

45. Lemon, J. *Plotrix: An R Package for Various Plotting Functions* (Version R package 3.8–1, 2006).

46. Bruvo, R., Michiels, N., D'souza, T. & Schulenburg, H. A simple method for the calculation of microsatellite genotype distances irrespective of ploidy level. *Mol. Ecol.* **13**, 2101–2106. https://doi.org/10.1111/j.1365-294X.2004.02209.x (2004).

47. Grünwald, N., Everhart, S., Knaus, B. & Kamvar, Z. Best practices for population genetic analyses. *Phytopathology* **107**, 1000–1010. https://doi.org/10.1094/PHYTO-12-16-0425-RVW (2017).

48. Jombart, T. & Ahmed, I. adegenet 1.3–1: New tools for the analysis of genome-wide SNP data. *Bioinformatics* **27**, 3070–3072. https://doi.org/10.1093/bioinformatics/btr521 (2011).

49. Jombart, T., Devillard, S. & Balloux, F. Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. *BMC Genet.* **11**, 9. https://doi.org/10.1186/1471-2156-11-94 (2010).

50. Cullingham, C., Cooke, J. & Coltman, D. Effects of introgression on the genetic population structure of two ecologically and economically important conifer species: Lodgepole pine (*Pinus contorta* var. *latifolia*) and jack pine (Pinus banksiana). *Genome* **56**, 577–585. https://doi.org/10.1139/gen-2013-0071 (2013).

51. Diniz-Filho, J. *et al.* Mantel test in population genetics. *Genet. Mol. Biol.* **36**, 475–485. https://doi.org/10.1590/S1415-475720130004000002 (2013).

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

53. Vegan: Community ecology package v. R package version 2.5–3 (R package version 2.5–3). (2018).

54. Excoffier, L., Smouse, P. & Quattro, J. Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics* **131**, 479–491 (1992).

55. Cornuet, J., Ravigné, V. & Estoup, A. Inference on population history and model checking using DNA sequence and microsatellite data with the software DIYABC (v1.0). *BMC Bioinform.* **11**, 401–411. https://doi.org/10.1186/1471-2105-11-401 (2010).

56. Cornuet, J. *et al.* DIYABC v2.0: A software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. *Bioinformatics* **30**, 1187–1189. https://doi.org/10.1093/bioinformatics/btt763 (2014).

57. Dickson, J. In *Silvics of North America* Vol. 2 (eds Burns, R. & Honkala, B.) 266–269 (United States Department of Agriculture-Forest Service, 1990).

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

59. Petit, R. *et al.* Glacial refugia: Hotspots but not melting pots of genetic diversity. *Science* **300**, 1563–1565 (2003).

60. David, R. & Hamann, A. Glacial refugia and modern genetic diversity of 22 western North American tree species. *Proc. R. Soc. B Biol. Sci.* **282**, 20142903. https://doi.org/10.1098/rspb.2014.2903 (2015).

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

62. Bialozyt, R., Ziegenhagen, B. & Petit, R. Contrasting effects of long distance seed dispersal on genetic diversity during range expansion. *J. Evol. Biol.* **19**, 12–20. https://doi.org/10.1111/j.1420-9101.2005.00995.x (2006).

63. Petit, R. Early insights into the genetic consequences of range expansions. *Heredity* **106**, 203–204. https://doi.org/10.1038/hdy.2010.60 (2011).

64. Dubreuil, M. *et al.* Genetic effects of chronic habitat fragmentation revisited: Strong genetic structure in a temperate tree, *Taxus baccata* (Taxaceae), with great dispersal capability. *Am. J. Bot.* **97**, 303–310. https://doi.org/10.3732/ajb.0900148 (2010).

65. Hamrick, J., Godt, M. & Sherman-Broyles, S. In *Population Genetics of Forest Trees* Vol. 42 (eds Adams, W., Strauss, S., Copes, D. & Griffin, A) 95–124 (Springer, Dordrecht, 1992).

66. Hamrick, J. & Godt, M. Effects of life history traits on genetic diversity in plant species. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* **351**, 1291–1298 (1996).

67. Spaulding, H. & Rieske, L. The aftermath of an invasion: Structure and composition of central appalachian hemlock forests following establishment of the hemlock woolly adelgid, *Aelges tsugae*. *Biol. Invasions* **12**, 3135–3143. https://doi.org/10.1007/s10530-010-9704-0 (2010).

68. Hadziabdic, D. *et al.* Analysis of genetic diversity in flowering dogwood natural stands using microsatellites: The effects of dogwood anthracnose. *Genetica* **138**, 1047–1057. https://doi.org/10.1007/s10709-010-9490-8 (2010).

69. Marquardt, P., Echt, C., Epperson, B. & Pubanz, D. Genetic structure, diversity, and inbreeding of eastern white pine under different management conditions. *Can. J. For. Res.* **37**, 2652–2662 (2007).

70. Potter, K. *et al.* Widespread inbreeding and unexpected geographic patterns of genetic variation in eastern hemlock (*Tsuga canadensis*), an imperiled North American conifer. *Conserv. Genet.* **13**, 475–498. https://doi.org/10.1007/s10592-011-0301-2 (2012).

71. Thammina, C., Kidwell-Slak, D., Lura, S. & Pooler, M. SSR markers reveal the genetic diversity of asian *Cercis* taxa at the US National Arboretum. *HortScience* **52**, 498–502. https://doi.org/10.21273/hortsci11441-16 (2017).

72. Chang, C., Bongarten, B. & Hamrick, J. Genetic structure of natural populations of black locust (*Robinia pseudoacacia* L.) at Coweeta, North Carolina. *J. Plant Res.* **111**, 17–24. https://doi.org/10.1007/BF02507146.pdf (1998).
73. Marquardt, P. & Epperson, B. Spatial and population genetic structure of microsatellites in white pine. *Mol. Ecol.* **13**, 3305–3315. https://doi.org/10.1111/j.1365-294X.2004.02341.x (2004).
74. Victory, E., Glaubitz, J., Rhodes-Jr, O. & Woeste, K. Genetic homogeneity in *Juglans nigra* (Juglandaceae) at nuclear microsatellites. *Am. J. Bot.* **93**, 118–126. https://doi.org/10.3732/ajb.93.1.118 (2006).
75. Hadziabdic, D. *et al.* Genetic diversity of flowering dogwood in the Great Smoky Mountains National Park. *Tree Genet. Genomes* **8**, 855–871. https://doi.org/10.1007/s11295-012-0471-1 (2012).
76. Nybom, H. Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. *Mol. Ecol.* **13**, 1143–1155. https://doi.org/10.1111/j.1365-294X.2004.02141.x (2004).
77. Donselman, H. Variation in native populations of eastern redbud (*Cercis canadensis* L.) as influenced by geographic location [USA]. In *Proceedings, of the Florida State Horticulture Society* Vol. 89. 370–373 (1976).
78. Dirr, M. *Manual of Woody Landscape Plants: Their Identification, Ornamental Characteristics, Culture, Propagation and Uses* (Stipes Publishing Co, Champaign, 1990).
79. Fritsch, P., Schiller, A. & Larson, K. Taxonomic implications of morphological variation in *Cercis canadensis* (Fabaceae) from Mexico and adjacent parts of Texas. *Syst. Bot.* **34**, 510–520. https://doi.org/10.1600/036364409789271254 (2009).
80. Nevo, E. *et al.* Drought and light anatomical adaptive leaf strategies in three woody species caused by microclimatic selection at evolution canyon, Israel. *Israel J. Plant Sci.* **48**, 33–46 (2000).
81. Fritsch, P. *et al.* Leaf adaptations and species boundaries in North American *Cercis*: Implications for the evolution of dry floras. *Am. J. Bot.* **105**, 1577–1594. https://doi.org/10.1002/ajb2.1155 (2018).
82. Raulston, J. Redbud. *Am. Nurseryman* **171**, 39–51 (1990).
83. Robertson, K. *Cercis*: The redbuds. *Arnoldia* **36**, 37–49 (1976).
84. Davis, C., Fritsch, P., Li, J. & Donoghue, M. Phylogeny and biogeography of *Cercis* (Fabaceae): Evidence from nuclear ribosomal ITS and chloroplast ndhF sequence data. *Syst. Bot.* **27**, 289–302. https://doi.org/10.1043/0363-6445-27.2.289 (2002).
85. Hopkins, M. In *Rhodora* Vol. 44 (eds M Fernald, C Eatherby, L Griscom, & S Marris) 193–211 (New England Botanical Club, Inc., 1942).
86. Griffin, J., Ranney, T. & Pharr, D. Heat and drought influence photosynthesis, water relations, and soluble carbohydrates of two ecotypes of redbud *(Cercis canadensis)*. *J. Am. Soc. Hortic. Sci.* **129**, 497–502. https://doi.org/10.21273/JASHS.129.4.0497 (2004).
87. Fritsch, P. & Cruz, B. Phylogeny of *Cercis* based on DNA sequences of nuclear ITS and four plastid regions: Implications for transatlantic historical biogeography. *Mol. Phylogenet. Evol.* **62**, 816–825. https://doi.org/10.1016/j.ympev.2011.11.016 (2012).
88. Chung, M., Chung, M., Oh, G. & Epperson, B. Spatial genetic structure in a *Neolitsea sericea* population (Lauraceae). *Heredity* **85**, 490–497. https://doi.org/10.1046/j.1365-2540.2000.00781.x (2000).
89. Dean, D. *et al.* Analysis of genetic diversity and population structure for the native tree *Viburnum rufidulum* occurring in Kentucky and Tennessee. *J. Am. Soc. Hortic. Sci.* **140**, 523–531. https://doi.org/10.21273/JASHS.140.6.523 (2015).
90. Hagler, J., Mueller, S., Teuber, L., Machtley, S. & Van-Deynze, A. Foraging range of honey bees, *Apis mellifera*, in alfalfa seed production fields. *J. Insect Sci.* **11**, 144. https://doi.org/10.1673/031.011.14401 (2011).
91. Pasquet, R. *et al.* Long-distance pollen flow assessment through evaluation of pollinator foraging range suggests transgene escape distances. *Proc. Natl. Acad. Sci.* **105**, 13456–13461 (2008).
92. Hayden, W. Redbud seedpods hold surprises. *Bull. Virginia Native Plant Soc.* **32**, 1–6 (2013).
93. Schnabel, A., Laushman, R. & Hamrick, J. Comparative genetic structure of two co-occurring tree species, *Maclura pomifera* (Moraceae) and *Gleditsia triacanthos* (Leguminosae). *Heredity* **67**, 357–364. https://doi.org/10.1038/hdy.1991.99 (1991).
94. Nakanishi, A., Tomaru, N., Yoshimaru, H., Manabe, T. & Yamamoto, S. Effects of seed- and pollen-mediated gene dispersal on genetic structure among *Quercus salicina* saplings. *Heredity* **102**, 182–189. https://doi.org/10.1038/hdy.2008.101 (2008).
95. Vekemans, X. & Hardy, O. New insights from fine-scale spatial genetic structure analyses in plant populations. *Mol. Ecol.* **13**, 921–935. https://doi.org/10.1046/j.1365-294X.2004.02076.x (2004).
96. Gonzales, E., Hamrick, J., Smouse, P., Trapnell, D. & Peakall, R. The impact of landscape disturbance on spatial genetic structure in the Guanacaste tree, *Enterolobium cyclocarpum* (Fabaceae). *J. Hered.* **101**, 133–143. https://doi.org/10.1093/jhered/esp101 (2009).
97. Post, D. Change in nutrient content of foods stored by eastern woodrats (*Neotoma floridana*). *J. Mammal.* **73**, 835–839 (1992).
98. Surrency, D. & Owsley, C. (ed. Natural Resources Conservation Service United States Department of Agriculture) 146 (United States Department of Agriculture, Natural Resources Conservation Service, 2001).
99. Wakeland, B. & Swihart, R. Ratings of white-tailed deer preferences for woody browse in Indiana. *Proceedings of the Indiana Academy of Science* **118**, 96–101 (2009).
100. Wright, V., Fleming, E. & Post, D. Survival of *Rhyzopertha dominica* (Coleoptera, Bostrichidae) on fruits and seeds collected from woodrat nests in Kansas. *J. Kansas Entomol. Soc.* **63**, 344–347 (1990).
101. Sullivan, J. (ed. Forest Service U.S. Department of Agriculture, Rocky Mountain Research Station) (U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. Fire Sciences Laboratory, 1994).
102. Weir, B. & Ott, J. Genetic data analysis II. *Trends Genet.* **13**, 379 (1997).
103. Magni, C., Ducousso, A., Caron, H., Petit, R. & Kremer, A. Chloroplast DNA variation of *Quercus rubra* L. in North America and comparison with other Fagaceae. *Mol. Ecol.* **14**, 513–524. https://doi.org/10.1111/j.1365-294X.2005.02400.x (2005).
104. Peterson, B. & Graves, W. 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, doi:https://doi.org/10.1111/jbi.12621 (2016).
105. Shaw, J. & Small, R. Chloroplast DNA phylogeny and phylogeography of the North American plums (*Prunus* subgenus *Prunus* section *Prunocerasus*, Rosaceae). *Am. J. Bot.* **92**, 2011–2030. https://doi.org/10.3732/ajb.92.12.2011 (2005).
106. Rowe, K., Heske, E., Brown, P. & Paige, K. Surviving the ice: Northern refugia and postglacial colonization. *Proc. Natl. Acad. Sci.* **101**, 10355–10359 (2004).
107. Graignic, N., Tremblay, F. & Bergeron, Y. Influence of northern limit range on genetic diversity and structure in a widespread North American tree, sugar maple (*Acer saccharum* Marshall). *Ecol. Evol.* **8**, 2766–2780. https://doi.org/10.1002/ece3.3906 (2018).
108. Bemmels, J., Knowles, L. & Dick, C. Genomic evidence of survival near ice sheet margins for some, but not all, North American trees. *Proc. Natl. Acad. Sci.* **116**, 8431–8436. https://doi.org/10.7302/Z2JS9NNG (2019).
109. Jia, H. & Steven, R. Fossil leaves and fruits of *Cercis* L. (Leguminosae) from the Eocene of western North America. *International Journal of Plant Sciences* **175**, 601–612, doi:https://doi.org/10.1086/675693 (2014).
110. Kraemer, M. & Favi, F. Emergence phenology of *Osmia lignaria* subsp *lignaria* (Hymenoptera: Megachilidae), its parasitoid *Chrysura kyrae* (Hymenoptera: Chrysididae), and bloom of *Cercis canadensis*. *Environ. Entomol.* **39**, 351–358. https://doi.org/10.1603/en09242 (2010).
111. USDA. Census of horticultural specialties. Volume 3 AC-12-SS-3, Washington, DC (2014).

Acknowledgements

Author contributions

D.H., W.K., and R.T. conceived and designed the experiments including the major conceptual ideas and proof outline. All authors assisted in sample collection and preparation. M.O. carried out the experiments and S.B. contributed to the processes. M.O., M.N., S.B., and D.H. troubleshoot technical details of the experiments. M.O., M.N., J.Z., and D.H. contributed to data analyses. SE provided technical advice on data preparation and analyses. All authors contributed to the interpretation of the results, manuscript writing, and editing. All authors provided critical feedback to shape the experiments, analyses, and finally produce the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-01020-z.

Correspondence and requests for materials should be addressed to D.H.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.