NRC Research Press

ARTICLE

Genetic variation patterns of American chestnut populations at EST-SSRs

Oliver Gailing and C. Dana Nelson

Abstract: The objective of this study is to analyze patterns of genetic variation at genic expressed sequence tag – simple sequence repeats (EST-SSRs) and at chloroplast DNA markers in populations of American chestnut (*Castanea dentata* Borkh.) to assist in conservation and breeding efforts. Allelic diversity at EST-SSRs decreased significantly from southwest to northeast along the Appalachian range, suggesting repeated founder events during postglacial migration. Comparatively high allelic diversity in Ontario, northwest of the Appalachian range, suggested more recent long-distance dispersal. Clinal variation of allele frequencies along the Appalachian axis was also in accordance with postglacial colonization from one refugium southwest of the Appalachian range. We observed clustering of the northwestern population from Ontario with southwestern populations and sharing of a rare chloroplast haplotype among western populations across the whole latitudinal range. This pattern is consistent with a divergence of postglacial migration routes and higher levels of more recent potentially human-mediated gene exchange between populations west of the Appalachian range. Population pairs east and west of the Appalachian axis showed pronounced allele frequency differences over a small geographic range. These patterns of genetic variation should be considered when sampling reproductive material for conservation and breeding.

Key words: EST-SSRs, *Castanea dentata*, microsatellites, clinal variation, conservation.

Résumé : Cette étude avait pour objectif d'analyser les patrons de variation génétique des EST-SSR géniques et des marqueurs de l'ADN chloroplastique des populations de châtaigniers d'Amérique (*Castanea dentata* Borkh.) afin de soutenir les efforts de conservation et de reproduction. La diversité allélique des EST-SSR diminuait significativement de sud-ouest au nord-est le long de la chaîne des Appalaches, suggérant l'existence d'événements fondateurs répétés durant la migration postglaciaire. Une diversité allélique comparativement élevée dans la chaîne des Appalaches du nord-ouest de l'Ontario suggérait une dispersion sur de longues distances plus récente. La variation clinale des fréquences alléliques le long de l'axe appalachien était aussi en accord avec la colonisation postglaciaire à partir d'un refuge au sud-ouest de la chaîne des Appalaches. Les auteurs ont observé un regroupement de la population du nord-ouest de l'Ontario avec les populations du sud-ouest et le partage d'un haplotype chloroplastique rare parmi les populations de l'ouest, à travers toute la portée latitudinale. Ce patron est cohérent avec une divergence des routes de migration postglaciaire et des niveaux plus élevés d'échanges géniques plus récents, impliquant potentiellement l'humain, entre les populations de l'ouest de la chaîne des Appalaches. Les paires de populations de l'est et de l'ouest de l'axe appalachien présentaient des différences prononcées quant aux fréquences alléliques sur une petite étendue géographique. Ces patrons de variation génétique devraient être pris en considération lors de l'échantillonnage du matériel reproductif à des fins de conservation et de reproduction. [Traduit par la Rédaction]

Mots-clés : EST-SSR, *Castanea dentata*, microsatellites, variation clinale, conservation.

Introduction

The American chestnut (*Castanea dentata* Borkh.) was a dominant tree species in eastern North American forest ecosystems. The introduction of the fungal disease called chestnut blight, caused by *Chryphonectria parasitica* from Asia in the late 19th century, reduced American chestnut to a small understory shrub reproducing mainly vegetatively by root sprouting (Kubisiak et al. 1997;

Received 22 December 2016. Accepted 10 February 2017.

O. Gailing.* Michigan Technological University, School of Forest Resources and Environmental Science, 1400 Townsend Drive, Houghton, MI 49931, USA.
C.D. Nelson. USDA Forest Service, Southern Research Station, Southern Institute of Forest Genetics, 23332 Success Road, Saucier, MS 39574, USA; Forest Health Research and Education Center, University of Kentucky, 730 Rose Street, Lexington, KY 40546, USA.
Corresponding author: Oliver Gailing (email: ogailing@mtu.edu).
*Present address: University of Goettingen, Faculty of Forest Sciences and Forest Ecology, Forest Genetics and Forest Tree Breeding, Buesgenweg 2, 37077 Goettingen, Germany.

Botany **95**: 799–807 (2017) dx.doi.org/10.1139/cjb-2016-0323

Published at www.nrcresearchpress.com/cjb on 22 February 2017.

Table 1. Sample locations and haplotype frequencies.

					Haplotype			
Name	Abbreviation	*N*	Latitude	Longitude	H1	H2	H3	H4
Murphy_NC	MNC	32	35.05	84.01	11	0	0	2
Asheville_BR	ABR	32	35.46	82.10	0	9	1	0
Kentucky_PC	KPC	32	37.50	83.51	11	0	0	0
Maryland_MD	MMD	32	39.37	79.07	8	2	0	0
Pennsylvania_YC	PYC	32	39.48	76.59	10	0	0	0
Ontario_CA	OCA	32	43.08	80.30	9	1	0	0
New York_UC	NYUC	32	41.44	74.13	10	0	0	0
Portland_CT	PCT	32	41.35	72.37	10	0	0	0
Massachusetts_A	MA	32	42.22	72.31	11	0	0	0
Portland_CT	PCT	32	41.35	72.37	10	0	0	0

Note: Decimal coordinates are shown for latitude and longitude. Populations are listed from southwest to northeast along the Appalachian axis (Fig. 1).

Anagnostakis 2001). Considerable progress has been made to develop blight-resistant chestnuts for restoration purposes, using primarily genetic-marker-assisted backcross breeding to incorporate blight resistance from *Castanea mollissima* Blume into a *C. dentata* genetic background (Hebard 2005; Clark et al. 2016). Restoration programs are more likely to be successful if hybrid American chestnut populations are genetically diverse and locally adapted. For this purpose, genetic variation of the American chestnut parents used in the breeding program should be captured from many individuals originating from different geographic regions and climatic zones. The conservation of genetic variation in fitness-related traits (adaptive genetic variation) is crucial for the successful restoration of American chestnut because the species' reintroduction is threatened by other biotic (e.g., *Phytophtora cinnamomi*) and abiotic stressors (Anagnostakis 2001; Rhoades et al. 2003). While growing genomic resources and gene-based markers are becoming available for American chestnut and related species (Barakat et al. 2012; Bodénès et al. 2012; Kubisiak et al. 2013), genetic variation at genic microsatellites (expressed sequence tag – simple sequence repeat (EST-SSR) markers) has not yet been analyzed in natural populations of American chestnut.

Recolonization of suitable habitats after the last glacial period had a strong effect on patterns of genetic variation in American chestnut (Davis 1983; Li and Dane 2013). Pollen records and distribution of genetic diversity at nuclear and chloroplast markers suggest refugial areas along the Gulf Coast and a relatively slow northward migration with an arrival in Conneticut and southern Ontario only about 2000 years ago (Davis 1983; Li and Dane 2013). Analysis of genetic variation patterns at anonymous random amplified polymorphic DNA (RAPD) and nuclear microsatellite markers, revealed the highest genetic diversity and number of rare alleles in southwestern populations, and clinal variation in the number of rare alleles and allele frequencies along the Appalachian axis (Kubisiak and Roberds 2006). No pronounced regional pattern was detected at these markers, and the observed pattern was interpreted as consistent with a single metapopulation in which genetic drift played a significant role after postglacial expansion of the species to the north and northeast. Cluster analysis using the unweighted pair group method with arithmetic mean (UPGMA) and principal component analysis (PCA) based on 20 allozymes revealed some regional structure, with one southern population and groups of south-central, north-central, and northern Appalachian populations (Huang et al. 1998). However, no statistical tests were performed to confirm or quantify regional genetic structure.

The aim of this study was to analyze genetic variation and differentiation at EST-SSRs with known genetic map locations and random (even) genomic distribution and at chloroplast markers, in a significant part of the species distribution range. Regional genetic structure was analyzed by Bayesian analysis of population structure (Pritchard et al. 2000), principal coordinate analysis (PCoA), and neighbor-joining trees with bootstrap resampling. To analyze clinal patterns of genetic diversity, genetic variation parameters were correlated with geographic location (longitude, latitude, relative location on the Appalachian axis).

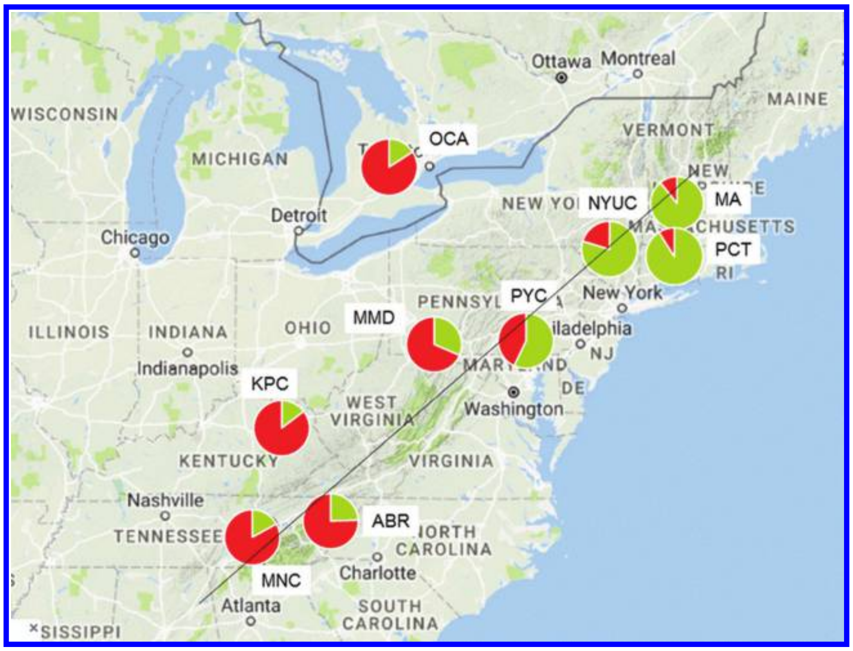
The specific objectives of the study were: (*i*) to identify patterns of genetic diversity and differentiation at genic and chloroplast microsatellite makers in American chestnut populations, and (*ii*) to associate these genetic variation patterns with geographic location.

Materials and methods

Plant materials

A total of nine populations (~32 trees per population or site) that covered a large part of the species' distribution range (Table 1; Fig. 1) were analyzed in our study. These samples (frozen tissue collected from dormant buds or expanded leaves) were originally collected as part of the study reported by Kubisiak and Roberds (2006). For this study, DNA was isolated using a cetyltrimethylammonium bromide (CTAB)-based method, as described in Kubisiak et al. (1997). The intergenic spacer

Fig. 1. Geographic locations of *Castanea dentata* populations and ancestry in genetic clusters as identified in STRUCTURE (see Fig. 3). Map data © Google 2016. [Colour online.]



trnT (UGA)–*trnL* (UAA) of the chloroplast genome was used to distinguish the widespread *C. dentata* chloroplast (cp) DNA haplotype from haplotypes characteristic of the related species *Castanea pumila* (Kubisiak and Roberds 2006). Genetic variation patterns at anonymous RADP markers and at six nuclear microsatellites gave no evidence for two distinct species in the sample set (Kubisiak and Roberds 2006); however, the occurrence of interspecific hybrids and later generation introgressive forms could not be excluded.

Marker analyses

A total of 25 EST-SSRs developed and genetically mapped in *Castanea mollissima* (Supplementary data, Table S1[1]) (Kubisiak et al. 2013) were tested for amplification and polymorphism in 24 *Castanea dentata* samples from eight populations. Seventeen markers that amplified single polymorphic loci were selected for the population analysis (Supplementary data, Table S2). Gene annotations were obtained using BLASTN 2.231 (Zhang et al. 2000) by comparing the EST contigs to the NCBI data base. A multiplex PCR touchdown protocol was developed that allowed us to analyze up to four markers with non-overlapping size ranges in one PCR reaction (multiplex 1: CmSI0031, CmSI0391, CmSI0600, CmSI0678; multiplex 2: CmSI0559, CmSI0561, CmSI0608, CmSI0611; multiplex 3: CmSI0396, CmSI0495, CmSI0527, CmSI0683; multiplex 4: CmSI0327, CmSI0537, CmSI0551; multiplex 5: CmSI0049, CmSI0594). The touchdown program in the Biometra TProfessional Thermocycler (Jena) was as follows: initial denaturation at 95 °C for 15 min, 10 touchdown cycles of 1 min at 94 °C, 1 min at 60 °C (–1 °C per cycle), and 1 min at 72 °C, followed by 25 cycles at 94 °C for 1 min, 50 °C for 1 min, and 72 °C for 1 min, and a final extension at 72 °C for 20 min. The PCR mix consisted of 0.2 μL (5 U/μL) HOTFIREPol DNA polymerase (Solis BioDyne, Estonia), 1.5 μL 10× reaction buffer B, 1.5 μL $MgCl_2$ (25 mmol/L), 0.2 μL (5 picomole/μL) tailed forward primer, 0.5 μL (5 picomole/μL) pig-tailed reverse primer (Kubisiak et al. 2013), 1.5 μL dye-labelled (6-FAM) M13 primer (5 picomole/μL), 1 μL dNTPs (2.5 mmol/L each dNTP), 2 μL DNA (ca. 0.6 ng/μL), and 5 μL H_2O. PCR products were separated on an ABI 3130xl Genetic Analyzer (Applied Biosystems), and alleles were called using GeneMapper version 4.0 (Applied Biosystems) after careful visual inspection.

Chloroplast microsatellites *ccmp1*, *ccmp2*, *ccmp3*, *ccmp4*, *ccmp5*, *ccmp6*, *ccmp7*, *ccmp10* (Weising and Gardner 1999), *ucd4*, *udt1*, and *udt4* (Deguilloux et al. 2003) were tested for amplification and polymorphism in two randomly selected samples of each of the nine populations. Polymorphic markers *ccmp3*, *ccmp4*, *ccmp5*, *udt1*, and *udt4* were amplified in a total of 10 randomly selected samples from each population. PCR protocols essentially followed the protocols described previously (Weising and Gardner 1999; Deguilloux et al. 2003); however, for *ucd4*, *udt1*, and *udt4* a touchdown protocol (Gailing et al. 2009) was used with an initial denaturation at 95 °C for 15 min, eight touch-down cycles of 1 min at 94 °C, annealing at 53 °C for 1 min (–1 °C per cycle), elongation at 72 °C for

[1] Supplementary data are available with the article through the journal Web site at http://nrcresearchpress.com/doi/suppl/10.1139/cjb-2016-0323.

1 min, followed by 33 cycles at an annealing temperature of 45 °C.

Data analysis

Genetic variation in populations was assessed as the number of alleles per locus (N_a), number of private alleles N_{pr}, number of locally common alleles found in ≤25% (N_{25}), and ≤50% (N_{50}) of populations, observed heterozygosity (H_o), and expected heterozygosity (H_e) (Nei 1973) in the program GeneAlEx version 6.41 (Peakall and Smouse 2006). The inbreeding coefficient (F) was calculated as $(H_e - H_o)/H_e$ in GeneAlEx. Significant deviations from Hardy–Weinberg proportions were tested using probability tests for each locus and population, with default Markov chain parameters (dememorization number, 1000; batches, 100; iterations per batch, 1000) in GENEPOP version 4.2 (Raymond and Rousset 1995). Genetic differentiation among populations and between population pairs was calculated as F_{ST} in GeneAlEx, and the significance of population differentiation was tested with the exact G test in GENEPOP with default Markov chain parameters for individual loci. Fisher's exact test was applied to test for significant genetic differentiation among populations and between population pairs across all 17 loci. Cavalli-Sforza's chord genetic distance (Cavalli-Sforza and Edwards 1967) and Nei's DA genetic distance (Nei et al. 1983) between populations were assessed in POPULATIONS 1.2.30 (Langella 1999). These distances are frequently used for microsatellites because they perform well in the reconstruction of phylogenetic trees from microsatellites (Takezaki and Nei 1996). Neighbor-joining trees were calculated with 1000 bootstrap replications on loci in POPULATIONS, and trees were visualized in TreeViewX (Page 1996). Genotypic linkage disequilibrium (LD) was tested for each pair of loci in each of the nine populations with the log likelihood statistic in GENEPOP. Significant LD is reported at the 5% level after Bonferroni correction (Rice 1989).

PCoA was performed in GeneAlEx to represent Nei's unbiased genetic distance (Nei 1978) between populations. Linear regression analyses were performed between principal coordinates (PCoA1, PCoA2) and longitude/latitude in WinSTAT (Fitch 2006). Bayesian analysis of population structure in the program STRUCTURE version 2.3.4 (Pritchard et al. 2000) was used to determine the number of genetic clusters (K) and to assign individual samples to genetic clusters following the approach described previously (Lind and Gailing 2013). Specifically, we performed five independent runs for K = 1 to K = 9 with a burn-in period of 30 000 and 100 000 replications for each K. The admixture model with correlated allele frequencies, without prior information on population identity, was applied. The most probable value for K was determined by plotting the logarithmized probabilities [Pr(X|K)] against K and selecting the value for K where ln[Pr(X|K)] plateaued. Mantel tests of geographic distances against genetic distances (F_{ST}, Nei's unbiased genetic distance) were conducted in GeneAlEx.

Table 2. Genetic variation within populations.

Pop.	N_a	N_{pr}	N_{25}	N_{50}	H_o	H_e	F
MNC	5.824	0.824	0.353	1.000	0.517	0.531	0.033
ABR	4.706	0.118	0.471	0.824	0.451	0.468	0.084
KPC	4.706	0.235	0.353	0.824	0.481	0.498	0.050
MMD	4.471	0.000	0.235	0.882	0.493	0.475	−0.047
PYC	4.471	0.059	0.235	0.824	0.490	0.475	−0.033
OCA	5.294	0.529	0.471	1.059	0.477	0.517	0.084
NYUC	4.059	0.000	0.118	0.647	0.513	0.489	−0.058
PCT	3.706	0.000	0.118	0.588	0.439	0.396	−0.101
MA	3.882	0.059	0.118	0.765	0.499	0.489	−0.034
Mean	4.569	0.203	0.275	0.824	0.484	0.482	−0.002

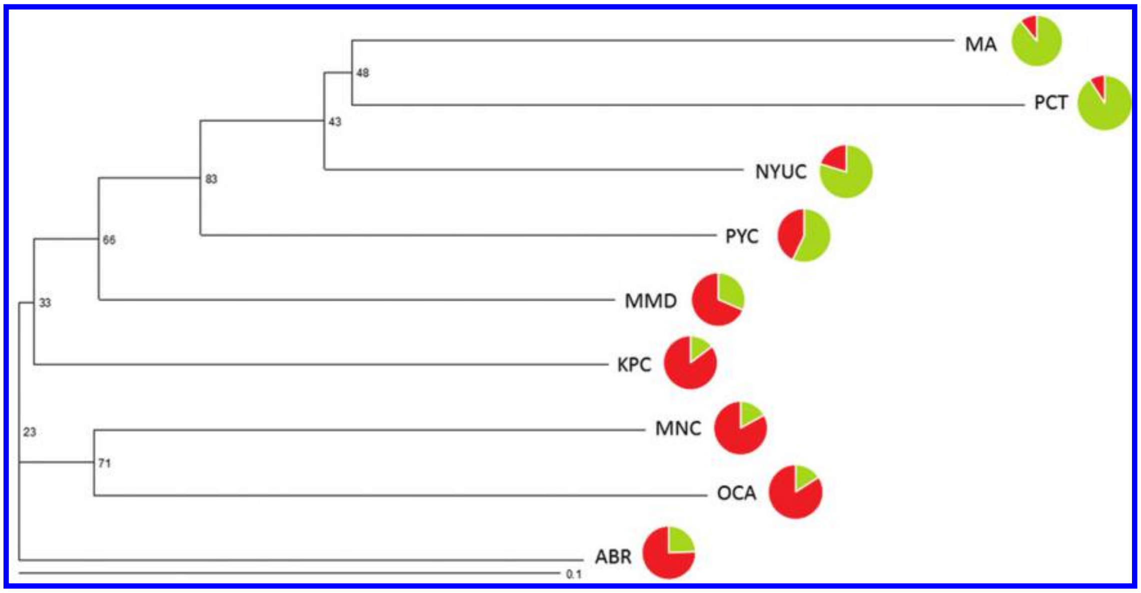
Note: N_a, number of alleles per locus; N_{pr}, number of private alleles; N_{25}, number of locally common alleles that are found in ≤25% of populations; N_{50}, number of locally common alleles that are found in ≤50% of populations; H_o, observed heterozygosity; H_e, expected heterozygosity; F, inbreeding coefficient; $F = (H_e - H_o)/H_e$. Populations are grouped from southwest to northeast along the Appalachian axis.

Correlations between genetic variation parameters H_o, H_e, N_a, N_{pr}, N_{25}, N_{50}, and ancestry according to genetic assignment in STRUCTURE against longitude, latitude, and relative geographic location along the Appalachian axis were calculated in WinSTAT. To determine the location of populations on the Appalachian axis, a reference line was drawn between southwestern and northeastern populations, excluding the northwestern population OCA. The relative position of populations on this axis was determined by connecting populations to the axis with perpendicular lines according to Kubisiak and Roberds (2006). Since the northwestern population OCA grouped with the southernmost population, based on phylogenetic analyses (see below), no position along the axis was assigned to this population.

Results

Genetic variation varied among populations, with the highest variation in the southernmost population MNC (H_e = 0.531, N_a = 5.824), and the lowest variation in the northeastern population PCT (H_e = 0.396, N_a = 3.706) (Table 2). The number of private alleles was highest in the southernmost population MNC (N_{pr} = 0.824), and lowest in the northeastern populations PCT and NYUC, and in the central population MMD (N_{pr} = 0.000). The northernmost population OCA at the northwestern distribution edge of the species had the second highest number of private alleles (N_{pr} = 0.529) and the highest number of locally common alleles (N_{25} = 0.471, N_{50} = 1.059). No association of genetic variation (H_e, H_o, N_a, N_{pr}, N_{25}, N_{50}) with latitude was observed. Overall there was a significant decrease in the number of alleles N_a, N_{pr}, N_{25}, and N_{50} along the Appalachian range from southwest to northeast. Thus, the location of populations on the Appalachian axis (see Fig. 1, excluding population OCA with high values for N_{pr}, N_{25}, and N_{50} northwest of the axis)

Fig. 2. Neighbor-joining trees based on Cavalli-Sforza's (Cavalli-Sforza and Edwards 1967) genetic distance. Numbers at nodes indicate bootstrap support after 1000 bootstrap replicates. Pie charts show ancestry in the northeastern and western genetic clusters as identified by STRUCTURE (Fig. 3). [Colour online.]



was strongly associated with N_{25} (r = 0.906, p < 0.0001) and with N_a (r = 0.908, p < 0.0001) and was moderately associated with N_{50} (r = 0.767, p = 0.0131) and N_{pr} (r = 0.758, p = 0.0090). Consequently, N_a (r = 0.857, p = 0.0016), N_{25} (r = 0.857, p = 0.0016), N_{50} (r = 0.724, p = 0.0136), and N_{pr} (r = 0.758, p = 0.0090) showed a pronounced and significant decline from west to east (Supplementary data, Figs. S1 and S2).

Inbreeding coefficients across all markers were low (F ranged from −0.101 to 0.084, Table 2) and not consistently positive for any population. Significant deviations from Hardy–Weinberg proportions at the 5% level after Bonferroni corrections were found for CmSI0678 in populations MMD (F = −0.325, H_o = 0.906, H_e = 0.684) and NYUC (F = −0.169, H_o = 0.839, H_e = 0.717), for CmSI0611 in population ABR (F = 1, H_o = 0.000, H_e = 0.121), and for CmSI0594 in population MNC (F = 0.366, H_o = 0.094, H_e = 0.148). Differences in mean F values were detected among populations, with western populations showing positive values and northeastern populations showing negative values. In particular, the northwestern population OCA (with high allelic diversity) showed signatures of recent inbreeding (mean F = 0.084), with 12 out of the 17 markers showing positive F values ranging from 0.034 to 0.509 (all markers: F = −0.292 to 0.509). The northeastern population PCT (with low allelic diversity) showed no indication of recent inbreeding (F = −0.101) with 15 of the 17 markers showing negative F values ranging from −0.309 to −0.021 (all markers: F = −0.309 to 0.133).

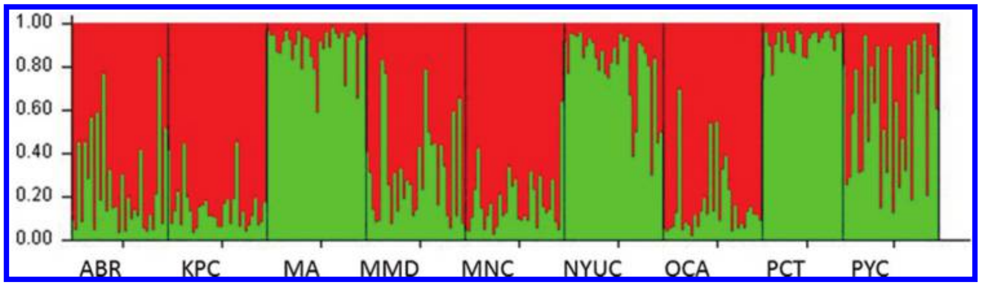
Out of a total of 1224 pairwise comparisons (136 per population), only four marker pairs showed significant LD at the 5% level after Bonferroni correction. No marker pair showed significant LD in more than one population.

Genetic differentiation

Pairwise F_{ST} ranged from 0.021 between the northwestern population OCA (Ontario) and the southernmost population MNC (North Carolina) to 0.058 between the northeastern population PCT and the southwestern population ABR (North Carolina) (Supplementary data, Table S3). The northeastern population PCT, which had the lowest within-population variation, was strongly differentiated from all other populations (F_{ST} = 0.036 to 0.058), but showed more pronounced differentiation from the western (F_{ST} = 0.044 to 0.058) than from the northeastern populations (F_{ST} = 0.036 to 0.043). Genetic differentiation among populations ranged from 0.027 to 0.110 for individual markers and genetic differentiation across all markers was 0.058 (Supplementary data, Table S4).

A Mantel test between geographic and genetic distances (Nei's unbiased genetic distance, F_{ST}) was significant for all populations (R^2 = 0.377, p < 0.0001) as well as for Appalachian populations alone (i.e., after exclusion of population OCA, R^2 = 0.537, p < 0.0001) as shown here for Nei's unbiased genetic distance. The northeastern populations MA, NYUC, and PCT were differentiated from the western populations as revealed by PCoA based on Nei's unbiased genetic distance (Supplementary data, Fig. S3). PCoA1 explained 62.62% of the variation and was significantly associated with longitude (R^2 = 0.860, p = 5.1×10^{-13}) (Supplementary data, Fig. S4). PCoA2 explained 17.23% of the variation. Phylogenetic trees based on Cavalli-Sforza's chord genetic distances and Nei's DA grouped the three northeastern populations MA, PCT and NYUC with populations PYC and MMD (from Pennsylvania and Maryland) with significant bootstrap support (Fig. 2; Supplementary data, Fig. S5). Resolution of

Fig. 3. Structure plots for $K = 2$ sorted by population. Ancestry in genetic clusters is shown for each individual. [Colour online.]



genetic relationships was low among the western populations. However, the southernmost population MNC (North Carolina) clustered with the northernmost population OCA (Ontario). Likewise, STRUCTURE analysis identified two genetic clusters and separated the northeastern populations (MA, PCT, NYUC) from the western populations (ABR, MNC, KPC, OCA), while PYC (Pennsylvania) and MMD (Maryland) occupied intermediate positions (Figs. 1 and 3). Generally, there was an increase of ancestry in the northeastern cluster from southwest to northeast (Figs. 1 and 2; Supplementary data, Fig. S5). The northernmost population OCA west of the Appalachian range in Ontario had a high ancestry coefficient in the southwestern cluster. Concordantly, ancestry in the northeastern genetic cluster was strongly correlated with longitude (R^2 = 0.936, p < 0.0001, Supplementary data, Fig. S6) but was not significantly associated with latitude. The northernmost population OCA had the highest deviation from the regression line showing an even lower ancestry in the northeastern cluster (15.9%) than expected based on the regression model (Supplementary data, Fig. S6). Overall allele frequencies showed a strong association with longitude. The average absolute value of correlation coefficients across all alleles at all loci was 0.415. Allele CmSI0600_275 showed the steepest decline in frequency from east to west (R^2 = 0.766, p = 0.0019) and was strongly associated with the relative position of populations on the Appalachian axis (R^2 = 0.751, p = 0.0053) (Supplementary data, Fig. S7). Overall, populations east of the Appalachian axis, including ABR (16.1%), PYC (24.2%), and the three northeastern populations NYUC, MA, PCT (28.2% to 58.0%), showed much higher frequencies of allele CmSI0600_275 than the western populations MNC (3.2%), KPC (3.2%), and OCA (0.0%) (Supplementary data, Fig. S7). The neighboring populations MMD (1.6%) and PYC (24.2%) in the central Appalachian region, and neighboring populations MNC (3.2%) and ABR (16.1%) in the southern Appalachian region showed pronounced frequency differences for this allele over a small geographic range.

Chloroplast analyses

The chloroplast microsatellites *ccmp3*, *ccmp4*, *ccmp5*, *udt1*, and *udt4* were polymorphic, resulting in four haplotypes (Supplementary data, Table S5). Haplotypes 1 and 2 (H1, H2) were very similar to each other, whereas haplotypes 3 and 4 (H3, H4) showed distinct variants at four of the five polymorphic markers (Table 1). Haplotype 1 was the predominant haplotype in most populations, and five out of the nine populations were fixed on this haplotype. In the southwestern population ABR, most samples were characterized by H2; haplotype 2 was also found in the central population MMD and in the northwestern population OCA. Haplotypes 3 and 4 were found at low frequency only in the two southwestern populations ABR and MNC.

Discussion

Genetic variation and differentiation patterns suggest one glacial refugium and divergence of postglacial remigration routes

Genetic measures of multiplicity, such as the number of alleles per locus and the number of locally common alleles, decreased significantly from southwest to northeast. Moreover, analyses of maternally inherited cpDNA revealed a decrease in haplotype diversity along the Appalachian axis; high haplotype diversity was evident in the southwestern range of the species whereas a single haplotype was present in northeastern chestnut populations (Kubisiak and Roberds 2006; Shaw et al. 2012; Li and Dane 2013). In the present study, the rare and distinct H3 and H4 haplotypes were found at low frequency only in the southwestern populations ABR and MNC, whereas the northeastern populations were fixed on the most common H1 haplotype.

In contrast to northeastern populations, the northernmost population west of the Appalachian range in Ontario (OCA) is characterized by high allelic diversity comparable to southwestern populations. This population groups with southwestern populations in Principal Coordinate and STRUCTURE analyses, and is most similar genetically to the southwestern population MNC (Fig. 3). In addition, the rare H2 haplotype, predominant in the southwestern ABR population, is shared among western populations across the whole latitudinal range. Chestnut colonized Ontario approximately at the same time (2000 years ago) as the northeastern range was colonized (Davis 1983). Hence, the high allelic diversity in OCA and pronounced differentiation between OCA and the northeastern populations was likely the result of several factors, including the divergence of postglacial remigration routes, effective reproductive isolation between the northwestern OCA population and northeastern populations, and higher (potentially human-mediated) migration rates in south–north direction than along the Appalachian axis. Likewise, genetic variation patterns in populations of European chestnut (*Castanea sativa* Mill.) in Spain suggested human-mediated distribution of reproductive material (Martin et al. 2012).

Thus, the strong association of allelic diversity with longitude, but not with latitude, is likely a reflection of historical migration/gene flow including founder events along the southwest–northeast migration axis and divergence of migration routes to the north and northeast.

In addition to a decrease in allelic diversity along the Appalachian axis, a pattern of clinal variation of allele frequencies along the Appalachian axis is consistent with the slow and continuous migration of American chestnut along the Appalachian axis from one major glacial refugium southwest of the Appalachian range, likely close to the Gulf Coast. This pattern of clinal variation is in accordance with observations at non-genic microsatellite, anonymous RAPD markers and one isozyme marker (Huang et al. 1998; Kubisiak and Roberds 2006).

However, the grouping of the northwestern population OCA with southwestern populations suggests not only a higher migration rate in the south–north direction (west of the Appalachian range) than in the southwest-northeast direction, but also restricted gene flow between populations west and east of the Appalachian range, especially for northern populations. One allele (275) at CmSI0600, in particular, showed a steep increase in frequency from west to east at a small geographic range and along the Appalachian axis with highly elevated frequencies in northeastern populations. These patterns of genetic variation have to be considered for in situ and ex situ conservation of reproductive material.

Signatures of inbreeding and balancing selection

The pattern of clinal variation observed for allelic diversity was not found for expected (H_e) and observed (H_o) heterozygosities, but comparable levels of gene diversity were observed in most populations except the northeastern population PCT. Levels of gene diversity (H_e) are affected by gene flow within and among populations (Hamrick and Godt 1996), but could also be affected by balancing selection (Hamrick et al. 1979; Ziehe et al. 1989). Thus inbreeding values (F) were moderately positive in southwestern populations and in the northwestern population OCA, and moderately negative for most northeastern and central Appalachian populations. Signatures of recent inbreeding were especially evident in population OCA (F = 0.084), for which 12 out of the 17 markers showed positive F values, whereas for the northeastern population PCT, 15 out of 17 markers showed negative F values, suggesting no inbreeding and potentially balancing selection favoring heterozygous genotypes (overdominance hypothesis; Ziehe and Roberds 1989; Charlesworth and Charlesworth 2010).

Signatures of inbreeding in population OCA are in accordance with more recent effects on gene flow patterns as result of forest management or differential mortality. The excess of heterozygotes in the northeastern population PCT may be related to the duration and frequency of vegetative reproduction. Many studies have shown that individual heterozygosity is positively correlated with growth (e.g., reviewed in Mitton and Grant 1984) and individuals with high individual heterozygosity may be favored in vegetatively reproducing stands. Indeed, an excess of heterozygotes at isozyme loci was observed in four *C. dentata* populations from south Virginia, and heterozygosity was positively associated with growth (Stilwell et al. 2003). Human-mediated management before the American chestnut pandemic and the change from sexual to vegetative reproduction as a result of chestnut blight infection likely affected genetic variation by changing gene flow patterns and selection regimes. Because American chestnut is now largely reproducing by root sprouting, future depletion of the gene pool is expected if opportunities for sexual reproduction are not increased by restoration efforts with blight-resistant genotypes (Kubisiak and Roberds 2006).

Conclusions

Most of the genetic variation (H_e, Nei 1973) at genic EST-SSRs is present within populations, i.e., about 94% of the variation can be sampled on average within populations. This estimate is very similar to the one obtained at nongenic SSRs and anonymous RAPDs (Huang et al. 1998; Kubisiak and Roberds 2006). However, a significant decrease in allelic diversity was observed from southwest to northeast. Decrease in allelic diversity within populations can negatively affect the long-term evolutionary potential of populations in changing environmental conditions (England et al. 2003). Overall, allele frequencies showed a strong association with longitude and population pairs east and west of the Appalachian axis showed pronounced allele frequency differences over a small geographic range. These patterns of genetic variation should be considered when sampling reproductive material for conservation and breeding. Additional populations should be sampled to perform genome-wide analyses including genes with potential role in local adaptation.

Acknowledgements

Partial funding for this study was provided by The American Chestnut Foundation. We thank Oleksandra Dolynska for help with the lab work during Oliver Gailing's research visit to the Department of Forest Genetics and Forest Tree Breeding at the Georg-August University of Göttingen. The first author is thankful for the opportunity provided by Reiner Finkeldey to work on this proj-

ect during a research visit in 2015 at the Department of Forest Genetics and Forest Tree Breeding.

References

Anagnostakis, S.L. 2001. The effect of multiple importations of pests and pathogens on a native tree. Biol. Invasions, **3**: 245–254. doi:10.1023/A:1015205005751.

Barakat, A., Staton, M., Cheng, C.H., Park, J., Yassin, N.B.M., Ficklin, S., Yeh, C.C., Hebard, F., Baier, K., Powell, W., Schuster, S.C., Wheeler, N., Abbott, A., Carlson, J.E., and Sederoff, R. 2012. Chestnut resistance to the blight disease: insights from transcriptome analysis. BMC Plant Biol. **12**: 38. doi:10.1186/1471-2229-12-38. PMID:22429310.

Bodénès, C., Chancerel, E., Gailing, O., Vendramin, G.G., Bagnoli, F., Durand, J., Goicoechea, P.G., Soliani, C., Villani, F., Mattioni, C., Koelewijn, H.P., Murat, F., Salse, J., Roussel, G., Boury, C., Alberto, F., Kremer, A., and Plomion, C. 2012. Comparative mapping in the Fagaceae and beyond with EST-SSRs. BMC Plant Biol. **12**: 153. doi:10.1186/1471-2229-12-153. PMID:22931513.

Cavalli-Sforza, L.L., and Edwards, A.W. 1967. Phylogenetic analysis. Models and estimation procedures. Am. J. Hum. Genet. **19**: 233–257. PMID:6026583.

Charlesworth, B., and Charlesworth, D. 2010. Elements of evolutionary genetics. Roberts and Company Publishers, Greenwood Village, Colo.

Clark, S.L., Schlarbaum, S.E., Saxton, A.M., and Hebard, F.V. 2016. Establishment of American chestnuts (*Castanea dentata*) bred for blight (*Cryphonectria parasitica*) resistance: influence of breeding and nursery grading. New Forests, **47**: 243–270. doi:10.1007/s11056-015-9512-6.

Davis, M.B. 1983. Quaternary history of deciduous forests of eastern North America and Europe. Ann. Mo. Bot. Gard. **70**: 550–563. doi:10.2307/2992086.

Deguilloux, M.-F., Dumolin-Lapegue, S., Gielly, L., Grivet, D., and Petit, R.J. 2003. A set of primers for the amplification of chloroplast microsatellites in *Quercus*. Mol. Ecol. Notes, **3**: 24–27. doi:10.1046/j.1471-8286.2003.00339.x.

England, P.R., Osler, G.H.R., Woodworth, L.M., Montgomery, M.E., Briscoe, D.A., and Frankham, R. 2003. Effects of intense versus diffuse population bottlenecks on microsatellite genetic diversity and evolutionary potential. Conserv. Genet. **4**: 595–604. doi:10.1023/A:1025639811865.

Fitch, R. 2006. WinSTAT for Excel. The statistics add-in for Microsoft Excel. R. Fitch Software.

Gailing, O., Wachter, H., Rogge, M., Heyder, J., and Finkeldey, R. 2009. Chloroplast DNA analyses of very old, presumably autochthonous *Quercus robur* L. stands in North-Rhine Westphalia. Allgemeine Forst und Jagdzeitung, **180**: 221–227.

Hamrick, J.L., and Godt, M.J.W. 1996. Effects of life history traits on genetic diversity in plant species. Philos. Trans. R. Soc. Ser. B. **351**: 1291–1298. doi:10.1098/rstb.1996.0112.

Hamrick, J.L., Linhart, Y.B., and Mitton, J.B. 1979. Relationships between life history characteristics and electrophoretically detectable genetic variation in plants. Annu. Rev. Ecol. Syst. **10**: 173–200. doi:10.1146/annurev.es.10.110179.001133.

Hebard, F. 2005. The backcross breeding program of the American Chestnut Foundation. Journal of the American Chestnut Foundation, **19**: 55–77. doi:10.17660/ActaHortic.2014.1019.20.

Huang, H.W., Dane, F., and Kubisiak, T.L. 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. doi:10.2307/2446368. PMID:21684985.

Kubisiak, T., and Roberds, J.H. 2006. Genetic structure of American Chestnut populations based on neutral DNA markers. *In* Restoration of American Chestnut to forest lands. National Park Service, Washington, D.C. pp. 109–122.

Kubisiak, T.L., Hebard, F.V., Nelson, C.D., Zhang, J.S., Bernatzky, R., Huang, H., Anagnostakis, S.L., and Doudrick, R.L. 1997. Molecular mapping of resistance to blight in an interspecific cross in the genus *Castanea*. Phytopathology, **87**: 751–759. doi:10.1094/PHYTO.1997.87.7.751. PMID: 18945098.

Kubisiak, T.L., Nelson, C.D., Staton, M.E., Zhebentyayeva, T., Smith, C., Olukolu, B.A., Fang, G.C., Hebard, F.V., Anagnostakis, S., Wheeler, N., Sisco, P.H., Abbott, A.G., and Sederoff, R.R. 2013. A transcriptome-based genetic map of Chinese chestnut (*Castanea mollissima*) and identification of regions of segmental homology with peach (*Prunus persica*). Tree Genet. Genomes, **9**: 557–571. doi:10.1007/s11295-012-0579-3.

Langella, O. 1999. Populations, 1.2.30. CNRS UPR9034.

Li, X.W., and Dane, F. 2013. Comparative chloroplast and nuclear DNA analysis of *Castanea* species in the southern region of the USA. Tree Genet. Genomes, **9**: 107–116. doi:10.1007/s11295-012-0538-z.

Lind, J., and Gailing, O. 2013. Genetic structure of *Quercus rubra* L. and *Q. ellipsoidalis* E. J. Hill populations at gene-based EST-SSR and nuclear SSR markers. Tree Genet. Genomes, **9**: 707–722. doi:10.1007/s11295-012-0586-4.

Martin, M.A., Mattioni, C., Molina, J.R., Alvarez, J.B., Cherubini, M., Herrera, M.A., Villani, F., and Martin, L.M. 2012. Landscape genetic structure of chestnut (*Castanea sativa* Mill.) in Spain. Tree Genet. Genomes, **8**: 127–136. doi:10.1007/s11295-011-0427-x.

Mitton, J.B., and Grant, M.C. 1984. Associations among protein heterozygosity, growth rate, and developmental homeostasis. Annu. Rev. Ecol. Syst. **15**: 479–499. doi:10.1146/annurev.es.15.110184.002403.

Nei, M. 1973. Analysis of gene diversity in subdivided populations. Proc. Natl. Acad. Sci. U.S.A., **70**: 3321–3323. doi:10.1073/pnas.70.12.3321. PMID:4519626.

Nei, M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics, **83**: 583–590. PMID:955405.

Nei, M., Tajima, F., and Tateno, Y. 1983. Accuracy of estimated phylogenetic trees from molecular data. 2. Gene frequency data. J. Mol. Evol. **19**: 153–170. PMID:6571220.

Page, R.D.M. 1996. TreeView: An application to display phylogenetic trees on personal computers. Comput. Appl. Biosci. **12**: 357–358. PMID:8902363.

Peakall, R., and Smouse, P.E. 2006. GENEALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol. Ecol. Notes, **6**: 288–295. doi:10.1111/j.1471-8286.2005.01155.x.

Pritchard, J.K., Stephens, M., and Donnelly, P. 2000. Inference of population structure using multilocus genotype data. Genetics, **155**: 945–959. PMID:10835412.

Raymond, M., and Rousset, F. 1995. GENEPOP (Version 1.2): population genetics software for exact tests and ecumenicism. J. Hered. **86**: 248–249. doi:10.1093/oxfordjournals.jhered.a111573.

Rhoades, C.C., Brosi, S.L., Dattilo, A.J., and Vincelli, P. 2003. Effect of soil compaction and moisture on incidence of phytophthora root rot on American chestnut (*Castanea dentata*) seedlings. For. Ecol. Manage. **184**: 47–54. doi:10.1016/S0378-1127(03)00147-6.

Rice, W.R. 1989. Analyzing tables of statistical tests. Evolution, **43**: 223–225. doi:10.2307/2409177.

Shaw, J., Craddock, J.H., and Binkley, M.A. 2012. Phylogeny and phylogeography of North American *Castanea* Mill. (Fagaceae) using cpDNA suggests gene sharing in the southern Appalachians (*Castanea* Mill., Fagaceae). Castanea, **77**: 186–211.

Stilwell, K.L., Wilbur, H.M., Werth, C.R., and Taylor, D.R. 2003. Heterozygote advantage in the American chestnut, *Castanea* *dentata* (Fagaceae). Am. J. Bot. **90**: 207–213. doi:10.3732/ajb.90.2.207. PMID:21659110.

Takezaki, N., and Nei, M. 1996. Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. Genetics, **144**: 389–399. PMID:8878702.

Weising, K., and Gardner, R.C. 1999. A set of conserved PCR primers for the analysis of simple sequence repeat polymorphisms in chloroplast genomes of dicotyledonous angiosperms. Genome, **42**: 9–19. doi:10.1139/g98-104. PMID:10207998.

Zhang, Z., Schwartz, S., Wagner, L., and Miller, W. 2000. A greedy algorithm for aligning DNA sequences. J. Comput. Biol. **7**: 203–214. doi:10.1089/10665270050081478. PMID:10890397.

Ziehe, M., and Roberds, J.H. 1989. Inbreeding depression due to overdominance in partially self-fertilizing plant populations. Genetics, **121**: 861–868. PMID:17246494.

Ziehe, M., Gregorius, H.-R., Glock, H., Hattemer, H.H., and Herzog, S. 1989. Gene resources and gene conservation in forest trees: general concepts. *In* Genetic Effects of Air Pollutants in Forest Tree Populations. *Edited by* F. Scholz, H.-R. Gregorius, and D. Rudin. Springer, Berlin, Germany. pp. 173–185.