American Journal of Botany 93(3): 377–388. 2006.

ANCESTRAL CHLOROPLAST POLYMORPHISM AND HISTORICAL SECONDARY CONTACT IN A BROAD HYBRID ZONE OF *AESCULUS* (SAPINDACEAE)[1]

JENNIFER L. MODLISZEWSKI,[2] DAVID T. THOMAS,[2] CHUANZHU FAN,[3] DANIEL J. CRAWFORD,[4] CLAUDE W. DEPAMPHILIS,[5] AND QIU-YUN (JENNY) XIANG[2,6]

[2]Department of Botany, North Carolina State University, Raleigh, North Carolina 27695-7612 USA; [3]Department of Ecology and Evolution, University of Chicago, Chicago, Illinois 60637 USA; [4]Department of Ecology and Evolutionary Biology and the Natural History Museum and Biodiversity Research Center, University of Kansas, Lawrence, Kansas 66045-2106 USA; and [5]Department of Biology and Huck Institutes of Life Sciences, Pennsylvania State University, University Park, Pennsylvania 16802 USA

Knowledge regarding the origin and maintenance of hybrid zones is critical for understanding the evolutionary outcomes of natural hybridization. To evaluate the contribution of historical contact vs. long-distance gene flow in the formation of a broad hybrid zone in central and northern Georgia that involves *Aesculus pavia*, *A. sylvatica*, and *A. flava*, three cpDNA regions (*matK*, *trnD-trnT*, and *trnH-trnK*) were analyzed. The maternal inheritance of cpDNA in *Aesculus* was confirmed via sequencing of *matK* from progeny of controlled crosses. Restriction site analyses identified 21 unique haplotypes among 248 individuals representing 29 populations from parental species and hybrids. Haplotypes were sequenced for all cpDNA regions. Restriction site and sequence data were subjected to phylogeographic and population genetic analyses. Considerable cpDNA variation was detected in the hybrid zone, as well as ancestral cpDNA polymorphism; furthermore, the distribution of haplotypes indicates limited interpopulation gene flow via seeds. The genealogy and structure of genetic variation further support the historical presence of *A. pavia* in the Piedmont, although they are at present locally extinct. In conjunction with previous allozyme studies, the cpDNA data suggest that the hybrid zone originated through historical local gene flow, yet is maintained by periodic long-distance pollen dispersal.

Key words: *Aesculus*; cpDNA inheritance; hybrid zone; phylogeography; Pleistocene; Sapindaceae; secondary contact; southeastern United States.

Hybrid zones provide a natural setting for the observation of evolutionary processes, such as adaptation, speciation, and introgression (Arnold, 1997; Avise, 2004). Although hybridization was once regarded by some as simply "evolutionary noise" (e.g., Wagner, 1969, 1970), recent views consider hybridization and hybrid zones as vehicles for the creation of novel species and adaptations, as well as mechanisms for augmenting genetic diversity, and for either strengthening or reducing reproductive isolating barriers between species (e.g., Rieseberg and Wendel, 1993; Arnold, 1997; Rieseberg, 1997). Hybridization and introgression have highly influenced the distribution, diversity, and occurrence of species observed today, given their evolutionary significance and prevalence (Stace, 1987; Rieseberg and Wendel, 1993; Ellstrand et al., 1996). Awareness of the genetic structure and introgression within hybrid zones and those species involved in their formation is critical to predicting the potential evolutionary outcomes of natural hybridization. Therefore, a necessary step toward this goal is determining the genetic composition of hybrid zones via detailed multilocus analyses involving independent loci from both nuclear and cytoplasmic genomes (McCauley, 1995; Ouborg et al., 1999; Avise, 2004). Traditional population genetics often interprets genetic similarity between populations as a result of gene flow. However, genetic similarity between taxa may also be attributable to shared common ancestry. Thus, it is important to combine traditional population genetic analyses with a phylogeographic approach (e.g., Avise et al., 1987; Schaal et al., 1998; Hewitt, 2001) to determine the relative contributions of shared ancestry and genetic exchange among species in hybrid zones. Although the phylogeographic approach is typically applied to the study of intraspecific variation, in a complex of hybridizing species, taxonomic barriers often become indistinct, rendering the phylogeographic approach a useful tool for analyzing the geographic patterns of genetic variation within a complex of recently diverged species (e.g., García-Paris et al., 2000; Beheregay et al., 2004). In a hybrid zone, alleles (or haplotypes) shared between parental species can be a result of gene introgression via interspecific gene flow, recency of common ancestry of the hybridizing species, or a combination of both of these phenomena. Furthermore, alleles shared between parental species and hybrid populations may be caused by both historical and current gene flow that can be both local and long distance. Although distinguishing these alternative processes is not an easy task, it is essential to the understanding of hybrid zone evolution and may be achieved via synthesis of evidence from gene genealogy and patterns of genetic variation of loci that differentially track the histories of dispersal. In flowering plants, pollen and seeds are the typical

[1] Manuscript received 1 June 2005; revision accepted 24 October 2005.

The authors thank C. Chan of the Holden Arboretum and S. Wiegrefe of the Morton Arboretum for materials used in the cpDNA inheritance analysis. We thank T. Lasseigne and the J. C. Raulston Arboretum for allowing us to perform initial crossing experiments. The study was supported by a Faculty Research grant from Idaho State University and a Faculty Research and Development grant from NCSU, both awarded to J. Q. Y. X.

[6]Author for correspondence (e-mail: jenny_xiang@ncsu.edu); phone: 919-515-2728; fax: 919-515-3436

dispersal agents, with the latter more vagile than the former in most plants.

In the genus *Aesculus* L. (Sapindaceae), the seeds are large and heavy (Schopmeyer, 1974), thus it is not likely that seeds will be distributed across broad geographic ranges. In contrast, the pollen may be dispersed long distances via the ruby-throated hummingbird (dePamphilis and Wyatt, 1989). This discrepancy between dispersal distances of pollen and seed allows for the elucidation of the relative contribution of pollen vs. seeds to the genetic architecture of the hybrid zone via a combination of loci from nuclear and cpDNA genomes. Nuclear markers will be distributed by both pollen and seed, while only seeds, given the maternal inheritance of chloroplasts, will disperse cpDNA markers.

The genus *Aesculus* is comprised of trees and shrubs that are widely cultivated for their showy, large inflorescences and dense foliage. The genus consists of 13–19 species discontinuously distributed in four separate regions: eastern Asia, Europe, western North America, and eastern North America (Hardin, 1957a; Fang, 1981). Four species of the genus in eastern North America constitute sect. *Pavia* (Mill.) Person. Hybridization among three species of sect. *Pavia*, *A. pavia* L., *A. flava* Ait., and *A. sylvatica* Bartr., has resulted in the creation of a hybrid zone in central to northern Georgia and adjacent areas (Hardin, 1957b; dePamphilis and Wyatt, 1989). This hybrid zone is notable for its broad width (~200 km) and geographic and genetic asymmetry. Hybrids between *A. pavia* and *A. sylvatica* occur within the geographic range of *A. sylvatica*, but nearly 145 km from the nearest populations of *A. pavia* (Hardin, 1957b; dePamphilis and Wyatt, 1989, 1990). The natural distribution of *A. sylvatica* is limited to the deciduous and pine forests of the Piedmont from the southern limits of Virginia to northwestern Alabama, while *A. pavia* occurs in the mixed pine and deciduous forests of the southern Coastal Plain; *A. flava* is distributed throughout deciduous forests in the Appalachian Mountain range and northward to the Ohio River Valley (Hardin, 1957c; Fig. 1). To explain the asymmetrical distribution of hybrids, Hardin (1957b, 1957c) proposed that species of sect. *Pavia* were once allopatric in the Appalachians and evolved divergently and that hybridization occurred during a period of secondary contact as a result of range expansion of *A. pavia* during the interglacial warm periods of the Pleistocene. The present-day absence of *A. pavia* from the hybrid zone is posited to be a result of southward range reduction of the species from the Piedmont and subsequent localized extinction of *A. pavia* in the Piedmont.

According to the leading edge theory (Hewitt, 1993), the climatic oscillations of the Pleistocene led to the creation of refugia during periods of cooler temperatures (Hewitt, 1996, 2000, 2001). Populations that survived in refugia migrated and expanded their ranges during periods of warmer climates, resulting in a subsequent reduction in genetic diversity in the newly colonized regions. As a consequence of the range expansions and contractions of newly diverged species, hybrid zones lacking one or both parental species could be formed (Remington, 1968; Hewitt, 1996) because hybrids were able to survive in environments that were no longer suitable for one or both of the parental species. Thus, the extinction of *A. pavia* in the hybrid zone could have been caused by environmental selection favoring hybrids and discriminating against *A. pavia* parentals in the Piedmont region of northern Georgia. Remington (1968) identified several areas, or suture zones, where range expansions and contractions were likely to bring together previously allopatric species. One such zone is located in the southeastern United States.

dePamphilis and Wyatt (1989, 1990) conducted an allozyme and pollination biology study of *Aesculus* in the hybrid zone. Their allozyme data were concordant with the morphological data, with both confirming the broad width and genetic asymmetry of the hybrid zone. Although there were no fixed species-specific alleles, computer simulations and admixture analyses were used to provide evidence that multilocus genotypes and overall population structures were indicative of hybridization, in agreement with earlier morphological studies. dePamphilis and Wyatt (1989) posited that the genetic presence of *A. pavia* in the hybrid zone reflects long-distance pollen dispersal of *A. pavia* by ruby-throated hummingbirds (*Archilochus colubris*). The ruby-throated hummingbird is a common pollinator of all three *Aesculus* species, and it migrates from the Coastal Plain to the Piedmont each year during the flowering season of *Aesculus* (dePamphilis and Wyatt, 1989).

The two proposed hypotheses, historical secondary contact (HSC) vs. recurrent long-distance gene flow (LDF), may be distinguished by the current genetic composition of hybrid populations. If hybridization occurred during the Pleistocene, followed by local extinction of *A. pavia*, the presence of *A. pavia* and *A. flava* cpDNA haplotypes in the Piedmont and hybrid zone would provide strong evidence for the hypothesis of HSC between presently allopatric species. In contrast, hybridization by recent long-distance pollen dispersal should result in some individuals in hybrid populations containing mostly *A. pavia* nuclear genes, while chloroplast genes of *A. pavia* (if inherited maternally) should be extremely rare or absent in the zone, given the probable limited dispersal distance of the large and heavy *Aesculus* seeds (Schopmeyer, 1974). Hybrid populations containing nuclear markers predominantly from one species and the cytotype of another species would reveal the nature of the contribution of each

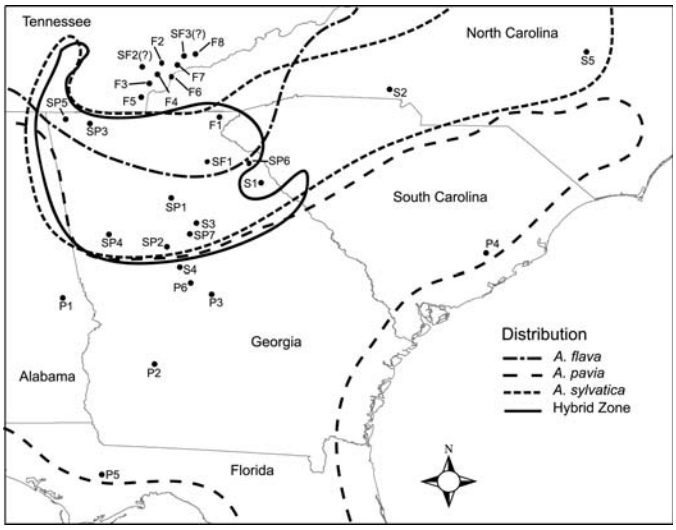

Fig. 1. Map of the distribution in southeastern USA of populations of *Aesculus pavia*, *A. flava*, *A. sylvatica*, *A. sylvatica* × *pavia*, and *A. sylvatica* × *flava* included in this study. The approximate location of the *Aesculus* hybrid zone and taxa is indicated, as delineated by dePamphilis and Wyatt (1989). The SF populations indicated by a question mark are uncertain in their identification as hybrid or as *A. flava*.

species to the hybrid zone, whether it be through pollen dispersal, seed dispersal, or both. Thus, evidence from the cpDNA genome, when compared to the previously published allozyme analysis of dePamphilis and Wyatt (1989, 1990) will allow us to further assess the HSC hypothesis. In the current study, we present data from three regions of cpDNA in order to examine and confirm the mechanisms that have contributed to the origin and maintenance of the hybrid zone.

Although cpDNA is generally inherited maternally in angiosperms (Harris and Ingram, 1991; Birky, 2001), instances of nonmaternal inheritance have been documented (e.g., Sewell et al., 1993). Therefore, it is critical to first verify the mode of cpDNA inheritance in *Aesculus* so that observed geographic patterns of cpDNA haplotypes may then be attributed to seed dispersal. Thus, our objectives were to (1) determine the mode of cpDNA inheritance in *Aesculus*, (2) examine the pattern of cpDNA variation in the hybrid zone, and (3) compare results of the present cpDNA analysis to the nuclear allozyme analysis of dePamphilis and Wyatt (1989, 1990) for the purpose of inferring the evolutionary history of the *Aesculus* hybrid zone.

MATERIALS AND METHODS

Sampling and DNA extraction—Leaf samples of *A. pavia*, *A. sylvatica*, *A. flava*, and hybrids between *A. pavia* and *A. sylvatica* and *A. flava* and *A. sylvatica* were collected from 29 natural populations, both within and outside the boundaries of the hybrid zone (Fig. 1). Species and putative hybrids were identified based on the diagnostic features of Hardin (1957c). Three to 26 individuals were sampled per population, with most population samples consisting of 10 individuals (Table 1). A total of 248 individuals were analyzed. Samples were collected from both "pure" allopatric populations of *A. flava*, *A. pavia*, and *A. sylvatica* and from populations sympatric and parapatric to the hybrid zone (Fig. 1). DNA was extracted from fresh or silica-dried leaf materials using the mini-prep method of Cullings (1992), with modifications described in Xiang et al. (1998). Additionally, fresh leaf material was obtained from parents and F1 progeny of controlled crosses of *Aesculus* spp. for 17 crosses, with one F1 analyzed per cross (Table 2), from crosses performed at the Holden Arboretum (Kirtland, Ohio, USA).

Chloroplast inheritance—DNA sequences of the chloroplast *matK* gene were compared between parental species and their progenies for 17 crosses to determine the mode of cpDNA inheritance in *Aesculus* (Table 2). PCR amplification and sequencing of the gene were performed using primers developed by Sang et al. (1997). Amplification was carried out on a Peltier Thermal Cycler (PTC-100, Bio-Rad, Hercules, California, USA) with a 50-μl reaction volume, using approximately 80 ng template DNA (3 μl), 5 μl 10× Mg-free buffer, 6 μl 25 mM $MgCl_2$, 8 μl 2.5 mM each dNTP, 2.0 μl 10 mg/ml BSA, 0.3 μl *Taq* DNA polymerase (5 U/μl), 0.5 μl 20 μM matK1F (forward) and matK3R (reverse) primer, and 24.7 μl sterile deionized water. Hot-start PCR conditions were as follows: an initial 6 min at 96°C, followed by the addition of *Taq* DNA polymerase, 30 cycles of 40 s at 94°C, 1 min at 45°C, and 2 min 30 s at 72°C, followed by a final cycle of 40 s at 94°C, 1 min at 45°C, and 4 min at 72°C. Polymerase chain reaction products were purified by precipitation using 20% PEG (polyethylene glycol)/2.5 mol NaCl/L sterile deionized water, as done in Morgan and Soltis (1993). Purified PCR products were sequenced with the ABI Prism dRhodamine Terminator Cycle Sequencing Ready Reaction kit (Applied Biosystems, Foster City, California, USA) on a PTC-100 following the procedure of Fan and Xiang (2003). Sequences were aligned visually, and nucleotide site mutations and indels were determined visually as well. Polymorphic sites between progeny and parentals were visually inspected on chromatographs for errors in base calling. A total of 1372 base pairs (bp), spanning the majority of the *matK* gene, were compared.

cpDNA restriction site variation—Three cpDNA regions were analyzed for all populations: the *matK* gene (1830 bp), the *trnD* (tRNA-Val-GUC)–*trnT* (tRNA-Gly-GGU) (1415 bp) intergenic spacer, and the *trnH* (tRNA-Val-GUG)–*trnK* (tRNA-Phe-UUU) intergenic spacer (2024 bp). Amplification of the *matK* gene was performed using primers matK1F (forward) and matK1R (reverse) developed by Sang et al. (1997), as described. Amplification of the *trnD-trnT* and *trnH-trnK* intergenic spacers was performed with primers trnD-F and trnT-R for the *trnD-trnT* region and trnH-F and trnK-R for the *trnH-trnK* region (Demesure et al., 1995). PCRs were conducted in a 50-μl reaction volume as described, except with an annealing temperature of 50°C for *trnD-trnT* and 63°C for *trnH-trnK*.

PCR products of a subset of representative samples were subjected to restriction site analysis with 10 enzymes for all regions (*Taq*I, *Hha*I, *Mse*I, *Rsa*I, *Hsp*92II, *Msp*I, *Hin*dIII, *Eco*RI, *Bam*HI, and *Bsr*BI) and two additional enzymes for the *matK* gene (*Bsr*I and *Bst*NI) to screen for species-specific markers. Based on the results of the screening, *Bsr*I, *Msp*I, *Bst*NI, and *Taq*I were used for *matK*, *Rsa*I and *Mse*I for the *trnD-trnT* region, and *Bsr*BI for the *trnH-trnK* region. DNA samples from all populations were then PCR-amplified, and restriction digests were performed based on screening results. Digestion was performed with 20-μl reaction volumes containing 5–15 μl of PCR products for 3 to 3.5 h of incubation at enzyme specific temperatures, in separate reactions for each enzyme. The restriction digests were subsequently electrophoresed on a 0.7% agarose/0.7% synergel gel for 3.5 to 4 h at 100 V. Gels were stained with ethidium bromide and digitally photographed using a UV transilluminator and imaging system (Kodak Digital Science 1D, Rochester, New York, USA). The cpDNA haplotype for each sample was determined based on the combined chloroplast banding pattern at the three loci.

Each restriction site haplotype was subsequently sequenced for all regions. All sequences used in this study, including those used for determining the inheritance of chloroplasts in *Aesculus*, were submitted to GenBank (accession nos. AY968606–AY68671, Appendix). For the sequencing of *matK*, the primers matK3F, matK3R, and matK2R (Sang et al., 1997) were used in addition to matK1F and matK1R. For *trnD-trnT*, the primers of Demesure et al. (1995) as well as two internal primers designed for *Aesculus*, trnDT-F2 (5′-TGCCTCCTTGAAAGAGAGATG-3′) and trnDT-R2 (5′-CCGTTCGCAGA TTTTCAGAT-3′) were used. For sequencing of *trnH-trnK*, the primers of Demesure et al. (1995) and three internal primers designed for *Aesculus*, trnH2F (5′-ACTCGTATACACGAAGATCG-3′), trnH3F (5′-CTTATAGCCC CGTGTCAACC-3′), and trnK2R (5′-TGAACCCGTTTCTGGATCTC-3′) were used. *Aesculus glabra,* the fourth species of the monophyletic sect. *Pavia* (Xiang et al., 1998), was included in the study to serve as the outgroup for rooting of the haplotype genealogy.

Genealogy and nucleotide diversity—Haplotype sequences were aligned using ClustalX (Thompson et al., 1997), with some manual modifications. The aligned sequences were utilized in nucleotide diversity analysis and genealogy reconstruction. Genetic diversity was estimated with DnaSP 4.00 (Rozas et al., 2003) for each locus. Indices calculated include the total number of mutations (Eta), the average number of nucleotide substitutions between two sequences per site (π, Nei, 1987), and the average number of nucleotide differences between two sequences (κ, Tajima, 1983). The genealogy of haplotypes was estimated using two approaches, phylogenetic analysis and construction of a minimum spanning tree. Modeltest 3.6 (Posada and Crandall, 1998) was used to determine the model that best fit the evolution of the concatenated sequence data. The neighbor-joining (NJ) algorithm (Saitou and Nei, 1987) implemented in PAUP* (version 4.04b10; Swofford, 2003) was used to construct a tree based on the maximum likelihood distance method with the general-time-reversible (GTR) model (Rodríguez et al., 1990) indicated to best fit the data according to Modeltest results. Support for the tree was estimated using bootstrap analysis with 10 000 replicates (Felsenstein, 1985). The Bayesian analysis was conducted with the program MrBayes v. 3.0 (Huelsenbeck and Ronquist, 2001), while burn-in was determined using the program Tracer version 1.1 (Jamin and Lautenbaucher, 1993). The Bayesian analysis was run for 100 000 generations, well past the convergence point of 5700 generations, with trees sampled and saved every 100 generations. The trees were loaded into PAUP* (Swofford, 2003) to construct a majority rule consensus tree after removing the first 57 trees estimated prior to the convergence point. Frequency values of the trees serve as estimates of the posterior probability of nodes (Huelsenbeck and Ronquist, 2001). A minimum spanning tree (Excoffier and Smouse, 1994) was reconstructed for the haplotype sequences using the program MINSPNET packaged within Arlequin ver. 2.000 (Schneider et al., 2000) to compare with the genealogy derived from phylogenetic analysis.

Population genetic analyses—A data matrix of band presence/absence data from restriction site patterns was constructed to estimate *F* statistics and to examine the molecular variation using Arlequin ver. 2.000 following Schneider et al. (2000). F_{ST} values estimated by Arlequin were used to perform an

TABLE 1. Location of the populations of *Aesculus* that were sampled, along with the number of individuals sampled in the analysis, and the haplotypes observed in each population.

Taxon	Population	Location	No. individuals	Haplotype
A. flava (F)	F1[a]	Rabun Co., GA	8	10
	F2	Blount Co., TN (GSMNP)	3	4[b]
	F3	Blount Co., TN (GSMNP)	10	1[b], 3, 4[b]
	F4	Blount Co., TN (GSMNP)	10	1[b], 4[b]
	F5	Blount Co., TN (GSMNP)	8	1[b]
	F6	Blount Co., TN (GSMNP)	2	1[b]
	F7	Border of Blount Co., TN and Swain Co., NC	8	1[b]
	F8	Sevier Co., TN (GSMNP)	10	18
A. pavia (P)	P1[a]	Lee Co., AL	10	8[b,c]
	P2[a]	Lee Co., GA	10	5, 6[c]
	P3[a]	Twiggs Co., GA	5	9
	P4	Berkley Co., SC (FMNF)	11	20
	P5	Liberty Co., FL	11	15
	P6	Bibb Co., GA	10	14[c]
A. sylvatica (S)	S1[a]	Elbert Co., GA	11	7, 8[b,c]
	S2[a]	Gastonia, NC	9	19
	S3[a]	Jasper Co., GA	10	8[b,c]
	S4[a]	Monroe Co., GA	10	8[b,c]
	S5	Wake Co., NC	11	16
A. sylvatica × *flava*	SF1[a]	Banks Co., GA	10	11, 12, 13
(SF)	SF2?	Blount Co., TN (GSMNP)	7	17
	SF3?	Sevier Co., TN (GSMNP)	2	2
A. sylvatica × *pavia*	SP1[a]	Gwinnet Co., GA	5	6[c]
(SP)	SP2[a]	Butts Co., GA	11	8[b,c]
	SP3[a]	Catoosa Co., GA	10	21
	SP4[a]	Coweta Co., GA	10	8[b,c]
	SP5[a]	Dade Co., GA	5	6[c]
	SP6[a]	Hart Co., GA	15	8[b,c]
	SP7	Jasper Co., GA	10	8[b,c], 14[c]

Note: Identification of SF2 and SF3 from GSMNP was uncertain. The order of species in the hybrid populations does not imply pollen flow. AL, Alabama; FL, Florida; FMNF, Francis Marion National Forest; GA, Georgia; GSMNP, Great Smoky Mountains National Park; NC, North Carolina; SC, South Carolina; TN, Tennessee.

[a] Populations used in dePamphilis and Wyatt studies (1989, 1990).

[b] Haplotypes shared among populations.

[c] Haplotypes shared between hybrids and parental species.

analysis of molecular variance (AMOVA) (Weir and Cockerham, 1984; Excoffier et al., 1992; Weir, 1996). Several grouping strategies were implemented for the purpose of exploring the partitioning of genetic variation, with the first set including three groups, separated by physiogeographic region: the Coastal Plain, Piedmont, and Appalachian Mountains. For the second approach, the Piedmont was further divided into two groups, the Piedmont and the hybrid zone, with all other groups remaining the same. To test whether variation was partitioned among species, AMOVA was performed using both the parental species and hybrid taxa as groups, as well as with only the parental species as groups. Individual analyses for each parental and hybrid taxa were also performed separately, but these tests were lacking in statistical power, due to low sample size. To test for the effects of non-independence of restriction site bands from different enzymes, AMOVA was also performed with data from individual enzymes. The results did not differ significantly from the AMOVA results from data combined from all enzymes, thus are not presented here. Pairwise estimates of F_{ST} and the number of migrants per generation, *Nm*, (Slatkin, 1985) were estimated with Arlequin ver. 2.000 (Schneider et al., 2000) with 10 000 permutations and a significance level of $\alpha = 0.05$. Estimates of *Nm* were calculated according to the derivation of *Nm* for haploid genomes (Birky et al., 1989), where $Nm = 0.5(1/(F_{ST}-1))$.

TABLE 2. List of interspecific crosses among *Aesculus* spp. used in the analysis for inheritance of chloroplasts, and the number of polymorphisms in the *matK* gene among parental species.

Parental species (♀ × ♂)	No. crosses	No. polymorphisms
A. glabra var. *arguta* × *A. parviflora*	1	17
A. glabra var. *arguta* × *A. pavia*	1	3
A. glabra var. *arguta* × *A. sylvatica* var. *pubescens*	1	4
A. hippocastanum × *A. turbinata*	2	4
A. parviflora f. *serotina* × *A. parviflora*	1	5
A. pavia × *A. glabra* var. *arguta*	3	3
A. pavia × *A. flava*	1	1
A. pavia × *A. parviflora* f. *serotina*	1	18
A. pavia × *A. sylvatica* var. *pubescens*	2	7
A. pavia × *A. turbinata*	1	13

Note: The maternal species (♀, receiving the pollen) is listed first, and the paternal species (♂, pollen donor) is listed second. The number of polymorphisms between parental species is listed for each cross, with all of the progeny having the maternal sequence for each polymorphism.

RESULTS

Haplotype distribution and diversity—Restriction site analysis of the three loci, *matK*, *trnD-trnT*, and *trnH-trnK* identified 21 unique haplotypes (Appendix S1, see Supplemental Data accompanying online version of this article). A haplotype was designated as unique if the combined restriction fragment-banding pattern from all three loci was distinguishable from every other haplotype. In other words, each haplotype had at least one restriction site or length polymorphism distinguishing it from all other haplotypes. Seven haplotypes (5, 6, 8, 9, 14, 15, 20) were identified in *A. pavia*,

with six of them detected only in that species (5, 6, 9, 14, 15, 20). Four haplotypes (7, 8, 16, 19) were detected in *A. sylvatica*, of which three were apparently unique to the species (7, 16, 19). Five haplotypes (1, 3, 4, 10, 18) were found in *A. flava*, all of which were *A. flava* specific. Additionally, one haplotype (21) was found in only one *A. pavia* × *sylvatica* population (SP3), and five (2, 11, 12, 13, 17) were found only in *A. sylvatica* × *flava* populations (SF1–3). Two of the *A. pavia*-specific haplotypes (6, 14) occur in hybrid populations between *A. pavia* and *A. sylvatica* (SP1, SP5, and SP7), while no *A. flava*- or *A. sylvatica*-specific haplotypes were found in any hybrid populations. One haplotype (8) was common among *A. pavia*, *A. sylvatica*, and *A. pavia* × *sylvatica* hybrids, and widespread in the hybrid zone and surrounding areas (Fig. 2). Haplotypes 1, 3, 4, 17, and 18 were restricted to the Appalachian Mountains, while other haplotypes (i.e., 5, 7, 9, 10, 11, 12, 13, 15, 16, 19, 20, and 21) were unique to single populations throughout the Piedmont and Coastal Plain. Five haplotypes (1, 4, 6, 8, and 14) occurred in more than one population. Most populations are fixed for a single haplotype, with the exception of the six polymorphic populations, F3, F4, P2, S1, SF1, and SP7.

Within the 1860-bp *matK* gene, 25 sites were variable. The 1464-bp *trnD-trnT* intergenic spacer yielded a total of 26 variable sites, and 54 sites were variable in the *trnH-trnK* spacer. Nucleotide diversity indices averaged over all three loci (Table 3) reveal that, among species, *A. sylvatica* has the highest genetic diversity, followed by *A. flava*. It must be noted that estimates of nucleotide diversity of species may be influenced by sampling bias. Only small parts of the large distributional ranges of *A. pavia* and *A. flava* were sampled, whereas populations of *A. sylvatica* were sampled throughout its much less extensive species range (Fig. 1). When diversity is partitioned among regions, the Piedmont (which includes *A. sylvatica* and the hybrid zone) has the highest nucleotide diversity across all diversity indices, followed by the Appalachians. The diversity indices estimated for the hybrid zone include data from all individuals found within the hybrid

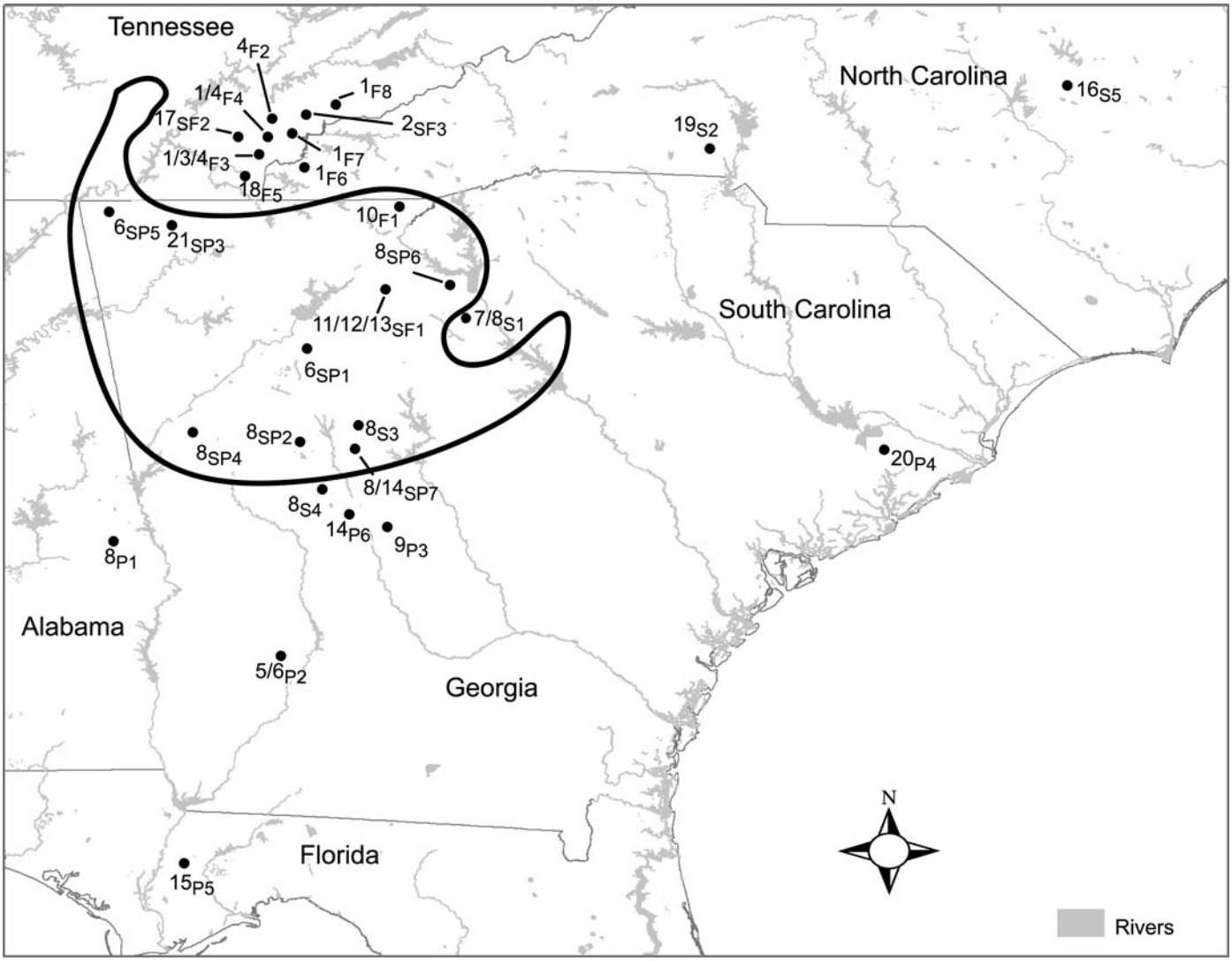

Fig. 2. Map of the distribution of 21 chloroplast haplotypes identified in *Aesculus*, with the population indicated as a subscript of the haplotype: F, *A. flava*; P, *A. pavia*; S, *A. sylvatica*; SF, *A. sylvatica* × *flava*; SP, *A. sylvatica* × *pavia*, and population designated as in Table 1. Haplotypes in the same population are separated by a backslash.

TABLE 3. Nucleotide and haplotype diversity indices, compared among geographic regions and *Aesculus* species estimated based on the haplotype sequences combined from *matK*, *trnD-trnT*, and *trnH-trnK*.

Region	No. haplotypes	$\pi \times 10^{-3}$	κ	Eta
Mountains	7	4.96	7.84	19.33
Piedmont	10	6.82	10.47	30.33
Hybrid zone	8	5.73	8.95	21.33
Coastal Plain	7	4.86	7.82	18.67
All	21	6.08	9.32	35.00
Species				
A. flava	5 (5)	5.97	9.47	19.33
A. pavia	7 (6)	4.86	7.82	18.67
A. sylvatica	4 (3)	8.23	12.94	24.33

Note: Nucleotide diversity measures include π, the average number of substitutions per site between two sequences; κ, the average number of nucleotide differences between two sequences; and Eta, the total number of mutations. Numbers in parentheses indicate the number of species-specific haplotypes.

zone. Nucleotide diversity indices for the Coastal Plain are identical to values for *A. pavia* because only *A. pavia* occurs in the Coastal Plain.

Genealogy—Phylogenetic analyses conducted with the concatenated DNA sequence data reveal little taxonomic or geographic structure. The Bayesian tree (not presented) is similar to the NJ tree (Fig. 3), with both revealing the following groups of haplotypes: strongly supported clades h_1-h_{18}-h_{15} from *A. flava*, *A. pavia*, and hybrids between *A. sylvatica* and *A. flava* (94% bootstrap value) and h_{10}-h_{11}-h_{13}-h_{12} from *A. flava* and hybrids between *A. sylvatica* and *A. flava* (100% bootstrap value), within which the latter three haplotypes are found only in hybrids of *A. sylvatica* and *A. flava* and form a clade with 97% bootstrap value (Fig. 3a). Three large clades with support of >50% bootstrap values were recognized (clades I–III in Fig. 3a). Each clade contains a mixture of *A. pavia*, *A. flava*, *A. sylvatica*, and hybrids. Clade I contains haplotypes 1, 18, 15, 16, and 21 from populations F3 to F8 from the Appalachians, P5 from the Florida panhandle, S5 from central North Carolina, and SP3 from the northern region of the hybrid zone (Fig. 3a). Clade II contains haplotypes 10–13, 19, and 20 from populations F1 and SF1 from northern Georgia, P4 haplotype from the coast of South Carolina, and S2 from western North Carolina. Clade III contains haplotypes 2–4, 6, 8, 14, and 17 from population F3 and questionable hybrid populations SF3 and SF2 in the Appalachians, along with P1, P2, P6, SP1, SP2, SP4, SP6, SP7, S1, S3, and S4 distributed throughout northern and central Georgia and eastern Alabama. The genealogical pattern from phylogenetic analysis (Fig. 3a) is largely concordant with the minimum spanning tree (Fig. 4) with the exception of the placement of haplotypes 17 and 1. Haplotype 17 is grouped with haplotype 2 in the phylogeny (Fig. 3a), but the two are separated in the minimum spanning tree (Fig. 4). The underlying cause for this discrepancy is unclear, but may result from the inherent assumption made in the construction of the minimum spanning tree that more prevalent and shared haplotypes should be closely grouped on the minimum spanning tree (Excoffier and Smouse, 1994). Because the minimum spanning tree is also designed to incorporate allele frequencies and population structure, the rarity of haplotype 2, which is found in two individuals in one population, is most likely affecting its placement on the minimum spanning tree. Nonetheless, the overall lack of spatial and taxonomic structure can be seen in both estimates of genealogical relationships. The minimum spanning tree suggests that three of the sampled haplotypes are ancestral (8, 6, 11) based on their internal positions (Castello and Templeton, 1994). Haplotype 8 is positioned internally and is connected to more haplotypes than other internal haplotypes on the minimum spanning tree (Fig. 4). This haplotype is also widely distributed among populations (Fig. 2) and potentially represents the most ancestral haplotype of those surveyed.

Population genetic analyses—Results of the AMOVA indicate that there is little cpDNA differentiation among regions and species (Tables 4, 5), congruent with the results of the genealogy (Fig. 3a). For instance, haplotypes from the same species or geographic regions do not group together in the same clade (Fig. 3a) nor are they closely positioned on the minimum spanning tree (Fig. 4). In all geographic groupings, most of the variation was found within regions and among populations (Table 4). For taxonomic groupings, the majority of genetic variation was partitioned among populations of species or hybrid taxa (Table 5). The high level of variation found among populations suggests that populations are relatively highly differentiated from each other, in agreement with the finding that most populations are apparently fixed for a single haplotype. In contrast, the low level of variation among species is congruent with the finding of lack of species structure from the genealogy analysis (Figs. 3, 4). The lack of geographic structure of haplotypes (Fig. 2) is also reflected in the low levels of variation partitioned among regions in the AMOVA analysis (Table 4).

All pairwise F_{ST} estimates were significant at the $\alpha = 0.05$ level. Pairwise F_{ST} estimates, along with the estimates of Nm (Table 6) revealed that the number of migrants per generation is greater than that would be expected ($Nm = 1$) if allele frequencies were governed solely by random genetic drift (Wright, 1931). Furthermore, more migration events seem to be occurring or have occurred between *A. sylvatica* and hybrid populations than between *A. pavia* and hybrid populations, or between *A. flava* and hybrid populations. This is expected given the close geographic proximity of *A. sylvatica* populations to hybrid populations.

DISCUSSION

Evidence of ancestral chloroplast polymorphism and of historical secondary contact via seeds—Results from the controlled hybridization study (Appendix S2, see Supplemental Data accompanying online version of this article) confirm the maternal inheritance of cpDNA in *Aesculus*. This evidence permits us to determine the maternal parent of hybrid populations and to trace the pattern of seed flow. Given the purported limited dispersal distance of *Aesculus* seeds as contrasted with the potential for long-distance pollen dispersal, we expect to see a genealogy and genetic structure in which haplotypes cluster according to species or geographic regions and a pattern of genetic variation showing significant population structure. This pattern was not seen; the polymorphisms do not cluster according to taxonomic groups or geographic regions (i.e., Appalachian Mountains, Piedmont, and Coastal Plain) (Figs. 3, 4; Tables 4, 5). This lack of structure suggests either ancestral cpDNA polymorphism in sect. *Pavia* (e.g., divergent chloroplast DNA haplotypes within

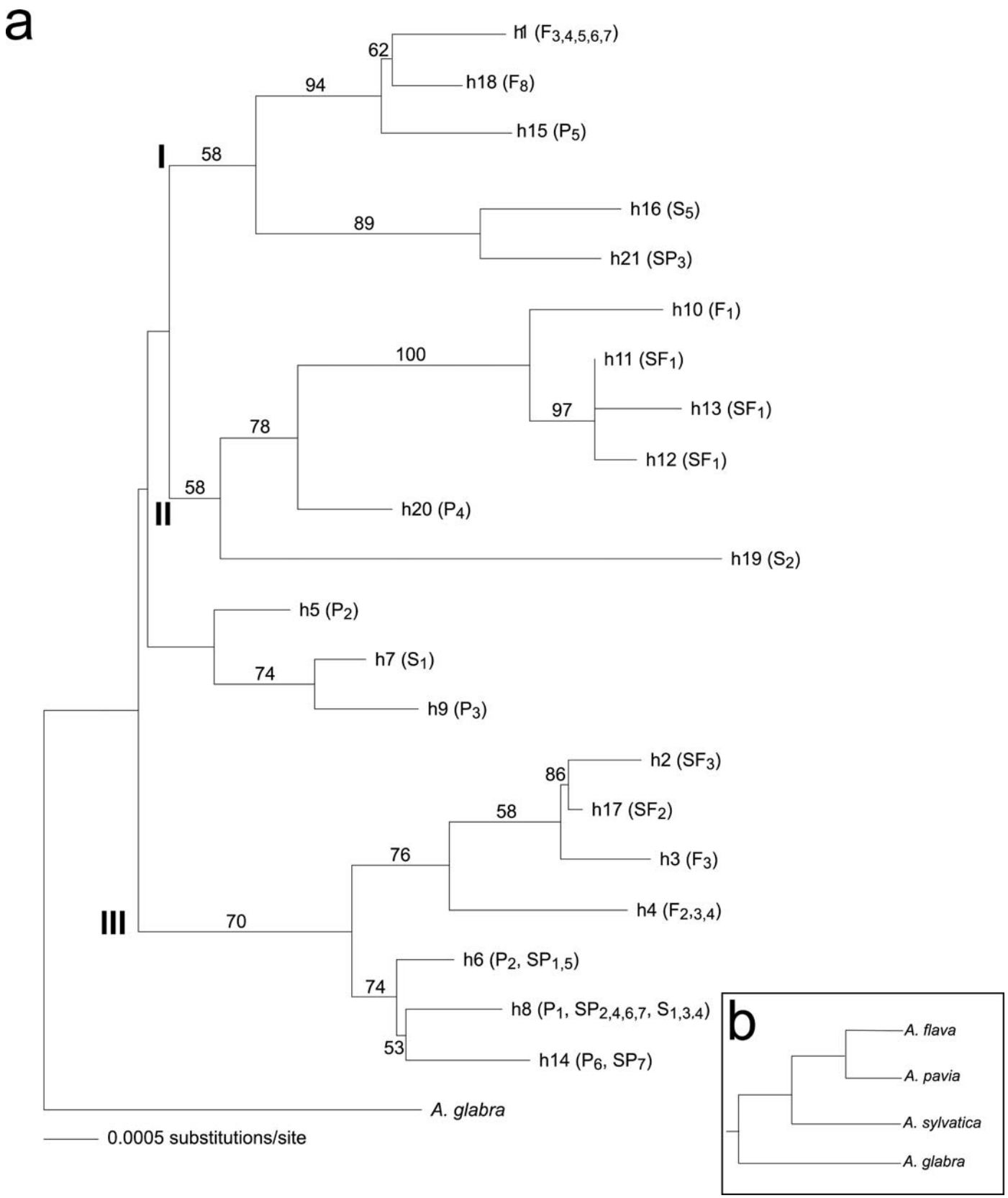

Fig. 3. (a) Neighbor-joining tree constructed from *matK*, *trnD-trnT*, and *trnH-trnK* sequence data for individual haplotypes. Bootstrap values are indicated above the branch. Taxa are indicated after haplotype: *Aesculus pavia* (P), *A. sylvatica* (S), *A. flava* (F), *A. sylvatica* × *flava* (SF) and *A. sylvatica* × *pavia* (SP). Three major clades have been indicated: I, II, and III. (b) Representation of morphology (Forest et al., 2001) and ITS/morphology phylogeny of Xiang et al. (1998).

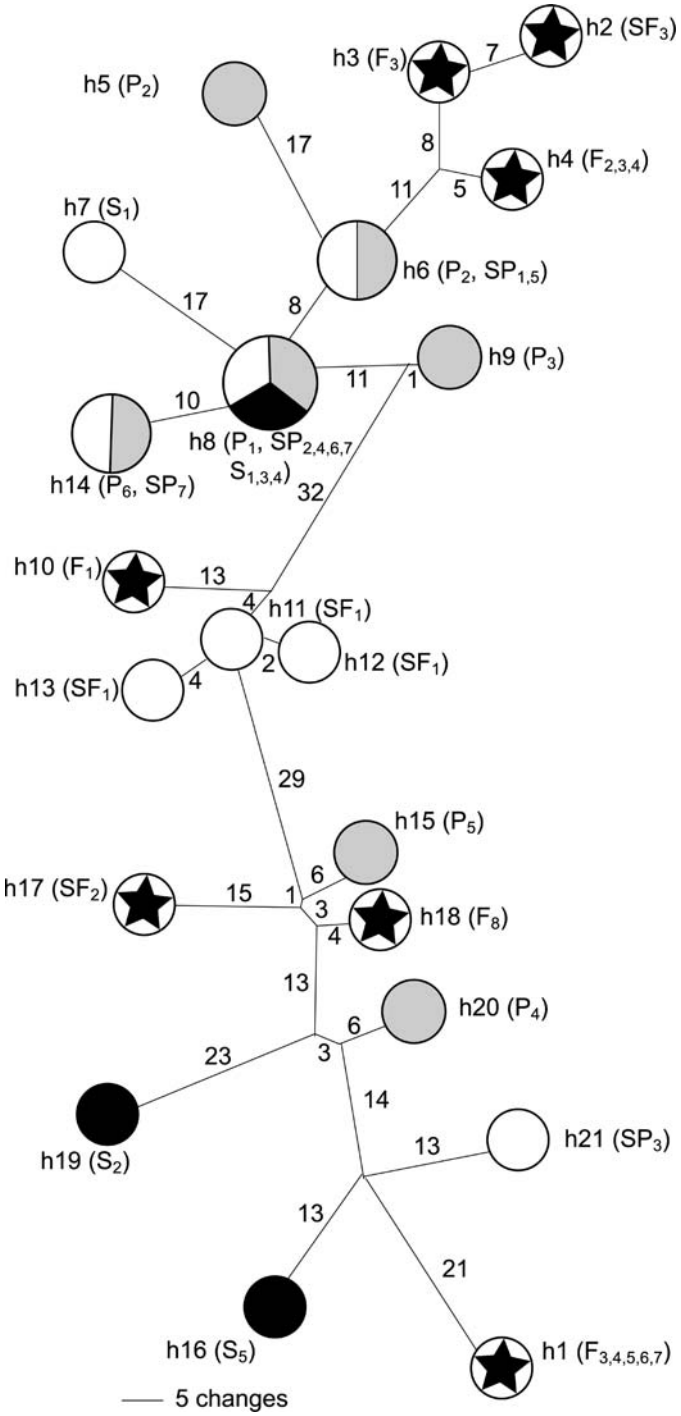

Fig. 4. Minimum spanning tree of restriction fragment length polymorphism haplotypes identified in *Aesculus* with corresponding populations indicated. Haplotypes are coded on the basis of geographical region: hybrid zone (white), Coastal Plain (gray), Piedmont (black, not within hybrid zone), and Appalachian Mountains (star). Numbers next to branches indicate the number of unobserved haplotypes between observed haplotypes.

species were inherited from a polymorphic common ancestor of the species in sect. *Pavia*), or extensive ancient cpDNA capture among species. The ancestral polymorphism hypothesis seems more plausible based on congruencies of the species phylogeny (Fig. 3b) to clades I and II of the cpDNA genealogy (Fig. 3a). Within clades I and II (Fig. 3a), it appears that *A. sylvatica* diverged first, followed by *A. pavia* and then *A. flava*. This branching order is consistent with that found in the ITS/morphology phylogenies (Fig. 3b) constructed by Xiang et al. (1998) and Forest et al. (2001). The branching order of clade III is not clearly consistent with the ITS/morphology species phylogeny. However, based on the presence of three major clades, each containing all three species of *Aesculus* in the NJ tree (Fig. 3a), it is reasonable to conclude that at least three different cytotypes were present in the most recent common ancestor of *A. sylvatica*, *A. pavia*, and *A. flava*. Chloroplast capture would have occurred randomly and thus concordance of plastid and ITS phylogenies is not expected. Additionally, invoking a hybridization scenario to explain the present geographical distribution of haplotypes necessitates the assumption of extensive ancient gene flow via seeds across a broad geographic scale. Otherwise, it is difficult to account for the grouping of widely disjunct populations from different species into a single clade (Fig. 3a; P5-F8-S5; P4-F1-S2). The results of the AMOVA analysis indicate that, while there is little differentiation at the broadest spatial scale (i.e., Appalachian Mountains, Piedmont, and Coastal Plain), populations within these regions have a high degree of genetic isolation. This is congruent with the expectation of limited seed dispersal and ancestral polymorphism of cpDNA in each species. The isolation of cpDNA among populations is further corroborated in the distributional pattern of haplotypes; most populations are fixed for one haplotype, and most haplotypes are not shared among populations (Fig. 2, Table 1). The lack of spatial structure across species (Fig. 2, Table 5) and physiogeographic provinces (Table 4) further confirms that lineage sorting and ancestral chloroplast polymorphism are affecting the pattern and structure of variation in the zone. The combined results of the AMOVA and genealogy further support the occurrence of ancestral polymorphism of cpDNA and limited seed dispersal on a broad spatial scale (i.e., across physiogeographic provinces) as well as on a less extensive, between-population scale. Ancestral polymorphism of cpDNA has resulted in an inflated estimate of the migration rates of seeds (Table 6), which is derived from F_{ST} values. The inflated estimate illustrates an inherent limitation of F statistics, which cannot distinguish between gene flow and shared common ancestry because they are based solely on genetic similarities and do not take into consideration population history. In the case of the *Aesculus* hybrid zone, the occurrence of particular cytotypes among species and different spatial regions is likely a result of gene flow, while the sharing of closely related, but not identical, cytotypes among species and regions is the product of shared common ancestry.

Under the HSC hypothesis, we expect to observe that some *A. pavia* × *sylvatica* populations share haplotypes with *A. pavia* or have cpDNA haplotypes grouped with *A. pavia* in the genealogy. The sharing of haplotype 6 between *A. pavia* (P2) and *A. pavia* × *sylvatica* populations (SP1 and SP5) and haplotype 14 between *A. pavia* (P6) and *A. pavia* × *sylvatica* (SP7) (Fig. 3a, clade III) strongly supports the HSC hypothesis and further implies that *A. pavia* was the maternal parent of these hybrid populations. This evidence indicates that *A. pavia* was historically present in the hybrid zone and had contact with *A. sylvatica*. Admixture analysis of allozyme loci (dePamphilis and Wyatt, 1990) indicated that the nuclear genome of the hybrid populations SP1 and SP5 is composed mainly of *A.*

TABLE 4. Results of the analysis of molecular variance (AMOVA) of *Aesculus* populations for the entire southeastern United States among physiogeographic regions.

Source of variation	df	SS	Variance components	% of total variance	*P*
CP + P + MTN					
Among regions	2	14.497	0.048	9.99	0.009
Among populations within regions	26	85.255	0.382	80.24	<0.001
Within populations	219	10.184	0.047	9.77	<0.001
Total	247	109.936	0.476		
CP + P + HZ + MTN					
Among regions	3	21.763	0.065	13.63	0.002
Among populations within regions	25	77.989	0.366	76.64	<0.001
Within populations	219	10.184	0.065	9.73	<0.001
Total	247	109.936	0.496		

Note: CP, Coastal Plain; P, Piedmont; MTN, the Appalachian Mountains; and HZ, the hybrid zone.

sylvatica alleles, suggesting that while *A. pavia* was the maternal parent of these populations, the *A. pavia* nuclear genes in these hybrid populations have been diluted through backcrosses to *A. sylvatica* by way of localized pollen dispersal.

Haplotype 8 is shared by most hybrid *A. sylvatica* × *pavia* populations (SP 2, 4, 6, 7); it is also found in populations P1 and S1, 3 and 4. Thus the maternal parent of these hybrid populations may be *A. pavia*, *A. sylvatica*, or both. Given that haplotype 8 is closely related to haplotype 6 and haplotype 14 of *A. pavia*, it appears likely that haplotype 8 originated from *A. pavia* but that the cpDNA donor of haplotype 8 to the hybrid populations most likely varies, depending on the distance of hybrid populations to potential donor populations. Allozyme data (dePamphilis and Wyatt, 1990) suggests a contribution of both *A. pavia* and *A. sylvatica* to the nuclear genome of hybrid individuals in these populations. Additionally, the widespread distribution of haplotype 8 in the hybrid zone suggests that it may be favored by exogenous environmental conditions in the Piedmont.

The remaining hybrid population (SP3) is closely related to *A. sylvatica* (S5) (Fig. 3a), suggesting that *A. sylvatica* is the cpDNA donor to this population. However, a large proportion (0.676) of SP3 contains *A. pavia* nuclear alleles (dePamphilis

TABLE 5. Results of analysis of molecular variance (AMOVA), grouped according to species (*Aesculus flava*, *A. sylvatica*, and *A. pavia*) and hybrid taxa (*A. sylvatica* × *pavia* and *A. sylvatica* × *flava*).

Source of variation	df	SS	Variance components	% of total variance	*P*
Parental species and hybrid taxa					
Among species	4	21.947	0.039	8.40	0.021
Among populations within species	24	77.804	0.381	81.64	<0.001
Within populations	219	10.184	0.047	9.96	<0.001
Total	247	109.935	0.467		
Parental species only					
Among species	2	11.924	0.036	7.45	0.037
Among populations within species	16	57.173	0.414	84.82	<0.001
Within populations	146	5.509	0.038	7.73	<0.001
Total	164	74.606	0.488		
A. flava					
Among populations	7	17.893	0.345	82.61	<0.001
Within populations	51	3.700	0.073	17.39	NS
Total	58	21.593	0.417		
A. pavia					
Among populations	5	21.973	0.482	96.33	<0.001
Within populations	49	0.900	0.018	3.67	NS
Total	54	22.873	0.500		
A. sylvatica					
Among populations	4	17.307	0.423	95.53	<0.001
Within populations	46	0.909	0.020	4.47	NS
Total	50	18.216	0.443		
A. sylvatica × *flava*					
Among populations	2	4.147	0.347	66.47	<0.001
Within populations	16	2.800	0.175	33.53	NS
Total	18	6.947	0.522		
A. sylvatica × *pavia*					
Among populations	6	16.484	0.303	90.22	<0.001
Within populations	57	1.875	0.033	9.78	NS
Total	63	18.359	0.333		

Note: NS indicates that the *P*-value was non-significant.

TABLE 6. Pairwise estimates of *Nm* and F_{ST} values between *Aesculus* species and hybrids.

Species (*Nm*)	*A. pavia*	*A. sylvatica*	*A. flava*	*A. p.* × *s.*	*A. s.* × *f.*
A. pavia (0.00953)		0.148	0.221	0.184	0.186
A. sylvatica (0.01168)	2.880		0.279	0.126	0.253
A. flava (0.05262)	1.760	1.291		0.351	0.262
A. p. × *s.* (0.02710)	2.217	3.457	0.924		0.343
A. s. × *f.* (0.12617)	2.190	1.478	1.406	0.958	

Note: All F_{ST} values are significant on the $\alpha = 0.05$ level. F_{ST} values are indicated in the upper diagonal, while *Nm* estimates are indicated in the lower diagonal. *Nm* values for within species are indicated in parentheses. *f., flava*; *p., pavia*; *s., sylvatica*.

and Wyatt, 1990), suggesting that *A. pavia* has contributed significantly to this population via long-distance pollen dispersal.

A possible alternative explanation to secondary contact for the aforementioned observations of shared chloroplast haplotypes between species, of the ancestral positions of hybrid haplotypes (h8, 11, Fig. 4) and for the presence of each species on all major clades in the NJ tree is that the hybrid zone is a zone of primary intergradation of a single species currently in the process of speciation. Primary intergradation would occur as the zone was formed in response to selection across continuous populations of the species prior to speciation (Harrison, 1990, 1993). Evidence from morphology, hybrid fitness values, and DNA sequence divergence between the species strongly supports the secondary contact hypothesis and refutes the primary intergradation hypothesis. The combination of morphological features distinguishing the four species (Hardin, 1957c; dePamphilis and Wyatt, 1989) and evidence of lower relative fitness in the form of reduced pollen viability for hybrids in contrast to the "pure species" (dePamphilis and Wyatt, 1989) offer compelling testimony for secondary contact rather than primary intergradation. According to a calibrated ITS clock of $1.722 +- 0.21 \times 10^{-9}$ nucleotide substitutions per site per year (Xiang et al., 1998), *A. flava* and *A. sylvatica* diverged about 10.2 mya, while *A. flava* and *A. pavia* diverged ~5.1 mya, suggesting that they had diverged long before the climatic oscillations of the Pleistocene (~1.8 mya).

Possible long-distance downstream seed dispersal—dePamphilis and Wyatt (1989) theorized that the genetic presence of *A. flava* in the hybrid zone was due to dispersal of seeds downstream. However, the terrain of the Appalachian Mountains and the flow of rivers in the southeastern United States (Fig. 2) must also be taken into account. *Aesculus flava* seeds that fall on the western side of the mountain range will flow west-southwest into the Mississippi River and downstream into the Gulf of Mexico, thus bypassing the central and eastern portions of the hybrid zone. Seeds falling on the eastern side of the mountain will fall into rivers such as the Nantahala that flow into central and eastern portions of the hybrid zone. However, populations situated farther north in North Carolina would be most likely to drop their seeds into rivers flowing into the Atlantic Ocean, rather than through the hybrid zone. While it is possible that gene flow by pollen dispersal may occur between populations of *A. flava* situated throughout the mountains, claims of gene flow via seeds are unsubstantiated. The hybrid *A. sylvatica* × *flava* population (SF1) in the hybrid zone is not closely related to those in the mountains (Fig. 3a), suggesting that long-distance pollen dispersal on a broad geographic scale can not be verified from the results of this study. In the dePamphilis and Wyatt (1989) study, populations of *A. flava* sampled in northern Georgia could have contributed to the hybrid zone through dispersal of seeds downstream because they are situated on the eastern slope of the Appalachian Mountains. In our study, all populations of *A. flava* sampled, with the exception discussed next, were situated on the westward side of the Appalachian Mountains. In contrast, any close affinity of *A. flava* haplotypes with those in the central and eastern parts of the hybrid zone or with *A. sylvatica* and *A. pavia* detected in our study may be due to the past sympatry of these species, most likely in the Piedmont. One possible exception to this generalization is the population of *A. flava* in Rabun County in northern Georgia, where its seeds could fall into rivers in the Nantahala watershed. In accordance with this argument, one hybrid specific haplotype (11 in SF1) is located at an internal position on the haplotype network, giving rise to the haplotypes found in a "pure" *A. flava* population (10 in F1, Fig. 4). This suggests that the donor population of this haplotype was once present in the hybrid zone and has become extinct or was not sampled, or that seed dispersal has occurred across a more limited geographic scale.

Conclusions—Our analysis of genealogical relationships, spatial distribution of cpDNA haplotypes, measures of diversity, and partitioning of cpDNA variation has enabled us to address key issues regarding the origin and present structure of the *Aesculus* hybrid zone. Our results suggest ancestral polymorphism of the chloroplast genome of *Aesculus* sect. *Pavia* and abundant cpDNA diversity in the *Aesculus* hybrid zone. A small number of haplotypes observed in hybrid populations may no longer be present in parental populations. The observed patterns of genetic variation and sharing of chloroplast haplotypes between *A. pavia* and hybrid populations support Hardin's (1957c) hypothesis of historical secondary contact between *A. pavia* and *A. sylvatica* in the Piedmont. In addition to the evidence from allozyme studies (dePamphilis and Wyatt, 1989, 1990), which support the long-distance dispersal hypothesis, the cumulative evidence supports the idea that secondary contact of the species during the Pleistocene, followed by both local gene flow from *A. sylvatica* and long-distance gene flow via *A. pavia* pollen, were the formative and maintaining mechanisms for the *Aesculus* hybrid zone. Although it is not uncommon for hybrid zones to have been formed via secondary contact in the Pleistocene (Barton and Hewitt, 1985), evidence of mechanisms for their stability is often lacking (Harrison, 1990). If ruby-throated hummingbirds are in fact responsible for the long-distance pollen dispersal events, it is also possible that the ongoing annual hummingbird migration continues to be an active force in maintaining and perhaps even in expanding the hybrid zone in *Aesculus* and other plant species in the eastern United States. Results from our study suggest that the relationship between vertebrate

migration patterns and the genetic structure of plant species should be investigated further to gain insight into the forces that shape the genetic architecture of hybrid zones.

LITERATURE CITED

ARNOLD, M. L. 1997. Natural hybridization and evolution. Oxford University Press, Oxford, UK.

AVISE, J. C. 2004. Molecular markers, natural history, and evolution, 2nd ed. Sinauer, Sunderland, Massachusetts, USA.

AVISE, J. C., J. ARNOLD, R. M. BALL, E. BERMINGHAM, T. LAMB, J. E. NEIGEL, C. A. REEB, AND N. C. SAUNDERS. 1987. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics* 18: 489–522.

BARTON, N. H., AND G. M. HEWITT. 1985. Analysis of hybrid zones. *Annual Review of Ecology and Systematics* 16: 113–148.

BEHEREGAY, L. B., J. P. GIBBS, N. HAVILL, T. H. FRITTS, J. R. POWELL, AND A. CACCONE. 2004. Giant tortoises are not so slow: rapid diversification and biogeographic consensus in the Galápagos. *Proceedings of the National Academy of Sciences, USA* 101: 6514–6519.

BIRKY, C. W. 2001. The inheritance of genes in mitochondria and chloroplasts: laws, mechanisms, and models. *Annual Review of Genetics* 35: 125–148.

BIRKY, C. W., P. FUERST, AND T. MARUYAMA. 1989. Organelle gene diversity under migration, mutation and drift: equilibrium expectations, approach to equilibrium, effects of heteroplasmic cells, and comparisons to nuclear genes. *Genetics* 121: 613–627.

CASTELLO, J., AND A. R. TEMPLETON. 1994. Root probabilities for intraspecific gene trees under neutral coalescent theory. *Molecular Phylogenetics and Evolution* 3: 102–113.

CULLINGS, K. W. 1992. Design and testing of a plant-specific PCR primer for ecological and evolutionary studies. *Molecular Ecology* 1: 233–240.

DEMESURE, B., N. SODZI, AND R. J. PETIT. 1995. A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants. *Molecular Ecology* 4: 129–131.

DEPAMPHILIS, C. W., AND R. WYATT. 1989. Hybridization and introgression in buckeyes (Aesculus: Hippocastanaceae): a review of the evidence and a hypothesis to explain long-distance gene flow. *Systematic Botany* 14: 593–611.

DEPAMPHILIS, C. W., AND R. WYATT. 1990. Electrophoretic confirmation of interspecific hybridization in Aesculus (Hippocastanaceae) and the genetic structure of a broad hybrid zone. *Evolution* 44: 1295–1317.

ELLSTRAND, N. C., R. WHITKUS, AND L. H. RIESEBERG. 1996. Distribution of spontaneous plant hybrids. *Proceedings of the National Academy of Sciences, USA* 93: 5090–5093.

EXCOFFIER, L., P. E. SMOUSE, AND J. M. QUATTRO. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131: 479–491.

EXCOFFIER, L., AND P. E. SMOUSE. 1994. Using allele frequencies and geographic subdivision to reconstruct gene trees within a species: molecular variance parsimony. *Genetics* 136: 343–359.

FAN, C., AND J. Q.-Y. XIANG. 2003. Phylogenetic analyses of Cornales based on 26S rRNA and combined 26S rDNA-*matK*-*rbcL* sequence data. *American Journal of Botany* 90: 1357–1372.

FANG, W. P. 1981. Hippocastanaceae. *In* W. P. Fang [ed.], Flora republicae popularis sinicae, 274–289. Science Press, Beijing, China.

FELSENSTEIN, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39: 783–791.

FOREST, F., J. N. DROUIN, R. CHAREST, L. BROUILLET, AND A. BRUNEAU. 2001. A morphological phylogenetic analysis of *Aesculus* L. and *Billia* Peyr. (Sapindaceae). *Canadian Journal of Botany* 79: 154–169.

GARCÍA-PARIS, M., D. A. GOOD, G. PARRA-OLEA, AND D. B. WAKE. 2000. Biodiversity of Costa Rican salamanders: implications of high levels of genetic differentiation and phylogeographic structure for species formation. *Proceedings of the National Academy of Sciences, USA* 97: 1640–1647.

HARDIN, J. W. 1957a. A revision of the American Hippocastanaceae, II. *Brittonia* 9: 173–195.

HARDIN, J. W. 1957b. Studies in the Hippocastanaceae. III. A hybrid swarm in the buckeyes. *Rhodora* 59: 45–51.

HARDIN, J. W. 1957c. Studies in the Hippocastanaceae. IV. Hybridization in *Aesculus*. *Rhodora* 59: 185–203.

HARRIS, S. A., AND R. INGRAM. 1991. Chloroplast DNA and biosystematics: the effects of intraspecific diversity and plastid transmission. *Taxon* 40: 393–412.

HARRISON, R. G. 1990. Hybrid zones: windows on evolutionary processes. *In* J. Antonovics and D. Futuyama [eds.], Oxford surveys in evolutionary biology, 71–128. Oxford University Press, Oxford, UK.

HARRISON, R. G. 1993. Hybrids and hybrid zones: historical perspective. *In* R. G. Harrison [ed.], Hybrid zones and the evolutionary process, 3–12. Oxford University Press, Oxford, UK.

HEWITT, G. M. 1993. After the ice: *Parallelus* meets *Erythropus* in the Pyrenees. *In* R. G. Harrison [ed.], Hybrid zones and the evolutionary process, 140–164. Oxford University Press, Oxford, UK.

HEWITT, G. M. 1996. Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society* 58: 247–276.

HEWITT, G. M. 2000. The genetic legacy of the Quaternary ice ages. *Nature* 405: 907–913.

HEWITT, G. M. 2001. Speciation, hybrid zones and phylogeography—or seeing genes in space and time. *Molecular Ecology* 10: 537–549.

HUELSENBECK, J., AND F. RONQUIST. 2001. MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics* 17: 754–755.

JAMIN, M., AND M. E. LAUTENBACHER. 1993. TRACER, version 1.1: a Mathematica package for gamma-algebra in arbitrary dimensions. *Computer Physics Communications* 74: 265–268.

MCCAULEY, D. E. 1995. The use of chloroplast DNA polymorphism in studies of gene flow in plants. *Trends in Ecology and Evolution* 10: 198–202.

MORGAN, D. R., AND D. E. SOLTIS. 1993. Phylogenetic relationships among members of Saxifragaceae sensu lato based on *rbc*L sequence data. *Annals of the Missouri Botanical Garden* 80: 631–660.

NEI, M. 1987. Molecular evolutionary genetics. Columbia University Press, New York, New York, USA.

OUBORG, N. J., Y. PIQUOT, AND J. M. VAN GROENENDAEL. 1999. Population genetics, molecular markers, and the study of dispersal in plants. *Journal of Ecology* 87: 551–568.

POSADA, D., AND K. A. CRANDALL. 1998. Modeltest: testing the model of DNA substitution. *Bioinformatics* 14: 817–818.

REMINGTON, C. 1968. Suture-zones of hybrid interaction between recently joined biotas. *Evolutionary Biology* 2: 321–428.

RIESEBERG, L. H. 1997. Hybrid origins of plant species. *Annual Review of Ecology and Systematics* 28: 359–389.

RIESEBERG, L. H., AND J. F. WENDEL. 1993. Introgression and its consequences in plants. *In* R. G. Harrison [ed.], Hybrid zones and the evolutionary process, 71–109. Oxford University Press, Oxford, UK.

RODRÍGUEZ, F., J. L. OLIVER, A. MARÍN, AND J. R. MEDINA. 1990. The general stochastic model of nucleotide substitution. *Journal of Theoretical Biology* 142: 485–501.

ROZAS, J., J. C. SÁNCHEZ-DELBARRIO, X. MESSEGUER, AND R. ROZAS. 2003. DnaSP, DNA polymorphism analysis by the coalescent and other methods. *Bioinformatics* 19: 2496–2497.

SAITOU, N., AND M. NEI. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4: 406–425.

SANG, T., D. J. CRAWFORD, AND T. F. STUESSY. 1997. Chloroplast DNA phylogeny, reticulate evolution and biogeography of *Paeonia* (Paeoniaceae). *American Journal of Botany* 84: 1120–1136.

SCHAAL, B. A., D. A. HAYWORTH, K. M. OLSEN, J. T. RAUSCHER, AND W. A. SMITH. 1998. Phylogeographic studies in plants: problems and prospects. *Molecular Ecology* 7: 465–474.

SCHNEIDER, S., D. ROESSLI, AND L. EXCOFFIER. 2000. Arlequin ver. 2.000:

a software for population genetics data analysis. Genetics and Biometry Laboratory, University of Geneva, Geneva, Switzerland.

SCHOPMEYER, C. S. 1974. Seeds of woody plants in the United States. USDA handbook no. 450, Washington D.C., USA.

SEWELL, M. M., Y. L. QIU, C. R. PARKS, AND M. W. CHASE. 1993. Genetic evidence for trace paternal transmission of plastids in *Liriodendron* and *Magnolia* (Magnoliaceae). *American Journal of Botany* 80: 854–858.

SLATKIN, M. 1985. Rare alleles as indicators of gene flow. *Evolution* 39: 53–65.

STACE, C. A. 1987. Hybridization and the plant species. *In* K. M. Urbanska [ed.], Differentiation patterns in higher plants, 115–127. Academic Press, New York, New York, USA.

SWOFFORD, D. L. 2003. PAUP*: phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer, Sunderland, Massachusetts, USA.

TAJIMA, F. 1983. Evolutionary relationships of DNA sequences in finite populations. *Genetics* 105: 437–460.

THOMPSON, J. D., T. J. GIBSON, F. PLEWNIAK, F. JEANMOUGIN, AND D. G. HIGGINS. 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* 24: 4876–4882.

WAGNER, W. H. 1969. The role and taxonomic treatment of hybrids. *Bioscience* 19: 785–789.

WAGNER, W. H. 1970. Biosystematics and evolutionary noise. *Taxon* 19: 146–151.

WEIR, B. S. 1996. Genetic data analysis II: methods for discrete population genetic data. Sinauer, Sunderland, Massachusetts, USA.

WEIR, B. S., AND C. C. COCKERHAM. 1984. Estimating *F*-statistics for the analysis of population structure. *Evolution* 38: 1358–1370.

WRIGHT, S. 1931. Evolution in Mendelian populations. *Genetics* 16: 97–159.

XIANG, J. Q.-Y., D. J. CRAWFORD, A. D. WOLFE, Y. -C. TANG, AND C. W. DEPAMPHILIS. 1998. Origin and biogeography of *Aesculus* L. (Hippocastanaceae): a molecular phylogenetic perspective. *Evolution* 52: 988–997.

APPENDIX. Voucher information, GenBank accession numbers, source, and origin for taxa used for (A) cpDNA inheritance and for (B) restriction site analysis in this study. A dash (—) indicates that the information is missing or was not sampled. Source indicated as cultivated (Cult.), a cross between cultivated individuals (Cult. cross), or wild-collected (wild). Origin indicates arboretum from which fresh material was obtained; HA = Holden Arboretum, JCRA = J.C. Raulston Arboretum, MA = Morton Arboretum. For taxa sampled in the restriction site analysis, the county of origin is listed in parentheses after the accession number. In some instances, multiple individuals from a particular taxa were sampled for each locus. For each species, commas separate accession numbers referring to the same locus, while semicolons separate loci.

A) cpDNA inheritance

Taxon; *matK* (accession numbers listed here if multiple individuals of the same species were sampled); Source; Origin (accession number listed here if only one individual was sampled).

Aesculus flava[a] Ait.; AY968638; wild; HA 84-570.

A. glabra Willd. **var. *arguta*** (Buckl.); AY968624; Cult.; HA 58-522.

A. hippocastanum L.; AY968630 (LCSHB), AY968631 (64-251); Cult.; HA.

A. parviflora Walt.; AY968627 (74-72A), AY968628 (68-62A); Cult.; HA.

A. parviflora* f. *serotina Rehd.; AY968629; Cult.; HA 69-360.

***A. pavia* 'Humilis'**; AY968625; Cult.; HA 55-917.

***A. pavia* 'Humilis'**; AY968626; Cult.; HA 75-237.

A. sylvatica Bartr. **var. pubescens**; AY968637; Cult.; HA 62-265.

A. turbinata Blume; AY968632 (236-82), AY968633 (664-56); Cult.; MA.

A. glabra* var. *arguta* × *A. parviflora; AY968634; Cult. cross; HA 96017.

***A. glabra* var. *arguta* × *A. pavia* 'Humilis'**; AY968641; Cult. cross; HA 96016.

A. glabra* var. *arguta* × *A. sylvatica* var. *pubescens; AY968645; Cult. cross; HA 97115.

A. hippocastanum* × *A. turbinata; AY968636; Cult. cross; HA 98122.

A. hippocastanum* × *A. turbinata; AY968648; Cult. cross; HA 97131.

A. parviflora* f. *serotina* × *A. parviflora; AY968644; Cult. cross; HA 96047.

A. pavia* 'Humilis' × *A. flava; AY968647; Cult. cross; HA 97130.

A. pavia* 'Humilis' × *A. glabra* var. *arguta; AY968639; Cult. cross; HA 94055.

A. pavia* 'Humilis' × *A. glabra* var. *arguta; AY968640; Cult. cross; HA 95025.

A. pavia* 'Humilis' × *A. glabra* var. *arguta; AY968642; Cult. cross; HA 96024.

A. pavia* 'Humilis' × *A. parviflora* f. *serotina; AY968635; Cult. cross; HA 96020.

A. pavia* 'Humilis' × *A. sylvativa* var. *pubescens; AY968643; Cult. cross; HA 96032.

A. pavia* 'Humilis' × *A. sylvatica* var. *pubescens; AY968646; Cult. cross; HA 97129.

A. pavia* 'Humilis' × *A. turbinata; AY968649; Cult. cross; HA 97132.

B) Restriction site analysis

Taxon; *matK* (county, state); *trnD-trnT* (county); *trnH-trnK* (county, state); Source; Origin (when applicable).

A. flava Ait.; AY968668 (Blount Co., TN), AY968670 (Rabun Co., GA); AY968610 (Blount Co., TN), AY968612 (Blount Co., TN), AY968613 (Border of Blount Co., TN and Swain Co., NC), AY968617 (Rabun Co., GA); AY968650 (Blount Co., TN), AY968656 (Rabun Co., GA), AY968657 (Sevier Co., TN); wild.

A. glabra Willd. **var. *arguta*** (Buckl.); AY968671; AY968623; AY968663; Cult.; JCRA 960612.

A. pavia L.; AY968667 (Lee Co., AL); AY968606 (Berkley Co., SC), AY968608 (Bibb Co., GA), AY968607 (Liberty Co., FL), AY968620 (Twiggs Co., GA); AY968653 (Berkley Co., SC) AY968651 (Lee Co., AL), AY968654 (Liberty Co., FL), AY968652 (Twiggs Co., GA); wilds.

A. sylvatica Bartr.; AY968665 (Gastonia, NC), AY968666 (Wake Co., NC); AY968616 (Gastonia, NC), AY968615 (Jasper Co., GA), AY968614 (Wake Co., NC); AY968658 (Gastonia, NC), AY968655 (Wake Co., NC); wild.

***Aesculus* sp. (*flava* × *sylvatica*)**; —; AY968619 (Banks Co., GA), AY968621 (Banks Co., GA), AY968622 (Banks Co., GA), AY968611 (Blount Co., TN); AY968659 (Banks Co., GA), AY968660 (Banks Co., GA), AY968661 (Banks Co., GA), AY968664 (Blount Co., TN); wild.

***Aesculus* sp. (*pavia* × *sylvatica*)**; AY968669 (Hart Co., GA); AY968618 (Dade Co., GA), AY968609 (Jasper Co., GA); AY968662 (Catoosa Co., GA); wild.

[a] Possible hybrid between *A. flava* and *A. pavia*