Kubisiak, T.L. and J.H. Roberds. 2005. Genetic structure of American chestnut populations based on neutral DNA markers. In, proc. of conf. on restoration of American chestnut to forest lands, Steiner, K.C. and J.E. Carlson (eds.).

GENETIC STRUCTURE OF AMERICAN CHESTNUT POPULATIONS BASED ON NEUTRAL DNA MARKERS

T. L. Kubisiak and J. H. Roberds
USDA Forest Service, Southern Research Station, Southern Institute of Forest Genetics,
23332 Hwy 67, Saucier, MS 39574 USA (tkubisiak@fs.fed.us)

Abstract: Microsatellite and RAPD markers suggest that American chestnut exists as a highly variable species, even at the margins of its natural range, with a large proportion of its genetic variability occurring within populations (~95%). A statistically significant proportion also exists among populations. Although genetic differentiation among populations has taken place, no disjunct regional pattern of variation exists. A cline in allele frequencies and number of rare alleles occurs along the Appalachian axis, with the highest levels of gene diversity and the greatest numbers of rare alleles being found in southwestern populations. Population pairwise estimates of genetic distance are significantly associated with the geographic distance between populations. Geographically proximate populations are slightly more genetically similar than geographically distant populations. Genetic variability in American chestnut follows a pattern consistent with the hypothesis of a single metapopulation in which genetic drift plays a major evolutionary role. Results of this study are based on neutral genetic loci and do not necessarily reflect genetic differentiation for adaptive genes or gene complexes. Therefore, in order to assure that most of the variation at these genes is also captured in conservation and breeding endeavors, sampling should focus on collecting a fairly large number of individuals from each of several geographic areas.

Keywords: *Castanea dentata* / SSR / RAPD / genotypic diversity / haplotype diversity

INTRODUCTION

The American chestnut (*Castanea dentata* Borkh.) was once one of the most important timber and nut-producing tree species in eastern North America (U.S. Census Bureau 1908). Its native range extended from southern Maine and Ontario in the north to Georgia, Alabama and Mississippi in the south (Sargent 1905). The species now exists primarily as stump sprouts across this entire range, the victim of a devastating canker disease. The disease, chestnut blight, is caused by an exotic fungal pathogen now known systematically as *Cryphonectria parasitica* (Barr 1979). After more than half a century of blight, numerous living stems of American chestnut still exist in the understory of upland forests in the mid-Appalachians (Stephenson et al. 1991). Prolific stump sprouting has enabled American chestnut to persist, but as sexual reproduction is infrequent, its gene pool will likely face serious erosion when old root systems fail to produce sprouts and perish.

Because resistance to *C. parasitica* is low or lacking in American chestnut, Burnham (1981) proposed the use of a classical backcross breeding program to develop blight resistant timber-type trees. Adopting this methodology as their charter, the non-profit philanthropic organization The American Chestnut Foundation (TACF) has since developed a vigorous backcross breeding program designed to introgress the resistance of Chinese chestnut (*C. mollissima* Blume) into American chestnut (Hebard 1994; Kubisiak et al. 1997). TACF's initial efforts focused on American chestnut trees in southwest Virginia, but the goal is to restore the species throughout its entire native range. Thus, information regarding the amount and distribution of molecular genetic variation in American chestnut might help to better determine the number of breeding locations that will be needed across the species range.

Previously, little was known about how genetic variability is distributed across the landscape that comprises the natural range of this species. In an exploratory examination of genetic variability for American chestnut, Huang et al. (1998) obtained results with allozyme and random amplified polymorphic DNA (RAPD) markers that suggest as many as four regional metapopulations might exist. However, hierarchical AMOVA was not performed to quantify this putative regional component, nor were statistical tests employed to test for significant differences. Since that research was completed, the magnitude, significance, and patterns of regional structure have been the subjects of much discussion and debate (F.V. Hebard, P. Sisco, and G. Miller personal communication). Given the importance of regional structure in regards to breeding blight resistant regionally adapted American chestnut, we felt compelled to embark on a more thorough examination of genetic variation in American chestnut using microsatellite and RAPD markers.

Here we report results obtained from an analysis of genetic structure for populations of American chestnut occurring over a significant portion of its natural range. We assayed six microsatellite and 19 RAPD markers and based our analysis on allele and haplotype frequency variation observed for these neutral loci. Our objective for this research was to secure a more detailed and complete understanding of population structure for American chestnut. In the following sections we describe genetic differentiation patterns observed within and among populations and report estimates of diversity parameters associated with microsatellite and RAPD loci segregating in American chestnut. Finally, we compare our results to patterns of variability previously reported for neutral markers in American chestnut as well as in other tree species.

MATERIALS AND METHODS

Population sampling and DNA extraction

A rangewide sampling of expanded leaves or dormant buds of American chestnut were collected at 22 sites across its natural range (refer to Figure 1). Most of the samples were collected from sites in State or National Forests, but a few sites were located on private land holdings. Each sample was assigned a unique ID and sent to the Southern Institute of Forest Genetics in Saucier, Mississippi for DNA extraction and analysis. Total nucleic acids were isolated from tree tissues as described in Kubisiak et al. (1997).

Species evaluation

A panel of DNAs consisting of eight American chestnut (one from each of eight different sites sampled for this study), six Chinese chestnut (trees from USDA import #'s 70315, 104061, 78626, 104014, 104015, and 104016), seven Henry chinkapin (*C. henryi* Rehder & Wils.) (trees from USDA import # 104058, the Nanjing Botanical Garden, Nanjing, Peoples Republic of China (PRC), and the Wuhan Institute of Botany, Wuhan, PRC), four Seguin chestnut (*C. seguinii* Dode) (trees from USDA import # 70317), seven European chestnut (*C. sativa* Mill.) (including trees from the Caucasus Mountains of southern Russia, Bursa, Turkey, and the Black Forest in Germany), and eight Alleghany chinkapin (*C. pumila* Mill.) (Harrison County, Mississippi) were amplified using the polymerase chain reaction (PCR) and a chloroplast-specific primer pair (a, b) as described in Taberlet et al. (1991).

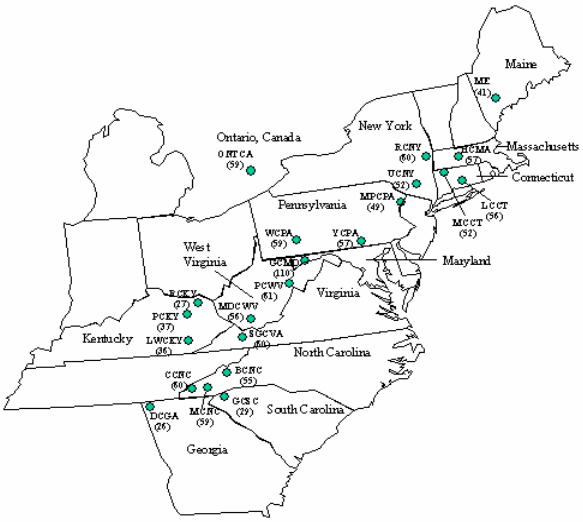

Figure 1. Map of the geographic origin of the 22 *Castanea dentata* Borkh. populations sampled in this investigation. The number in parentheses refers to the number of trees sampled at each location.

Microsatellite PCR amplification and detection

Primer sequences and PCR conditions for microsatellite loci developed in European chestnut (*C. sativa*) were obtained from the literature (Marinoni et al. 2003). Primer sequences for microsatellite loci developed in white oak (*Quercus alba* L.) were obtained from A. David and D. Wagner at the University of Kentucky. For each microsatellite, the forward primer was 5'-end labeled with one of three fluorescent dyes to facilitate detection using the Applied Biosystems 3100 Genetic Analyzer and the GENESCAN® version 3.7 fragment analysis software (Applied Biosystems, Inc. Foster City, CA). Microsatellites were PCR amplified and the products post-PCR multiplexed by color and size whenever possible. Allele sizes were determined by including the GENESCAN®-500[TAMRA] internal size standard in each sample lane. The data were scored using GENOTYPER® version 3.7 (Applied Biosystems, Inc. Foster City, CA).

RAPD PCR amplification and detection

RAPD amplification and detection was based on the protocols reported in Kubisiak et al. (1997). RAPD fragments were identified by the manufacturer primer code corresponding to the primer responsible for their amplification, followed by a subscript four digit number indicating the approximate fragment size in base pairs. Markers were chosen based on the intensity of amplification (only intensely amplified bands were scored) and the absence of co-migrating DNA fragments. All markers were found to conform to Mendelian expectations based on their inheritance in at least one of four different interspecific chestnut pedigrees.

Data analysis

A search for common/redundant multi-locus genotypes and haplotypes was performed (Excoffier and Slatkin 1995). Populations were tested for Hardy-Weinberg proportions using both χ^2 and G^2 tests (Weir 1990). Allele frequencies for each population were computed and estimates obtained for effective number of alleles per locus (A_e), Nei's (1972) measures of gene diversity (h), Nei's (1978) unbiased

measure of genetic distance (D), Michalakis and Excoffier's (1996) genetic differentiation measure (Φ_{ST}) for the microsatellite loci, and Nei's (1987) genetic differentiation measure (G_{ST}) for RAPD loci using the software program ARLEQUIN version 2.001 (Schneider et al. 2000) and POPGENE version 1.31 (Yeh et al. 1997). In addition, χ^2 and G^2 tests were calculated to test homogeneity of allele frequencies among populations. For microsatellite analysis in ARLEQUIN, alleles were coded assuming a step-wise mutation model. Associations between allele frequency and latitude or longitude were first studied using the PROC STEPWISE procedure in SAS version 8.01(SAS, 1999). A variable was only added to the model if its F-statistic was significant at the 5% level. Once added, any variable that did not have a F-value significant at the 5% level was deleted from the model. Associations between the observed number of alleles per locus, number of rare alleles per locus (rare alleles are those with frequencies less than 0.05 computed across all populations), effective number of alleles per locus, gene diversity and latitude or longitude were also studied. In order to further investigate any apparent clinal trends, a composite dependent variable (CDV) was computed that combined both latitude and longitude. First, a reference line was drawn between the southwestern most and northeastern most populations. Then, perpendicular lines were drawn that connected the various populations to this line. Distances (converted into kilometers) along the reference line to the population perpendiculars were used as values for CDV. Genetic distance (D) and among population differentiation were calculated for each pair of populations and associations with geographic distance were investigated using the PROC REG procedure in SAS. Genetic associations existing among populations were first studied using unweighted pair-group mean analysis (UPGMA) based on the matrix of Nei's genetic distance, and then by principal components analysis (PCA) conducted on allele frequency data using the PROC PRINCOMP procedure in SAS.

RESULTS

Putative species identification

Primers that amplified the intergenic spacer region between *trn*T (UGU) and the *trn*L (UAA) 5' exon of the chloroplast genome (primers a and b: 5'-CATTACAAATGCGATGCTCT-3' and 5'-TCTACCGATTTCGCCATATC-3', respectively; Taberlet et al. 1991) were found to uniquely differentiate American chestnut chloroplast DNA from all other *Castanea* (chestnut and chinkapin) species. Based on DNA sequence data (data courtesy F. Dane and P. Lang of Auburn University) this primer pair was found to amplify a band 857 base pairs (bp) in length in American chestnut, and bands ranging from 942 to 945 bp in all other *Castanea* species including the native chinkapin (both *C. pumila* var. *alleghaniensis* and *C. pumila* var. *ozarkensis*). Much of the size difference observed between American chestnut and the other *Castanea* species was due to two unique deletions (one 12 bp and the other 75 bp in length) contained within this region of the American chestnut chloroplast genome. A larger sampling of native chinkapin (specifically *C. pumila*; var. *alleghaniensis* - 48 trees) has yet to show the presence of these large deletions.

Based on the phenotype observed for this marker, of the 1158 chestnut trees sampled for this study 165 (14.2%) were eliminated from further analysis as they did not have the smaller chloroplast band characteristic of American chestnut. These 165 trees were collected from nine different sample sites. Four of the nine sites had very few suspect trees. One site had to be completely eliminated from the study as all trees sampled were found to be suspect. Four sites had to be pooled with the most geographically proximate site in the same state as a large number of suspect trees were found. In total, as many as 993 trees from 18 different sample sites were available for analysis of genetic variation in American chestnut.

Microsatellite-based genetic differentiation

Data describing the microsatellite loci used in our analyses are presented in Table 1. Considerable variation was displayed by the 6 loci. Five of the six loci had very little missing data and were thus used to search for common or redundant multilocus genotypes (MLGs) and haplotypes. Based on these five loci, only five redundant MLGs were observed. Each redundant MLG was only found to occur twice. Based on the same five loci, 114 of 1603 estimated haplotypes were found to occur more than once either within or across populations, but there was no apparent geographic trend to their distribution.

Table 1. Microsatellite and RAPD primer sequence, repeat type, allele size, and number of unique alleles identified in samples collected from 18 populations of *Castanea dentata* Borkh. located throughout the species natural range in eastern North America.

Locus	Primer Sequence 5'-3'	Repeat type	Allele size (bp)	Number of unique alleles
Microsatellites				
*Cs*CAT01[a]	F[b]:AGAATGCCCACTTTTGCA R:CTCCCTTATGGTCTCG	$(AC)_nAT(AC)_n$	167-211	31
*Cs*CAT14	F:GAGGTTGTTGTTCATCATTAC R:ATCTCAAGTCAAAAGGTGTC	$(AC)_n$	121-151	15
*Cs*CAT15	F:TCTGCGACCTCGAAACCGA R:CTAGGGTTTTCATTTCTAG	$(AG)_n$	115-141	15
*Qa*CA022	F:AACAATAGGAGTTGGTTTGAG R:GTTAGGGTTTGGAAAATAGGA	$(AC)_n$	160-188	13
*Qa*GA068	F:GCTTTTCTTTCCAGGGCTAC R:GTGGGACAGTGAGGCAGAG	$(AG)_n$	156-192	17
*Qa*GA209	F:CAAGCAGTATTGTTTTATCTC R:GTTGCCCCTGTGAACTAC	$(AG)_n$	227-265	15
RAPDs				
106	CGTCTGCCCG	NA	500	2
			525	2
			650	2
			700	2
			800	2
184	CAAACGGCAC	NA	450	2
			1150	2
			1800	2
213	CAGCGAACTA	NA	900	2
		NA	1000	2
225	CGACTCACAG	NA	800	2
			1450	2
237	CGACCAGAGC	NA	825	2
			1000	2
			1250	2
423	GGGTCTCGAA	NA	600	2
			875	2
500	TTGCGTCATG	NA	775	2
514	CGGTTAGACG	NA	575	2

[a]Locus names beginning with *Cs* were derived from *Castanea sativa* (Marinoni et al. 2003) and those beginning with *Qa* were derived from *Quercus alba* (sequences courtesy of A. David and D. Wagner). RAPD primer sequences were obtained from J. Hobbs at the University of British Columbia, BC, Canada.
[b]F=forward primer, and R=reverse primer

The expected genotype frequencies at all loci, and in all populations, conformed to Hardy-Weinberg expectations, except for locus *Qa*GA209 in population PCKY that showed a significant excess of homozygotes. Allele frequencies for alleles at greater than 10% frequency over all populations, plus those found to be significantly associated with latitude and/or longitude, are displayed by population in

Table 2. All six single-locus contingency χ^2 analyses as well as G^2 tests for homogeneity of allele frequency across populations indicated significant ($p<0.05$) departures from homogeneity.

Table 2. Microsatellite allele frequencies estimated for 18 populations of *Castanea dentata* Borkh. located throughout the species natural range in eastern North America. Only those alleles at frequencies greater than 0.1 over all populations and those alleles significantly associated with latitude or longitude (identified in italic and ***bold-italic***, respectively) are presented.

Locus and				west								east					
Allele (bp)	CCNC	BCNC	GCSC	PCKY	RCKY	SGCVA	ONTCA	PCWV	GCMD	WCPA	YCPA	MPCPA	UCNY	LCCT	RCNY	HCMA	MCCT	ME
CcCAT01																		
182	*0.0303*	*0.0098*	*0.0517*	*0.0096*	*0.0000*	*0.0424*	*0.0000*	*0.0067*	*0.0000*	*0.0000*	*0.0000*	*0..0000*	*0.0000*	*0.0357*	*0.0000*	*0.0179*	*0.0200*	*0.0000*
186	***0.0455***	***0.0098***	***0.0172***	***0.1058***	***0.1154***	***0.0508***	***0.0172***	***0.0467***	***0.0000***	***0.0179***	***0.0000***	***0.0294***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***
187	***0.0000***	***0.0196***	***0.0000***	***0.0000***	***0.0000***	***0.0339***	***0.0603***	***0.0267***	***0.0693***	***0.0714***	***0.0357***	***0.0980***	***0.1489***	***0.0536***	***0.1379***	***0.2232***	***0.0900***	***0.1125***
190	0.0455	0.1765	0.1034	0.0577	0.0769	0.0593	0.0690	0.0867	0.1832	0.0714	0.0804	0.1275	0.1702	0.2143	0.3448	0.0357	0.1000	0.1500
191	***0.0000***	***0.0000***	***0.0517***	***0.0096***	***0.0000***	***0.0000***	***0.0603***	***0.0133***	***0.0990***	***0.0089***	***0.0179***	***0.0196***	***0.0957***	***0.1429***	***0.1034***	***0.0446***	***0.0500***	***0.0625***
192	***0.1439***	***0.1569***	***0.0862***	***0.1346***	***0.0769***	***0.0593***	***0.0517***	***0.0600***	***0.0545***	***0.0089***	***0.0089***	***0.0000***	***0.1064***	***0.0089***	***0.0690***	***0.0089***	***0.0000***	***0.0125***
194	0.0985	0.1275	0.0690	0.0769	0.2115	0.0763	0.1638	0.1467	0.0693	0.1607	0.3036	0.0980	0.0957	0.2679	0.0776	0.0625	0.0700	0.0875
196	0.3409	0.1373	0.2241	0.2692	0.1538	0.1864	0.1897	0.1933	0.3218	0.3929	0.3929	0.2647	0.2872	0.2411	0.1466	0.1429	0.0800	0.1500
200	***0.0682***	***0.0588***	***0.0345***	***0.0481***	***0.0962***	***0.1102***	***0.0862***	***0.0400***	***0.0050***	***0.0000***	***0.0000***	***0.0490***	***0.0000***	***0.0089***	***0.0259***	***0.0089***	***0.0000***	***0.0375***
202	*0.0076*	*0.0098*	*0.0690*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0067*	*0.0050*	*0.0000*	*0.0268*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*
CsCAT14																		
129	0.1045	0.1373	0.1552	0.1458	0.3077	0.1780	0.2500	0.2403	0.2892	0.2632	0.3661	0.1389	0.3043	0.1455	0.2500	0.1792	0.4314	0.1750
133	0.3284	0.6078	0.4483	0.4792	0.4808	0.5000	0.4310	0.4091	0.4069	0.3158	0.3661	0.3796	0.2147	0.6000	0.5690	0.4717	0.3725	0.5125
143	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0259*	*0.0000*	*0.0196*	*0.0000*	*0.0000*	*0.0000*	*0.0217*	*0.0000*	*0.0086*	*0.0566*	*0.0000*	*0.0125*
145	***0.0075***	***0.0098***	***0.0000***	***0.0104***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***
CsCAT15																		
115	*0.0104*	*0.0185*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*
117	*0.0104*	*0.0926*	*0.1250*	*0.0135*	*0.1538*	*0.0833*	*0.0000*	*0.1346*	*0.0309*	*0.0122*	*0.0167*	*0.0132*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0357*
121	*0.0104*	*0.1111*	*0.0750*	*0.1622*	*0.1538*	*0.0833*	*0.1129*	*0.0385*	*0.2037*	*0.2805*	*0.0333*	*0.1316*	*0.2273*	*0.1383*	*0.3021*	*0.1410*	*0.0500*	*0.1786*
127	*0.0521*	*0.0556*	*0.0500*	*0.0270*	*0.0769*	*0.0333*	*0.0000*	*0.0385*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0758*	*0.0000*	*0.0000*	*0.0128*	*0.0000*	*0.0000*
133	0.2708	0.1111	0.1250	0.1486	0.1154	0.1000	0.2097	0.2500	0.1852	0.1951	0.1333	0.2368	0.1818	0.1064	0.1042	0.2436	0.2000	0.2500
137	*0.0521*	*0.1481*	*0.0000*	*0.0405*	*0.0385*	*0.0500*	*0.1935*	*0.1250*	*0.2037*	*0.0854*	*0.2000*	*0.1316*	*0.1212*	*0.2766*	*0.3646*	*0.3333*	*0.1000*	*0.2500*
139	***0.1042***	***0.0370***	***0.0000***	***0.1081***	***0.0769***	***0.1000***	***0.0806***	***0.1154***	***0.0370***	***0.0854***	***0.0667***	***0.1184***	***0.0152***	***0.0319***	***0.0000***	***0.0128***	***0.0125***	***0.0357***
QaCA022																		
170	0.4141	0.6058	0.4655	0.2885	0.2500	0.2500	0.2586	0.2123	0.5727	0.5603	0.3482	0.3627	0.4688	0.4286	0.4741	0.2411	0.2000	0.2683
172	***0.1797***	***0.2308***	***0.2414***	***0.2596***	***0.1538***	***0.2679***	***0.1810***	***0.2808***	***0.1591***	***0.1207***	***0.1964***	***0.3824***	***0.3125***	***0.3661***	***0.3362***	***0.3036***	***0.5100***	***0.3049***
174	*0.0312*	*0.0192*	*0.0345*	*0.0192*	*0.0385*	*0.0446*	*0.0086*	*0.0137*	*0.0000*	*0.0086*	*0.0089*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0122*
182	0.0234	0.0288	0.0172	0.0769	0.0962	0.2232	0.2672	0.1438	0.0955	0.0517	0.1786	0.0686	0.0312	0.1071	0.0862	0.0893	0.1000	0.1463
QaGA068																		
164	*0.0758*	*0.0000*	*0.1379*	*0.0943*	*0.0400*	*0.0000*	*0.0000*	*0.0065*	*0.0140*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0122*
166	0.2955	0.1250	0.0517	0.0943	0.0400	0.0455	0.0259	0.0649	0.0374	0.0882	0.0818	0.0283	0.1979	0.2455	0.1810	0.1071	0.0769	0.0732
168	0.1515	0.0865	0.1379	0.2642	0.4400	0.0273	0.0345	0.2857	0.1869	0.1176	0.0545	0.0755	0.2083	0.1727	0.3190	0.3214	0.1250	0.1220
170	0.2273	0.2019	0.1034	0.2358	0.1800	0.2455	0.1207	0.1753	0.1589	0.0098	0.2091	0.3774	0.1771	0.0818	0.1724	0.2946	0.0673	0.3537
172	0.0530	0.1250	0.2069	0.0943	0.0200	0.3455	0.2931	0.3312	0.3224	0.3824	0.3909	0.2547	0.1458	0.0909	0.0603	0.1339	0.0577	0.2073
174	0.0909	0.0962	0.0517	0.0566	0.1600	0.1182	0.0086	0.0325	0.0935	0.0490	0.1636	0.2453	0.0938	0.2636	0.0690	0.0714	0.2115	0.0732
180	***0.0076***	***0.0096***	***0.0345***	***0.0943***	***0.0400***	***0.0273***	***0.0086***	***0.0260***	***0.0000***	***0.0098***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0000***	***0.0179***	***0.0000***	***0.0000***
QaGA209																		
233	*0.0682*	*0.0510*	*0.0517*	*0.0648*	*0.1000*	*0.0424*	*0.2368*	*0.0395*	*0.0490*	*0.0259*	*0.1071*	*0.1951*	*0.1383*	*0.1071*	*0.0741*	*0.2411*	*0.0119*	*0.1098*
235	***0.0152***	***0.1633***	***0.1897***	***0.2870***	***0.2000***	***0.1271***	***0.2193***	***0.2171***	***0.1961***	***0.1293***	***0.0446***	***0.0976***	***0.0745***	***0.1071***	***0.0556***	***0.0625***	***0.1667***	***0.0488***
241	0.4545	0.3161	0.3276	0.4722	0.4800	0.2288	0.3421	0.2961	0.5784	0.5517	0.6786	0.3293	0.4362	0.3482	0.3889	0.2411	0.6190	0.4390
243	*0.1439*	*0.0204*	*0.0690*	*0.0463*	*0.0400*	*0.1949*	*0.0439*	*0.1447*	*0.0490*	*0.0086*	*0.0625*	*0.0122*	*0.0000*	*0.0000*	*0.0000*	*0.0357*	*0.0000*	*0.0244*
249	***0.0152***	***0.1429***	***0.0345***	***0.0000***	***0.0000***	***0.1441***	***0.0702***	***0.0855***	***0.0539***	***0.1293***	***0.0179***	***0.2683***	***0.1064***	***0.0804***	***0.1111***	***0.2589***	***0.1667***	***0.1098***
251	***0.0303***	***0.0000***	***0.0000***	***0.0000***	***0.0200***	***0.0085***	***0.0000***	***0.0132***	***0.0392***	***0.0086***	***0.0000***	***0.0000***	***0.1277***	***0.1607***	***0.1296***	***0.0357***	***0.0000***	***0.0244***
255	*0.0303*	*0.0204*	*0.0862*	*0.0000*	*0.0400*	*0.0339*	*0.0000*	*0.0329*	*0.0147*	*0.0000*	*0.0357*	*0.0244*	*0.0000*	*0.0179*	*0.0093*	*0.0000*	*0.0000*	*0.0000*

Differentiation statistics computed over all populations are shown in Table 3. All single-locus estimates of among population differentiation (Φ_{ST}) were found to be significantly different from random expectations. Based on stepwise regression analysis, at least one allele at all six of the microsatellite loci were found to be significantly ($p<0.05$) associated with latitude or longitude (see markers in *italic* and ***bold italic***, respectively in Table 2). A visual inspection of allele frequencies across the sample sites shows a northeast-southwest trend. Allele frequencies tend to be either low in the northeast and high in the southwest, or vice versa. Due to this apparent trend, we again performed regression analysis using the composite dependent variable (CDV) that combined both latitude and longitude. Results of these regression analyses are displayed in Table 4. Frequencies of alleles at all six loci were found to significantly vary with the value for CDV. An example of these changes are illustrated in Figure 2 for two alleles, one with allele frequency increasing with CDV distance and the other with allele frequency decreasing. The number of rare alleles per locus was also found to be significantly associated with the CDV for three of the six microsatellite loci (Table 4). A visual inspection of the number of rare alleles across sample sites again shows a northeast-southwest trend, with higher numbers of rare alleles being harbored in southwestern populations. Several loci were also found to be significantly associated with CDV based on the number of unique alleles, effective number of alleles, and gene diversity (Table 4). As a general trend, there appears to be slightly more alleles and hence more effective numbers of alleles and

slightly higher levels of gene diversity in southwest populations than in those located in the northeast (Figure 2).

Table 3. Summary of genetic diversity descriptive statistics for six microsatellite loci segregating in 18 populations of *Castanea dentata* Borkh. located throughout the species natural range in eastern North America.

Locus	Sample Size	n_a[a]	n_e	h	h_o	Φ_{ST}	Nm
*Cs*CAT01	1974	31	9.222	0.892	0.844	0.097	4.655
*Cs*CAT14	1974	15	3.779	0.735	0.710	0.029	16.741
*Cs*CAT15	1336	15	8.519	0.883	1.000[c]	0.032	15.125
*Qa*CA022	1998	13	4.198	0.762	0.730	0.046	10.370
*Qa*GA068	1982	17	7.144	0.860	0.786	0.030	16.167
*Qa*GA209	1936	15	4.456	0.776	0.705	0.034	14.206
Mean	1870	17.667	6.220	0.818	0.755[d]	0.048[b]	12.877
St. Dev		6.653	2.379	0.068	0.059		

[a] n_a = observed number of alleles, n_e = effective number of alleles, and h = Nei's (1978) gene diversity, h_o = observed heterozygosity, Φ_{ST} = Michalakis and Excoffier's (1996) measure of among population differentiation, and Nm = number of migrants exchanged between populations per generation
[b] Mean Φ_{ST} was estimated by summing variance components across loci
[c] observed heterozygosity for this locus was equal to one as the second allele for all trees amplifying only one apparent microsatellite allele was scored as unknown or missing data
[d] Mean and St. Dev. do not include h_o for locus *Cs*CAT15

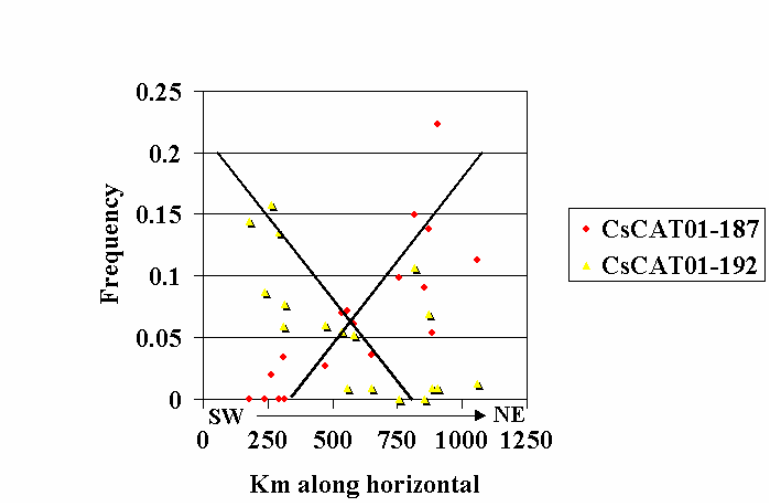

Figure 2. Plot of allele frequency by composite dependent variable (CDV) expressed in units of kilometers along horizontal.

Table 4. Summary of regression analyses for significant associations (Pr>F<0.05) between allele frequency, number of rare alleles, observed number of alleles, effective number of alleles, and gene diversity and a composite dependent variable (CDV) expressed in units of kilometers.

Frequency		Regression		
Locus	Allele (bp)	equation	R^2	Pr>F
*Cs*CAT01	186	Y=0.07264-0.00005024*CDV	0.391	0.0055
*Cs*CAT01	187	Y=-0.03971+0.00011178*CDV	0.659	<0.0001
*Cs*CAT01	191	Y=-0.01470+0.00006155*CDV	0.397	0.0051
*Cs*CAT01	192	Y=0.13460-0.00008108*CDV	0.490	0.0012
*Cs*CAT01	200	Y=0.08059-0.00004559*CDV	0.330	0.0127
*Cs*CAT14	145	Y=0.00572-0.00000444*CDV	0.304	0.0177
*Cs*CAT15	117	Y=0.10464-0.00006849*CDV	0.335	0.0118
*Cs*CAT15	127	Y=0.05820-0.00003688*CDV	0.337	0.0115
*Cs*CAT15	137	Y=-0.01573+ 0.00017672*CDV	0.559	0.0004
CsCAT15	139	$Y=0.08369-0.00000003*CDV^2$	0.248	0.0353
*Qa*CA022	172	$Y=0.19220+0.00000007*CDV^2$	0.371	0.0073
*Qa*CA022	174	$Y=0.06282-0.00009197*CDV+0.00000003*CDV^2$	0.744	<0.0001
*Qa*GA068	164	$Y=0.14814-0.00025546*CDV+0.00000011*CDV^2$	0.523	0.0039
*Qa*GA068	180	Y=0.04318-0.00002958*CDV	0.309	0.0166
*Qa*GA209	235	$Y=0.18213-0.00000005*CDV^2$	0.267	0.0282
*Qa*GA209	243	Y=0.12080-0.00007540*CDV	0.352	0.0095
*Qa*GA209	249	Y=0.01543+0.00008946*CDV	0.257	0.0318
*Qa*GA209	251	$Y=0.00268+0.00000003*CDV^2$	0.231	0.0436
*Qa*GA209	255	Y=0.04579-0.00002821*CDV	0.317	0.0150
106	525	Y=0.77195-0.00011743*CDV	0.426	0.0045
225	800	Y=-0.07714-0.00033161*CDV	0.494	0.0017
237	1000	Y=0.85531-0.00008368*CDV	0.251	0.0405
237	1250	$Y=0.40626-0.00086307*CDV+0.00000037*CDV^2$	0.624	0.0011
Number of Rare Alleles[a]				
Locus		Regression equation	R^2	Pr>F
*Cs*CAT01		Y=14.18493-0.00490*CDV	0.289	0.0213
*Cs*CAT15		Y=5.13625-0.00331*CDV	0.515	0.0008
*Qa*GA068		$Y=8.91562-0.01293*CDV+0.00000513*CDV^2$	0.615	0.0008
All loci		$Y=58.40985-0.07272*CDV+0.00003079*CDV^2$	0.593	0.0012
Observed Number of Alleles[b]				
Locus		Regression equation	R^2	Pr>F
*Cs*CAT01		$Y=0.65276-0.00079907*CDV+0.00000036*CDV^2$	0.548	0.0026
*Cs*CAT15		$Y=0.44260-0.00050452*CDV+0.00000022*CDV^2$	0.512	0.0046
*Qa*CA068		$Y=0.47837-0.00060844*CDV+0.00000027*CDV^2$	0.517	0.0043
Effective Number of Alleles				
Locus		Regression equation	R^2	Pr>F
CsCAT15		Y=8.79012-0.00212*CDV	0.525	0.0007
Gene Diversity				
Locus		Regression equation	R^2	Pr>F
CsCAT15		$Y=0.87519-0.00000003*CDV^2$	0.463	0.0019

[a]number of rare alleles = number of rare alleles in population/number of individuals in population
[b]observed number of alleles = number of observed alleles in population/number of individuals in population

Estimates of genetic distance (D) between pairwise comparisons of populations based on all six loci varied from a low of 0.062 to a high of 0.372, averaging 0.206. Similarly computed pairwise identity estimates ranged from 0.689 to 0.940, yielding a mean of 0.814. Pairwise estimates of genetic distance were significantly (p=0.0011) associated with the geographic distance between paired populations. However, only a small proportion of the variation found among populations was explained by this dependent variable (R^2=0.069). Estimates of genetic differentiation (Φ_{ST}) between pairwise comparisons of populations varied from a low of -0.003, to a high of 0.156, and averaged 0.048 across loci. These estimates were not significantly associated with geographic distance between the paired populations. Thus, populations in close geographic proximity tend to have slightly higher genetic identities than those

more geographically distant. Single-locus, as well as multi-locus, UPGMA based on genetic distance and PCA based on allele frequencies computed over all sample sites did not reveal patterns of differentiation consistent with regional structure. Geographically proximate sample sites did not group together, and group membership varied from locus to locus.

RAPD-based genetic differentiation

Data describing the RAPD loci used in our analyses are presented in Table 1. In all populations studied, genotypic frequencies observed for microsatellite loci did not significantly deviate from Hardy-Weinberg expectations. Assuming then that the RAPD loci we investigated also have genotypes distributed in Hardy-Weinberg proportions, we can estimate their allele frequencies from observed frequencies for the homozygous null genotypes. Allele frequencies estimated using this approach are displayed by population in Table 5. Sixteen of 19 single-locus contingency χ^2 and G^2 tests for heterogeneity of allele frequencies across populations were found to be significant ($p<0.05$).

Table 5. Band-present RAPD allele frequencies for 19 loci assayed from samples collected in 17 populations of *Castanea dentata* Borkh. located throughout the species natural range in eastern North America. Alleles significantly associated with latitude or longitude are identified in italic and ***bold italic***, respectively.

Locus	CCNC	BCNC	GCSC	PCKY	RCKY	SGCVA	ONTCA	PCWV	GCMD	WCPA	YCPA	MPCPA	UCNY	RCNY	HCMA	MCCT	ME
106_{0500}	0.1308	0.1722	0.0211	0.0658	0.0426	0.1056	0.1762	0.2929	0.1728	0.0839	0.0887	0.0780	0.2421	0.1220	0.1982	0.0364	0.0917
106_{0525}	***0.3084***	***0.1611***	***0.1340***	***0.2138***	***0.1584***	***0.1244***	***0.1136***	***0.1921***	***0.0547***	***0.1036***	***0.0887***	***0.0780***	***0.0324***	***0.1835***	***0.0000***	***0.0000***	***0.1056***
106_{0650}	0.7446	0.6400	0.8000	0.8652	0.5918	1.0000	0.8093	0.8830	0.6181	0.6220	0.8159	0.8419	0.7083	0.7980	0.8110	0.7327	0.7261
106_{0700}	0.0000	0.0474	0.1835	0.0187	0.0646	0.0513	0.0090	0.0138	0.0000	0.0009	0.0087	0.0382	0.0435	0.0000	0.0272	0.0364	0.0126
106_{0800}	0.7051	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.8821	1.0000	0.6727	0.8143	0.8419	1.0000	0.7474	1.0000	1.0000	1.0000
184_{0450}	*0.0221*	*0.0973*	*0.0000*	*0.0090*	*0.0000*	*0.0433*	*0.0000*	*0.0344*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0087*	*0.0000*	*0.0126*
184_{1150}	0.0435	0.0973	0.1340	0.0646	0.1056	0.0980	0.0770	0.0859	0.0786	0.0299	0.0691	0.0000	0.0000	0.2175	0.0087	0.0000	0.0126
184_{1800}	0.1495	0.3061	0.1340	0.1443	0.1835	0.3622	0.2230	0.2289	0.3064	0.1034	0.2042	0.2929	0.3386	0.0632	0.1982	0.3453	0.3481
213_{0900}	***0.2421***	***0.1368***	***0.2929***	***0.1798***	***0.0911***	***0.2254***	***0.1972***	***0.1835***	***0.1815***	***0.1679***	***0.0728***	***0.0968***	***0.2421***	***0.3353***	***0.6526***	***1.0000***	***0.2745***
213_{1000}	0.3477	0.4059	0.4084	0.4606	0.4599	0.2362	0.2112	0.3280	0.3660	0.2279	0.2867	0.5412	0.4947	0.2462	0.2689	1.0000	0.3206
225_{0800}	***0.7446***	***1.0000***	***1.0000***	***1.0000***	***1.0000***	***1.0000***	***1.0000***	***0.8821***	***0.7621***	***0.6331***	***0.4702***	***0.5918***	***0.6703***	***0.4161***	***0.5044***	***0.8333***	***0.6508***
225_{1450}	0.0022	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0897	0.0000
237_{0825}	0.3386	0.2672	0.2546	0.4117	0.2517	0.2362	0.1310	0.1780	0.2327	0.2279	0.1943	0.4689	0.1340	0.0426	0.3841	0.1590	0.2510
237_{1000}	***0.0871***	***0.0093***	***0.1416***	***0.1611***	***0.2789***	***0.0084***	***0.0417***	***0.0921***	***0.0483***	***0.1120***	***0.0177***	***0.0129***	***0.0871***	***0.0105***	***0.0262***	***0.0000***	***0.0247***
237_{1250}	*0.3084*	*0.3914*	*0.3511*	*0.2302*	*0.2517*	*0.1633*	*0.0632*	*0.3175*	*0.1035*	*0.1914*	*0.0823*	*0.0801*	*0.0211*	*0.1248*	*0.1906*	*0.0382*	*0.2510*
423_{0600}	0.1403	0.1835	0.0426	0.1158	0.0835	0.1633	0.1313	0.2536	0.2341	0.2352	0.2277	0.0267	0.1882	0.3773	0.0903	0.2632	0.1377
423_{0875}	0.0109	0.0392	0.0675	0.0180	0.0408	0.0426	0.0190	0.0598	0.0652	0.0801	0.0267	0.0823	0.0114	0.0742	0.0267	0.0000	0.0392
500_{0775}	0.0000	0.0000	0.0000	0.0000	0.0000	0.0084	0.0000	0.0066	0.0246	0.0000	0.0000	0.0000	0.0572	0.0000	0.0457	0.0000	0.0123
514_{0575}	0.1972	0.3412	0.3406	0.2799	0.1282	0.1734	0.2459	0.2494	0.3614	0.3140	0.2277	0.1029	0.0585	0.5337	0.6026	0.2735	0.3753

Differentiation statistics computed over all populations are presented in Table 6. Estimates of among population differentiation (G_{ST}) were found to be significantly greater than zero at 14 of the 19 loci. Based on our stepwise regression analysis, allele frequencies at six of the 19 RAPD loci were significantly ($p<0.05$) associated with latitude or longitude (markers in *italic* and ***bold-italic***, respectively in Table 5). As was observed for the microsatellite loci, a visual inspection of allele frequencies across the sample sites showed a northeast-southwest trend. Again, we performed regression analysis using the CDV. Four loci were found to be significantly associated with the CDV (Table 4). At all four loci, band-present allele frequencies were higher in southwest populations than in those from the northeast.

Estimates of genetic distance (D) between pairwise combinations of populations computed across loci varied from a low of 0.003, to a high of 0.144, with a mean value of 0.037. Similarly computed pairwise identity estimates ranged from 0.866 to 0.997, with a mean of 0.964. Unlike the microsatellite data, pairwise estimates of genetic distance were not significantly ($p=0.0571$) associated with the geographic distance between paired populations and neither were pairwise estimates of genetic differentiation. As was the case for the microsatellite loci, single-locus or multi-locus UPGMA computed from RAPD genetic distances, or PCA based on RAPD allele frequencies, did not reveal differentiation patterns suggestive of regional structure.

Table 6. Summary of genetic diversity descriptive statistics for 19 RAPD loci assayed from samples collected in 17 populations of *Castanea dentata* Borkh. located throughout the species natural range in eastern North America.

Locus	Sample Size	n_e	h	G_{ST}	Nm
106_{0500}	845	1.321	0.243	0.049	9.802
106_{0525}	845	1.273	0.214	0.056	8.443
106_{0650}	849	1.552	0.356	0.062	7.628
106_{0700}	844	1.051	0.049	0.056	8.471
106_{0800}	843	1.178	0.151	0.184	2.222
184_{0450}	883	1.030	0.029	0.046	10.364
184_{1150}	881	1.142	0.124	0.050	9.413
184_{1800}	878	1.567	0.362	0.047	10.080
213_{0900}	794	1.535	0.348	0.067	6.914
213_{1000}	801	1.818	0.450	-0.006	2000.0
225_{0800}	808	1.598	0.374	-0.008	2000.0
225_{1450}	810	1.010	0.010	-0.336	2000.0
237_{0825}	871	1.578	0.366	0.060	7.863
237_{1000}	873	1.126	0.112	0.081	5.709
237_{1250}	869	1.416	0.294	0.081	5.642
423_{0600}	861	1.426	0.299	0.053	8.889
423_{0875}	858	1.090	0.082	0.016	30.536
500_{0775}	870	1.021	0.021	0.031	15.406
514_{0575}	858	1.700	0.414	0.093	4.899
Mean	850	1.339	0.226	0.036	9.517[b]
St. Dev		0.258	0.148		

[a] n_e = effective number of alleles, and h = Nei's (1978) gene diversity, G_{ST} = Nei's (1987) measure of among population differentiation, and Nm = number of migrants exchanged between populations per generation
[b] Mean excludes estimates for loci 213_{1000}, 225_{0800}, and 225_{145}

DISCUSSION

One of our main concerns in this investigation was inclusion of trees that are not pure American chestnut. Inappropriate trees include interspecific hybrids or pure species other than American chestnut, especially the native congener species chinkapin (*Castanea pumila*). Inclusion of such contaminants could have inflated our estimates of genetic diversity, especially in populations containing the non-American chestnut samples, as well as clouded true patterns of genetic variability. Chloroplast DNA sequence variations have been widely used to investigate interspecific relationships among plant species (Palmer et al. 1988, Clegg et al. 1991) because they evolve slowly. We identified a chloroplast-specific marker (primers a and b; Taberlet et al. 1991) that quickly differentiates American chestnut chloroplast DNA from all other *Castanea* species, including the native *C. pumila*. Unfortunately, maternal inheritance of chloroplasts precludes our ability to distinguish interspecific hybrids of maternal American chestnut origin. As a result, our sample set might still contain some interspecific hybrids, however, the number should be small as most collections were made in either State Forests or National Forests where non-native *Castanea* species do not extensively occur.

Our results demonstrate that high levels of microsatellite and RAPD variability exist in American chestnut, and that most of this variation occurs within local populations (95.2% and 96.4%, respectively). These results are comparable to observations made in other long-lived, outcrossing, woody plant species (Hamrick and Godt 1990; Hamrick et al. 1992), where as a rule, greater than 90% of the variation occurs within populations. Our results are also consistent with previous observations of allozyme variability in

C. sativa and American chestnut, where 90% of the diversity was reported to exist within populations (Pigliucci et al. 1990; Huang et al. 1998). Whereas only scant evidence for a cline in allele frequency variation (alleles at 1 of 14 polymorphic allozyme loci) was previously reported for American chestnut (Huang et al. 1998), our results clearly demonstrate that a cline in allele frequencies and number of rare alleles exists along the Appalachian axis. Clinal variation of allele frequencies along latitudinal and longitudinal gradients has been reported for a number of tree species (Lagercrantz and Ryman 1990; Zanetto and Kremer; Leonardi and Menozzi 1995, Tomaru et al. 1997), including *C. sativa* (Pigliucci et al. 1990; Villani et al. 1991; Villani et al. 1992; Villani et al. 1994). The main proposition set forth to explain this phenomenon is that geographical variation in allele frequencies resulted from post-glacial migration and founding events. Such processes are consistent with the patterns of variability we observed for American chestnut. The highest levels of gene diversity and the greatest numbers of rare alleles are found in the southwestern portion of its range. This suggests that its glacial refugium existed in the southeastern U.S., perhaps extending southward into the Gulf Coastal plain of present day Mississippi and Alabama. As a general finding, American chestnut still exists as a highly variable species, even at the margins of its natural range, with a large proportion of its genetic variability occurring within populations. Furthermore, existence of the clinal pattern of variation implies that extensive gene flow took place among populations before the spread of chestnut blight.

Although most of the genetic variation found in American chestnut occurs within local populations, a statistically significant proportion exists among populations. Magnitudes of the Φ_{ST} and G_{ST} estimates obtained in our investigation are slightly lower than those reported for American chestnut by Huang et al. (1998). In this research we used a chloroplast-specific marker to identify trees that were not pure American chestnut and excluded these individuals. However, Huang et al. (1998) did not take precautionary measures to identify aberrant specimens. Inclusion of such individuals in some samples will tend to inflate levels of among population differentiation. Although our estimates of among population differentiation might be considered low, Φ_{ST} values obtained for all six microsatellite loci and G_{ST} values obtained for 14 of the 19 RAPD loci studied indicate that populations significantly differ in allele frequency. Moreover, population pairwise estimates of genetic distance, based on microsatellite haplotype frequencies, were shown to be significantly associated with the geographic distance between populations. Thus we conclude that geographically proximate populations are slightly more genetically similar than geographically distant populations. These findings lead us to conclude that although long distance gene flow was possible in the past, it was infrequent enough to allow genetic differentiation to take place.

From UPGMA and PCA analyses, it is evident that regional differentiation did not occur in American chestnut. Geographically proximal populations did not group together, and group make-up differed across loci. In contrast, Huang et al. (1998) concluded that a somewhat weak and incomplete pattern of regional differentiation exists, based largely on latitudinal differences. Although the results obtained by UPGMA and PCA of the allozyme data were interpreted as being suggestive of regional structure, hypothetical regional effects were not quantified by means of a hierarchical AMOVA, nor were statistical tests employed to detect differences. Because of our more comprehensive sampling of the natural range (18 populations versus 12), larger sample sizes collected (average 55 trees per population versus 22 trees), and elimination of suspect samples (i.e. trees that did not have the characteristic American chestnut chloroplast haplotype), we believe the results obtained in this investigation represent a more accurate picture of population structure in American chestnut.

Our findings clearly demonstrate that American chestnut still exists as a highly variable species throughout its entire native range. In spite of this high variability, we must point out that the results of this study represent variability existing in the pre-blighted forest, and caution that unless measures are taken to restore American chestnut and enhance opportunities for it to sexually reproduce, this species will likely face serious erosion of its gene pool as root systems fail to produce sprouts and die. Along

these lines, results of this study can be used as a baseline in the future for assessing the degree and rapidity of such a decline.

Taking into account the differentiation observed at these loci, no disjunct regional pattern of variation exists. Prior to introduction of the blight, genetic variability in American chestnut followed a pattern consistent with the hypothesis of a single metapopulation where genetic drift played a major evolutionary role. Currently, approximately 95% of the neutral genetic variation of the species can be captured by sampling within any one population. However, the results of this study are based on neutral genetic loci and do not necessarily reflect genetic differentiation for adaptive genes or gene complexes. Therefore, in order to assure that most of the variation produced by these genes is also captured in conservation and breeding endeavors, sampling should focus on collecting a fairly large number of individuals (50 to 100 or more) from each of several geographic areas.

LITERATURE CITED

Barr, M.E. 1979. The diaporthales in North America with emphasis on gnomonia and its segregants. Mycologia Memoir 7, J. Cramer, Lehre, Germany.

Burnham, C.R. 1981. Blight-resistant American chestnut: There's hope. Plant Dis. 65:459-460.

Clapper, R.B. 1954. Chestnut breeding techniques and results. II. Inheritance of characters, breeding for vigor, and mutations. J. Hered. 45:201-208.

Clegg, M.T., G.H. Learn, and E.M. Goldberg. 1991. Molecular evolution of chloroplast DNA. P. 135-149. in Evolution at the molecular level, Selander, R.K., A.G. Clark, and T.S. Whittam (eds.) Sinauer Associates Inc., Sunderland, MA.

Excoffier, L., and M. Slatkin. 1995. Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. Mol. Biol. Evol. 12:921-927.

Hamrick, J.L., and M.J. Godt. 1990. Allozyme diversity in plant species. P. 43-63. in Plant population genetics, breeding, and genetic resources, Brown, H.D., M.T. Clegg, A.L. Kahler, and B. Weir (eds.). Sinauer Associates Inc., Sunderland, MA.

Hamrick, J.L., M.J. Godt, and S.L. Sherman-Broyles. 1992. Factors influencing levels of genetic diversity in woody plant species. P. 95-124. in Population genetics of forest trees. Adams, W.T., S.H. Strauss, D.L. Copes, and A.R. Griffin (eds.). Kluwer Academic Publishers, London.

Hebard, F.V. 1994. The American Chestnut Foundation breeding plan: beginning and intermediate steps. J. Am. Chestnut Found. 8(1):21-28.

Huang, H., F. Dane, and T.L. Kubisiak. 1998. Allozyme and RAPD analysis of the genetic diversity and geographic variation in wild populations of the American chestnut (Fagaceae). Am. J. Bot. 85:1013-1021.

Kubisiak, T.L., F.V. Hebard, C.D. Nelson, J. Zhang, R. Bernatzky, H. Huang, S.L. Anagnostakis, and R.L. Doudrick. 1997. Molecular mapping of resistance to blight in an interspecific cross in the genus *Castanea*. Phytopathology 87:751-759.

Lagercrantz, U., and N. Ryman. 1990. Genetic structure of Norway spruce (*Picea abies*): concordance of morphological and allozyme variation. Evolution 44:38-53.

Leonardi, S., and P. Menozzi. 1995. Genetic variability of *Fagus sylvatica* L. in Italy: the role of postglacial recolonization. Heredity 75:35-44.

Marinoni, D., A. Akkak, G. Bounous, K. Edwards, and R. Botta. 2003. Development and characterization of microsatellite markers in *Castanea sativa* (Mill.). Mol. Breed. 11(2):127-136.

Michalakis, Y., and L. Excoffier. 1996. A generic estimation of population subdivision using distances between alleles with special reference for microsatellite loci. Genetics 142:1061-1064.

Nei, M. 1972. Genetic distance between populations. Am. Natur. 106:283-292.

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

Nei, M. 1987. Analysis of gene diversity in subdivided populations. P. 187-192. in Molecular Evolutionary Genetics. Columbia University Press, New York.

Palmer, J.D., R.K. Jansen, H.J. Michaels, M.W. Chase, and J.R. Manbart. 1988. Chloroplast DNA variation and plant phylogeny. Ann. Miss. Bot. Gar. 75:1180-1206.

Pigliucci, M., S. Benedettelli, and F. Villani. 1990. Spatial patterns of genetic variability in Italian chestnut (*Castanea sativa*). Can. J. Bot. 68:1962-1967.

Sargent, C.S. 1905. Manual of trees of North America. Houghton Mifflin Co., Boston and New York. p. 220-222.

SAS Institute Inc. 1999. SAS/STAT Users Guide, Version 8. SAS Institute, Inc., Cary, NC. 3884 p.

Schneider, S., D. Roessli, and L. Excoffier. 2000. A software for population genetics data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland.

Stephenson, S.L., H.S. Adams, and M.L. Lipford. 1991. The present distribution of chestnut in the upland forest communities of Virginia. Bull. Torrey Bot. Club 118:24-32.

Taberlet, P., L. Gielly, G. Patou, and J. Bouvet. 1991. Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Mol. Biol. 17:1105-1109.

Tomaru, N., T. Mitsutsuji, M. Takahashi, Y. Tsumura, K. Uchida, and K. Ohba. 1997. Genetic diversity in *Fagus crenata* (Japanese beech): influence of the distributional shift during the late-Quaternary. Heredity 78:241-251.

U.S. Census Bureau. 1908. The lumber cut of the United States, 1907. For. Products 2:1-53.

Villani, F., M. Pigliucci, S. Benedettelli, and M. Cherubini. 1991 Genetic differentiation among Turkish chestnut (*Castanea sativa* Mill.) populations. Heredity 66:131-136.

Villani, F., M. Pigliucci, and M. Cherubini. 1994. Evolution of *Castanea sativa* Mill. in Turkey and Europe. Genetic Research, Cambridge. 63:109-166.

Villani, F., M. Pigliucci, M. Lauteri, and M. Cherubini. 1992. Congruence between genetic, morphometric, and physiological data on differentiation of Turkish chestnut (*Castanea sativa*). Genome 35:251-256.

Weir, B.S. 1990. Methods for discrete population genetic data. P. 110-113 in Genetic data analysis. Sinauer Associates Inc., Sunderland, Mass.

Yeh, F.C., R-C. Yang, T.J. Boyle, Z-H. Ye, and J.X. Mao. 1997. POPGENE, the user-friendly shareware for population genetic analysis. Molecular Biology and Biotechnology Centre, University of Alberta, Canada.

Zanetto, A., and A. Kremer. 1995. Geographical structure of gene diversity in *Quercus petraea* (Matt.) Liebl. I. Monolocus patterns of variation. Heredity 75:506-517.

ACKNOWLEDGEMENTS

The authors thank Glen Johnson and Charles Burdine for their skilled technical assistance with DNA extraction, PCR amplification, microsatellite marker detection, and data acquisition; Andy David and Dave Wagner for the white oak microsatellite primer sequences; Gabriele Baccaro and Roberto Botta for releasing the European chestnut microsatellite primer sequences prior to publication of their manuscript; and Sandra Anagnostakis; Dave Armstrong, Glen Beaver, Robert Bernatzky; Mary Bunch, Peter Carson, Hill Craddock, Mark Double, Fred Hebard, Craig Hibben, E. Kenneth James, Michael Kluempke, Jeff Lewis, Paul Sisco, Bob Summersgill, Wayne Swank, Melissa Thomas-VanGundy, Wells Thurber, Cathy Townsend, Stan Webb and Eric Weisse for their assistance in collecting chestnut samples for this study.