Tree Genetics & Genomes (2017) 13: 105
DOI 10.1007/s11295-017-1189-x

ORIGINAL ARTICLE

Population isolation results in unexpectedly high differentiation in Carolina hemlock (*Tsuga caroliniana*), an imperiled southern Appalachian endemic conifer

Kevin M. Potter[1] · **Angelia Rose Campbell**[2] · **Sedley A. Josserand**[3] · **C. Dana Nelson**[3,4] · **Robert M. Jetton**[5]

Received: 16 December 2016 / Revised: 14 August 2017 / Accepted: 10 September 2017 / Published online: 18 September 2017

Abstract Carolina hemlock (*Tsuga caroliniana* Engelm.) is a rare conifer species that exists in small, isolated populations within a limited area of the Southern Appalachian Mountains of the USA. As such, it represents an opportunity to assess whether population size and isolation can affect the genetic diversity and differentiation of a species capable of long-distance gene flow via wind-dispersed pollen and seed. This information is particularly important in a gene conservation context, given that Carolina hemlock is experiencing mortality throughout its range as a result of infestation by hemlock wooly adelgid (*Adelges tsugae* Annand), an exotic insect. In this study, 439 Carolina hemlock trees from 29 areas (analyzed as populations) were sampled, representing an extensive range-wide sampling of the species. Data from 12 polymorphic nuclear microsatellite loci were collected and analyzed for these samples. The results show that populations of Carolina hemlock are extremely inbred (F_{IS} = 0.713) and surprisingly highly differentiated from each other (F_{ST} = 0.473) with little gene flow (N_m = 0.740). Additionally, most populations contained at least one unique allele. This level of differentiation is unprecedented for a North American conifer species. Numerous genetic clusters were inferred using two different clustering approaches. The results clearly demonstrate that, existing as a limited number of small and isolated populations, Carolina hemlock has insufficient gene flow to avoid widespread genetic drift and inbreeding, despite having the capacity to disperse pollen and seed relatively long distances by wind. These results have important conservation implications for this imperiled species.

Keywords Disjunct populations · Gene conservation · Inbreeding · Microsatellite · Population isolation · Rare species

Communicated by S. C. González-Martínez

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11295-017-1189-x) contains supplementary material, which is available to authorized users.

✉ Kevin M. Potter
kpotter@ncsu.edu

[1] Department of Forestry and Environmental Resources, North Carolina State University, 3041 Cornwallis Road, Research Triangle Park, Raleigh, NC 27709, USA

[2] Department of Forestry and Environmental Resources, North Carolina State University, Campus Box 8008, Raleigh, NC 27695-8008, USA

[3] Southern Institute of Forest Genetics, USDA Forest Service, Southern Research Station, 23332 Success Road, Saucier, MS 39574, USA

[4] Forest Health Research and Education Center, University of Kentucky, Lexington, KY 40546, USA

[5] Camcore, Department of Forestry and Environmental Resources, North Carolina State University, Campus Box 8008, Raleigh, NC 27695-8008, USA

Introduction

Carolina hemlock (*Tsuga caroliniana* Engelm.) is a rare conifer with a limited geographic range within the Appalachian mountains of the USA. It occurs in an area of approximately 465 km by 165 km in North Carolina, South Carolina, Tennessee, Virginia, and Georgia. Carolina hemlock stands tend to be small (area and numbers of trees) and isolated (separated by 2 to 20 km), often occurring on exposed rocky outcrops at elevations between 600 and 1500 m (Jetton et al. 2008a; Humphrey 1989). This is true even in the core of the range (Fig. 1), where populations are small and discrete rather than continuously distributed across the landscape, as well as

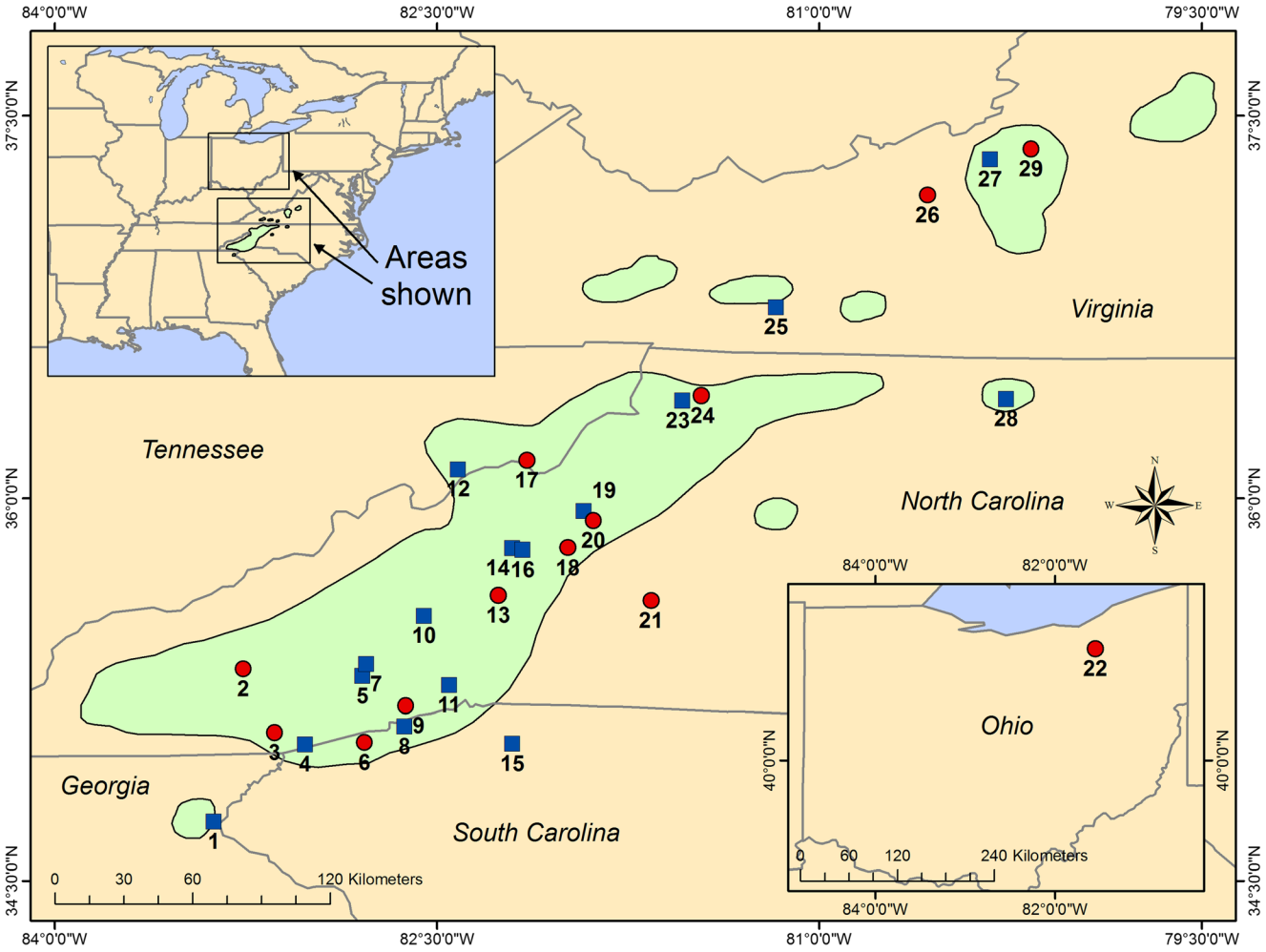

Fig. 1 Sampled populations of Carolina hemlock (*Tsuga caroliniana*). See Table 1 for population information. Blue populations (squares) are those for which collections of seed for gene conservation have been conducted. Seed had not been collected from red populations (circles) as of summer 2015

in a handful of more isolated populations that are peripheral to the center of the species distribution.

Because Carolina hemlock clearly exists as a set of populations that are both isolated and limited in size, we might expect that the species would bear the signature of small-population processes such as inbreeding and genetic drift (Young et al. 1996), including low within-population genetic variation and high between-population differentiation (Willi et al. 2006). This pattern might be particularly evident in populations that are peripheral to the core of the species distribution, where neutral marker studies have, on average, found a decline in diversity and an increase in differentiation in plants (Eckert et al. 2008). In fact, genetic diversity analyses of tree species have found this to be the case across several continents (e.g., Kitamura et al. 2015; Mattioni et al. 2017; Pandey and Rajora 2012). On the other hand, Carolina hemlock is a conifer with both pollen and seeds that are dispersed by wind (Godman and Lancaster 1990; Coladonato 1993). Species with effective long-distance dispersal of pollen and seeds are able to achieve high gene flow and, as a result, often have high genetic variability within species and populations (Hamrick et al. 1992; Hamrick and Godt 1996), but low differentiation among populations. Wind, in particular, is expected to be an effective pollen dispersal mechanism (Govindaraju 1988). For example, another conifer endemic to the Southern Appalachian region, Fraser fir (*Abies fraseri* (Pursh.) Poir), also exists in a limited number of isolated populations but has been found to have relatively little differentiation among populations, possibly as a result of frequent long-distance pollen dispersal (Potter et al. 2008).

Patterns of genetic variation, and their causes, have important implications for the genetic conservation of Carolina hemlock, which is a slow-growing, stress-tolerant, and late-successional species that often occurs in pure stands (Rentch et al. 2000; Jetton et al. 2008a). Like its eastern North America congener, eastern hemlock (*T. canadensis* (L.) Carr.), it is under severe threat from the hemlock wooly adelgid (*Adelges tsugae* Annand) (HWA), an invasive pest introduced from Japan. Once infested with HWA, which feed at the base of hemlock needles regardless of a tree's age and produce two

clonally replicated generations per year, Carolina hemlocks often die in as little as 4 years, though some trees have survived infestation for more than 10 years (Havill et al. 2011). Conditions that place stress on the trees, such as drought and extreme temperatures, can make them more vulnerable to HWA infestation and death (Quimby 1996), and it is possible that future climate changes may interact with and exacerbate the impacts of HWA (Lemieux et al. 2011). Individual Carolina hemlock trees may be less susceptible to HWA than the more widespread eastern hemlock (Jetton et al. 2008b), but the species as a whole may be at greater risk due to its rarity and limited range (Havill et al. 2011). In fact, Carolina hemlock is ranked as Near Threatened on the IUCN Red List (Farjon 2013). The loss of both hemlock species is expected to have numerous ecosystem impacts, including potentially drastic changes to forest community structure, understory composition, nutrient cycling, soil pH, hydrologic processes, microclimate, and wildlife assemblages (Orwig and Foster 1998; Jenkins et al. 1999; Brooks 2001; Kizlinski et al. 2002; Ford and Vose 2007; Allen et al. 2009; Yorks et al. 2000; Quimby 1996).

As a result of these concerns, ex situ gene conservation of hemlock genetic resources has been the focus of a collaborative effort between the US Department of Agriculture (USDA) Forest Service and Camcore, an international tree breeding and conservation program in the Department of Forestry and Environmental Resources at North Carolina State University. This project involves the collection of seeds from trees within populations across the ranges of both Carolina hemlock and eastern hemlock for long-term storage in seed banks and for the establishment of protected conservation seed orchards inside and outside the ranges of the species (Jetton et al. 2008a; Jetton et al. 2010). Genetic material from seed banks and conservation plantings will be made available for eventual restoration of degraded or extirpated natural populations. These ex situ collections should incorporate as much broad adaptability and genetic diversity as possible, given that genetic diversity provides a basis for adaptation and resilience to other sources of environmental stress and change, which is particularly important given the growing number, variety, and frequency of stress exposures to tree species (Schaberg et al. 2008). In the absence of information about adaptive genetic variation, a good understanding of the range-wide genetic structure of Carolina hemlock can be quantified using neutral molecular markers. Such an assessment should include an analyses of whether significant genetic diversity differences exist between populations that have or have not yet been the focus of a gene conservation seed collection effort.

Additionally, Carolina hemlock presents an opportunity to assess whether population size and isolation can affect the genetic diversity and differentiation of a species capable of long-distance gene flow via wind-dispersed pollen and seed. Since Carolina hemlock populations are generally small and isolated from each other, we hypothesize that little gene flow occurs among them and, as a result, they are highly differentiated, with levels of differentiation associated with distance among populations. We further hypothesize that populations disjunct from the core of the species range possess particularly low levels of genetic variation and high levels of differentiation compared to more centrally located populations.

To test these hypotheses, we used highly polymorphic microsatellite markers developed for Carolina hemlock (Josserand et al. 2008) to assess genetic variation and genetic structure across the range of the species. Specifically, we used 12 nuclear microsatellite markers to genotype 439 trees from 29 populations of Carolina hemlock to (1) assess the relationship between population isolation and genetic diversity and differentiation, (2) evaluate genetic variation in peripheral disjunct versus core range populations; and (3) compare genetic variation between populations that have and have not yet been the focus of gene conservation seed collections.

Materials and methods

Sample collection and DNA extraction

Twenty-nine populations of Carolina hemlock were sampled throughout the geographic range of the species (Fig. 1 and Table 1). This included a population located at Ritchie Ledges in Cuyahoga Valley National Park in northeast Ohio, but it is unclear whether this population is planted or natural. The samples were collected during the summer and autumn of 2013. The 439 trees included in this study represent the most extensive and thorough sampling of Carolina hemlock of which we are aware. From each tree sampled, four 6-in branch tips (one from each cardinal direction on the tree) were cut and bagged. Twenty trees per population were sampled where possible, but some populations were too small to do so. In these cases, we sampled as many trees as were available. Four populations encompassed fewer than five trees and represented all the accessible trees in those populations. These were excluded from most population-level analyses. When possible, a distance of at least 50 m was maintained between sampled trees to reduce the likelihood of sampling neighbors, which are more likely to be closely related due to short-distance seed dispersal (Brown and Hardner 2000). This distance is consistent with established gene conservation strategies employed by Camcore in Carolina hemlock seed collections (Jetton et al. 2007). The number of sampled trees infested by HWA was recorded in each population; HWA was present in 27 of the 29 study populations. Some low-infestation populations had been treated chemically for HWA by insecticides including imidacloprid.

Populations were classified as disjuncts if they are located outside the edge of continuous main range of the Carolina

Table 1 Identification number, location, sample size, coordinates, elevation in meters, and hemlock wooly adelgid (HWA) prevalence for the populations included in the microsatellite analysis of Carolina hemlock

ID	Population	County, state	N	Lat.	Lon.	Elev. (m)	HWA
1	Tallulah Gorge[a]	Rabun County, GA	2	34.73	− 83.38	457	33%
2	Kelsey Tract	Macon County, NC	2	35.33	− 83.26	1305	100%
3	Whiteside Mountain	Macon County, NC	4	35.08	− 83.14	1407	100%
4	Whitewater Falls	Transylvania County, NC	2	35.04	− 83.02	804	50%
5	Looking Glass	Transylvania County, NC	20	35.30	− 82.79	971	85%
6	Roundtop Mountain	Pickens County, SC	20	35.04	− 82.78	929	35%
7	Cradle of Forestry	Transylvania County, NC	13	35.35	− 82.78	994	30%
8	Caesar's Head	Greenville County, SC	18	35.11	− 82.63	909	25%
9	Dupont	Henderson County, NC	19	35.19	− 82.62	779	85%
10	Biltmore Estate	Buncombe County, NC	19	35.54	− 82.55	680	15%
11	Carl Sandburg Home	Henderson County, NC	16	35.27	− 82.45	717	35%
12	Cliff Ridge	Unicoi County, TN	19	36.11	− 82.42	673	35%
13	Kitsuma Peak	McDowell County, NC	16	35.62	− 82.26	942	90%
14	Carolina Hemlocks CG	Yancey County, NC	20	35.81	− 82.20	841	10%
15	Table Rock State Park[a]	Pickens County, SC	11	35.04	− 82.20	896	9%
16	Crabtree	Avery County, NC	15	35.80	− 82.16	1146	40%
17	Iron Mountain	Carter County, TN	20	36.15	− 82.15	880	50%
18	Dobson Knob	Burke/McDowell County, NC	18	35.81	− 81.99	965	100%
19	Linville Falls	Avery County, NC	20	35.95	− 81.92	1013	25%
20	Hawskbill Mountain	Burke County, NC	14	35.91	− 81.89	1073	100%
21	South Mountains State Park[a]	Burke County, NC	20	35.60	− 81.66	840	95%
22	Cuyahoga Valley[a]	Cuyahoga County, OH	17	41.24	− 81.55	310	0%
23	Bluff Mountain[a]	Ashe County, NC	20	36.38	− 81.54	1381	30%
24	Mount Jefferson[a]	Ashe County, NC	8	36.40	− 81.46	1314	0%
25	Cripple Creek[a]	Wythe County, VA	20	36.75	− 81.17	921	100%
26	Kentland Farm[a]	Montgomery County, VA	19	37.19	− 80.58	592	70%
27	Sinking Creek[a]	Montgomery County, VA	20	37.33	− 80.33	997	100%
28	Hanging Rock[a]	Stokes County, NC	20	36.39	− 80.27	504	90%
29	Dragon's Tooth[a]	Montgomery County, VA	7	37.37	− 80.17	681	100%

HWA prevalence of hemlock wooly adelgid

[a] Disjunct population

hemlock distribution as defined by Little (1971) (Fig. 1); the two exceptions were the Bluff Mountain (no. 23) and Mount Jefferson (no. 24) populations, which are within the main range as defined by Little but are approximately 50 km from the nearest recorded population within Little's main range, besides each other. Most of the populations in the study were naturally occurring, but a few populations may have been planted. Specifically, historical records and conversations with local land managers indicate that populations at Biltmore Estate and Cuyahoga Valley are likely to be entirely planted while several trees at the Carl Sandburg Home National Historical Site were planted from seeds originating from naturally occurring trees on or near the property.

The foliage samples were kept cold and shipped to the USDA Forest Service Southern Institute of Forest Genetics in Saucier, Mississippi, where they were frozen and later thawed for DNA extraction. The DNeasy 96 Plant Mini Kit (Qiagen, Chatsworth, California, USA) was used to extract DNA from the foliage samples. The extracted DNA was then quantified using a NanoDrop 8000 spectrophotometer (Thermo Scientific Inc.).

Microsatellite genotyping and analysis

Microsatellite markers are codominant, highly polymorphic, and highly variable markers (Echt et al. 1999; Kalia et al. 2011). Range-wide analyses using these markers can be used to test hypotheses regarding recent population processes such as genetic bottlenecks and gene flow, to provide data for understanding phylogeographic patterns, and to identify areas harboring high levels of genetic variation (e.g., Dvorak et al. 2009; Potter et al. 2015). A set of 35 dinucleotide and

trinucleotide microsatellite primer pairs isolated from Carolina hemlock (Josserand et al. 2008) were screened using a diverse subset of trees (22 samples from 22 populations) to develop a panel of markers for this study. Fifteen markers, each informative, consistently amplifying, and easy-to-score, were selected for the study (Table 2).

The DNA extracted from the foliage samples was diluted and amplified using the protocols described in Josserand et al. (2008). Dilutions were prepared from the samples at a concentration of 5 ng of DNA per microliter, in low TE buffer. The DNA was dried onto 96-well PCR plates. Klear Taq KBiosiences, Invitrogen 10× PCR buffer, $MgCl_2$, and dNTP was added to the plates, as well as the appropriate dyes and primers. These plates were placed on thermocyclers to amplify the DNA through PCR reactions. Two different control samples and a blank well were included with every 96-well plate.

Once the amplification was complete, ABI plates were prepared with a 1:10 dilution of the PCR product, Hi-dye formamide, and 600 LIZ (Applied Biosystems) size standard. These plates were run on an ABI Biosystems 3130 Genetic Analyzer (Applied Biosystems). The marker data were scored using GeneMapper 4.0 software (Applied Biosystems).

Genetic variation analyses

Inbreeding and null allele assessment Allele calls from the 15 microsatellite loci were exported from GeneMapper to conduct analyses of genetic variation across loci and at the population level. Preliminary analyses suggested the existence of high inbreeding coefficients in these loci, which could lead to overestimates of null allele frequencies; at the same time, it was possible that the inbreeding coefficient estimates may have been biased upwards by the presence of non-trivial null allele frequencies (Campagne et al. 2012). We therefore used the population inbreeding model (PIM)-based approach of Chybicki and Burczyk (2009), an expectation-maximization (EM) algorithm in the program INest 2.0, to estimate null allele frequencies for each locus while taking into account the potential presence of inbreeding within a sample population. We found that three loci (TcSI_075, TcSI_085, and TcSI_087) had excessively high (> 0.2) null allele estimated frequencies (Table 2), so these were not included in any species-level or population-level analyses.

We used INest 2.0 to calculate inbreeding coefficients (F_{IS}) for the species and for each locus while taking estimated null allele frequencies into account (Chybicki and Burczyk 2009). Specifically, we determined the relative importance of inbreeding and null alleles by implementing a Bayesian Individual Inbreeding Model (IIM), which estimates a Deviance Information Criterion (DIC) using a Gibbs sampler. INest 2.0 compares models that include estimated null alleles, inbreeding, genotyping error, and combinations of these three. A significant inbreeding effect is supported when the model with the lowest DIC contains all three. Once the model with the lowest DIC is selected, the program calculates the mean inbreeding coefficient across loci.

Hardy Weinberg equilibrium and population differentiation assessment Fisher's exact tests for Hardy Weinberg equilibrium were performed using GENEPOP 4.2 (Raymond and Rousset 1995) for each locus and population, based on 100 batches and 1000 iterations. The MULTTEST procedure in SAS 9.4 (SAS Institute Inc. 2013) was then used to calculate *q*-values (*p* values adjusted for the false discovery rate associated with multiple comparisons). FSTAT 2.3.9.2 (Goudet 2002) was used to test for linkage disequilibrium between pairs of loci, based on 1560 permutations and adjusted for multiple comparisons.

The "exclusion null alleles" (ENA) method in FreeNA (Chapuis and Estoup 2007) was run with 50,000 replicates to calculate among population differentiation (F_{ST}) across loci, overall, and between all pairs of populations, while accounting for estimated null alleles, and to generate a matrix of population-pairwise Cavalli-Sforza chord distances (D_C) (Cavalli-Sforza and Edwards 1967) that also accounted for estimated null alleles. Inter-population gene flow (N_m) using the private allele method, and corrected for sample size, was estimated using GENEPOP 4.2.

We used GenAlEx 6.41 to conduct an analysis of molecular variance (AMOVA) (Excoffier et al. 1992; Huff et al. 1993) to determine the partitioning of diversity among and within populations, with the significance of the variance components determined with 999 permutations. This method generates Φ_{PT}, which estimates the proportion of the total variance that is partitioned among populations (Huff 1997; Excoffier et al. 1992) and is analogous to F_{ST} (Peakall and Smouse 2006). We also calculated per-locus estimates of Jost's D (Jost 2008), D_{est}, an additional measure of genetic differentiation across all populations of Carolina hemlock, using SMOGD, version 1.2.5 (Crawford 2010). Jost's D is a statistic designed to avoid mathematical inconsistencies of other common statistics used to evaluate genetic differentiation (Jost 2008), and is particularly useful for highly polymorphic markers such as nuclear microsatellites. We calculated the arithmetic means of D_{est} across the Carolina hemlock loci and calculated the 95% confidence interval for this mean, using the confidence intervals of each locus from 1000 bootstrap replicates.

Population-level genetic diversity estimation GenAlEx 6.41 (Peakall and Smouse 2012) generated a list of unique (private) alleles (A_U) present in each population and calculated the mean alleles per locus (A). To account for unequal population sample sizes, we used the package HP-Rare (Kalinowski 2005) to produce unbiased population-level estimates of the mean number of private alleles (A_{UR}) and alleles per locus (A_R) using rarefaction (standardized to 20 genes per

Table 2 Description of the 15 *Tsuga caroliniana* (from Josserand et al. 2008) nuclear microsatellite loci used in the study, with size range, measures of genetic variation, inbreeding, deviation from Hardy-Weinberg equilibrium, and estimated null allele frequency for each

Locus	size range	*A*	H_O	H_E	A_E	D_{est}	F_{ST}	F_{IS}	HWE	Null
TcSI_012	270–290	9	0.044	0.069	1.07	0.004	0.088	0.354	[a]	0.036
TcSI_030	227–229	2	0.189	0.276	1.38	0.168	0.378	0.315	[a]	0.001
TcSI_044	170–184	6	0.105	0.482	1.93	0.371	0.587	0.782	[a]	0.155
TcSI_057	160–164	5	0.055	0.473	1.90	0.427	0.700	0.884	[a]	0.079
TcSI_060	121–130	6	0.064	0.451	1.82	0.28	0.545	0.858	[a]	0.121
TcSI_062	284–327	10	0.152	0.677	3.10	0.546	0.470	0.776	[a]	0.040
TcSI_066	236–262	6	0.105	0.486	1.95	0.402	0.521	0.785	[a]	0.035
TcSI_074	193–207	8	0.114	0.568	2.31	0.442	0.496	0.799	[a]	0.067
TcSI_080	242–260	9	0.183	0.761	4.18	0.658	0.459	0.760	[a]	0.043
TcSI_083	276–304	15	0.147	0.709	3.44	0.604	0.460	0.793	[a]	0.038
TcSI_089	138–164	8	0.316	0.551	2.23	0.319	0.319	0.427	[a]	0.000
TcSI_677	238–262	11	0.212	0.727	3.66	0.497	0.358	0.708	[a]	0.063
Total		*95*					*0.473*	*0.713*	[a]	
Mean		*7.92*	*0.141*	*0.519*	*2.41*	*0.393*	*0.448*	*0.687*		*0.056*
Loci not included because of high estimated null frequencies										
TcSI_075	198–216	7	0.109	0.643	2.80	0.542	0.581	0.831	[a]	0.276
TcSI_085	304–310	4	0.039	0.234	1.31	0.180	0.507	0.832	[a]	0.317
TcSI_087	301–313	9	0.152	0.526	2.11	0.435	0.414	0.711	[a]	0.232

Null: estimated proportion of null alleles (Brookfield 1996). Only the first 12 loci were included in analyses because of the high frequency of estimated null alleles in the other three

A alleles per locus, H_O observed heterozygosity, H_E expected heterozygosity, A_E effective alleles, D_{est} estimate of Jost's D, F_{ST} among-population variation, F_{IS} inbreeding coefficient, *HWE* Hardy-Weinberg exact test of heterozygote deficiency

[a] $q < 0.05$ using false discovery rate adjustment

population). For each population, GenAlEx generated allele frequencies; observed (H_O) and expected (H_E) heterozygosities; fixation index (($H_E - H_O$)) / H_E) (F), with substantial positive values indicating inbreeding and substantial negative values indicating excess of heterozygosity; and percent polymorphic loci (P_P).

Population-level comparisons Using the UNIVARIATE procedure in SAS 9.4 (SAS Institute Inc. 2013), we calculated within-group population means for several genetic diversity metrics: A_R, A_{UR}, P_P, H_O, H_E, F, mean pairwise D_C with all other populations, and mean pairwise F_{ST} with all other populations, (1) for populations disjunct from or existing within the main range of the species (10 and 15, respectively) and (2) for populations having or not having a completed seed collection for gene conservation at the time of the study (14 and 11, respectively). To test the null hypothesis that there was no significant difference between the means of each pair of groups (core vs. disjunct and seed collection vs. no seed collection), we conducted an exact two-sample Wilcoxon rank-sum test using the NPAR1WAY procedure in SAS, with 10,000 Monte Carlo runs generating p values. We then employed the MULTTEST procedure to calculate q-values. Additionally, we used the CORR procedure in SAS to test for correlations between the diversity statistics and three geographic characteristics of the populations: latitude, longitude, and elevation. We calculated Spearman correlations because several of the variables did not meet the assumptions of normality. The means comparison and correlation analyses encompassed 25 populations, as those with fewer than seven samples were excluded.

We used the isolation by distance (IBD) (Bohonak 2002) Web services version 3.23 (Jensen et al. 2005), with 10,000 randomizations, to perform Mantel tests assessing whether there was a significant relationship between pairwise geographic distances and both population pairwise F_{ST} estimates and chord genetic distances, with and without these values being log-transformed.

Genetic bottleneck assessment To test whether any Carolina hemlock populations had experienced population bottlenecks in the recent past, we used Bottleneck 1.2.02 (Piry et al. 1999) to compute the difference, averaged over loci, between actual heterozygosity and the heterozygosity that would be expected if the population were in mutation-drift equilibrium. An excess of heterozygosity is expected to be consistent with a recent population bottleneck, while a deficiency of heterozygotes suggests recent population expansion without

immigration (Cornuet and Luikart 1996; Karhu et al. 2006). We used a two-phase model (TPM) of microsatellite mutation, which is an intermediate between the single mutation model (Kimura and Ohta 1978) and the infinite allele model (Kimura and Crow 1964). In keeping with the presumed model for microsatellites (Piry et al. 1999), the parameter settings consisted of 95% single-step mutations and 5% multiple-step changes, with 12% variance in multistep mutations. Significance of heterozygosity excess or deficiency was evaluated with a one-sided Wilcoxon sign-rank test using 5000 simulation iterations. Since only one population exhibited heterozygosity excess, we reported p values from tests of heterozygosity deficiency (H_{def}).

Genetic structure analyses

We used the neighbor-joining (NJ) method (Saitou and Nei 1987) to generate a phylogram depicting likely evolutionary relationships among the sampled populations of Carolina hemlock. The NJ algorithm is a robust method for constructing trees from genetic distances (Mihaescu et al. 2009). The phylogram was computed from allelic frequencies using chord genetic distance (D_C) (Cavalli-Sforza and Edwards 1967), which does not require assumptions about the model under which microsatellites mutate and is considered superior to most others in phylogenetic tree topology construction over short spans of evolutionary time (Takezaki and Nei 1996; Libiger et al. 2009). We used the SEQBOOT, GENDIST, NEIGHBOR, and CONSENSE components of PHYLIP 3.695 (Felsenstein 2005) to generate the NJ phylogram. Bootstrapping was employed with 1000 replicates to generate confidence estimates associated with the topology of the phylogram.

Bayesian assignment tests are useful for detecting population genetic structure. STRUCTURE 2.3.4 is a model-based Bayesian clustering method that analyzes marker-based genotypes for each individual tree. It is used to determine the number and composition of genetic clusters that best describe the population structure (Pritchard et al. 2000). We ran STRUCTURE 2.3.4 using the admixture model, with a burn in length of 50,000 replicates and 500,000 Markov chain Monte Carlo (MCMC) replicates after burn-in. We did this twice, once without sample locations used as prior information, and once with. Including sample locations in the analysis aids in the detection of clusters, especially when fewer data are available (Hubisz et al. 2009). With each of these parameter sets, we ran STRUCTURE 20 times for each possible maximum number of genetic clusters (K) from 1 to 20. We then used STRUCTURE HARVESTER (Earl and Vonholdt 2012) to calculate the ΔK statistic of Evanno et al. (2005). For the analysis incorporating sample locations, the ΔK statistic revealed a dominant peak at K = 2, with smaller peaks at K = 3, 12, 17, and 19, suggesting the possibility of substructure occurring within each of two strongly differentiated clusters (Evanno et al. 2005). The analysis without sample locations revealed a single dominant peak at K = 2 (Supplementary Fig. 1). The results from the 20 iterations of K = 2, 3, and 12 using sample locations, and of K = 2 without sample locations, were imported into CLUMPP 1.1.2 (Jakobsson and Rosenberg 2007) to generate averaged Q matrices of individual and population posterior cluster probabilities using the greedy algorithm and the G′ pairwise matrix similarity statistic. We then imported the population matrix into ArcMap 10.1 (ESRI 2012) to generate maps showing the geographic distribution of the clusters (K = 2, 3, 12) throughout the range of Carolina hemlock.

STRUCTURE and other similar Bayesian clustering methods assume Hardy-Weinberg equilibrium within populations (Putman and Carbone 2014). Because exact tests for Hardy-Weinberg equilibrium indicated a significant deficit of heterozygotes for all loci included in this study (see below), we also employed discriminant analysis of principal components (DAPC) to estimate the number of genetic clusters in the microsatellite data. DAPC is a versatile clustering method that does not rely on a particular population genetics model and is therefore free of assumptions about Hardy-Weinberg equilibrium and linkage disequilibrium (Jombart et al. 2010). Using the *adegenet* package (Jombart 2008), version 2.0.1, in R version 3.3.0 (R Core Team 2016), we applied K-means clustering of K from 1 to 80 after transforming the raw allelic data with a principal component analysis (PCA). We then inspected a resulting graph of Bayesian Information Criterion (BIC) values for each K to determine which number of clusters best described the data, maximizing variation between clusters and minimizing variation within them. The BIC values declined sharply from K = 1 to K = 13, with a slight "elbow" at about K = 14 and a less sharp decline until flattening at about K = 30 and then increasing after K = 40. Selecting K = 14, we next performed discriminant analysis to determine membership probabilities within each of the clusters, and then calculated the proportion of overall genetic cluster presence probability for each population, based on the probability of cluster membership for individuals in the population. We used ArcMap 10.1 (ESRI 2012) to map the geographic distribution of the clusters throughout the species range.

Results

Species-level microsatellite results

The 12 microsatellite loci included in the analysis averaged 7.92 alleles per locus across the 439 samples of Carolina hemlock, ranging from a minimum of two alleles (*TcSI_030*) to a maximum of 15 (*TcSI_083*) (Table 2). The species exhibited moderate expected heterozygosity (mean of 0.519 across loci)

and low observed heterozygosity (mean 0.141), and exact tests for Hardy-Weinberg equilibrium indicated a significant deficit of heterozygotes for all loci. The significant and highly positive inbreeding coefficient (F_{IS}) of 0.713 (95% confidence interval 0.684–0.742) indicates a deficit of heterozygotes and a very high level of inbreeding. The average estimated proportion of null alleles across loci was 0.056 after the removal of three loci with unacceptably high estimated null frequencies (see the "Methods" section). The results of the INest 2.0 analysis indicate that inbreeding is a more important factor than null alleles, given that DIC values for the models including inbreeding, null alleles, and genotyping errors (16,162.69), and including only inbreeding and genotyping errors (16,194.37), were much lower than the model including null alleles and genotyping errors (17,102.74). No linkage disequilibrium was apparent between any pairs of loci after adjusting the p value for multiple comparisons.

The three assessments of among-population microsatellite differentiation consistently estimated high levels of genetic differentiation among populations. The F_{ST} analysis, for example, estimated an extremely high amount of genetic differentiation among rather than within populations. F_{ST} values were obtained both adjusted and unadjusted for estimated null alleles, but no significant difference existed between the two (F_{ST} across loci, adjusted for nulls = 0.473, 95% confidence interval 0.361–0.535; F_{ST} across loci, not adjusted for nulls = 0.493, 95% confidence interval 0.370–0.558). The F_{ST} values for individual loci ranged from 0.088 (TcSI_012) to 0.70 (TcSI_057), with most between 0.300 and 0.600 (Table 2). Similarly, the AMOVA results revealed that half of the microsatellite variance (Φ_{PT} = 0.500) was partitioned among populations, with the rest occurring within populations (Supplementary Table 1). Inter-population gene flow (Nm) was estimated at 0.740, or less than one migrant per generation. The mean D_{est} across loci was 0.393 (95% confidence interval: 0.377–0.409), ranging across loci from 0.004 (TcSI_012) to 0.658 (TcSI_080) (Table 2).

Population-level genetic variation and differentiation

The 29 Carolina hemlock populations averaged 2.15 microsatellite alleles per locus (A) and 2.09 alleles per locus when standardized for sample size (A_R) (Table 3). Nearly two-thirds of the populations (18) contained at least one unique allele (A_U), including one (no. 18) that had four, and three others (nos. 3, 11, and 12) that had three each. In general, populations with the highest A_R and A_U are located in the core of the species range in North Carolina (Fig. 2a, b). The same general pattern was the case with percent polymorphic loci (P_P) (Table 3). Polymorphism was relatively low, with a mean across populations of 63.8%.

The mean observed heterozygosity (H_O) across Carolina hemlock populations (0.130) was less than the mean expected heterozygosity (H_E) (0.233) (Table 3). Nearly all populations were significantly out of Hardy-Weinberg equilibrium, with the exception of one population in the northern portion of the species' range (no. 24). In addition, this and another northern disjunct population (no. 29) were the only adequately sampled populations that were not inbred (Fig. 2c). Widespread inbreeding was apparent given a positive mean F of 0.346 across the populations, and by the fact that nearly all populations had positive F values. The populations that were the most genetically distinct, based on mean pairwise chord genetic distance (D_C) between each population and the 28 others (Table 3 and Fig. 2d), were in the extreme northern (nos. 22, 27, 29) and southern (nos. 2, 7, 9, and 15) parts of the species distribution. Mean pairwise D_C averaged across populations was 0.457, and the mean population pairwise F_{ST} was 0.470. Based on pairwise D_C, the Cuyahoga Valley population in northern Ohio (no. 22), which may have been planted, was most closely related to three populations near the middle of the species range (nos. 17, 19, and 20).

Carolina hemlock did not exhibit the excess of heterozygosity expected following a recent genetic bottleneck but rather a significant heterozygosity deficiency (p = 0.0004), indicating a possible relatively recent population expansion without immigration (Cornuet and Luikart 1996; Karhu et al. 2006). One population (no. 21), exhibited the significant excess of heterozygosity expected following a recent genetic bottleneck, while seven populations had significant ($p < 0.05$) heterozygosity deficiency associated with population expansion; all but two of these (no. 22 and no. 23) are located in the core of the Carolina hemlock distribution (Table 3).

Neighbor-joining phylogram and cluster assignment

The consensus neighbor-joining phylogram of D_C genetic distance among populations (Fig. 3) did not reveal any statistically significant broad-scale evolutionary patterns within Carolina hemlock but did show high bootstrap support for the clustering of some neighboring populations, such as no. 5 with no. 7 in the southern part of the species distribution and no. 14 and no. 16 in the central part. A well-supported cluster of northern populations included the grouping of no. 24 and no. 26 with especially high bootstrap support, along with no. 28 and no. 23.

The STRUCTURE analysis inferred the possible existence of genetic substructuring within Carolina hemlock. The ΔK method (Evanno et al. 2005) most strongly supported the possibility of two genetic clusters in the species (both when population locations were included as prior information and when they were not), although it also reasonably inferred their further division into three 12, 17, and 19 clusters (with sample locations included) (Supplementary Fig. 1). With K = 2, one gene pool each occurred primarily in the northern and

Table 3 Measures of genetic variation for each of 29 populations of Carolina hemlock, based on 12 nuclear microsatellite loci

ID	Population	A	A_R	A_U	A_{UR}	P_P	H_O	H_E	F	HWE	Mean D_C	Mean F_{ST}	H_{def}
1	Tallulah Gorge	0.83	3.17	1	0.17	0.0	0.000	0.000	–	–	0.468	0.545	.
2	Kelsey Tract	1.00	1	0	0.00	0.0	0.000	0.000	–	–	0.509	0.657	.
3	Whiteside Mountain	2.42	2.42	3	0.33	83.3	0.194	0.401	0.478	[a]	0.454	0.535	0.423
4	Whitewater Falls	1.25	1.25	0	0.00	25.0	0.208	0.115	− 0.778	–	0.431	0.315	.
5	Looking Glass	2.08	1.73	2	0.25	66.7	0.084	0.125	0.171	[a]	0.466	0.593	0.006
6	Roundtop Mountain	1.92	1.82	1	0.42	66.7	0.109	0.215	0.433	[a]	0.466	0.496	0.527
7	Cradle of Forestry	1.33	1.31	0	0.00	33.3	0.007	0.044	0.739	[a]	0.487	0.424	0.031
8	Caesar's Head	2.58	2.54	2	0.50	75.0	0.141	0.385	0.562	[a]	0.418	0.349	0.898
9	Dupont	2.50	2.18	0	0.58	83.3	0.193	0.268	0.171	[a]	0.533	0.669	0.042
10	Biltmore Estate	3.08	2.72	1	0.50	91.7	0.209	0.367	0.436	[a]	0.401	0.316	0.062
11	Carl Sandburg Home	3.33	2.96	3	0.67	91.7	0.174	0.368	0.518	[a]	0.458	0.585	0.051
12	Cliff Ridge	2.50	2.69	3	0.50	75.0	0.263	0.328	0.200	[a]	0.443	0.368	0.455
13	Kitsuma Peak	2.33	2.24	1	0.17	83.3	0.122	0.344	0.696	[a]	0.448	0.558	0.615
14	Carolina Hemlocks CG	2.33	2.02	0	0.25	66.7	0.195	0.224	0.172	[a]	0.454	0.640	0.027
15	Table Rock State Park	1.75	1.73	0	0.17	41.7	0.114	0.159	0.330	[a]	0.564	0.637	0.313
16	Crabtree	2.42	2.25	2	0.33	91.7	0.147	0.291	0.635	[a]	0.407	0.430	0.160
17	Iron Mountain	2.42	2.19	0	0.42	91.7	0.047	0.250	0.754	[a]	0.433	0.419	0.008
18	Dobson Knob	3.67	3.38	4	0.67	100.0	0.143	0.476	0.685	[a]	0.440	0.287	0.311
19	Linville Falls	2.50	2.22	1	0.33	75.0	0.221	0.292	0.283	[a]	0.423	0.401	0.150
20	Hawksbill Mountain	3.00	2.97	0	0.67	91.7	0.143	0.428	0.701	[a]	0.463	0.418	0.449
21	South Mountains State Park	2.17	2.08	1	0.33	75.0	0.229	0.373	0.366	[a]	0.435	0.377	0.999
22	Cuyahoga Valley	2.42	2.28	1	0.33	83.3	0.104	0.284	0.635	[a]	0.480	0.481	0.042
23	Bluff Mountain	2.17	1.95	2	0.25	58.3	0.109	0.197	0.331	[a]	0.452	0.507	0.039
24	Mount Jefferson	1.33	1.33	0	0.08	33.3	0.125	0.100	− 0.175	ns	0.472	0.571	0.156
25	Cripple Creek	1.50	1.44	1	0.08	50.0	0.051	0.129	0.533	[a]	0.442	0.365	0.578
26	Kentland Farm	1.42	1.31	0	0.17	33.3	0.093	0.067	0.028	[a]	0.415	0.356	0.156
27	Sinking Creek	2.17	1.85	1	0.33	75.0	0.056	0.193	0.478	[a]	0.479	0.556	0.082
28	Hanging Rock	2.75	2.45	2	0.75	91.7	0.209	0.299	0.288	[a]	0.429	0.307	0.062
29	Dragon's Tooth	1.17	1.17	0	0.00	16.7	0.071	0.049	− 0.316	–	0.475	0.480	0.250
	Mean	*2.15*		*1.1*	*0.32*	*63.8*	*0.130*	*0.233*	*0.346*		*0.457*	*0.470*	*0.265*

A mean alleles per locus, A_R mean alleles per locus standardized by rarefaction, A_U unique (private) alleles, A_{UR} unique alleles standardized by rarefaction, P_P percent of polymorphic loci, H_O mean observed heterozygosity, H_E mean expected heterozygosity, F mean fixation index across loci, *HWE* Hardy-Weinberg exact test of heterozygote deficiency, *Mean* D_c mean pairwise chord distance (Cavalli-Sforza and Edwards 1967) with all other populations, *Mean* F_{ST} mean pairwise differentiation with all other populations, H_{def} p value for test of heterozygote deficiency

[a] $q < 0.05$ using false discovery rate adjustment

southern halves of the species range, with a zone of admixture in the middle (Fig. 4a and Supplementary Fig. 2). The exceptions were two populations at the southern end of the range with very small sample sizes (no. 1 with $n = 2$ and no. 3 with $n = 3$). With K = 3, one cluster (Gene Pool K3.1) predominated in the northern part of the range and the other two clusters predominated throughout the southern part, with Gene Pool K3.2 most common in the southwest and Gene Pool K3.3 more common in the southeast (Fig. 4b). With K = 12, the range of Carolina hemlock was a mosaic of gene pools (Fig. 4c). Populations that shared high proportions of gene pools were consistent with those clustered in the neighbor-joining phylogram (Fig. 3), including Biltmore Estate (no. 10), Carolina Hemlocks Campground (no. 14), and Crabtree (no. 16) in Gene Pool K14.2; Bluff Mountain (no. 23), Mt. Jefferson (no. 24), Kentland Farm (no. 26), and Hanging Rock (no. 28) in Gene Pool K14.4; Caesar's Head (no. 8), Dupont (no. 9), Carl Sandburg Home (no. 11), and Table Rock State Park (no. 15) in Gene Pool K14.11; and Looking Glass (no. 5) and Cradle of Forestry (no. 7) in Gene Pool K14.5. The Cuyahoga Valley population (no. 22), consists mostly of northern gene pools in K = 2, 3, and 12, although it encompasses relatively high proportions of southern gene pools as well (Fig. 4 and Supplementary Fig. 2).

The *adegenet* clustering analysis, meanwhile, suggests that the samples could be reasonably grouped into between about

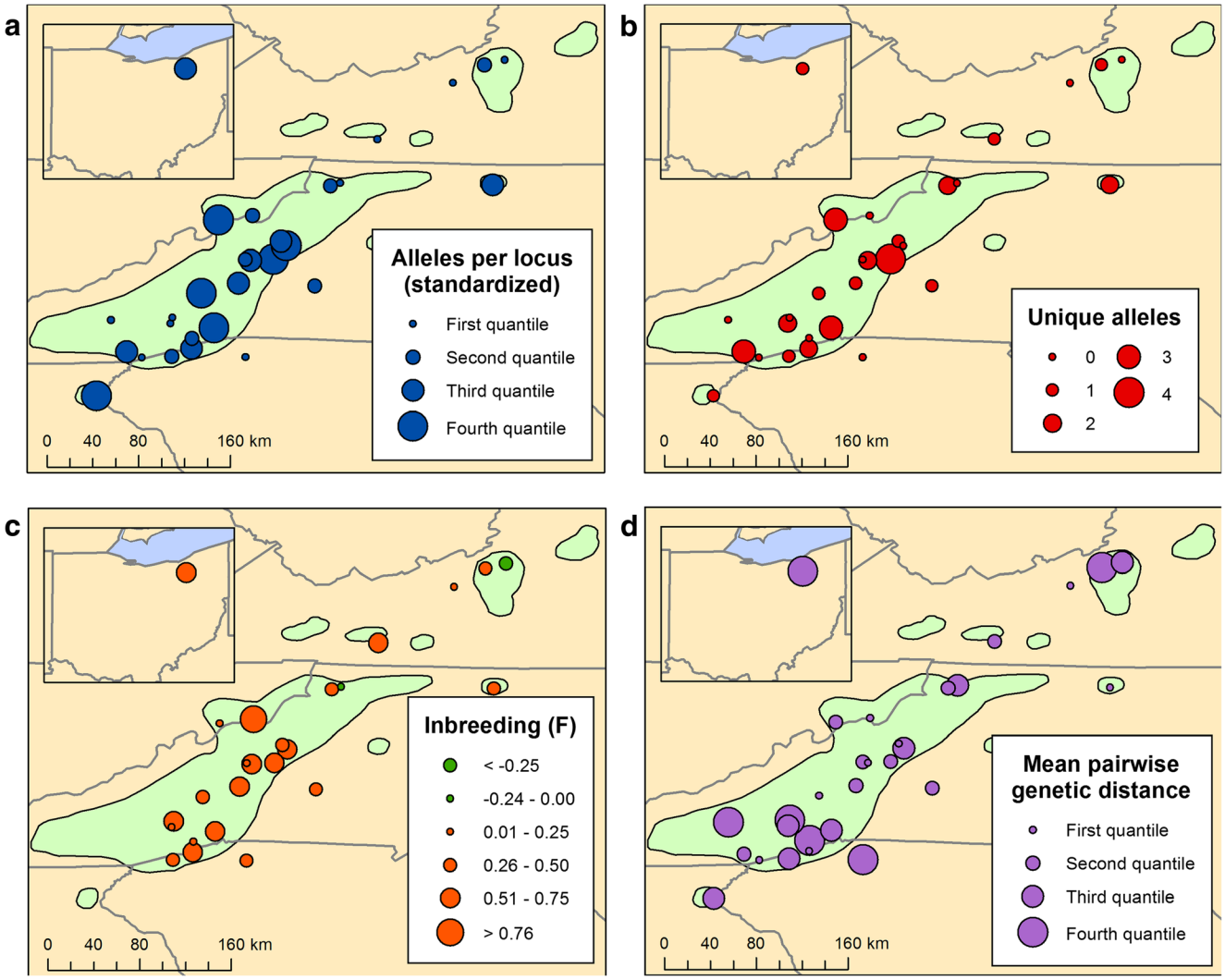

Fig. 2 Carolina hemlock population classifications of **a** alleles per locus (A), **b** unique alleles (A_U), **c** inbreeding coefficient (F_{IS}), and **d** mean pairwise chord distance (D_C), based on 16 polymorphic nuclear microsatellite loci

14 and 30 clusters. We selected the small end of this range. When mapped, the *adegenet* gene pools had similar geographic distributions to STRUCTURE; for example, northern populations no. 24, no. 26, and no. 28 and southern populations no. 2, no. 5, and no. 7 consisted of similar gene pools (Fig. 5). The Cuyahoga Valley population (no. 22) contained mainly of gene pools from the southern and central parts of the Carolina hemlock range.

Group comparisons and correlations with geographic variables

Several standard measures of genetic variation were significantly different between Carolina hemlock populations located in the core range of the species compared to those disjunct from it, even when applying the more conservative q-value to account for multiple comparisons (Table 4). Specifically, compared to disjunct populations, core-range populations had higher standardized allelic richness, standardized number of unique alleles per locus, polymorphism, and expected heterozygosity. Disjunct populations, meanwhile, were less inbred than core populations, as measured by the fixation index F, though this was not the case at $\alpha \leq 0.05$ for q. No significant differences existed between populations which had or had not yet been the focus of a gene conservation seed collection effort (Table 4).

We found no statistically significant correlations between any population-level genetic diversity measure and latitude, longitude, or elevation (results not shown). At the same time, the Mantel test in IBD (Bohonak 2002) demonstrated a significant positive correlation between pairwise population geographic distance and genetic distance: $r = 0.517$ when log-transformed and 0.389 when not ($p < 0.0001$ for both). These correlations increased to 0.565 and 0.573 when the Cuyahoga Valley population was removed. Meanwhile, we also detected a significant correlation between pairwise

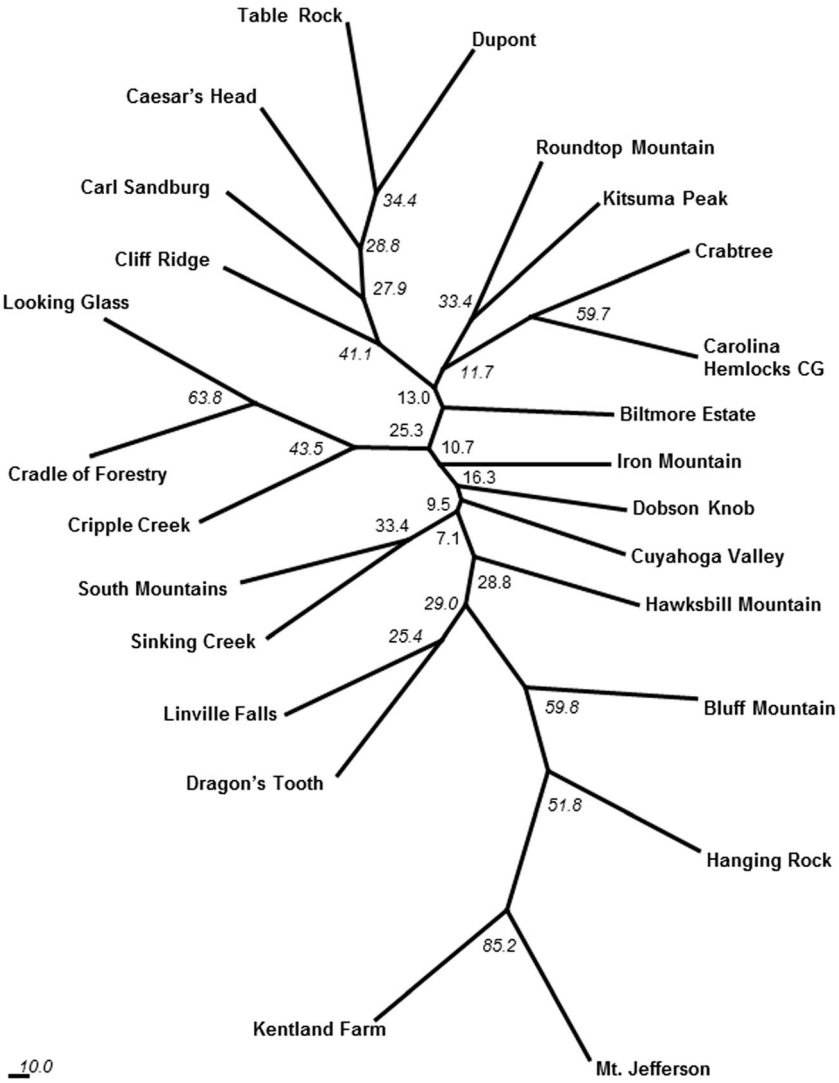

Fig. 3 Consensus neighbor-joining phylogram depicting D_C genetic distance (Cavalli-Sforza and Edwards 1967) among 25 populations of Carolina hemlock. The values represent the percent bootstrap support for the nodes over 1000 replicates

geographic distance and F_{ST} estimates, with r = 0.433 ($p < 0.0001$) when these values were log-transformed, and r = 0.322 (p = 0.0041) when they were not. Again, these increased to 0.474 and 0.536, respectively, when Cuyahoga Valley was removed.

Discussion

Our range-wide microsatellite assessment of Carolina hemlock demonstrates that this Southern Appalachian endemic has low genetic diversity, is highly inbred, and consists of populations that are highly differentiated from each other. Populations outside the core of the species range are characterized by particularly low variation and high differentiation. These findings have important implications for the genetic conservation of Carolina hemlock.

On one hand, the low levels of genetic diversity and high levels of inbreeding in Carolina hemlock are not surprising, given that a species consisting of a limited number of relatively small and isolated populations is expected to be susceptible to genetic drift and inbreeding (Young et al. 1996). At the same time, the level of population differentiation in the species is extremely high, with almost half of the genetic differentiation occurring among rather than within populations (F_{ST} = 0.473, Φ_{PT} = 0.500, D_{est} = 0.393). While species with naturally disjunct ranges are likely to have higher genetic diversity partitioned among populations than those with continuous ranges (Hamrick 2004), we would expect that Carolina hemlock might be able to overcome the geographical isolation of its populations through long-distance dispersal of its seeds and, especially, its pollen. Instead, this level of differentiation seems unprecedented among North American conifers with wind-dispersed seed and pollen, even compared to those that exist in a relatively small number of isolated populations. This includes bristlecone fir (*Abies bracteata* (D. Don) D. Don ex Poit.), with 8% of differentiation among six populations using 30 isozyme loci (Ledig et al. 2006); Brewer

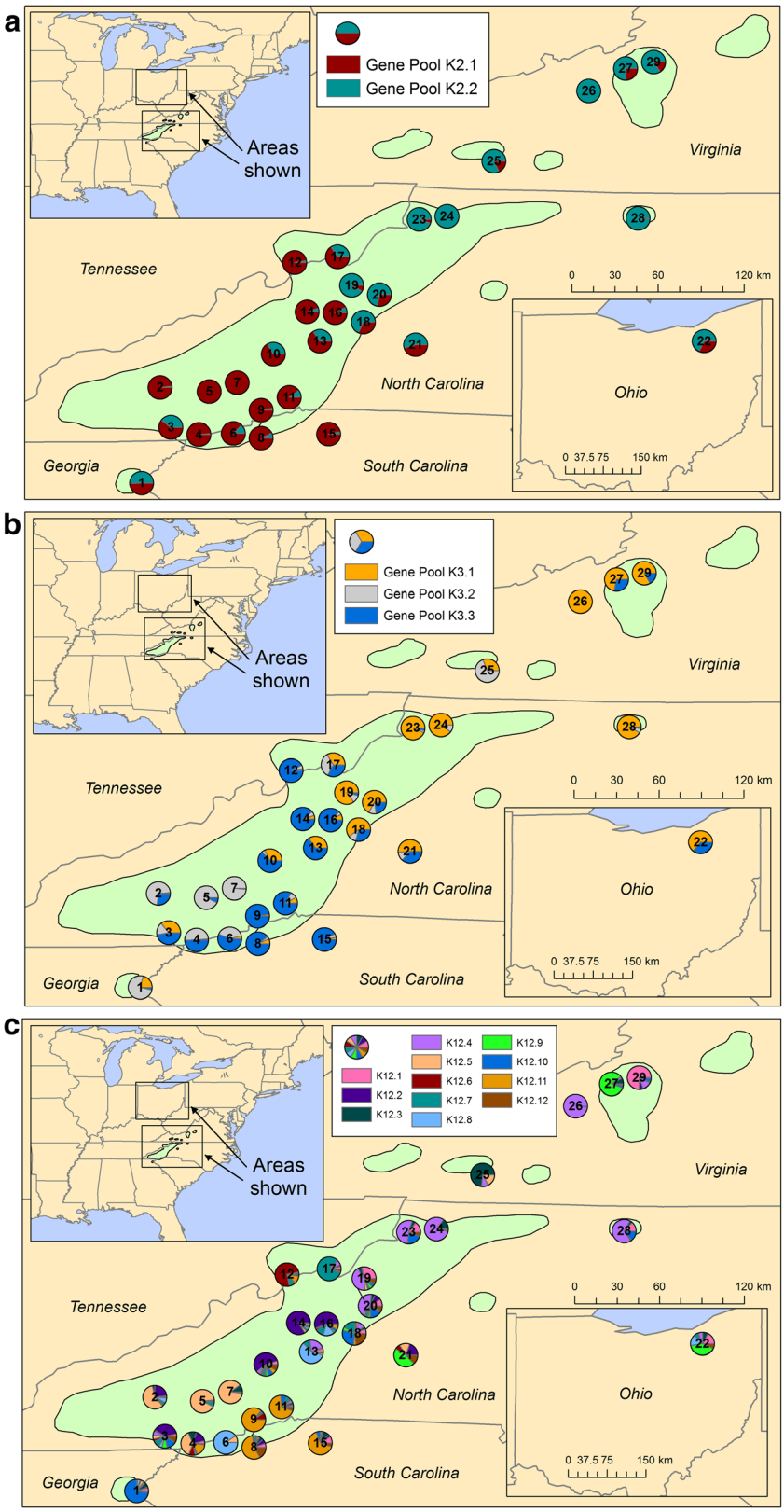

Fig. 4 The proportion, within each Carolina hemlock population, of inferred ancestry from the **a** K = 2, **b** K = 3, and **c** K = 12 genetic clusters inferred using Structure 2.3.3 (Pritchard et al. 2000), with the analysis incorporating sample locations. See Table 1 for population information

spruce (*Picea breweriana*, *S. Watson*) with 15.7% (10 populations, 26 isozymes) (Ledig et al. 2005); Rocky Mountain bristlecone pine (*Pinus aristata* Englem.) with 13.1% (16 populations, 21 isozymes) (Schoettle et al. 2012); and Monterey pine (*P. radiata* D. Don) with 14% (5 populations, 19 microsatellite loci) (Karhu et al. 2006). Two conifer species

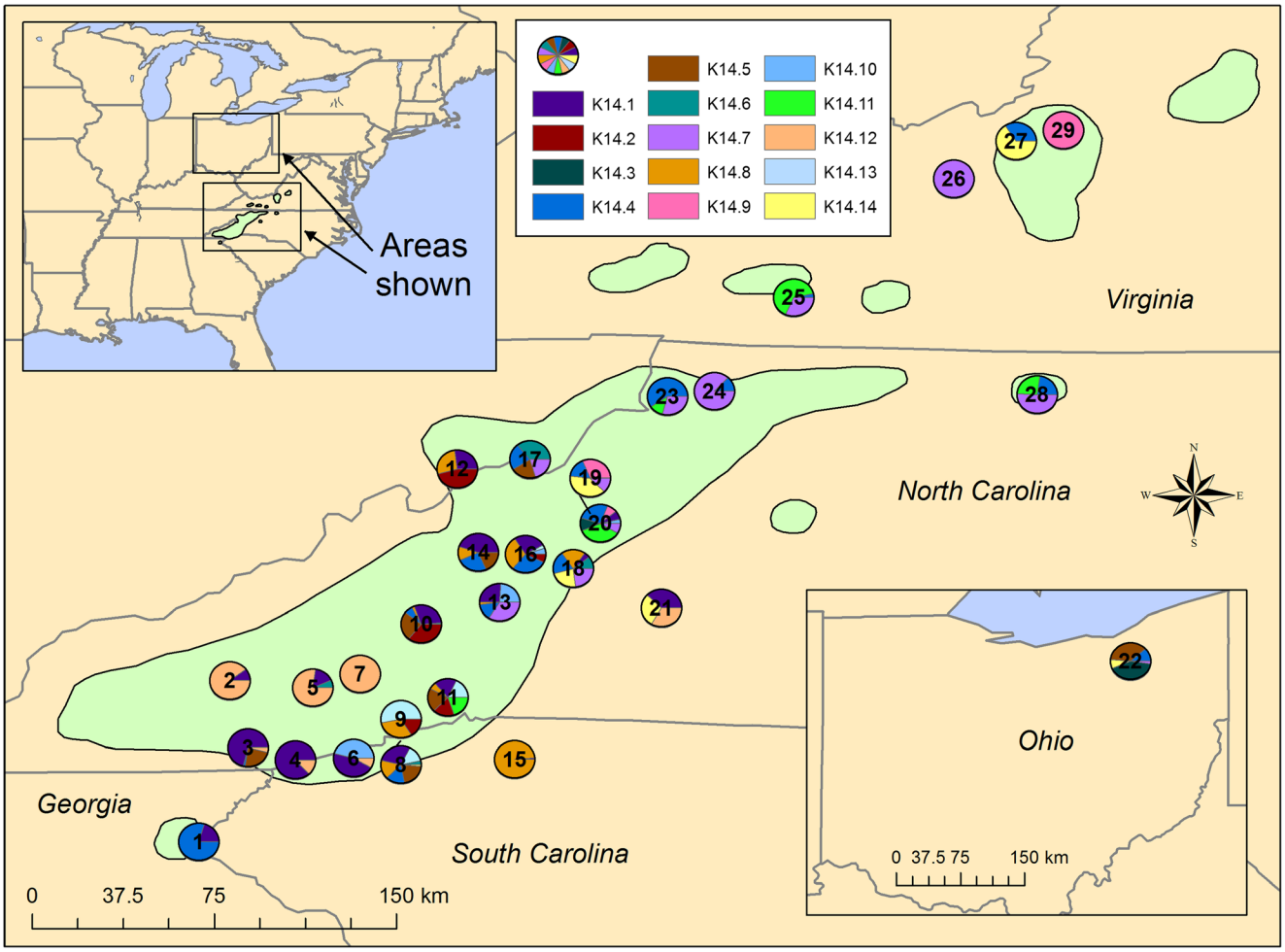

Fig. 5 The proportion, within each Carolina hemlock population, of inferred ancestry from the K = 14 genetic clusters inferred using the R package *adegenet* (Jombart 2008). See Table 1 for population information

endemic to the Appalachian Mountains also exhibited less among-population differentiation than Carolina hemlock: Fraser fir with almost none of its genetic variation among populations according to an analysis of eight microsatellites (Potter et al. 2008), and Table Mountain pine (*P. pungens* Lamb.), with 13.5% in an allozyme study of 20 populations (Gibson and Hamrick 1991). Bishop pine (*P. muricata* D. Don) in California was the range-restricted conifer with the next highest proportion of among-population differentiation compared to Carolina hemlock (based on five populations and 32 isozymes), but this was still only 32% (Wu et al. 1999). The genetic diversity of Carolina hemlock was markedly lower than that of eastern hemlock (Potter et al. 2012; Lemieux et al. 2011) and mountain hemlock (*T. mertensiana* [Bong.] Carr.) (Johnson et al. 2017), and also demonstrated much higher levels of among-population differentiation and number of population-level unique alleles.

Our results strongly suggest that Carolina hemlock is below the threshold level of gene flow necessary to avoid genetic drift and inbreeding in isolated populations (Ellstrand 1992). Our results further estimate fewer than a single interpopulation migrant per generation in this species (N_m = 0.740), many times fewer than that of its fellow Southern Appalachian endemic Fraser fir (N_m = 9.77) (Potter et al. 2008). The specific reasons for this dearth of interpopulation gene exchange beg further investigation. It is worth noting, however, that the largest Carolina hemlock populations are moderate in size (ca. 200 ha) while the smallest encompass only a handful of trees (Jetton et al. 2008a). Additionally, Carolina hemlock seeds, while winged, are the largest in its genus in North America (Barbour et al. 2008), and the dispersal distance of the smaller eastern hemlock seeds is typically limited to tree height because of their small wings (Godman and Lancaster 1990). It seems highly unlikely that Carolina hemlock seeds often cover the long distances between populations. Wind-borne conifer pollen, meanwhile, may be able to travel those distances (Ziegenhagen et al. 2004; Liepelt et al. 2002), and the significant positive correlation between pairwise population geographic distance and genetic distance in this study suggests that this occurs to some

Table 4 Comparison between means of genetic variation statistics for populations disjunct from or existing within the main range of the species and having or not having a completed seed collection for gene conservation at the time of the study

	Mean	SD	Min	Max	Mean	SD	Min	Max	Differences of means	
									p	*q*
(a) Disjunct status	Disjunct (n = 10)				Core (n = 15)					
Allelic richness, standardized (A_R)	1.76	0.44	1.17	2.45	2.35	0.54	1.31	3.38	**0.009**	**0.041**
Unique alleles, standardized (A_{UR})	0.25	0.21	0	0.75	0.42	0.20	0	0.67	**0.021**	**0.041**
Percent loci polymorphic (P_P)	55.83	24.86	16.67	91.67	78.89	16.63	33.33	100.00	**0.011**	**0.041**
Observed heterozygosity (H_O)	0.116	0.06	0.051	0.229	0.147	0.067	0.007	0.263	0.091	0.122
Expected heterozygosity (H_E)	0.185	0.106	0.049	0.373	0.294	0.113	0.044	0.476	**0.018**	**0.041**
Inbreeding (F)	0.25	0.309	− 0.316	0.635	0.477	0.227	0.171	0.754	**0.045**	0.073
Mean pairwise D_C	0.464	0.042	0.415	0.564	0.449	0.033	0.401	0.533	0.173	0.198
Mean pairwise F_{ST}	0.464	0.108	0.307	0.637	0.464	0.12	0.287	0.669	0.456	0.456
(b) Ex situ conservation status	Seed collection (n = 14)				No seed collection (n = 11)					
Allelic richness, standardized (A_R)	2.13	0.50	1.31	2.96	2.09	1.17	1.25	3.38	0.382	0.500
Unique alleles, standardized (A_{UR})	0.35	0.21	0	0.75	0.35	0.23	0	0.67	0.500	0.500
Percent loci polymorphic (P_P)	70.24	18.98	33.33	91.67	68.94	28.16	16.67	100.00	0.391	0.500
Observed heterozygosity (H_O)	0.141	0.075	0.007	0.263	0.125	0.052	0.047	0.229	0.247	0.500
Expected heterozygosity (H_E)	0.243	0.106	0.044	0.385	0.259	0.143	0.049	0.476	0.392	0.500
Inbreeding (F)	0.405	0.179	0.171	0.739	0.362	0.382	− 0.316	0.754	0.435	0.500
Mean pairwise D_C	0.452	0.041	0.401	0.564	0.460	0.032	0.415	0.533	0.214	0.500
Mean pairwise F_{ST}	0.463	0.120	0.307	0.640	0.465	0.109	0.287	0.669	0.478	0.500

Values of p and q significant at $\alpha \leq 0.05$ are in bold

SD standard deviation, *min* minimum, *max* maximum, *p p* value, *q* false discovery rate adjusted *p* value

degree in Carolina hemlock. At the same time, hemlock pollen is particularly susceptible to desiccation (Nienstaedt and Kriebel 1955), and the small size and widely dispersed nature of the Carolina hemlock populations therefore may make long-distance inter-population gene flow a rare event, as pollen-mediated gene exchange in forest trees exhibits a high probability of dispersal at local scales that decreases rapidly with distance (Ellstrand 1992).

Genetic composition of marginal populations

Neutral molecular marker studies have found that, on average, within-population genetic diversity declines and among-population genetic differentiation increases from the center of plant species distributions to their peripheries (Eckert et al. 2008). Recent studies indicate that this pattern occurs in tree species across several continents. For example, genetic diversity was highest in central populations of European sweet chestnut (*Castanea sativa* Mill.), while private allelic richness was highest in the eastern part of the range and in some isolated western populations (Mattioni et al. 2017). Pedunculate oak (*Quercus robur* L.), meanwhile, had high differentiation among populations at its northern range margin in northern Europe (Pohjanmies et al. 2016). In the western USA, isolated disjunct populations of ponderosa pine (*P. ponderosa* Douglas ex Lawson) had lower genetic variation by some measures than central-range populations and greater genetic differentiation (Potter et al. 2015), a pattern that was repeated in eastern white cedar (*Thuja occidentalis* L.) in eastern Canada (Pandey and Rajora 2012) and Guatemalan fir (*Abies guatemalensis* Rehder) (Rasmussen et al. 2010). In Asia, northern disjunct populations of Japanese beech (*Fagus crenata* Blume) similarly have low expected heterozygosity and allelic richness and high genetic differentiation, possibly the result of ongoing northward expansion (Kitamura et al. 2015). The distribution of Carolina hemlock includes several peripheral disjunct populations along its southern, eastern, and northern edges. We sampled 10 of these for this study, although one (Tallulah Gorge, no. 1, a population encompassing a total of three accessible trees) was not included in our statistical comparisons because of its small size. Peripheral populations of Carolina hemlock may be of particular conservation concern because, in addition to being more likely to differ genetically, geographic outlier populations may be more vulnerable to environmental change as a result of lower levels of genetic variation (Yanchuk and Lester 1996).

Peripheral Carolina hemlock populations encompass significantly less genetic variation than core-range populations by nearly all measures (Table 4). This included the number of population-level unique alleles, unlike eastern hemlock

(Potter et al. 2012). Eastern hemlock, however, has a large area of generally spatially continuous occurrence at the core of its range, stretching from the Southern Appalachians northeast to the Maritime Provinces of Canada and then west into the Great Lakes region. Carolina hemlock, meanwhile, consists of relatively small and distinct populations even at the core of its range, apparently with very little interpopulation gene flow. Also, unlike eastern hemlock (Potter et al. 2012), the Carolina hemlock core populations have a significantly greater degree of inbreeding than marginal populations.

Regional patterns of genetic diversity and composition

The current distribution of Carolina hemlock, as well as the patterns of evolutionary relationships and genetic diversity within the species, was likely influenced by periodic episodes of glaciation in the late Quaternary period. During these episodes, the species would have endured in refuges, likely at lower latitudes and elevations (Hewitt 1996, 2000; Bennett and Provan 2008), from which the species would have subsequently spread during periods of warming (Lascoux et al. 2004; McLachlan et al. 2005). The results of species range-wide genetic assessments using polymorphic, codominant nuclear markers such as microsatellites can help to infer tree species Pleistocene refugia and hypothesized routes of post-Pleistocene colonization (O'Connell et al. 2008; Heuertz et al. 2004; Boys et al. 2005). Specifically, we would expect that populations closer to the refuge location will exhibit greater genetic diversity, while populations farther from the refuge will exhibit less (Comes and Kadereit 1998; Taberlet et al. 1998). For Carolina hemlock, we hypothesized that Pleistocene refugia were located toward the southern end of the species range, and thus that this area should harbor greater diversity currently.

While we detected no correlations between genetic diversity and latitude, longitude, or elevation, other evidence may shed light on the phylogeographic history of Carolina hemlock. For example, populations with the highest values for several measures of genetic variation, including alleles per locus and unique alleles, tend to cluster along the southeastern edge of the core range in North Carolina and South Carolina (Fig. 2a, b). Additionally, the geographic distribution of the STRUCTURE- and *adegenet*-inferred gene pools may be revealing. First, the K = 2 and K = 3 inferred genetic clusters may suggest that there were two or three refuges in which the species persisted during the Pleistocene, one each in the north and southeast and possibly one in the southwest. The further substructuring into 12 or more gene pools may have resulted from subsequent isolation of populations. Interestingly, although each of the K = 3 gene pools tend to dominate in a different region (especially Gene Pool 3.2 in the southwest and Gene Pool 3.1 in the northeast), most populations encompass a combination of the three gene pools, despite the fact that the populations apparently have very low levels of recent inter-population gene flow. This may suggest that at some point during the evolutionary history of Carolina hemlock, populations were less isolated and inter-population gene flow was greater. Another explanation for this pattern is the possible admixture of distinct gene pools in secondary contact zones (Durand et al. 2009), potentially following migration from separate Pleistocene refugia. Of particular interest in this regard are populations toward the center of the range that contain large proportions of all three gene pools, including Dobson Knob (no. 18), Hawksbill Mountain (no. 20), and South Mountains State Park (no. 21). In the K = 12 analyses, populations toward the center of the species range also tend to encompass a larger number of gene pools than do the populations at the southern and northern extremities (Fig. 4b and 5).

The results of this study also elucidate the genetic origins of likely planted populations of Carolina hemlock. These encompass the Biltmore Estate population (no. 10), which is almost certainly planted; the Carl Sandburg Home population (no. 11), which has some planted and some naturally regenerated trees; and the Cuyahoga Valley population (no. 22), which is suspected to be planted but may be natural. The gene pool composition of the Biltmore Estate population is most similar to the nearby Kitsuma Peak (no. 13) population in the K = 3 analysis (Fig. 4b), to nearby northern populations with the STRUCTURE K = 12 analysis (Fig. 4c), and to populations immediate north and south of it with the *adegenet* K = 14 analysis (Fig. 5). This suggests that the planted Biltmore Estate trees came from a relatively nearby source. Similarly, the Carl Sandburg Home population has gene pool composition similar to that of some of its nearest geographic neighbors including Caesar's Head (no. 8), Dupont (no. 9), and Table Rock (no. 15) (Figs. 4 and 5), with which it is also clustered in the neighbor-joining phylogram. Therefore, planted Carolina hemlock in this population likely also came from local sources. The Cuyahoga Valley population, meanwhile, is far outside the known natural range of Carolina hemlock. If it were natural, we would expect it to be highly differentiated from the other populations and to have a genetic composition highly similar to the more northerly populations. Its composition, however, encompasses a large proportion of southern gene pools. At the same time, the mean pairwise chord distance between Cuyahoga Valley and all other populations is only slightly above average (Table 3), additionally indicating that this is likely not a naturally occurring population. Finally, the Bottleneck analysis demonstrated that the Carolina hemlock at Cuyahoga Valley recently experienced a population expansion, potentially an indication of natural regeneration within a stand of trees planted there.

Gene conservation implications

The results of this first comprehensive range-wide Carolina hemlock genetic diversity study can help guide conservation

decision-making for this imperiled tree species. Given that genetic diversity is essential because it provides a basis for adaptation and resilience to environmental stress and change (Schaberg et al. 2008), an important goal will be to maintain Carolina hemlock genetic material with broad adaptability and high levels of genetic diversity, both in natural stands (in situ), where possible, and off-site (ex situ, e.g., in seed and pollen banks). A number of such gene conservation efforts are already underway. Currently, the USDA Forest Service and Camcore are collaborating in a program aimed at conserving the genetic resources of both Carolina hemlock and eastern hemlock (Jetton et al. 2013). This effort has led to the establishment of seed reserves and protected seedling orchards. So far, seeds have been collected from 168 trees in 24 populations throughout the range of Carolina hemlock, with the collected seeds placed in cold storage at two germplasm repositories in the USA, the USDA Agriculture Research Service's National Center for Genetic Resources Preservation in Fort Collins, Colorado, and the Camcore Seed Bank in Raleigh, North Carolina. Additionally, two conservation plantings of Carolina hemlock have been established in North Carolina, where they are managed to prevent HWA infestation, and in Brazil and Chile, outside the range of the adelgid (Jetton et al. 2013). These resources can be utilized in efforts to breed HWA-resistant genotypes (Vose et al. 2013) through interspecific hybridization alone (Montgomery et al. 2009; Bentz et al. 2002) or through hybridization followed by backcross breeding similar to efforts currently underway to breed blight-resistant American chestnuts (Diskin et al. 2006). It is potentially fortuitous that Carolina hemlock is closely related to Asian hemlocks, because it is therefore more likely to hybridize successfully with species having resistance to HWA (Havill et al. 2008). Montgomery et al. (2009) showed that hybrids between Carolina hemlock and Chinese hemlock (*T. chinensis* (Franch.) Pritzel ex Diels.) were more resistant to HWA than was Carolina hemlock.

At the same time, silvicultural methods, chemical controls, and biological control are being investigated or applied to maintain hemlock species in forested landscapes (e.g., Vose et al. 2013). Biological control methods, such as releasing predatory non-native beetles to keep the HWA in check, are not yet feasible for widespread control (Salom et al. 2008). Chemical controls, such as treating trees with the systemic neonicitinoid insecticide imidacloprid, are able to effectively eliminate an infestation of HWA on individual trees or in small stands. However, it is not feasible to continually maintain chemical controls over the entire range of the species (McClure 1991; Cheah et al. 2004). Silvicultural strategies include removing "reservoir" trees that are likely to spread HWA. These methods are more effective for slowing the spread of the HWA than halting or reversing it (McClure 1990). Once effective measures to control the hemlock wooly adelgid have been developed and deployed, the ex situ conservation resources described above will provide the material for reintroduction or restoration of populations which have suffered mortality and/or loss of genetic variation.

The information about population genetic structure and diversity revealed in this study will greatly help in developing better targeted, more effective conservation efforts (Vose et al. 2013), including the prioritization of populations for seed collection, storage, and adelgid-resistant hemlock breeding (Hastings et al. 2017). The fact that nearly all populations of Carolina hemlock currently show high levels of inbreeding emphasizes the necessity of quickly and effectively acting to preserve the genetic diversity of the species. The high levels of differentiation detected among Carolina hemlock populations, and the commonness of alleles unique to populations, underscore the importance of ensuring that ongoing gene conservation efforts sample as many populations as possible (Echt et al. 2011). Clearly, Carolina hemlock exists primarily as a limited set of small populations with restricted inter-population gene flow. The loss of genetic diversity in small and isolated populations of trees, such as those of Carolina hemlock, may be associated with genetic drift and inbreeding (Jaramillo-Correa et al. 2009) and could reduce overall population fitness (Reed and Frankham 2003) and their capacity to adapt to environmental change (Willi et al. 2006). In general, the consequences of climate change for rare species and those occurring in isolated habitats may be severe, because their populations are likely to be less numerous and may be less well-connected or occur over narrow geographical regions (Jump and Penuelas 2005). Of particular concern for Carolina hemlock is that a small geographic range is strongly correlated with extinction (Stork et al. 2009; Brook et al. 2008), and that range-restricted species are among the first of which entire species have gone extinct due to recent climate change (Parmesan 2006). Given these facts, Carolina hemlock gene conservation efforts should focus on areas containing high genetic variation, including allelic richness and heterozygosity, possessing unique alleles; and encompassing multiple and/or rare gene pools. Examples include populations along the eastern edge of the range core.

In addition, Carolina hemlock gene conservation efforts should incorporate some representation of peripheral populations, given that they sometimes encompass rare gene pools (Figs. 4b and 5). In general, peripheral forest tree populations are of high conservation interest, especially in the context of climate change, because they are where the most significant evolutionary changes may occur, they face the greatest extinction risk, they may represent the source of colonizing trees in leading edges, and they may provide genetic novelty to other parts of the existing species range (Fady et al. 2016). These populations may be among the most at risk because genetic variation and response to selection are thought to be positively correlated with population size (Hamrick et al. 1992; Willi et al. 2006; Hamrick and Godt 1996), and because individuals

in small populations have lower fitness as a result of environmental stress and inbreeding, which can substantially increase the probability of population extinction in changing environments (Willi et al. 2006). At the same time, differential adaptive pressures, genetic drift, and mutation may result in even greater differentiation among reproductively isolated populations, potentially leading to eventual speciation (Slatkin 1987).

Meanwhile, no significant genetic diversity differences exist between Carolina hemlock populations that have and have not been subject to ex situ seed collection (Table 4). Therefore, previous and ongoing gene conservation efforts have likely done a reasonably good job of sampling the genetic diversity of Carolina hemlock, as was suggested by a previous study using amplified fragment length polymorphisms (Potter et al. 2010).

Acknowledgements This work derives in part from Angelia Campbell's 2014 Master of Science thesis in the Department of Forestry and Environmental Resources at North Carolina State University ("Assessing Carolina Hemlock (*T. caroliniana* Engelm.) Population Genetic Structure and Diversity in the Southern Appalachians Using Microsatellite Markers"). We thank Craig Echt (Southern Institute of Forest Genetics) for his laboratory advice and assistance, Andy Whittier and Steve Galehouse for their invaluable help both planning and carrying out foliage collections, Julie Canavin for literature review assistance, Fred Gould and Jenny Xiang for their feedback and expertise, two anonymous reviewers for their insightful comments, and the following for allowing site access for field collections: Biltmore Estate, Blue Ridge Parkway, Carl Sandburg Home, Cherokee National Forest, Cuyahoga National Park, Georgia State Parks, George Washington/Jefferson National Forest, Mount Rogers National Recreation Area, Nantahala National Forest, North Carolina State Parks, Pisgah National Forest, South Carolina State Parks, The Nature Conservancy, and Virginia Tech. This work was supported in part through Cooperative Agreement 11-CA-11330126-139, Research Joint Venture Agreement 13-JV-11330110-043 and Cost Share Agreement 16-CS-11330110-038 between the US Department of Agriculture, Forest Service, Southern Research Station, and North Carolina State University, and Domestic Grant Agreement 12-DG-11083150-016 between the US Department of Agriculture, Forest Service, Forest Health Protection, and North Carolina State University. Mention of trade names does not imply endorsement or recommendation for use.

Compliance with ethical standards

Data archiving statement Microsatellite genotypes at all loci for all samples, as well as coordinates of all sampled trees, are archived at TreeGenes (https://dendrome.ucdavis.edu/tgdr/) as TGDR073.

References

Allen MC, Sheehan J, Master TL, Mulvihill R (2009) Responses to Acadian flycatchers (*Empidonax virescens*) to hemlock woolly adelgid (Adelges tsugae) infestation in Appalachian riparian forests. Auk 126(3):543–553. https://doi.org/10.1525/auk.2009.08073

Barbour JR, Ruth RH, Karrfalt RP (2008) *Tsuga* Carr. Hemlock. In: Bonner FT, Karrfalt RP (eds) The woody plant seed manual. Vol agriculture handbook 727. United States Department of Agriculture Forest Service, Washington, D.C., pp 1127–1139

Bennett KD, Provan J (2008) What do we mean by 'refugia? Quat Sci Rev 27(27–28):2449–2455. https://doi.org/10.1016/j.quascirev.2008.08.019

Bentz SE, Riedel LGH, Pooler MR, Townsend AM (2002) Hybridization and self-compatibility in controlled pollinations of eastern north American and Asian hemlock (*Tsuga*) species. J Arboric 28(4): 200–205

Bohonak AJ (2002) IBD (isolation by distance): a program for analyses of isolation by distance. J Hered 93(2):153–154. https://doi.org/10.1093/jhered/93.2.153

Boys J, Cherry M, Dayanandan S (2005) Microsatellite analysis reveals genetically distinct populations of red pine (*Pinus resinosa* Pinaceae). Am J Bot 92(5):833–841. https://doi.org/10.3732/ajb.92.5.833

Brook BW, Sodhi NS, Bradshaw CJA (2008) Synergies among extinction drivers under global change. Trends Ecol Evol 23(8):453–460. https://doi.org/10.1016/j.tree.2008.03.011

Brooks RT (2001) Effects of the removal of overstory hemlock from hemlock-dominated forests on eastern redback salamanders. For Ecol Manag 149(1–3):197–204. https://doi.org/10.1016/S0378-1127(00)00553-3

Brown AHD, Hardner CM (2000) Sampling the gene pools of forest trees for *ex situ* conservation. In: Young AG, Boshier D, Boyle TJ (eds) Forest conservation genetics: principles and practice. CABI Publishing, Collingwood, pp 185–196

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(6):975–982. https://doi.org/10.1111/1755-0998.12015

Cavalli-Sforza LL, Edwards AWF (1967) Phylogenetic analysis: models and estimation procedures. Evolution 21:550–570

Chapuis MP, Estoup A (2007) Microsatellite null alleles and estimation of population differentiation. Mol Biol Evol 24(3):621–631. https://doi.org/10.1093/molbev/msl191

Cheah C, Montgomery ME, Salom S, Parker BL, Costa S, Skinner M (2004) Biological control of hemlock woolly adelgid. FHTET-2004–04. United States Department of Agriculture, Forest Service, Forest Health Technology Enterprise Team, Morgantown, West Virginia

Chybicki IJ, Burczyk J (2009) Simultaneous estimation of null alleles and inbreeding coefficients. J Hered 100(1):106–113. https://doi.org/10.1093/jhered/esn088

Coladonato M (1993) *Tsuga caroliniana*. U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fire Sciences Laboratory http://www.fs.fed.us/database/feis/. Accessed 14 Sept 2016

Comes HP, Kadereit JW (1998) The effect of quaternary climatic changes on plant distribution and evolution. Trends Plant Sci 3(11):432–438. https://doi.org/10.1016/S1360-1385(98)01327-2

Cornuet JM, Luikart G (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. Genetics 144(4):2001–2014

Crawford NG (2010) SMOGD: software for the measurement of genetic diversity. Mol Ecol Resour 10(3):556–557. https://doi.org/10.1111/j.1755-0998.2009.02801.x

Diskin M, Steiner KC, Hebard FV (2006) Recovery of American chestnut characteristics following hybridization and backcross breeding to restore blight-ravaged Castanea dentata. For Ecol Manag 223(1–3):439–447. https://doi.org/10.1016/j.foreco.2005.12.022

Durand E, Jay F, Gaggiotti OE, Francois O (2009) Spatial inference of admixture proportions and secondary contact zones. Mol Biol Evol 26(9):1963–1973. https://doi.org/10.1093/molbev/msp106

Dvorak WS, Potter KM, Hipkins VD, Hodge GR (2009) Genetic diversity and gene exchange in *Pinus oocarpa*, a Mesoamerican pine with

resistance to the pitch canker fungus (*Fusarium circinatum*). Int J Plant Sci 170(5):609–626. https://doi.org/10.1086/597780

Earl DA, Vonholdt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour 4(2): 359–361. https://doi.org/10.1007/s12686-011-9548-7

Echt CS, Crane BS, Nelson CD (2011) Establishing restoration seed reserves in National Forest System seed orchards. In: Nelson CD, Rousseau RJ, Yuceer C (eds) Proceedings of the 31st Southern Forest Tree Improvement Conference. Biloxi, Mississippi, pp 46–49

Echt CS, Vendramin GG, Nelson CD, Marquardt P (1999) Microsatellite DNA as shared genetic markers among conifer species. Can J For Res 29(3):365–371. https://doi.org/10.1139/cjfr-29-3-365

Eckert CG, Samis KE, Lougheed SC (2008) Genetic variation across species' geographical ranges: the central-marginal hypothesis and beyond. Mol Ecol 17(5):1170–1188. https://doi.org/10.1111/j.1365-294X.2007.03659.x

Ellstrand NC (1992) Gene flow among seed plant populations. New For 6:241–256

ESRI (2012) ArcMap 10.1. Environmental Systems Research Institute Inc., Redlands, California

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14(8):2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131(2):479–491

Fady B, Aravanopoulos FA, Alizoti P, Mátyás C, Von Wühlisch G, Westergren M, Belletti P, Cvjetkovic B, Ducci F, Huber G (2016) Evolution-based approach needed for the conservation and silviculture of peripheral forest tree populations. For Ecol Manag 375:66–75. https://doi.org/10.1016/j.foreco.2016.05.015

Farjon A (2013) *Tsuga caroliniana*. IUCN red list of threatened species 2013: e.T34200A2850654. doi:https://doi.org/10.2305/IUCN.UK.2013-1.RLTS.T34200A2850654.en. Accessed 22 May 2017

Felsenstein J (2005) PHYLIP (phylogeny inference package), version 3.6. Department of Genome Sciences, University of Washington, Seattle

Ford CR, Vose JM (2007) *Tsuga canadensis* (L.) Carr. mortality will impact hydrologic processes in southern appalachian forest ecosystems. Ecol Appl 17(4):1156–1167. https://doi.org/10.1890/06-0027

Gibson JP, Hamrick JL (1991) Genetic diversity and structure in *Pinus pungens* (Table Mountain pine) populations. Can J For Res 21(5): 635–642

Godman RM, Lancaster K (1990) Eastern hemlock. In: Burns RM, Honkala BH (eds) Silvics of North America: 1. Conifers, vol 1. Agricultural handbook 654. United States Department of Agriculture Forest Service, Washington D.C

Goudet J (2002) FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3.2). version 2.9.3.2 edn. University of Lausanne, Lausanne

Govindaraju DR (1988) Relationship between dispersal ability and levels of gene flow in plants. Oikos 52(1):31–35

Hamrick JL (2004) Response of forest trees to global environmental changes. For Ecol Manag 197(1–3):323–335. https://doi.org/10.1016/j.foreco.2004.05.023

Hamrick JL, Godt MJW (1996) Effects of life history traits on genetic diversity in plant species. Philos Trans R Soc B 351(1345):1291–1298. https://doi.org/10.1098/rstb.1996.0112

Hamrick JL, Godt MJW, Sherman-Broyles SL (1992) Factors influencing levels of genetic diversity in woody plant species. New For 6:95–124

Hastings JM, Potter KM, Koch FH, Megalos M, Jetton RM (2017) Prioritizing conservation seed banking locations for imperiled hemlock species using multi-attribute frontier mapping. New For 48(2):301–316. https://doi.org/10.1007/s11056-017-9575-7

Havill NP, Campbell CS, Vining TF, LePage B, Bayer RJ, Donoghue MJ (2008) Phylogeny and biogeography of *Tsuga* (Pinaceae) inferred from nuclear ribosomal ITS and chloroplast DNA sequence data. Syst Bot 33(3):478–489. https://doi.org/10.1600/036364408785679770

Havill NP, Montgomery ME, Keena MA (2011) Hemlock woolly adelgid and its hemlock hosts: a global perspective. In: Onken B, Reardon R (eds) Implementation and status of biological control of the hemlock woolly Adelgid, vol FHTET-2011-04. U.S. Forest Service, Forest Health Technology Enterprise Team, Morgantown, pp 3–14

Heuertz M, Hausman JF, Hardy OJ, Vendramin GG, Frascaria-Lacoste N, Vekemans X (2004) Nuclear microsatellites reveal contrasting patterns of genetic structure between western and southeastern European populations of the common ash (*Fraxinus excelsior* L.) Evolution 58(5):976–988

Hewitt GM (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. Biol J Linn Soc 58(3):247–276. https://doi.org/10.1111/j.1095-8312.1996.tb01434.x

Hewitt GM (2000) The genetic legacy of the quaternary ice ages. Nature 405(6789):907–913

Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. Mol Ecol Resour 9(5):1322–1332. https://doi.org/10.1111/j.1755-0998.2009.02591.x

Huff DR (1997) RAPD characterization of heterogeneous perennial ryegrass cultivars. Crop Sci 37(2):557–564

Huff DR, Peakall R, Smouse PE (1993) RAPD variation within and among natural populations of outcrossing buffalograss (*Buchloe dactyloides* (Nutt.) Engelm.) Theor Appl Genet 86(8):927–934. https://doi.org/10.1007/bf00211043

Humphrey LD (1989) Life history traits of *Tsuga caroliniana* Engelm. (Carolina hemlock) and its role in community dynamics. Castanea 54(3):172–190

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23(14):1801–1806. https://doi.org/10.1093/bioinformatics/btm233

Jaramillo-Correa JP, Beaulieu J, Khasa DP, Bousquet J (2009) Inferring the past from the present phylogeographic structure of north American forest trees: seeing the forest for the genes. Can J For Res 39(2):286–307. https://doi.org/10.1139/x08-181

Jenkins JC, Aber JD, Canham CD (1999) Hemlock woolly adelgid impacts on community structure and N cycling rates in eastern hemlock forests. Can J For Res 29(5):630–645. https://doi.org/10.1139/cjfr-29-5-630

Jensen JL, Bohonak AJ, Kelley ST (2005) Isolation by distance, web service. BMC Genet 6:6. https://doi.org/10.1186/1471-2156-6-13

Jetton RM, Dvorak WS, Whittier WA (2008a) Ecological and genetic factors that define the natural distribution of Carolina hemlock in the southeastern United States and their role in *ex situ* conservation. For Ecol Manag 255(8–9):3212–3221. https://doi.org/10.1016/j.foreco.2008.01.032

Jetton RM, Hain FP, Dvorak WS, Frampton J (2008b) Infestation rate of hemlock woolly adelgid (Hemiptera: Adelgidae) among three north American hemlock (*Tsuga*) species following artificial inoculation. J Entomol Sci 43(4):438–442

Jetton RM, Whittier WA, Dvorak WS (2007) Guidelines for the collection of eastern (*Tsuga canadensis* [L.] Carr.) and Carolina (*Tsuga caroliniana* Engelm.) hemlock seed cones for *Ex Situ* gene conservation. Camcore, North Carolina State University, Raleigh

Jetton RM, Whittier WA, Dvorak WS, Rhea JR (2010) Status of gene conservation efforts for eastern and Carolina hemlock in the eastern United States. In: Onken B, Reardon R (eds) Fifth symposium on hemlock woolly Adelgid in the eastern United States, Asheville,

North Carolina, 2010. USDA Forest Service, Forest Health Technology Enterprise Team, Morgantown, pp 93–99

Jetton RM, Whittier WA, Dvorak WS, Rhea JR (2013) Conserved *ex situ* genetic resources of eastern and Carolina hemlock: eastern north American conifers threatened by the hemlock woolly adelgid. Tree Plant Notes 56(2):59–71

Johnson JS, Gaddis KD, Cairns DM, Konganti K, Krutovsky KV (2017) Landscape genomic insights into the historic migration of mountain hemlock in response to Holocene climate change. Am J Bot 104(3): 439–450. https://doi.org/10.3732/ajb.1600262

Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24(11):1403–1405. https://doi.org/10.1093/bioinformatics/btn129

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

Josserand SA, Potter KM, Echt CS, Nelson CD (2008) Isolation and characterization of microsatellite markers for Carolina hemlock (*Tsuga caroliniana*). Mol Ecol Resour 8(6):1371–1374. https://doi.org/10.1111/j.1755-0998.2008.02294.x

Jost L (2008) G_{ST} and its relatives do not measure differentiation. Mol Ecol 17(18):4015–4026. https://doi.org/10.1111/j.1365-294X.2008.03887.x

Jump AS, Penuelas J (2005) Running to stand still: adaptation and the response of plants to rapid climate change. Ecol Lett 8(9):1010–1020. https://doi.org/10.1111/j.1461-0248.2005.00796.x

Kalia RK, Rai MK, Kalia S, Singh R, Dhawan AK (2011) Microsatellite markers: an overview of the recent progress in plants. Euphytica 177(3):309–334. https://doi.org/10.1007/s10681-010-0286-9

Kalinowski ST (2005) HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic richness. Mol Ecol Notes 5(1):187–189. https://doi.org/10.1111/j.1471-8286.2004.00845.x

Karhu A, Vogl C, Moran GF, Bell JC, Savolainen O (2006) Analysis of microsatellite variation in *Pinus radiata* reveals effects of genetic drift but no recent bottlenecks. J Evol Biol 19(1):167–175. https://doi.org/10.1111/j.1420-9101.2005.00982.x

Kimura M, Crow JF (1964) Number of alleles that can be maintained in a finite population. Genetics 49(4):725–738

Kimura M, Ohta T (1978) Stepwise mutation model and distribution of allelic frequencies in a finite population. Proc Natl Acad Sci U S A 75(6):2868–2872

Kitamura K, Matsui T, Kobayashi M, Saitou H, Namikawa K, Tsuda Y (2015) Decline in gene diversity and strong genetic drift in the northward-expanding marginal populations of *Fagus crenata*. Tree Genet Genomes 11(3):36. https://doi.org/10.1007/s11295-015-0857-y

Kizlinski ML, Orwig DA, Cobb RC, Foster DR (2002) Direct and indirect ecosystem consequences of an invasive pest on forests dominated by eastern hemlock. J Biogeogr 29(10–11):1489–1503. https://doi.org/10.1046/j.1365-2699.2002.00766.x

Lascoux M, Palme AE, Cheddadi R, Latta RG (2004) Impact of ice ages on the genetic structure of trees and shrubs. Philos Trans R Soc B 359(1442):197–207. https://doi.org/10.1098/rstb.2003.1390

Ledig FT, Hodgskiss PD, Johnson DR (2005) Genic diversity, genetic structure, and mating system of brewer spruce (Pinaceae), a relict of the Arcto-tertiary forest. Am J Bot 92(12):1975–1986. https://doi.org/10.3732/ajb.92.12.1975

Ledig FT, Hodgskiss PD, Johnson DR (2006) Genetic diversity and seed production in Santa Lucia fir (*Abies bracteata*), a relict of the Miocene broadleaved evergreen forest. Conserv Genet 7(3):383–398. https://doi.org/10.1007/s10592-005-9049-x

Lemieux MJ, Beaulieu J, Bousquet J (2011) Chloroplast DNA polymorphisms in eastern hemlock: range-wide genogeographic analyses and implications for gene conservation. Can J For Res 41(5): 1047–1059. https://doi.org/10.1139/x11-005

Libiger O, Nievergelt CM, Schork NJ (2009) Comparison of genetic distance measures using human SNP genotype data. Hum Biol 81(4):389–406

Liepelt S, Bialozyt R, Ziegenhagen B (2002) Wind-dispersed pollen mediates postglacial gene flow among refugia. Proc Natl Acad Sci U S A 99(22):14590–14594. https://doi.org/10.1073/pnas.212285399

Little EL (1971) Atlas of United States trees. Volume 1. Conifers and important hardwoods. Washington, D.C.

Mattioni C, Martin MA, Chiocchini F, Cherubini M, Gaudet M, Pollegioni P, Velichkov I, Jarman R, Chambers FM, Paule L (2017) Landscape genetics structure of European sweet chestnut (*Castanea sativa* Mill): indications for conservation priorities. Tree Genet Genomes 13(2):39. https://doi.org/10.1007/s11295-017-1123-2

McClure MS (1990) Role of wind, birds, deer, and humans in the dispersal of hemlock woolly adelgid (Homoptera, Adelgidae). Environ Entomol 19(1):36–43

McClure MS (1991) Density-dependent feedback and population cycles in *Adelges tsugae* (Homoptera, Adelgidae) on *Tsuga canadensis*. Environ Entomol 20(1):258–264

McLachlan JS, Clark JS, Manos PS (2005) Molecular indicators of tree migration capacity under rapid climate change. Ecology 86(8): 2088–2098. https://doi.org/10.1890/04-1036

Mihaescu R, Levy D, Pachter L (2009) Why neighbor-joining works. Algorithmica 54(1):1–24. https://doi.org/10.1007/s00453-007-9116-4

Montgomery ME, Bentz SE, Olsen RT (2009) Evaluation of hemlock (*Tsuga*) species and hybrids for resistance to *Adelges tsugae* (Hemiptera: Adelgidae) using artificial infestation. J Econ Entomol 102(3):1247–1254

Nienstaedt H, Kriebel HB (1955) Controlled pollination of eastern hemlock. For Sci 1(2):115–120

O'Connell LM, Ritland K, Thompson SL (2008) Patterns of post-glacial colonization by western redcedar (*Thuja plicata*, Cupressaceae) as revealed by microsatellite markers. Botany 86(2):194–203. https://doi.org/10.1139/b07-124

Orwig DA, Foster DR (1998) Forest response to the introduced hemlock woolly adelgid in southern New England, USA. J Torrey Bot Soc 125(1):60–73. https://doi.org/10.2307/2997232

Pandey M, Rajora OP (2012) Genetic diversity and differentiation of core vs. peripheral populations of eastern white cedar, *Thuja occidentalis* (Cupressaceae). Am J Bot 99(4):690–699. https://doi.org/10.3732/ajb.1100116

Parmesan C (2006) Ecological and evolutionary responses to recent climate change. Annu Rev Ecol Evol Syst 37:637–669. https://doi.org/10.1146/annurev.ecolsys.37.091305.110100

Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol Ecol Notes 6(1):288–295. https://doi.org/10.1111/j.1471-8286.2005.01155.x

Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in excel. Population genetic software for teaching and research-an update. Bioinformatics 28(19):2537–2539. https://doi.org/10.1093/bioinformatics/bts460

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(4):502–503. https://doi.org/10.1093/jhered/90.4.502

Pohjanmies T, Elshibli S, Pulkkinen P, Rusanen M, Vakkari P, Korpelainen H, Roslin T (2016) Fragmentation-related patterns of genetic differentiation in pedunculate oak (*Quercus robur*) at two hierarchical scales. Silva Fenn 50(2):1510. 10.14214/Sf.1510

Potter KM, Frampton J, Josserand SA, Nelson CD (2008) Genetic variation and population structure in Fraser fir (*Abies fraseri*): a

microsatellite assessment of young trees. Can J For Res 38(8):2128–2137. https://doi.org/10.1139/X08-064

Potter KM, Hipkins VD, Mahalovich MF, Means RE (2015) Nuclear genetic variation across the range of ponderosa pine (Pinus ponderosa): phylogeographic, taxonomic, and conservation implications. Tree Genet Genomes 11 (3). doi:https://doi.org/10.1007/s11295-015-0865-y

Potter KM, Jetton RM, Dvorak WS, Frampton J, Rhea JR (2010) *Ex situ* seed collection represents genetic variation present in natural stands of Carolina hemlock. In: Onken B, Reardon R (eds) Fifth symposium on hemlock woolly adelgid in the eastern United States. USDA Forest Service, Forest Health Technology Enterprise Team, Asheville, pp 181–190

Potter KM, Jetton RM, Dvorak WS, Hipkins VD, Rhea R, Whittier WA (2012) Widespread inbreeding and unexpected geographic patterns of genetic variation in eastern hemlock (Tsuga canadensis), an imperiled north American conifer. Conserv Genet 13(2):475–498. https://doi.org/10.1007/s10592-011-0301-2

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155(2):945–959

Putman AI, Carbone I (2014) Challenges in analysis and interpretation of microsatellite data for population genetic studies. Ecol Evol 4(22):4399–4428. https://doi.org/10.1002/ece3.1305

Quimby JW (1996) Value and importance of hemlock ecosystems in the eastern United States. In: Salom SM, Tigner TC, Reardon RC (eds) Proceedings of the first hemlock woolly adelgid review, vol FHTET 96–10. U.S. Forest Service, Forest Health Technology Enterprise Team, Morgantown, pp 1–8

R Core Team (2016) R: a language and environment for statistical computing. R Foundation for Statistical Computing. http://www.R-project.org

Rasmussen KK, Strandby U, Kollmann J (2010) High genetic diversity within but limited differentiation among populations of the vulnerable Guatemalan fir. J Trop For Sci 22(3):247–259

Raymond M, Rousset F (1995) Genepop (Version-1.2): Population genetics software for exact tests and ecumenicism. J Hered 86(3):248–249

Reed DH, Frankham R (2003) Correlation between fitness and genetic diversity. Conserv Biol 17(1):230–237. https://doi.org/10.1046/j.1523-1739.2003.01236.x

Rentch JS, Adams HS, Coxe RB, Stephenson SL (2000) An ecological study of a Carolina hemlock (*Tsuga caroliniana*) community in southwestern Virginia. Castanea 65(1):1–8

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4(4):406–425

Salom SM, Kok LT, Lamb A, Jubb C, Onken B (2008) Biological control of hemlock woolly adelgid: what is it going to take to make it work? In: Onken B, Reardon R (eds) Proceedings of the fourth symposium on hemlock woolly adelgid in the eastern United States, vol FHTET 2008–01. U.S. Forest Service, Forest Health Technology Enterprise Team, Morgantown, pp 11–17

SAS Institute Inc. (2013) The SAS System for Windows, version 9.4. Cary, North Carolina

Schaberg PG, DeHayes DH, Hawley GJ, Nijensohn SE (2008) Anthropogenic alterations of genetic diversity within tree populations: implications for forest ecosystem resilience. Forest Ecol Manag 256(5):855–862. https://doi.org/10.1016/j.foreco.2008.06.038

Schoettle AW, Goodrich BA, Hipkins V, Richards C, Kray J (2012) Geographic patterns of genetic variation and population structure in *Pinus aristata*, Rocky Mountain bristlecone pine. Can J For Res 42(1):23–37. https://doi.org/10.1139/x11-152

Slatkin M (1987) Gene flow and the geographic structure of natural populations. Science 236(4803):787–792

Stork NE, Coddington JA, Colwell RK, Chazdon RL, Dick CW, Peres CA, Sloan S, Willis K (2009) Vulnerability and resilience of tropical forest species to land-use change. Conserv Biol 23(6):1438–1447. https://doi.org/10.1111/j.1523-1739.2009.01335.x

Taberlet P, Fumagalli L, Wust-Saucy AG, Cosson JF (1998) Comparative phylogeography and postglacial colonization routes in Europe. Mol Ecol 7(4):453–464. https://doi.org/10.1046/j.1365-294x.1998.00289.x

Takezaki N, Nei M (1996) Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. Genetics 144(1):389–399

Vose JM, Wear DN, Mayfield AE III, Nelson CD (2013) Hemlock woolly adelgid in the southern Appalachians: control strategies, ecological impacts, and potential management responses. For Ecol Manag 291:209–219. https://doi.org/10.1016/j.foreco.2012.11.002

Willi Y, Van Buskirk J, Hoffmann AA (2006) Limits to the adaptive potential of small populations. Annu Rev Ecol Evol Syst 37:433–458. https://doi.org/10.1146/annurev.ecolsys.37.091305.110145

Wu J, Krutovskii KV, Strauss SH (1999) Nuclear DNA diversity, population differentiation, and phylogenetic relationships in the California closed-cone pines based on RAPD and allozyme markers. Genome 42(5):893–908. https://doi.org/10.1139/gen-42-5-893

Yanchuk AD, Lester DT (1996) Setting priorities for conservation of the conifer genetic resources of British Columbia. For Chron 72(4):406–415

Yorks TE, Jenkins JC, Leopold DJ, Raynal DJ, Orwig DA (2000) Influences of eastern hemlock mortality on nutrient cycling. In: McManus KA, Shields KS, Souto DR (eds) Proceedings: symposium on sustainable management of hemlock ecosystems in eastern North America, vol Gen. Tech. Rep. NE-267. U.S. Department of Agriculture, Forest Service, Northeastern Forest Experiment Station, Newtown Square, pp 126–133

Young A, Boyle T, Brown T (1996) The population genetic consequences of habitat fragmentation for plants. Trends Ecol Evol 11(10):413–418. https://doi.org/10.1016/0169-5347(96)10045-8

Ziegenhagen B, Bialozyt R, Liepelt S (2004) Contrasting molecular markers reveal: gene flow via pollen is much more effective than gene flow via seeds. In: Werner D (ed) Biological resources and migration. Springer, Berlin, pp 239–251