Allozyme variation and recent evolutionary history of eastern hemlock (*Tsuga canadensis*) in the southeastern United States

K. M. Potter · W. S. Dvorak · B. S. Crane · V. D. Hipkins · R. M. Jetton · W. A. Whittier · R. Rhea

Abstract Eastern hemlock (*Tsuga canadensis* [L.] Carr.) is a widespread and ecologically important conifer species of eastern North America that is threatened by the hemlock woolly adelgid (*Adelges tsugae* Annand), a pest introduced into the United States from Asia in the 1920s. Information about the genetic composition of eastern hemlock is necessary to guide ex situ conservation efforts in the southeastern United States, where the species is expected to harbor relatively high amounts of genetic variation in areas of Pleistocene glacial refuge. Nineteen allozyme markers were used to quantify the genetic variation present in 20 eastern hemlock populations in the southeastern United States. Results indicate that the species has low levels of genetic diversity in the region compared to most other conifers, but greater population differentiation ($F_{ST} = 0.126$). Populations along the eastern periphery and in the Appalachian interior exhibited higher levels of diversity than those along the western periphery of its geographic range. The results suggest that the glacial refuge area for eastern hemlock was likely located east of the southern Appalachian Mountains, and indicate that ex situ conservation seed collections should be concentrated in these areas of higher diversity.

K. M. Potter (✉)
Department of Forestry and Environmental Resources, North Carolina State University, 3041 Cornwallis Road, Research Triangle Park, NC 27709, USA
e-mail: kpotter@ncsu.edu

W. S. Dvorak · R. M. Jetton · W. A. Whittier
Camcore, North Carolina State University, Campus Box 8008, Raleigh, NC 27695, USA

B. S. Crane
USDA Forest Service, Southern Region, 1720 Peachtree Road NW, Suite 816N, Atlanta, GA 30309, USA

V. D. Hipkins
USDA Forest Service, National Forest Genetic Electrophoresis Laboratory, 2480 Carson Road, Placerville, CA 95667, USA

R. Rhea
USDA Forest Service, Forest Health Protection, 200 W.T. Weaver Boulevard, Asheville, NC 28804, USA

Introduction

Eastern hemlock (*Tsuga canadensis* [L.] Carr.) is a long-lived conifer species distributed in eastern North America from Nova Scotia to Wisconsin and Minnesota, and south along the Appalachian Mountain chain into Alabama, Georgia, Tennessee, South Carolina, and North Carolina, with scattered populations in Indiana and western Ohio (Godman and Lancaster 1990; McWilliams and Schmidt 2000). It is considered a foundation forest species that stabilizes fundamental ecosystem processes, such as mediating soil moisture levels, regulating stream base-flows, and decreasing variation in stream temperature (Ellison et al. 2005). Eastern hemlock forests, however, are being devastated by the hemlock woolly adelgid (*Adelges tsugae* Annand) (McClure et al. 2003), an exotic insect which may cause the reduction or functional disappearance of hemlock from eastern forests in the coming decades (Orwig and Foster 1998; Ellison et al. 2005). This is likely to lead to severe ecological repercussions on forest composition and nutrient cycles (Orwig and Foster 1998; Jenkins et al. 1999; Kizlinski et al. 2002; Small et al. 2005), hydrologic processes in forest ecosystems (Ford and Vose 2007), fish and aquatic invertebrate community structure (Snyder et al. 2002; Ross et al. 2003), and wildlife assemblages (Brooks 2001; Tingley et al. 2002).

The conservation of gene diversity and genetic integrity is crucial for forest tree species or populations which face elimination (Rajora and Mosseler 2001). Ex situ conservation, such as that of eastern hemlock in the southeastern United States by the Camcore International Tree Conservation and Domestication program at North Carolina State University (Camcore 2005; Camcore 2006), aims to allow for the eventual restoration of degraded or eliminated populations while maintaining the species' ability to evolve to adapt to changing conditions. The effective conservation of eastern hemlock's genetic diversity, however, will first require the identification of existing population genetic structure, including the distribution of genetic variation within and among populations, the number of rare alleles, and levels of inbreeding (Li et al. 1992; Eriksson et al. 1993).

With the exception of rangewide research on the relationship between eastern hemlock seed source and seedling growth characteristics (Olson and Nienstaedt 1957; Nienstaedt and Olson 1961), previous population genetic studies have focused on the northern part of the species' distribution. An allozyme study by Zabinski (1992), for example, found little genetic variation across the range of eastern hemlock, but included only two populations from the Southeast, while an analysis of chloroplast DNA restriction fragment length polymorphisms (Wang et al. 1997) detected little differentiation among populations, but incorporated only three from the southern part of the hemlock range. Greater attention is warranted to the genetic variation of eastern hemlock in the southeastern United States, given that northern and southern seed sources of eastern hemlock exhibited significantly different expression of adaptively important traits associated with seedling development (Olson and Nienstaedt 1957; Nienstaedt and Olson 1961).

Additionally, the southeastern United States was likely the location of eastern hemlock refugia during the most recent glacial maximum of the Pleistocene epoch, 12,000–20,000 years ago (Davis 1981; Delcourt and Delcourt 1987). Regions hosting glacial refugia are generally expected to have higher genetic diversity than more recently

colonized areas (Hewitt 1996; Hewitt 2000). Among North American conifers, this appears to be the case in Virginia pine (*Pinus virginiana* Mill.) (Parker et al. 1997), Coulter pine (*P. coulteri* D.Don) (Ledig 2000), lodgepole pine (*P. contorta* ssp. *latifolia* Engelm.) (Cwynar and Macdonald 1987), and whitebark pine (*P. albicaulis* Engelm.) (Jorgensen and Hamrick 1997). Colonized areas may be more genetically diverse than refugial regions, however, after the rapid shift of a species' entire range causes the retreating rear edge to shrink and become severely bottlenecked (Hewitt 2000), or when a region has been colonized by more than one refuge (Petit et al. 2003).

The current study employed 19 allozyme markers to quantify the genetic variation present in eastern hemlock in the southeastern United States, and to better understand the recent evolutionary history of the species. Specifically, this project sampled ten populations across an altitudinal gradient in the southern Appalachian Mountains, which is the core of the eastern hemlock range in the region. Less genetic variation was expected among ten populations along the geographic periphery of the species. Along with results generated from other ongoing molecular marker analyses, these data also will be used to prioritize future eastern hemlock seed collection for ex situ conservation.

Materials and methods

Camcore staff collected foliar and winter bud samples from 20 eastern hemlock populations in the southeastern United States during February and March of 2006 (Fig. 1,
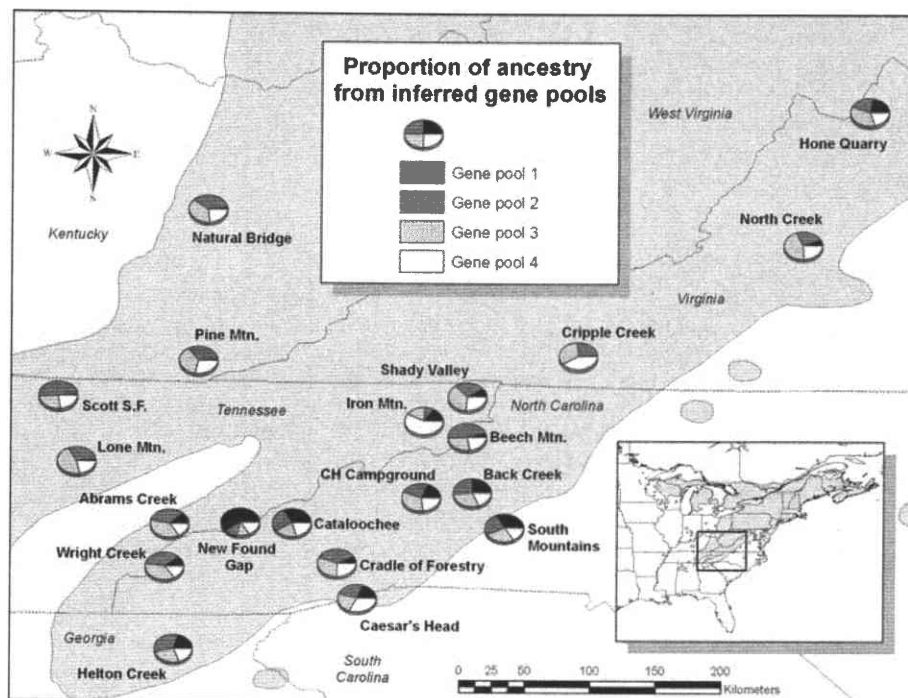

Fig. 1 The range of eastern hemlock, the location of the 20 eastern hemlock populations included in the analyses, and the proportion of ancestry of each population from the gene pools defined with the model-based clustering method of Structure (Pritchard et al. 2000)

Table 1). Ten populations were located in the interior (Appalachian Mountain) range of the species and were sampled on an altitudinal gradient. The remaining ten were located along the species' eastern and western geographic periphery, with eastern populations located east of the Blue Ridge/Appalachian escarpment, and western populations located along the Cumberland escarpment in Tennessee and Kentucky. Foliar and bud samples were collected from 20 trees per site (two samples per tree), from trees spaced at least 100 m apart in keeping with established gene conservation strategies, as the results of the study will guide future eastern hemlock seed collection efforts. This sampling strategy diversifies the seed source by avoiding the sampling of neighbors that may be related as a result of limited seed or pollen dispersal (Brown and Hardner 2000), which is an important consideration in eastern hemlock as most seeds fall within tree height (Godman and Lancaster 1990; Alvarez-Buylla et al. 1996; Jorgensen and Hamrick 1997). The sample size of 20 trees per population might under-represent the number of rare alleles in a population, but allowed for the inclusion of more populations than would have been possible with a larger sample size. The eight hundred samples from 400 trees were sent to the National Forest Genetic Electrophoresis Laboratory (NFGEL) in Placerville, California, for allozyme analysis.

National Forest Genetic Electrophoresis Laboratory staff stripped needles from both sides of 6 cm branches, placed them in a cold mortar, and ground them to a fine powder under liquid nitrogen. Approximately 0.5 ml of a Tris–HCl extraction buffer was added to the ground powder and the resulting slurry frozen at –80°C until electrophoresis, following NFGEL standard operating procedures. On the morning of the electrophoretic run, samples were thawed and absorbed onto three 3 mm paper wicks. Extracts were separated on 11% starch gels (Starch-Art) and stained for 12 enzymes resolved on three different buffer systems. A lithium borate electrode buffer (pH 8.3) was used with a Tris citrate gel buffer, pH 8.3, to resolve *LAP-1* and *2*, *PGI-1* and *2*, *PGM*, and *ME(7)*; a sodium borate electrode buffer (pH 8.0) was used with a Tris citrate gel buffer, pH 8.8, to resolve *UGPP*, *TPI-3*, *6PGD-1*, *2* and *3*, and *AAT-1*, *2* and *3*; a morpholine citrate electrode and gel buffer, pH 6.1, were used to resolve *MDH-1* and *3*, *IDH*, *EST*, and *SKD*. For quality control, all samples were scored twice and conflicts resolved. Additionally, 10% of samples were genotyped twice.

Diversity statistics were calculated using Popgene (Yeh et al. 2000), while the program Arlequin 3.0 (Excoffier et al. 2005) was used to conduct an analysis of molecular variance (AMOVA), which allows for the investigation of differentiation among populations and groups of populations by partitioning the total variation into components. Arlequin also generated pairwise F_{ST}-values for all pairs of populations. Nei's (Nei 1972) genetic distances were estimated for each pair of populations. These genetic distances were used to construct an unrooted consensus phenogram, bootstrapped over 1,000 replicates, using the SEQBOOT, GENDIST, NEIGHBOR, and CONSENSE components of PHYLIP 3.6 (Felsenstein 2005). Pearson correlation coefficients were calculated between population genetic data and geographic characteristics using the PROC CORR procedure in SAS 9.1 (SAS Institute Inc. 2003). Correlations between the genetic and characteristics of populations, including their geographic distances and their differences in elevation, were explored with Mantel tests using the Isolation by Distance 1.5 (IBD) software package (Bohonak 2002). Point-to-point distances were measured with ArcGIS 8.1 (ESRI 2001). The Bayesian clustering program Structure 2.0 (Pritchard et al. 2000) was used to search for hidden population structure, by assigning individual multilocus genotypes probabilistically to user-defined clusters or gene pools (*K*), achieving linkage equilibrium within clusters. The program was run three times for *K* = 1–8. The recommendations of the program's manual (Pritchard and Wen 2004) were followed, allowing the degree of

Table 1 Location, elevation, and allozyme genetic statistics for each of 20 eastern hemlock populations in analysis

Name	Location	Elev. (m)	A	A_e	A_p	P	H_o	H_e	F	$\#F_{ST}$
Abrams Creek	I	362	1.37 (0.68)	1.18 (0.47)	1	26.32	0.045 (0.125)	0.090 (0.188)	0.493	17
Back Creek	EP	382	1.37 (0.6)	1.18 (0.37)	0	31.58	0.061 (0.159)	0.101 (0.184)	0.553	14
Beech Mountain	I	949	1.32 (0.67)	1.13 (0.45)	0	21.05	0.029 (0.093)	0.054 (0.160)	0.458	9
Caesar's Head	I	935	1.37 (0.68)	1.19 (0.47)	0	26.32	0.037 (0.108)	0.095 (0.188)	0.674	8
Cataloochee	I	835	1.32 (0.58)	1.17 (0.45)	0	26.32	0.032 (0.095)	0.084 (0.186)	0.453	13
CH Campground	I	857	1.42 (0.69)	1.14 (0.35)	0	31.58	0.045 (0.149)	0.079 (0.158)	0.555	6
Cradle of Forestry	I	1,017	1.37 (0.68)	1.13 (0.42)	0	26.32	0.047 (0.151)	0.063 (0.156)	0.474	7
Cripple Creek	EP	787	1.21 (0.54)	1.09 (0.25)	0	15.79	0.034 (0.106)	0.052 (0.141)	0.255	13
Helton Creek	EP	723	1.42 (0.61)	1.17 (0.45)	1	36.84	0.032 (0.115)	0.088 (0.170)	0.694	7
Hone Quarry	EP	585	1.47 (0.77)	1.15 (0.34)	2	31.58	0.061 (0.186)	0.090 (0.168)	0.365	10
Iron Mtn.	I	758	1.32 (0.58)	1.09 (0.23)	0	26.32	0.045 (0.121)	0.062 (0.134)	0.226	18
Lone Mtn.	WP	335	1.21 (0.54)	1.08 (0.26)	0	15.79	0.032 (0.104)	0.048 (0.139)	0.263	13
Natural Bridge	WP	333	1.21 (0.54)	1.11 (0.42)	0	15.79	0.026 (0.103)	0.047 (0.155)	0.426	11
Newfound Gap	I	1,224	1.47 (0.77)	1.27 (0.5)	1	31.58	0.046 (0.121)	0.138 (0.231)	0.725	19
North Creek	EP	352	1.26 (0.56)	1.1 (0.32)	0	21.05	0.042 (0.140)	0.056 (0.145)	0.481	6
Pine Mtn.	WP	486	1.21 (0.54)	1.08 (0.27)	0	15.79	0.026 (0.103)	0.041 (0.131)	0.378	9
Scott State Forest	WP	442	1.16 (0.5)	1.07 (0.3)	0	10.53	0.021 (0.081)	0.033 (0.133)	0.178	13
Shady Valley	I	1,121	1.32 (0.58)	1.12 (0.36)	0	26.3	0.045 (0.118)	0.066 (0.154)	0.462	13
South Mountains	EP	419	1.37 (0.68)	1.16 (0.37)	0	26.32	0.050 (0.133)	0.088 (0.183)	0.301	12
Wright Creek	I	1,276	1.37 (0.68)	1.12 (0.33)	0	26.3	0.053 (0.144)	0.071 (0.156)	0.454	16
Mean		708.9	1.33 (0.9)	1.14 (0.05)	0.25	24.47	0.040 (0.011)	0.072 (0.025)	0.443	11.7

Location: *I* interior, *EP* eastern periphery, *WP* western periphery. *A* mean alleles per locus, A_e effective alleles per locus, A_p private alleles, *P* percent polymorphic loci, H_o observed heterozygosity, H_e expected heterozygosity, *F* fixation index (inbreeding coefficient), a measure of deviation from Hardy–Weinberg equilibrium, $\#F_{ST}$ number of statistically significant F_{ST}-values with other populations. Standard errors are in parentheses

admixture α to be inferred from the data, and set the distribution of allelic frequencies (λ) to 1. The burn-in was set to 50,000 replicates, and the Markov chain Monte Carlo to 10^6 replicates. The smallest number of gene pools K was selected after the log likelihood of the multilocus genotypic data $\ln(X|K)$ had plateaued at its highest level, as recommended by the program manual (Pritchard and Wen 2004). The fractions of ancestry from each inferred gene pool were calculated for each population and mapped.

The software package Bottleneck 1.2.02 (Cornuet and Luikart 1996; Piry et al. 1999) estimated the probability of genetic bottlenecks in the recent past for each population, and for hemlock throughout the Southeast. This program computes the difference, averaged over loci, between the actual heterozygosity and the heterozygosity expected from the number of alleles in a population assuming mutation-drift equilibrium. The program was used to run both the Wilcoxon test, which is the most powerful and robust test when fewer than 20 polymorphic loci are included in an analysis (Piry et al. 1999), and the standardized differences test (Cornuet and Luikart 1996). Positive values of T_2, generated by the standardized differences test, are indicative of a gene diversity excess caused by a recent reduction in effective population size, while negative values are consistent with a deficiency of gene diversity associated with a recent population expansion without immigration. The program generated estimates based on the two-phased model of mutation (TPM), which is an intermediate between the single mutation model (Kimura and Ohta 1978) and infinite alleles model (Kimura and Crow 1964), consisting mostly of one-step mutations but also including a small percentage (5%) of multiple-step changes (Piry et al. 1999).

Results

Eastern hemlock in the southeastern United States exhibits a low level of allozyme variation, as measured by percent of polymorphic loci, alleles per locus, and expected heterozygosity (Table 2). The allozyme analyses indicated an overall deficiency of heterozygotes in comparison to Hardy–Weinberg expectations at all levels of analysis, and in all populations, as indicated by the positive F inbreeding coefficient values (Tables 1, 2). Specifically, the positive inbreeding coefficient F_{IS} (of an individual relative to its own population) indicates the likely presence of inbreeding within the eastern hemlock populations included in the study. When populations were grouped by geographic location, western populations exhibited the lowest value for all genetic diversity statistics, while eastern populations had the greatest polymorphism, observed heterozygosity, and number of private alleles. The interior grouping had the highest number of effective alleles per locus, but was also the most inbred, as measured by F_{IS}, and the most differentiated, as measured by F_{ST} (Table 2). Within the populations, a statistically significant relationship existed between geographic location (where 1 = western periphery, 2 = interior, and 3 = eastern periphery) and observed heterozygosity ($r = 0.608$, $p = 0.005$), polymorphic loci ($r = 0.599$, $p = 0.005$), and number of alleles ($r = 0.532$, $p = 0.016$). Separately, population longitude was negatively correlated with observed heterozygosity ($r = -0.513$, $p = 0.021$), with populations farther east having higher values.

The proportion of variation residing among populations was 12.6% ($F_{ST} = 0.126$) according to the PopGene analysis (Table 2), and 10.9% ($F_{ST} = 0.109$) using the AMOVA method (Table 3). The AMOVA analysis also found that 89.2% of variation resides within populations, and that a statistically insignificant amount occurs among geographic groups (Table 3). Between pairs of geographic groups, the western group was the most

Table 2 Allozyme genetic statistics for each of three geographic groups of eastern hemlock populations in the southeastern United States, as well as for all 20 populations

Group	N	A	A_e	A_p	P	H_o	H_e	F	F_{IS}	F_{ST}	Nm	$\%F_{ST}$
Eastern periphery	119	1.74 (0.81)	1.14 (0.36)	3	52.63	0.047 (0.131)	0.083 (0.153)	0.566	0.398	0.071	3.292	40
Interior	192.9	1.63 (0.83)	1.17 (0.42)	2	42.11	0.042 (0.111)	0.091 (0.162)	0.638	0.457	0.144	1.483	71.11
Western periphery	79.1	1.32 (0.58)	1.09 (0.33)	0	26.32	0.026 (0.097)	0.044 (0.137)	0.368	0.363	0.052	4.606	16.67
All populations	391.1	1.84 (0.9)	1.15 (0.39)	–	57.9	0.040 (0.113)	0.080 (0.152)	0.426	0.426	0.126	1.735	–

N samples per group, A mean alleles per locus, A_e effective alleles per locus, A_p private alleles, P percent polymorphic loci, H_o observed heterozygosity, H_e expected heterozygosity, F fixation index (inbreeding coefficient), F_{IS} inbreeding coefficient of individuals relative to population, F_{ST} estimate of population differentiation within group, N_M estimated number of migrants per population per generation, $\%F_{ST}$ percent of pairwise between-population differences within the group that are statistically different from 0. Standard errors are in parentheses

Table 3 Results of analysis of molecular variance (AMOVA) of 20 eastern hemlock populations divided into three geographic groups, interior, eastern, and western

Source of variation	df	Variance component	Variation (%)	F-stat	p
Among groups (F_{CT})	2	0.005	0.8	0.007	0.195
Among populations within groups (F_{SC})	17	0.073	10.2	0.102	<0.001
Within populations	780	0.647	89.2		
Among populations (F_{ST})			10.9	0.109	<0.001

Significance level of variance components determined by 1,000 permutations

differentiated, with 3% of variation between the western periphery and the interior, and 2.2% between the western and eastern peripheries. The proportion of variation residing between the interior and eastern periphery was only 0.4%.

The eastern hemlock populations shared, on average, 98.9% genetic similarity, based on Nei's genetic identity (Nei 1972). This result is not surprising given estimates of inter-population migration (Nm) that varied from 1.48 per generation within the interior populations to 4.61 among the western peripheral populations (Table 2). The most differentiated population was Newfound Gap, the only population having statistically significant pairwise F_{ST}-values with every other population (Table 1). Newfound Gap, the population at the second highest elevation, was also the most diverse by several measures, including number of alleles per locus, effective alleles per locus, and expected heterozygosity. It was also only one of three populations with private alleles, along with Hone Quarry and nearby Abrams Creek. These three populations, along with Back Creek, Caesar's Head, and Helton Creek, consis-tently rated among the most diverse. All are eastern or interior populations. All of the western peripheral populations, meanwhile, consistently exhibited the lowest genetic diversity, with Scott State Forest having the lowest number of alleles and effective alleles, polymorphic loci, and observed and expected heterozygosity.

The elevation and latitude of populations were important in explaining differences in some genetic diversity statistics, although separating the respective influences of these geographic characteristics is difficult because populations farther south are somewhat more likely to be at higher elevations ($r = 0.406$, $p = 0.076$). Populations farther south had more effective alleles ($r = 0.423$, $p = 0.063$), higher polymorphism ($r = 0.405$, $p = 0.077$), greater expected heterozygosity ($r = 0.400$, $p = 0.083$), a greater percentage of statistically significant pairwise F_{ST} differences within their groups ($r = 0.466$, $p = 0.038$), and higher inbreeding coefficient F ($r = 0.447$, $p = 0.048$). Population elevation was positively cor-related with number of alleles ($r = 0.444$, $p = 0.05$), polymorphism ($r = 0.378$, $p = 0.1$), within-group pairwise F_{ST} differences ($r = 0.624$, $p = 0.003$), and inbreeding coefficient ($r = 0.401$, $p = 0.08$). Additionally, a Mantel test indicated the presence of a weak cor-relation between pairwise population elevation differences and pairwise F_{ST} differences ($r = 0.195$, $p = 0.042$) and between population elevation differences and Nei's pairwise genetic distances ($r = 0.192$, $p = 0.08$). The Mantel analyses found no correlation between interpopulation geographic distance and pairwise F_{ST}-values either for all of the popula-tions or within any of the three regional divisions.

The Structure analysis inferred the existence of four ancestral gene pools (K) in eastern hemlock populations of the southeastern United States, with the log-likelihood values for the data conditional on K, $\ln(X|K)$, increasing considerably from $K = 1$ to 4, then decreasing at higher K values (results not shown). Mapping the fractions of ancestry from each inferred gene pool for each population (Fig. 1) indicated that only three gene pools were common in the western peripheral populations (Scott State Forest, Natural Bridge,

Pine Mountain, and Lone Mountain), as well as in the interior Beech Mountain population and eastern Cripple Creek population. Gene pool 1 was most common in the Newfound Gap, Cataloochee, and South Mountains populations toward the southern edge of the eastern hemlock range. A consensus neighbor-joining phenogram based on Nei's genetic distance (Fig. 2) found limited support ($< \sim 500$ of the 1,000 replicates) for many of the population relationships. There was moderate support for the cluster of the southern populations at Newfound Gap, Cataloochee, Back Creek, and South Mountains, and some support for a cluster of Wright Creek, Abrams Creek, and Shady Valley in Tennessee.

Eastern hemlock in the southeastern United States did not exhibit the heterozygosity deficiency associated with a recent demographic bottleneck, nor did the eastern, interior and western geographical groupings within the region. Instead, eastern hemlock in the Southeast had an excess of heterozygosity associated with a recent population expansion, according to both the Wilcoxon and standardized difference tests in Bottleneck (Piry et al. 1999) (Table 4). The population-level Wilcoxon tests indicate that Newfound Gap is the only population that has recently experienced a genetic bottleneck.

Discussion

According to the allozyme results of this study, eastern hemlock in the southeastern United States exhibited considerably less genetic diversity than other conifers. The expected
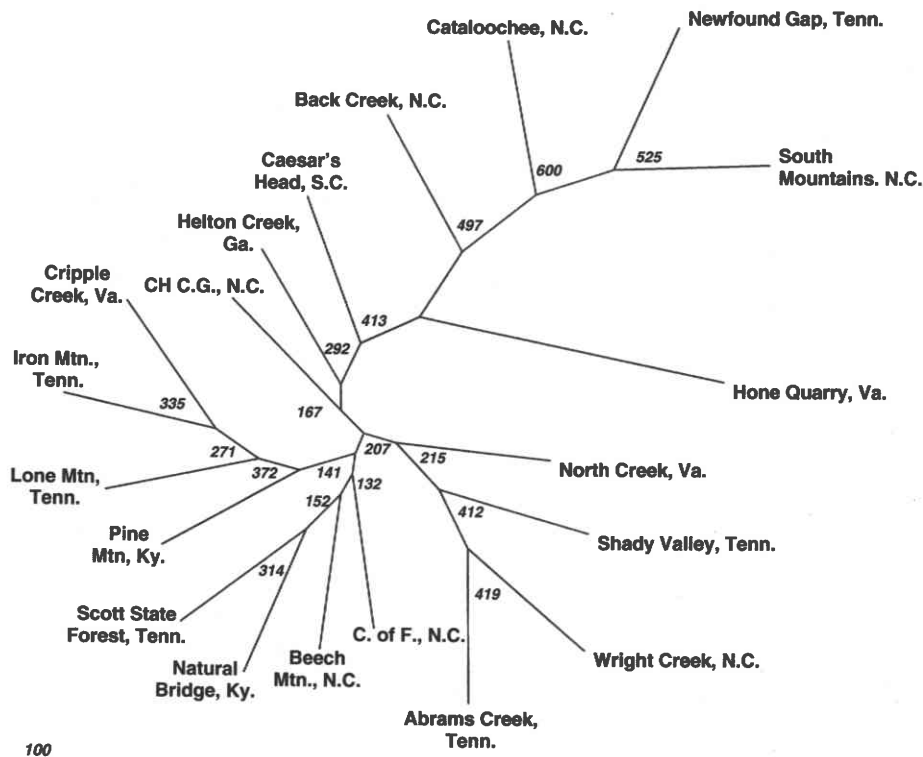

Fig. 2 Unrooted consensus neighbor-joining dendrogram depicting unbiased genetic distances (Nei 1978). The values represent the bootstrap support for the nodes over 1,000 replicates

Table 4 Measures of heterozygosity deficiency for eastern hemlock populations, using the program Bottleneck (Piry et al. 1999) assuming the two-phase model (TPM) of locus evolution

Group	T_2	p	Wilcoxon
All	−3.00	**	**
Eastern	−2.95	**	**
Interior	−1.68	*	*
Western	−1.42	*	*

T_2 bottleneck statistic from standardized difference test, with negative value indicating heterozygosity excess and a likely recent population expansion; Wilcoxon: statistical significance of Wilcoxon test for heterozygosity excess

*Significant at 0.1

**Significant at 0.01

heterozygosity of eastern hemlock was 0.072 compared to the average of 0.176 across 103 allozyme studies in Pinaceae species, while 57.9% of eastern hemlock loci were polymorphic, compared to 73% in the other Pinaceae species (Hamrick and Godt 1996). Also, while eastern hemlock had higher allozyme polymorphism than either mountain hemlock (*Tsuga mertensiana* Bong.) or western hemlock (*Tsuga heterophylla* [Raf.] Sarg.) in British Columbia (Ally et al. 2000; Wellman et al. 2003), eastern hemlock had lower expected heterozygosity values than both species, which were 0.087 and 0.142, respectively.

It is possible that the high inbreeding and low heterozygosity estimates in the current study result from a Wahlund effect caused by spatial sub-structuring of within-population demes (Alvarez-Buylla et al. 1996). It seems unlikely that a Wahlund effect would lead to large overestimates of inbreeding and large underestimates of heterozygosity, however, even though we did not employ a clustered sampling strategy to avoid population sub-structuring. Fine-scale genetic structure within plant populations is predominantly the result of limited pollen and seed dispersal (Cavers et al. 2005), so out-crossing tree species with effective long-distance dispersal mechanisms are expected to exhibit only weak fine-scale spatial genetic structure (Vekemans and Hardy 2004). As an out-crossing wind-pollinated species with very small seeds (Young and Young 1992), eastern hemlock is probably an efficient disperser of genes within and among stands. Additionally, the estimates of the number of inter-population migrants per generation (*Nm*) were high (Table 2), and while calculating this statistic from F_{ST} is problematic (Whitlock and McCauley 1999), the results suggest the potential for ample gene flow within and among populations.

This study offers insights into the recent evolutionary history of eastern hemlock in the southern portion of its current distribution. The results, when compared to past research, suggest that greater eastern hemlock genetic diversity resides in the southeastern United States than in the northern part of the species' range. Specifically, this study revealed a considerably greater level of polymorphism than in Zabinski's (1992) range-wide allozyme study of eastern hemlock, in which only one of ten loci was polymorphic. Of the five loci included in both analyses (*IDH*, *MDH-1*, *MDH-3*, *PGI-1*, and *PGI-2*), two were polymorphic in the Southeast (*IDH* and *PGI-2*), while none were polymorphic in the range-wide study (Zabinski 1992). The overall expected heterozygosity of Zabinski's study would be 0.037, or half that of the current Southeastern study, given that the expected heterozygosity of the single polymorphic locus was 0.37. These results support paleobotanical evidence that the refuge area for eastern hemlock during the peak of the

Wisconsinian glaciation was located in the Southeast (Cronin et al. 1981; Davis 1981; Delcourt and Delcourt 1987).

The current allozyme results also appear to shed additional light on where in the Southeast hemlock refuge regions may have been located. The presence of hemlock pollen in 25,000-year-old Carolina bay sediment in northeastern North Carolina (Whitehead 1973) and in 16,000-year-old pond sediments in northwest Georgia (Watts 1970), as well as the existence of a relict hemlock stand in the eastern Piedmont of North Carolina (Oosting and Hess 1956; Hardin and Cooper 1967), are all in keeping with the hypothesis that the refuge area was located in the Appalachians, on the coastal plain of the Carolinas and Virginia, or on the continental shelf (Cronin et al. 1981; Davis 1981). The discovery of hemlock pollen from 20,000 years ago in western Tennessee (Delcourt et al. 1980), however, led Delcourt and Delcourt (1987) to note the possibility that restricted hemlock populations existed at the time in a narrow latitudinal band of mixed conifer-northern hardwood forest stretching from the Mississippi valley to the Atlantic Ocean. The results of the current study, however, appear to support the hypothesis that eastern hemlock populations in the southeastern United States descended mostly or entirely from refugia located to the east of the southern Appalachians. Specifically, the eastern populations are most likely the closest to the glacial refugia, given their higher number of polymorphic loci and alleles per locus (Table 1) and the expectation that areas closer to a refuge will have higher genetic diversity than those more recently colonized (Hewitt 1996; Hewitt (2000). The high genetic variation present in these populations may, in fact, have resulted from the admixture of descendents from more than one glacial refuge, as Petit et al. (2003) found among 22 widespread European tree species. The four western peripheral populations, meanwhile, were the least diverse, and were presumably colonized later than the interior and eastern populations, and by fewer refugial lineages. This is consistent with Zabinski's (1992) results from the two southeastern populations in her analysis; her single Alabama population was monomorphic at the single locus polymorphic for the species, while the Joyce Kilmer Memorial Forest population in North Carolina had somewhat higher than average expected heterozygosity.

The results of the Structure cluster analysis for eastern hemlock in the Southeast (Fig. 1) infer the prevalence of four ancestral gene pools within and to the east of the Appalachians, but only three to the west. Of these, gene pool 1 seems most prevalent in the southern part of the range, suggesting that populations in this area may have descended at least in part from a single refuge. This is underscored by the observation that the only well-supported cluster in the neighbor-joining dendrogram (Fig. 2) encompasses the same populations: Newfound Gap, Cataloochee, Back Creek, and South Mountains. It seems unlikely that the western populations descended from a different refuge to the west of the Appalachians, given that they do not cluster closely in the neighbor-joining dendrogram, and that the three Structure gene pools represented in these populations are also common in the interior and eastern populations.

An allozyme study of Virginia pine detected a similar pattern to the current eastern hemlock results, leading the authors to conclude that this species moved northwestward across the Appalachians from glacial refugia located southeast of the mountain chain (Parker et al. 1997). The findings from the current study, as well as fossil data documenting the establishment of new hemlock stands in the upper Midwest (Parshall 2002), are in line with Hewitt's (2000) hypothesis that the leading edge of population expansion following climate warming would probably be by a few long-distance dispersers, resulting in lower heterozygosity and fewer alleles than in the refugial area. This genetic pattern was demonstrated for Coulter pine, where population mean heterozygosity and allele occurrence

decreased with latitude (Ledig 2000). The current eastern hemlock study shows a similar pattern, with population latitude negatively correlated with several genetic diversity measures; in other words, populations farther south had more effective alleles per locus, higher polymorphism, and greater expected heterozygosity. This may indicate that the southernmost populations are located at or near the site of one or more glacial refuge. The higher levels of inbreeding in these populations, and at Newfound Gap in particular, may have resulted from the decrease in interpopulation gene flow as the warming conditions shifted the main distribution of the hemlock range to the north by ~6,000 years ago (Delcourt and Delcourt 1987).

Another factor probably shaping the current genetic composition of eastern hemlock in the southeastern United States was a dramatic decline in the species that occurred 5,000 years ago in association with a pest infestation (Bhiry and Filion 1996; Fuller 1998). This caused a considerable decrease in the abundance of hemlock pollen in the fossil record, followed by a period of ~2,000 years before returning to its former level of abundance (Davis 1981; Delcourt and Delcourt 1987). This event may help explain why the current allozyme study revealed considerably less genetic variation in southern populations of eastern hemlock than in other conifers (Hamrick and Godt 1996), because local populations of the species probably passed through small demographic bottlenecks as the majority of hemlock trees died. This is, in fact, a plausible alternative explanation for the high inbreeding estimates for some of the southern populations, including Newfound Gap, the only population found by Bottleneck analysis to exhibit the heterozygosity deficiency consistent with a recent population bottleneck. However, the influence of the pest-caused mortality on eastern hemlock population genetics is likely to be smaller than that of long-term Pleistocene refugial and migration patterns. While eastern hemlock typically dropped to 10% dominance in central Appalachian forests following the dieback, the species apparently still existed in many places in reduced numbers and was, in fact, able to expand its range westward through northern Indiana and in Wisconsin (Delcourt and Delcourt 1987). These local bottlenecks would have reduced genetic diversity in many places, but at least some genetic variation and differentiation probably survived to be inherited by current eastern hemlock populations. At <2,000 years in duration, these recent bottlenecks were also much shorter than the periods of time over the course of the 100,000-year Wisconsinian glaciation during which eastern hemlock existed in glacial refugia. Interestingly, the Bottleneck analyses did not detect the heterozygosity deficiency associated with a recent demographic bottleneck (Table 4) in eastern hemlock in the Southeast, or in any of its three sub-regional divisions. In fact, the results strongly suggested a recent expansion in population size, without immigration, for the species and its three divisions. This may reflect the recovery of eastern hemlock from its severe size reduction 5,000–3,000 years ago, and the subsequent expansion of the species' importance that has continued into recent centuries (Delcourt and Delcourt 1987). The importance of eastern hemlock, however, is now decreasing in many places as a result of infestation by the exotic hemlock woolly adelgid (Orwig and Foster 1998; Kizlinski et al. 2002; Orwig et al. 2002).

Implications

Camcore and USDA Forest Service Forest Health Protection plan to collect and send seeds from many of the study populations to areas in the United States, Brazil and Chile, where the hemlock woolly adelgid does not occur. These seeds will be used to establish ex situ conservation areas to maintain the genetic diversity of eastern hemlock in case researchers

are unable to control hemlock mortality resulting from adelgid infestation. If this worst-case scenario occurs, genetic material from these ex situ eastern hemlock plantings could eventually be reintroduced into the southeastern United States to repopulate areas where this species has been devastated.

The results of this allozyme study suggest that ex situ conservation seed collections of eastern hemlock in the southeastern United States should be concentrated to the east of the southern Appalachians and in the high-elevation populations of the southernmost Appalachians in North Carolina, South Carolina, and Tennessee. The area of Great Smoky Mountains National Park, which includes the Newfound Gap, Cataloochee, and Abrams Creek populations, should be of particular interest, given the higher heterozygosity and number of rare alleles occurring in these populations. Areas of lower genetic diversity, such as in the western periphery of the eastern hemlock range in the Southeast, also should be included in collections, but at lower site selection intensity.

References

Ally D, El-Kassaby Y, Ritland K (2000) Genetic diversity, differentiation and mating system in mountain hemlock (*Tsuga mertensiana*) across British Columbia. For Genet 7(2):97–108

Alvarez-Buylla ER, Chaos A, Pinero D et al (1996) Demographic genetics of a pioneer tropical tree species: patch dynamics, seed dispersal, and seed banks. Evolution 50(3):1155–1166

Bhiry N, Filion L (1996) Mid-Holocene hemlock decline in eastern North America linked with phytophagous insect activity. Quat Res 45(3):312–320

Bohonak AJ (2002) IBD (isolation by distance): a program for analyses of isolation by distance. J Hered 93(2):153–154

Brooks RT (2001) Effects of the removal of overstory hemlock from hemlock-dominated forests on eastern redback salamanders. For Ecol Manage 149(1–3):197–204

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, Australia

Camcore (2005) 2005 Camcore Annual Report. Camcore International Tree Conservation and Domestication, Department of Forestry and Environmental Resources, North Carolina State University, Raleigh, NC

Camcore (2006) 2006 Camcore Annual Report. Camcore International Tree Conservation and Domestication, Department of Forestry and Environmental Resources, North Carolina State University, Raleigh, NC

Cavers S, Degen B, Caron H et al (2005) Optimal sampling strategy for estimation of spatial genetic structure in tree populations. Heredity 95(4):281–289

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

Cronin TM, Szabo BJ, Ager TA et al (1981) Quaternary climates and sea levels of the United States Atlantic Coastal Plain. Science 211(4479):233–240

Cwynar LC, Macdonald GM (1987) Geographical variation of lodgepole pine in relation to population history. Am Nat 129(3):463–469

Davis MB (1981) Quaternary history and the stability of forest communities. In: Shugart HH, Botkin DB, West DC (eds) Forest succession, concepts and applications. Springer, New York

Delcourt PA, Delcourt HR (1987) Long-term forest dynamics of the temperate zone. Springer, New York

Delcourt PA, Delcourt HR, Brister RC et al (1980) Quaternary vegetation history of the Mississippi embayment. Quat Res 13(1):111–132

Ellison AM, Bank MS, Clinton BD et al (2005) Loss of foundation species: consequences for the structure and dynamics of forested ecosystems. Front Ecol Environ 3(9):479–486

Eriksson G, Namkoong G, Roberds JH (1993) Dynamic gene conservation for uncertain futures. For Ecol Manage 62(1–4):15–37

ESRI (2001) ArcGIS 8.1. Environmental Systems Research Institute, Redlands, CA

Excoffier L, Laval G, Schneider S (2005) Arlequin ver. 3.0: an integrated software package for population genetics data analysis. Evol Bioinform Online 1:47–50

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

Fuller JL (1998) Ecological impact of the mid-holocene hemlock decline in southern Ontario, Canada. Ecology 79(7):2337–2351

Godman RM, Lancaster K (1990) Eastern hemlock. In: Burns RM, Barbara H. Honkala (eds) Silvics of North America: 1 Conifers. United States Department of Agriculture Forest Service, Washington, DC

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

Hardin JW, Cooper AW (1967) Mountain disjuncts in the eastern Piedmont of North Carolina. J Elisha Mitch Sci S 83(3):139–150

Hewitt GM (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. Biol J Linn Soc 58(3):247–276

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

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

Jorgensen SM, Hamrick JL (1997) Biogeography and population genetics of whitebark pine, *Pinus albicaulis*. Can J For Res 27(10):1574–1585

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 USA 75(6):2868–2872

Kizlinski ML, Orwig DA, Cobb RC et al (2002) Direct and indirect ecosystem consequences of an invasive pest on forests dominated by eastern hemlock. J Biogeogr 29(10–11):1489–1503

Ledig FT (2000) Founder effects and the genetic structure of Coulter pine. J Hered 91(4):307–315

Li P, MackKay J, Bousquet J (1992) Genetic diversity in Canadian hardwoods: implications for conservation. Forest Chron 68(6):709–719

McClure MS, Salom SM, Shields KS (2003) Hemlock woolly Adelgid. Forest Health Technology Enterprise Team. United States Department of Agriculture, Forest Service, Morgantown, WV

McWilliams WH, Schmidt TL (2000) Composition, structure and sustainability of hemock ecosystems in eastern North America. In: Proceedings of the symposium on sustainable management of hemlock ecosystems in Eastern North America, Durham, New Hampshire, United States Department of Agriculture, Forest Service

Nei M (1972) Genetic distance between populations. Am Nat 106(949):283–292

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89(3):583–590

Nienstaedt H, Olson JS (1961) Effects of photoperiod and source on seedling growth of eastern hemlock. Forest Sci 7(1):81–96

Olson JS, Nienstaedt H (1957) Photoperiod and chilling control growth of hemlock. Science 125(3246):492–494

Oosting HJ, Hess DW (1956) Microclimate and a relic stand of Tsuga canadensis in the lower piedmont of North Carolina. Ecology 37(1):28–39

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

Orwig DA, Foster DR, Mausel DL (2002) Landscape patterns of hemlock decline in New England due to the introduced hemlock woolly adelgid. J Biogeogr 29(10–11):1475–1487

Parker KC, Hamrick JL, Parker AJ et al (1997) Allozyme diversity in *Pinus virginiana* (Pinaceae): intra-specific and interspecific comparisons. Am J Bot 84(10):1372–1382

Parshall T (2002) Late holocene stand-scale invasion by hemlock (*Tsuga canadensis*) at its western range limit. Ecology 83(5):1386–1398

Petit RJ, Aguinagalde I, de Beaulieu JL et al (2003) Glacial refugia: hotspots but not melting pots of genetic diversity. Science 300(5625):1563–1565

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

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

Pritchard JK, Wen W (2004) Documentation for STRUCTURE software: version 2.0. Department of Human Genetics, University of Chicago, Chicago, IL

Rajora OP, Mosseler A (2001) Challenges and opportunities for conservation of forest genetic resources. Euphytica 118(2):197–212

Ross RM, Bennett RM, Snyder CD et al (2003) Influence of eastern hemlock (*Tsuga canadensis* L.) on fish community structure and function in headwater streams of the Delaware River basin. Ecol Freshw Fish 12(1):60–65

SAS Institute Inc. (2003) The SAS System for Windows, Version 9.1. Cary, NC

Small MJ, Small CJ, Dreyer GD (2005) Changes in a hemlock-dominated forest following woolly adelgid infestation in southern New England. J Torrey Bot Soc 132(3):458–470

Snyder CD, Young JA, Lemarie DP et al (2002) Influence of eastern hemlock (*Tsuga canadensis*) forests on aquatic invertebrate assemblages in headwater streams. Can J Fish Aquat Sci 59(2):262–275

Tingley MW, Orwig DA, Field R (2002) Avian response to removal of a forest dominant: consequences of hemlock woolly adelgid infestations. J Biogeogr 29(10–11):1505–1516

Vekemans X, Hardy OJ (2004) New insights from fine-scale spatial genetic structure analyses in plant populations. Mol Ecol 13(4):921–935

Wang C, Perlin MH, VanStockum RR et al (1997) Chloroplast DNA polymorphisms in Tsuga canadensis and Tsuga caroliniana. Can J For Res 27(9):1329–1335

Watts WA (1970) Full-glacial vegetation of northwestern Georgia. Ecology 51(1):17–33

Wellman H, Ritland C, Ritland K (2003) Genetic effects of domestication in western hemlock, *Tsuga heterophylla*. For Genet 10(5):229–239

Whitehead DR (1973) Late-Wisconsin vegetational changes in unglaciated eastern North America. Quat Res 3:621–631

Whitlock MC, McCauley DE (1999) Indirect measures of gene flow and migration: $F_{ST} \neq 1/(4N_m + 1)$. Heredity 82:117–125

Yeh FC, Yang R-C, Boyle TBJ et al (2000) PopGene32, microsoft windows-based freeware for population genetic analysis, version 1.32. Molecular Biology and Biotechnology Centre, University of Alberta, Edmonton, AB, Canada

Young JA, Young CG (1992) Seeds of woody plants in North America. Dioscorides Press, Portland, OR

Zabinski C (1992) Isozyme variation in eastern hemlock. Can J For Res 22(12):1838–1842