

CHLOROPLAST DNA PHYLOGENETICS OF THE NORTH AMERICAN CHESTNUTS
AND CHINQUAPINS (*CASTANEA* MILL., FAGACEAE)

By

Matthew Taylor Perkins

J. Hill Craddock
UC Foundation Robert M. Davenport
Professor in Biology
(Chair)

Eric M. O'Neill
Assistant Professor of Biology
(Committee Member)

Joey Shaw
UC Foundation Professor of Biology
(Committee Member)

Paul H. Sisco
Staff Geneticist, retired, The American
Chestnut Foundation
(Committee Member)

CHLOROPLAST DNA PHYLOGENETICS OF THE NORTH AMERICAN CHESTNUTS
AND CHINQUAPINS (*CASTANEA* MILL., FAGACEAE)

By

Matthew Taylor Perkins

A Thesis Submitted to the Faculty of the University of
Tennessee at Chattanooga in Partial
Fulfillment of the Requirements of the Degree
of Master of Science: Environmental Science

The University of Tennessee at Chattanooga
Chattanooga, Tennessee

December 2016

Copyright © 2016

By Matthew Taylor Perkins

All Rights Reserved

ABSTRACT

Evolutionary relationships and genetic structure of the North American *Castanea* were investigated using chloroplast DNA sequence data. Six plastome loci were PCR-amplified and sequenced in 77 accessions representing the three currently recognized North American *Castanea* species. Diagnostic morphological character states and a unique haplotype were shared among *C. pumila* and a plant tentatively identified as *C. dentata* in one sympatric site, suggesting past hybridization and chloroplast capture. Surprisingly, the cpDNA phylogeny did not agree with previous taxonomic treatments. The inability to distinguish between deep coalescence and interspecific hybridization as the causes of haplotype sharing makes phylogenetic reconstruction of the North American *Castanea* species difficult. Although non-D haplotypes were previously reported as diagnostic for *C. pumila* and hybrids, multiple non-D haplotype *C. dentata* were documented in the Southern Appalachians and Piedmont. The diversity of haplotypes observed in southern *C. dentata* populations provides further impetus to conserve *C. dentata* in the Southeast.

DEDICATION

This thesis is dedicated to the chestnut enthusiasts who have given their time and energy to conservation of the North American chestnuts and chinquapins.

ACKNOWLEDGMENTS

I am deeply indebted to my committee members for the guidance and support they have provided over the last two years. The mentorship and friendship provided by my major professor, Dr. J. Hill Craddock, has been critical to my success as a young researcher. Dr. Joey Shaw generously shared his time, lab space, resources, and expertise in botany and plant evolution throughout my time as a graduate student. Dr. Eric M. O'Neill introduced me to the field of conservation genetics and provided much-needed guidance when I encountered difficulties in the “wet-lab” work. Dr. Paul H. Sisco introduced me to molecular markers, gave me the opportunity to study at the Clemson University Genomics Institute, and has been my plant genetics “guru” over the last few years.

The completion of this project would not have been possible without the help of the following faculty and staff in the Dept. of Biology, Geology, and Environmental Science: Ms. Marketa Shuttters, Dr. Margaret Kovach, Dr. Jose Barbosa, Dr. Glenn P. Niemeyer, Dr. Cheryl Murphy, Ms. Kelly Locke, Prof. John Tucker, and Dr. Jennifer Boyd.

Dr. Tatyana Zhebentyayeva, of Clemson University, and Dr. Lisa W. Alexander, of the USDA, were important mentors throughout my time in graduate school. Their guidance in the fields of plant genomics, quantitative genetics, and plant breeding has allowed me to better understand my thesis topic.

I am particularly indebted to the people who guided me to most of the plants sampled in the present study (and in some cases provided lodging): Mr. David Morris, Mr. Will Calhoun,

Dr. Jack Agricola, Dr. Larry Brasher, Mr. Robert Simons, Mr. Clint Neel, Dr. Jim Lacefield, Dr. Gary Carver, Mr. Marty Schulman, Mr. Frank Cuchens, Mr. Ron Lance, Mr. Stan Hutto, Ms. Bonnie Dalzell, and Dr. Frederick V. Hebard.

The following organizations and agencies granted collecting permits and allowed me to sample *Castanea* on their properties: Nokuse Plantation, South Carolina Dept. of Parks, Recreation & Tourism, the Arkansas Natural Heritage Commission, Georgia Dept. of Natural Resources, the US Forest Service, and the Tennessee Dept. of Environment and Conservation.

The following herbaria loaned *Castanea* accessions for this study: the Arkansas Natural Heritage Commission Herbarium (ANHC), the Botanical Research Institute of Texas (BRIT), and the University of North Carolina Chapel Hill Herbarium (NCU).

The comradery of the Shaw Lab, the Craddock Lab, and the Clemson University Genomics Institute has been an important part of my graduate school experience. Conrad Blunck, Caleb A. Powell, Brooke Hadden, Jessica Tzeng, and Zach Irick volunteered their time to accompany me on sampling trips. Jesse Harris and I collaborated on DNA extractions, PCR, and sequencing of the Ozark Chinquapins used in this study. Fellow Environmental Science graduate student Danielle Floyd provided important help for the Network analysis.

I thank Mary Klinghard for her help with field sampling, creation of ArcGIS maps, reviewing my grant proposals, and emotional and intellectual support. I thank Cameron Perkins for help with field sampling, installing computer programs, and formatting data for analyses.

I gratefully acknowledge research funding and financial support from the following sources: the UTC Graduate School, a Dollywood Graduate Research Assistantship during my first year, a Foundation for the Carolinas grant to Dr. Paul H. Sisco, an Arkansas Natural Heritage Commission research grant to MTP and Dr. Joey Shaw, an American Chestnut

Foundation external grants program award to MTP and Dr. J. Hill Craddock, a UTC Provost Student Research Award to MTP, an Ozark Chinquapin Foundation award to Dr. Joey Shaw, an Association of Southeastern Biologists student travel award, the Southern Appalachian Botanical Society's Earl Core Student Research Award, an award from the UTC Department of Biology, Geology and Environmental Science for Outstanding Achievement as an Environmental Science Graduate Student, and the UTC Chapter of Sigma Xi's Graduate Student Research Award.

TABLE OF CONTENTS

ABSTRACT.....	iv
DEDICATION.....	v
ACKNOWLEDGMENTS	vi
LIST OF TABLES	xi
LIST OF FIGURES	xii
LIST OF ABBREVIATIONS.....	xiv
LIST OF SYMBOLS	xvi
CHAPTER	1
I. INTRODUCTION.....	1
Theoretical and Experimental Framework.....	2
Study Organisms: Overview of the North American <i>Castanea</i>	3
Leaf Morphology.....	4
Twig Morphology.....	5
Flower and Fruit Morphology	5
Habit	6
Biogeography of the North American <i>Castanea</i> species.....	6
Taxonomy of the North American <i>Castanea</i>	7
Evolution and Genetic Structure of the North American <i>Castanea</i>	9
Research Objectives.....	15
II. MATERIALS AND METHODS	17
Taxon Sampling.....	17
Laboratory Methods.....	18
DNA Extraction.....	18
Chloroplast DNA Sampling	19
Polymerase Chain Reaction and DNA Sequencing	20

Analyses	21
Alignments and Phylogenetic Analyses	21
Geographic Distribution of Haplotypes and Allele Frequencies.....	23
III. RESULTS AND DISCUSSION	24
Results.....	24
Description of Observed cpDNA Haplotypes	24
Phylogenetic Analyses	27
Geographic Distribution of cpDNA Polymorphisms	27
Discussion	28
Interspecific Hybridization: Molecular and Morphological Evidence	29
Systematic Questions in the North American Chinquapin Complex	30
IV. CONCLUSIONS AND FUTURE DIRECTIONS	34
Conclusions.....	34
Experimental Approaches for Future Studies	36
REFERENCES	41
APPENDIX.....	47
A. TABLES.....	47
B. FIGURES	54
VITA.....	75

LIST OF TABLES

1	Information for Ashe's collections annotated as Ozark Chinquapin.....	48
2	North American <i>Castanea</i> accessions genotyped for the present study.....	49
3	Summary of DNA sequence polymorphisms observed in six noncoding cpDNA regions in <i>C. dentata</i> , <i>C. pumila</i> , and <i>C. ozarkensis</i>	53

LIST OF FIGURES

1	Distribution of three North American <i>Castanea</i> species and sample sites of the present study.....	55
2	Lectotype of <i>Castanea alabamensis</i> Ashe.....	56
3	Dendrogram of allozyme divergence among five <i>Castanea</i> species from Dane et al. (2003).....	57
4	Allele frequencies in <i>C. dentata</i> , <i>C. pumila</i> , and <i>C. ozarkensis</i> populations sampled by Binkley (2008).....	58
5	cpDNA phylogenies recovered by Binkley (2008) (inset A) and by Shaw et al. (2012) (inset B).....	59
6	Circularized gene map of the plastid genome of <i>Castanea mollissima</i> (from Jansen et al. 2011) with the six loci sequenced for the present study.....	60
7	Median-joining network of 82 North American <i>Castanea</i> accessions.....	61
8	Allele frequencies in <i>C. dentata</i> populations sampled in the present study.....	62
9	The distinguishing SNP for the M haplotypes occurs within the <i>trnV-ndhC</i> intergenic spacer.....	63
10	Allele frequencies in <i>C. dentata</i> , <i>C. pumila</i> , and <i>C. ozarkensis</i> populations sampled in the present study.....	64
11	Results of Bayesian MCMC analysis for 83 accessions.....	65
12	Allele frequencies in <i>C. dentata</i> populations sampled by Binkley (2008).....	67
13	Allele frequencies in <i>C. pumila</i> populations sampled in the present study.....	68
14	Allele frequencies in <i>C. ozarkensis</i> populations sampled in the present study.....	69
15	Allele frequencies in <i>C. ozarkensis</i> populations sampled by Binkley (2008).....	70

16	Allele frequencies in <i>C. pumila</i> populations sampled by Binkley (2008).....	71
17	Allele frequencies in <i>C. dentata</i> , <i>C. pumila</i> , and <i>C. ozarkensis</i> populations sampled by Binkley (2008) and for the present study	72
18	Allele frequencies in <i>C. dentata</i> populations sampled by Kubisiak and Roberds (2006), by Binkley (2008), and for the present study.....	73
19	A specimen from 2015 field work near W.W. Ashe’s collection sites of <i>C. alabamensis</i>	74

LIST OF ABBREVIATIONS

ANHC, Arkansas Natural Heritage Commission Herbarium

BC₁, first-backcross generation

BC₃, third-backcross generation

bp, base pair(s)

cm, centimeter(s)

CMS, cytoplasmic male sterility

cpDNA, chloroplast DNA

DNA, deoxyribonucleic acid

dNTP, deoxyribonucleoside triphosphate

F₂, second filial generation

indel, insertion/deletion

km, kilometer

m, meter(s)

mL, milliliter(s)

mm, millimeter(s)

mtDNA, mitochondrial DNA

NCU, University of North Carolina Herbarium

nDNA, nuclear DNA

nt, nucleotide(s)

PCR, polymerase chain reaction

QTL, quantitative trait locus

RAPD, randomly amplified polymorphic DNA

rDNA, nuclear ribosomal DNA

SNP, single nucleotide polymorphism

SSR, simple sequence repeat

UCHT, University of Tennessee at Chattanooga Herbarium

UPGMA, unweighted pair group method with arithmetic averages

USDA, United States Department of Agriculture

μL , microliter(s)

LIST OF SYMBOLS

D , Nei's unbiased genetic distance

H_e , average expected (Hardy-Weinberg) heterozygosity

CHAPTER I

INTRODUCTION

Due to severe population declines caused by introduced pathogens and pests, the North American *Castanea* Mill. (Fagaceae) species are currently targets of extensive conservation efforts throughout their ranges (Jacobs et al. 2013). A number of approaches—including restoration breeding, deployment of hypovirulent strains of the chestnut blight fungus, genetic engineering, and translocation of seed and scionwood—have been used in attempts to conserve the germplasm resources of these species (Burnham 1988; Milgroom and Cortesi 2004; Zhang et al. 2013). Frequently stated goals of American Chestnut conservation efforts are to restore the species to its native range and to allow adaptive evolution to resume in these restored populations (Hebard et al. 2013). The attainment of such goals, however, requires a thorough knowledge of the distribution, genetic structure, and evolutionary relationships of the species in question. In the case of the North American *Castanea*, these requirements remain a challenge, as accurate species identification is suspected to be complicated by interspecific hybridization (Shaw et al. 2012), units of conservation remain poorly defined, and the taxonomic status of the morphologically and ecologically variable chinquapins (*C. pumila sensu lato*) is still a matter of debate.

The present thesis describes my attempt to contribute to the resolution of the above problems. In Chapter 1, I describe the phylogenetic, phylogeographic, and population genetic theory underlying this work. I also provide a review of empirical studies on the population

genetic structure of the North American *Castanea* species. In Chapter 2, I describe the methods employed. Chapter 3 contains results, including a plastid DNA phylogeny of the North American *Castanea*, and a discussion of findings from the present study. In Chapter 4, I provide a summary of my conclusions and recommend future avenues of research on natural populations of the North American *Castanea* species.

Theoretical and Experimental Framework

Chestnut genomes offer an appealing system for the study of interspecific hybridization and phylogeography in North American trees. Identification to species is often *but not always* straightforward, distributions of the three species contain both sympatric and allopatric portions, a wealth of genomic resources has been developed for Asian and European congeners (Kremer et al. 2012), and extensive studies have been conducted on hybridization and genetic variation in the closely related oaks, *Quercus* (reviewed by Petit et al. (2004)).

Resources of broader application are those that have been recently developed for studies of plant evolutionary relationships at lower taxonomic ranks. Such resources have been designed with the knowledge that noncoding regions of the chloroplast genome tend to evolve more rapidly than do coding regions (Gielly and Taberlet 1994). As a result, such variation in noncoding regions often occurs among individuals *within* species. It is thought that the discrepancy in sequence variation between coding and noncoding sequences is because noncoding regions of the plastome are under less functional constraint than are coding regions (Shaw et al. 2005). Therefore, the phylogenetic utility of many of these noncoding cpDNA regions has been studied (Shaw et al. 2005; Shaw et al. 2007; Shaw et al. 2014). Because of the phylogenetic and phylogeographic utility of noncoding DNA, results of a study employing such

data can be used to help inform conservation efforts. Of course, because such data are obtained from noncoding portions of the plastome, the application of such results to conservation must also be complemented by extensive taxonomic study.

Study Organisms: Overview of the North American *Castanea*

The genus *Castanea* Mill. (Fagaceae) is comprised of circa 8-10 species of trees and shrubs native to Asia, Europe, and eastern North America (Nixon 1997). Although taxonomic treatments vary, North America is currently thought to possess two or three of these species (Johnson 1988; Nixon 1997; Weakley 2015). In large portions of their distributions, the North American chestnuts and chinquapins were once considered important, both economically (Jaynes 1975; Payne et al. 1994; Craddock 2014) and ecologically (Paillet 1993; Paillet 2002; Foster and Faison 2014). Since the early 19th century, however, populations of North America's native *Castanea* species have been decimated by the introductions of three exotic invasive pests and pathogens (Anagnostakis 2001): *Phytophthora cinnamomi* Rands (causative agent of ink disease) in the early 19th century, *Cryphonectria parasitica* Barr (causative agent of chestnut blight) in the late 19th or early twentieth century, and *Dryocosmus kuriphilus* Yasumatsu (Chestnut Gall Wasp) in 1974.

The three lower taxa commonly recognized within the North American *Castanea* differ in several morphological characters. However, the prevalence of chestnut blight typically prevents the development of some of the characters most useful for species identification, namely, habit at maturity, flowers, and fruit (Shaw et al. 2012). The following section provides brief morphological descriptions of these taxa.

Leaf Morphology

American Chestnut, *Castanea dentata* (Marshall) Borkhausen: base cuneate (Nixon 1997); leaf blade narrowly obovate to oblanceolate (Nixon 1997), 90-300 × 30-100 mm (Nixon 1997); margins sharply serrate, each tooth triangular, gradually tapering to an awn often more than 2 mm (Nixon 1997); leaf apices acute or acuminate (Nixon 1997); abaxial lamina usually without stellate trichomes (Weakley 2015), appearing glabrous, but with minute glandular trichomes on lamina of younger plants (Nixon 1997; Weakley 2015), becoming essentially glabrous with age (Weakley 2015); sparse, simple trichomes on veins of the abaxial surface (Nixon 1997).

Allegheny Chinquapin, here treated as *Castanea pumila* (Linnaeus) P. Miller: base rounded to cordate (Nixon 1997); blade narrowly elliptic to narrowly obovate or oblanceolate (Nixon 1997), 40-210 × 20-80 mm (Nixon 1997); margins obscurely to sharply serrate, each abruptly acuminate tooth with awn usually less than 2 mm (Nixon 1997); leaf apices variable but usually not acuminate or long-acuminate (Johnson 1988); abaxial surfaces typically densely covered with appressed stellate or erect-woolly, whitish to brown trichomes, sometimes essentially glabrate, especially on shade leaves (Nixon 1997); veins of abaxial surface often minutely puberulent (Nixon 1997).

Ozark Chinquapin, here treated as *Castanea ozarkensis* W.W. Ashe: base rounded to slightly cordate or slightly cuneate (Nixon 1997); leaf blade narrowly obovate or oblanceolate, (40-)120-200(-260) × 30-100 mm (Nixon 1997); margins sharply serrate, each lateral vein terminating in cuneate or gradually acuminate tooth with awn usually more than 2 mm (Nixon 1997); leaf apex acute or acuminate (Nixon 1997); abaxial surface densely to sparsely covered

with appressed, whitish, minute, stellate trichomes, sometimes essentially glabrate, particularly on shade leaves, (Nixon 1997); veins glabrous or with a few simple trichomes (Nixon 1997).

Twig Morphology

American Chestnut, *Castanea dentata* (Marshall) Borkhausen: twigs stout; brown, essentially glabrous (Johnson 1988).

Allegheny Chinquapin, here treated as *Castanea pumila* (Linnaeus) P. Miller: twigs slender; brown, tan, or yellow-green; puberulent to tomentulose (Johnson 1988).

Ozark Chinquapin, here treated as *Castanea ozarkensis* W.W. Ashe: twigs stout; gray-brown; essentially glabrous (Johnson 1988).

Flower and Fruit Morphology

American Chestnut, *Castanea dentata* (Marshall) Borkhausen: cupule four-valved, enclosing three pistillate flowers/nuts (Nixon 1997); prickles dense on bur (Johnson 1988); nuts 18-25 × 18-25 mm, obovate, flattened on at least one side (Johnson 1988).

Allegheny Chinquapin, here treated as *Castanea pumila* (Linnaeus) P. Miller: cupule two-valved, enclosing one pistillate flower/fruit (Nixon 1997), however plants with two flowers/nuts per cupule are occasionally seen (Miller 1768; Sargent 1917; Fu and Dane 2003) and the occurrence of three nuts per cupule has been documented (M.T. Perkins, unpublished data); prickles remote to dense on bur (Johnson 1988); nuts 7-21 × 7-19 mm, conical, circular in cross section (Johnson 1988).

Ozark Chinquapin, here treated as *Castanea ozarkensis* W.W. Ashe: cupule two-valved, enclosing one pistillate flower/nuts (Nixon 1997); prickles remote to dense on bur (Johnson 1988); nuts 9-19 × 18-25 mm, conical, circular in cross section (Johnson 1988).

Habit

American Chestnut, *Castanea dentata* (Marshall) Borkhausen: formerly a large tree (to 30 m), frequently massive, now persisting mostly as multi-stemmed re-sprouts (sometimes to 5-10 m) because of destruction by chestnut blight (Nixon 1997).

Allegheny Chinquapin, here treated as *Castanea pumila* (Linnaeus) P. Miller: stoloniferous shrub, non-stoloniferous shrub, or tree (to 15 m) (Johnson 1988). Stoloniferous chinquapins of the southern Coastal Plain previously referred to as *C. paucispina* Ashe and *C. alnifolia* Nuttall are often 0.3 to 0.6 m high (Ashe 1926).

Ozark Chinquapin, here treated as *Castanea ozarkensis* W.W. Ashe: trees, occasionally shrubs, formerly often massive (to 20 m), now rarely more than 10 m, mostly persisting as re-sprouts following destruction by chestnut blight (Nixon 1997).

Biogeography of the North American *Castanea* species

The three commonly recognized North American *Castanea* species have generally allopatric distributions, although large regions of sympatry do exist (Figure 1). Tucker (1975) noted an intergradation of Ozark Chinquapin and Allegheny Chinquapin near the fall-line between Arkansas' Interior Highlands and Coastal Plain. Here, morphological intermediacy in both vegetative and reproductive characters made identification to species (or variety, in Tucker's treatment) difficult.

A second region of sympatry exists where the distributions of American Chestnut and Allegheny Chinquapin overlap in the Southern and Central Appalachians (Figure 1). For this region too, reports abound in the botanical literature of plants that are morphologically intermediate to the two co-occurring species (Dode 1908; Small 1933; Hardin and Johnson 1985; Johnson 1988). These morphologically intermediate individuals have resulted in the description of the hybrid taxon *Castanea* × *neglecta* Dode.

Finally, Johnson (1988) and W.H. Duncan (annotations on accessions housed at the University of North Carolina Chapel Hill Herbarium [NCU]) reported the disjunct occurrence of Ozark Chinquapin in north-central Alabama after examining Ashe's collections of *C. alabamensis* Ashe from the 1920s (Table 1, Figure 2). While Johnson (1988) concluded that these populations had been extirpated by chestnut blight, such a disjunction would make north-central Alabama a historical area of sympatry for American Chestnut, Allegheny Chinquapin, and Ozark Chinquapin.

Taxonomy of the North American *Castanea*

Although taxonomic study of the North American chestnuts and chinquapins was initiated over two centuries ago by Gronovius in his *Flora Virginica* (1739), disagreement regarding the number of species within the group still exists. Today, this disagreement is concerned with members of the *C. pumila* 'complex', hybrids among the chinquapins, and hybrids between the chinquapins and the American Chestnut. Jaynes (1975) gave an accurate assessment of their taxonomic status when he characterized the chinquapins as "an imprecisely defined group of shrubs and small trees found in the Southeast." As discussed above, however, there is generally agreement regarding the existence of three *taxa* within the North American

Castanea. These are the American Chestnut (*C. dentata*), the Allegheny Chinquapin (*C. pumila*), and the Ozark Chinquapin, the latter of which is treated either as a distinct species, *C. ozarkensis*, or as a variety of the Allegheny Chinquapin, *C. pumila* var. *ozarkensis*.

Contrasting views on the biological reality of the chinquapin taxa have been discussed by Johnson (1988), who recognized one species of chinquapin, *C. pumila*, and later authors (Nixon 1997; Weakley 2015), who recognized two species of chinquapin, *C. pumila* and *C. ozarkensis*. Johnson (1988) noted the existence of numerous morphologically intermediate individuals in the area of sympatry for *C. pumila* var. *pumila* and *C. pumila* var. *ozarkensis*, which he attributed to a combination of phenotypic plasticity and hybridization between Ozark and Allegheny Chinquapins. Because of this intergradation, Johnson argued that Ozark and Allegheny Chinquapins are appropriately considered varieties of the same species. In contrast, Nixon (1997) argued that because all *Castanea* species worldwide are interfertile, the occurrence of hybridization cannot be used to support the lumping of Ozark and Allegheny Chinquapins, unless one is willing to concede the existence of a single chestnut species worldwide.

While the most recent treatments recognize either two (Johnson 1988) or three (e.g., Weakley, 2015) species of North American *Castanea*, earlier treatments differed by recognizing a higher number of species in the “*C. pumila* complex” (Hardin and Johnson 1985). W.W. Ashe was certainly the most prolific author on this subject. Of the 28 new taxa or combinations published since Linnaeus’ description of *Fagus pumila* L. in 1753, 14 of these taxa were proposed by Ashe (Johnson 1988). Rather than dismiss Ashe’s taxonomic work as that of a “splitter”, I argue that Ashe’s descriptions are worth revisiting. For one, his descriptions of *Castanea* species in the Southeast took place just a few decades before the chestnut blight fungus swept through the entire range of these taxa (e.g., Ashe 1922; Ashe 1924; Ashe 1925, Ashe

1927). Therefore, from a conservation perspective, his account of morphological variation in pre-blight populations is valuable.

One of Ashe's descriptions has continued to have important repercussions in the botanical and conservation literature nearly a century after its publication. In 1925, Ashe published a description of *Castanea alabamensis*, a new species from northern and central Alabama (Figure 2). Prior to this publication, Ashe had thought that his chinquapin collections from this region represented a disjunct population of *C. ozarkensis* var. *arkansana* (Johnson 1988), which he had described earlier (Ashe 1923). Ashe's accessions of *C. alabamensis* are distinguished from most other chinquapins in the Southeast by their combination of an arborescent habit, glabrous abaxial leaf surfaces, ciliate leaf margins, and a single nut per bur. *C. alabamensis* was recognized as a distinct species in one important taxonomic work (Small 1933), but other authors (Camus 1929; Elias 1971; Little 1979) recognized it as a hybrid taxon, *C. × alabamensis*, due to its mosaic of features more typical of either *C. dentata* or *C. pumila*. Ashe's specimens of *C. alabamensis* were later determined by Johnson (1988) and W.H. Duncan to be a disjunct population of Ozark Chinquapin; however, Johnson (1988) concluded that these populations had been extirpated by chestnut blight. *Castanea alabamensis* is just one example of the many morphologically interesting synonyms that are now treated under *C. pumila sensu lato*. Clearly, more work is needed to verify the accuracy of these taxonomic conclusions and to determine their implications for conservation efforts.

Evolution and Genetic Structure of the North American *Castanea*

The evolutionary history of the North American *Castanea* has been investigated with a variety of molecular methods over the past two decades. Early work consisted of single-taxon

studies of enzyme and RAPD variation in *C. dentata* (Huang et al. 1998), in *C. pumila* var. *ozarkensis* (Dane et al. 1999), and in *C. pumila* var. *pumila* (Fu and Dane 2003). Based on analysis of 18 allozyme loci and 22 RAPD markers in 12 *C. dentata* localities sampled along the Appalachians from New York to Alabama, Huang et al. (1998) found a strong correlation between genetic distance and geographic distance ($r = -0.7077$, $P < 0.01$). Huang et al. (1998) also constructed dendrograms for the RAPD and allozyme datasets using the unweighted pair group method with arithmetic averages (UPGMA). They found that plants cluster into four broadly defined groups: a southernmost population in central Alabama, south-central Appalachian populations, north-central Appalachian populations, and northern Appalachian populations. In addition, the southernmost *C. dentata* site, in central Alabama, was found to have the highest genetic diversity of the 12 sample sites, as indicated by expected Hardy-Weinberg heterozygosity (H_e) of both RAPD and allozyme data, observed heterozygosity (H_o) of allozyme data, and effective number of alleles per locus (A_e) of both allozyme and RAPD data. This was postulated to be a result of a Pleistocene glacial refugium in the region (Huang et al. 1998). However, given the recent evidence of probable hybridization between *C. dentata* and *C. pumila* in the Southeast (Shaw et al. 2012), it is reasonable to suspect that introgressive hybridization may also partially explain Huang et al.'s result.

In their study of 12 Allegheny chinquapin sites in Florida, Mississippi, Alabama, Virginia, and Ohio (the latter planted), Fu and Dane (2003) found a high level of differentiation among the 12 subpopulations ($G_{st} = 0.30$), with a low amount of gene flow between subpopulations ($Nm = 0.57$). At several measures, Fu and Dane (2003) found higher levels of genetic variation in *C. pumila* var. *pumila* compared to other groups of plant species for which allozyme studies had been published; these groups included other woody species, other species

with a regional distribution, other wind-pollinated species, and other species with animal or gravity dispersed seeds. Interestingly, no trends in allele frequency distribution along the natural range were observed. An important conclusion of the three single-taxon studies of allozyme variation was that *C. dentata* was estimated to possess the lowest level of genetic variation (average expected Hardy-Weinberg heterozygosity, $H_e = 0.167$), while *C. pumila* var. *pumila* and *C. pumila* var. *ozarkensis* were estimated to have similar variation at allozyme loci ($H_e = 0.296$ and 0.272 , respectively) (Fu and Dane, 2003). The >100 year population bottleneck caused by chestnut blight was posited as an explanation for this result (Dane et al., 2003). It should be noted, however, that chestnut blight has also affected the Ozark and Allegheny Chinquapins, with this pandemic finally reaching the westernmost Allegheny Chinquapin populations in Texas in 1985 (Paillet 1993). Thus I think it necessary to investigate other possible explanations for the different estimates of H_e in the North American *Castanea* spp. Other possible explanations may be a relatively recent origin of *C. dentata* or the direction of gene flow among the North American *Castanea* spp. Moreover, for the allozyme analysis, seeds were sampled from all *Castanea* species, with the exception of *C. dentata*. Because of the rarity of flowering and fruiting *C. dentata*, dormant twigs with mature buds were sampled for allozymes. Future comparative studies should employ a standardized sampling and genotyping scheme to more effectively compare population genetic structure in the North American *Castanea* species.

The study by Dane et al. (2003) was also the most extensive effort to use allozymes to infer evolutionary relationships among the three North American taxa and their congeners in eastern Asia. In this study, the North American species were recovered as monophyletic (Figure 3). A divergence time between the North American clade and the Asian species was estimated at 10-13 mybp, with the most recent species exchange between Asia and North America occurring

via Beringia during the late Miocene to early Pliocene. However, given some of the problems encountered with sequence-based divergence time estimates up to that point (reviewed by Wen (1999)), Dane et al. (2003) cautioned against placing too much confidence on such an estimate.

Interestingly, among the North American taxa analyzed by Dane et al. (2003), highest values of Nei's (1978) genetic identity were observed between *C. pumila* var. *pumila* and *C. pumila* var. *ozarkensis* (0.930), while lowest identities were observed between *C. dentata* and *C. pumila* var. *pumila* (0.720). This finding suggests that *C. dentata* is sister to *C. pumila sensu lato*, and it can be interpreted to support the consideration of Ozark Chinquapin as a variety of the more widespread Allegheny Chinquapin. It should be noted, however, that the only naturally-occurring *C. pumila* var. *pumila* stands sampled were in the Gulf Coastal Plain and southwestern Virginia—two areas where Allegheny Chinquapins were later documented to possess higher cpDNA similarity with Ozark Chinquapins than with other Allegheny Chinquapins (Dane 2009; Shaw et al. 2012)

The studies of enzymatic and RAPD variation in the North American *Castanea* spp. were followed by an extensive study of microsatellite (or simple sequence repeat [SSR]) and RAPD variation in American chestnut by Kubisiak and Roberds (2003, 2006). Kubisiak and Roberds revisited Huang et al.'s (1998) hypothesis of four regional populations of *C. dentata* and increased sampling efforts to 1158 plants from 22 sites—mainly from the Appalachian portion of American Chestnut's distribution. By genotyping six microsatellite loci and 19 RAPD loci in each of their accessions, the authors found that most genetic variation occurs *within* populations (95.2% in SSRs and 96.4% in RAPD markers) (Kubisiak and Roberds 2006). This finding was consistent with that of Huang et al. (1998). In addition, allele frequencies at all six microsatellite loci were found to vary significantly with a composite dependent variable comprising latitude

and longitude; a similar association was also found for six of 19 RAPD loci. Moreover, a cline in rare alleles was observed from northeast to southwest along the Appalachian Mountains, with higher numbers of rare alleles and genetic diversity (as measured by Nei's h) being found in southwestern sample sites. To explain this pattern, Kubisiak and Roberds (2006) suggested that American Chestnut's Pleistocene glacial refugium was in the Southeast, with the former range extending into the Gulf Coastal Plain of Mississippi and Alabama.

Kubisiak and Roberds (2006), like Huang et al. (1998), used UPGMA clustering to assess relationships among populations, but unlike Huang et al. (1998), they did not find patterns of differentiation indicative of regional genetic structure; that is, populations did not cluster together based on their geographic origin. The authors noted, however, that this result was obtained using neutral genetic markers, and thus does not reflect differentiation for adaptive genes or gene complexes.

One caveat regarding Kubisiak and Roberds' findings was later discussed by Dane and Sisco (2014). Prior to SSR and RAPD genotyping, Kubisiak and Roberds used sequence data from the *trnT-trnL* intergenic spacer of cpDNA in an attempt to remove *C. pumila* and potential hybrids that were misidentified as *C. dentata* during sampling. In their screening panel, they found a 12-bp and 72-bp deletion in *C. dentata* that was not present in the samples they had of *C. pumila* and *Castanea* spp. from Europe and Asia. This process removed 165 of the 1158 plants sampled (14.2%). As a result, no *C. dentata* from Tennessee or Alabama—a large portion of the species' range—were used for population genetic analyses. Results of later studies (Dane 2009; Shaw et al. 2012; Li and Dane 2013) have shown that these deletions at *trnT-trnL* and other plastome loci are not reliable indicators of species identity in the group (Dane and Sisco 2014). For example, Dane (2009) later documented the purportedly diagnostic 12- and 72-bp deletions

in one *C. pumila* population from northeastern Georgia. Dane and Sisco (2014) also noted that in the studies of Shaw et al. (2012) and Li and Dane (2013), cpDNA haplotypes previously associated with *C. pumila* were found in several plants that were clearly identified as *C. dentata* in fruit and leaf morphology.

The earlier population genetic studies in the North American *Castanea* species were followed by multi-species phylogeography studies that employed noncoding cpDNA markers. The cpDNA phylogeography study by Binkley (2008) is arguably the most geographically extensive effort on the three species to date. Her first objective was to use a noncoding cpDNA marker at *trnV-ndhC* to test the hypothesis that morphologically intermediate plants from northwest Georgia (also referred to as the “Pocket chinquapins”) were the result of hybridization between *C. pumila* and *C. dentata*, i.e., that they represented the hybrid taxon *C. × neglecta*. She also surveyed *C. ozarkensis*, *C. pumila*, and *C. dentata* throughout their ranges (Figure 4). She found that the morphologically intermediate chinquapins from northwest Georgia possessed a haplotype that was not found in either putative parent species—she designated this haplotype M5. She recovered four broad haplotypic clades in her 233 accessions (Figure 5, inset A): D, which was only found in *C. dentata*; P, which was first found in *C. pumila*; O, which was primarily found in *C. ozarkensis*; and M, which was found in both *C. dentata* and *C. pumila* accessions. Importantly, she found that in multiple cases the same *trnV-ndhC* haplotype was found in both *C. pumila* and *C. dentata*. The O3 haplotype was shared among *C. ozarkensis* from Arkansas and *C. pumila* from southwestern Virginia, approximately 1,000 km to the east. The M4 haplotype was shared among *C. dentata* and *C. pumila* from Walker Co., Georgia. The P1 haplotype was shared among *C. pumila* from Tennessee and North Carolina and *C. dentata* from a different site in Tennessee. The M6 haplotype was shared among *C. pumila* from Georgia and

Alabama and *C. dentata* from a different site in Georgia. With the number of polymorphic characters in her dataset, she could not, however, determine the relative degree of similarity among the four clades. The inability to determine whether some shared haplotypes were basal or more derived meant that Binkley (2008) could not rule out the retention of ancestral polymorphism as the cause of shared haplotypes, so her conclusions of recent interspecific hybridization were tentative.

The next phase of this study employed six noncoding cpDNA loci in thirteen plants—one representative of each of the haplotypes found earlier, plus an extra representative of haplotype D2 (Shaw et al. 2012). Well supported relationships among the four clades were recovered (Figure 5, inset B): the O clade was the most basal, and it was sister to the clade that comprised the M, P, and D haplotypes; the M clade was also basal to the clade that comprised the D and P haplotypes; and the D and P clades were sister to each other. However, because only one representative of each haplotype was sequenced at five additional loci, it could not be determined whether the shared single-locus haplotypes corresponded to six-locus haplotypes that were also shared between species. An important question remains from these two studies: were certain *trnV-ndhC* shared among multiple species because of past interspecific hybridization, or deep coalescence?

Research Objectives

As evidenced by the above literature review, many important questions regarding the evolutionary relationships and taxonomy of the North American *Castanea* spp. remain. I investigated some of these questions for the present work. First, I tested the hypothesis of interspecific hybridization and chloroplast capture in the North American *Castanea* by sampling

sympatric *C. dentata* and *C. pumila* and sequencing them at six polymorphic noncoding chloroplast loci. I reasoned that a dataset comprised of Binkley's (2008) *trnV-ndhC* locus plus five loci additional loci would allow me to make a more confident conclusion of hybridization where I documented haplotype sharing between two co-occurring species. Second, I asked whether a phylogeny inferred from noncoding cpDNA agreed with current and historical taxonomic treatments of the group. Finally, I had the research goal of contributing to a broader phylogeographic study of the North American *Castanea*.

II. MATERIALS AND METHODS

Taxon Sampling

Leaf tissues were collected from a total of 379 individuals at 32 different sample sites. Field collections were made from populations representing the three currently recognized North American *Castanea* species (Table 2). Collections were made from populations where species occurred in allopatry and where species occurred in sympatry. I used the following characters to identify plants to species: the presence or absence of stellate trichomes (Nixon 1997), fruit morphology (Nixon 1997; Weakley 2015), twig morphology (Johnson 1988), and the presence or absence of a ciliate leaf margin (J.H. Craddock, pers. comm.). At sites where only a single species was observed, I attempted to sample 10 plants per site. At sites where *C. pumila* and *C. dentata* co-occurred, I attempted to sample 20–30 plants per site, to facilitate detection of chloroplast capture and calculation of population genetic statistics with SSR markers. Due to the rarity of these species, however, only a few individuals could be found in some sites. Some samples were sent by collaborators or received on loan from other institutions.

Because *Castanea* spp. from Europe and Asia are basal to the North American clade (Lang et al. 2007), the Chinese Chestnut, *C. mollissima* Blume, reference plastome sequence (Jansen et al. 2011) was used as the outgroup in phylogenetic analyses.

To address taxonomic hypotheses, samples were also collected from plants that could be keyed to one of the North American chinquapin synonyms (i.e., members of the ‘*C. pumila* complex’). Important morphotypes sampled were: *C. alnifolia*, a sub-shrub from the Coastal

Plain commonly called the Trailing Chinquapin; *C. alabamensis*, a chinquapin thought to be endemic to northern Alabama; and *C. floridana*, an arborescent chinquapin from the Coastal Plain. *C. alabamensis* has also been treated as a hybrid taxon, *C. × alabamensis* (Elias 1971), and as a disjunct of *C. ozarkensis* in north-central Alabama (Johnson 1988); therefore, these collections were also used to test this biogeographic hypothesis. In cases where I thought chinquapin samples corresponded to one of the *C. pumila* synonyms, I consulted treatments of taxonomic authors (e.g., Sargent 1919, for *C. alnifolia* var. *floridana* Sarg.) and the most recent taxonomic revision of the North American chinquapins (Johnson 1988).

For each plant sampled, an herbarium accession was taken and two or three of the youngest leaves available were frozen for DNA extraction.

Laboratory Methods

DNA Extraction

Due to time and funding constraints, 12 plants per site, if available, were selected from 20 sites (Figure 1) for DNA extraction and genotyping. DNA was extracted from leaves of 150 accessions using one of four protocols. Collections made during the spring of 2015 were extracted in Dr. Tatyana Zhebentyayeva's laboratory at the Clemson University Genomics Institute using a modification of the CTAB protocol employed by Kubisiak et al. (2013) (T. Zhebentyayeva, Clemson Univ., pers. comm.). Collections from Alabama and Maryland were extracted in Dr. Lisa Alexander's laboratory at the USDA Nursery Crop Research Station, McMinnville, TN, using either a modified alkaline lysis method or a CTAB method (Alexander 2016). In extractions employing the CTAB method of Alexander (2016), I added 3 μ L of β -mercaptoethanol immediately after disruption of leaf tissues in CTAB buffer. All other DNAs

were extracted using the Qiagen DNeasy Plant Mini Kit (Qiagen Corp., Valencia, CA) in Dr. Joey Shaw's laboratory at the University of Tennessee at Chattanooga. DNA quality and DNA concentration were evaluated by electrophoresis of genomic DNA on 1% agarose gels, amplification of a microsatellite marker in some samples, measurement using NanoDrop 8000 or 2000 spectrophotometers (Thermo Scientific), and measurement using a Qubit 2.0 Fluorometer (Thermo Fisher Scientific).

While DNA concentrations of samples extracted with the modified alkaline lysis method (Alexander, 2016) were high, sample purity ratios (260:280 nm and 260:230 nm) were poor, and extracts were brown and highly viscous, suggesting co-precipitation of polysaccharides and polyphenolics with DNA. Thus, only some plants from Alabama and no plants from Maryland were used for genotyping. Likewise, DNA extractions from the loaned Arkansas accessions often yielded low quality extracts, perhaps due to specimen age or method of preservation. I favor this explanation because the only six samples—of 24 total—to consistently yield PCR products had been stored in silica, whereas all others had been preserved using other methods. Thus, only six of these accessions were used for genotyping.

Chloroplast DNA Sampling

Because Shaw et al. (2014) recently estimated that four to eight of the most variable noncoding cpDNA loci will likely access most of the low-taxonomic discriminating power of the plastome, I chose to amplify the following six noncoding loci previously documented as polymorphic in the North American *Castanea* species (Shaw et al. 2012): 3' *trnV-ndhC*, *rpL16*, *trnS-trnG*, *rpL32-trnL*, *atpI-atpH*, *psbA-trnH*. Primer sequences of *psbA-trnH* and *rpL16* are described in Shaw et al. (2005). Primer sequences of *trnV-ndhC*, *rpL32-trnL*, *atpI-atpH*, and

trnS-trnG are described in Shaw et al. (2007). The positions of the six loci are shown on a map of the *Castanea mollissima* plastome on Figure 6.

Polymerase Chain Reaction and DNA Sequencing

Polymerase chain reaction (PCR) was performed using Eppendorf Mastercycler EP gradient S or Mastercycler personal thermocyclers (Westbury, NY) in 25- μ L volumes. The following reaction components were previously used by Shaw et al. (2012): 1 μ L of template DNA (10-100 ng), 1 \times *Ex Taq* buffer (TaKaRa), 200 μ mol/L each dNTP, 3.0 mmol/L MgCl₂, 0.1 μ mol/L each primer, and 1.25 units *Ex Taq* DNA polymerase (TaKaRa). Reactions included bovine serum albumin at a final concentration 0.2 μ g/ μ L. Thermal cycling of *trnV-ndhC*, *rpL16*, *rpL32-trnL*, *atpI-atpH*, and *trnS-trnG* followed the “*rpL16*” program of Shaw et al. (2005), which consists of the following steps: 80°C, 5 min.; 35 \times (95°C, 1 min.; 50°C, 1 min with a ramp of 0.3°C/s; 65°C, 5 min.); 65°C, 4 min. Thermal cycling of *psbA-trnH* followed the protocol of Taberlet et al. (1991), which consists of the following steps: 35 \times (94°C, 1 min.; 50°C, 1 min.; 72°C, 2 min.). Amplification success was determined by electrophoresis of PCR products in 1% agarose gels.

Unincorporated dNTPs and primers were removed with ExoSAP-IT (USB, Cleveland, Ohio), with one exception to the manufacturer’s protocol: 1 μ L, rather than 2 μ L, of ExoSAP-IT, were added to 5 μ L PCR products (J. Shaw, pers. comm.). Cycle sequencing was performed with the ABI Prism BigDye Terminator Cycle Sequencing Ready Reaction Kit, v. 3.1 (Perkin-Elmer/Applied Biosystems, Foster City, California). Sequencing reactions of the following four loci employed only one primer from each pair, as described by Shaw et al. (2005, 2007): *ndhC*, *trnH*, *rpL16F71*, and *trnS*. Because amplicons of *atpI-atpH* and *rpL32-trnL* are much longer than

800 bp in the study organisms, sequencing reactions of these loci required both the forward and reverse primers. Sequencing reaction products were purified by centrifugation through Sephadex G-50 (GE-Healthcare, Little Chalfont, United Kingdom) columns. Sequences were detected on an ABI 3730 capillary electrophoresis instrument (Applied Biosystems, Foster City, CA) at the University of Tennessee Genomics Core in Knoxville, TN.

Analyses

Alignments and Phylogenetic Analyses

Sequences were viewed and edited in Geneious v.9 and v.10 (Biomatters Ltd.). Sequences were aligned using the MUSCLE alignment plugin for Geneious. All polymorphic sites in the alignments were checked against the original chromatogram files to ensure correct base calls. A concatenated six locus dataset was constructed for the all North American accessions plus the *C. mollissima* outgroup. For some samples at the *trnV-ndhC* locus, a sequencing error was observed between positions 220-260 from the 3' end of *ndhC*. This sequencing error was first documented by Binkley (2008), who found it in nearly all of her 234 *trnV-ndhC* sequences. While multiple peaks were present at a few sites in this region, the sequencing error did not complicate alignment. Base calls could still be made visually, and my base calls for this small number of samples were checked against sequences of higher quality. Comparison of all *trnV-ndhC* sequences with the *C. mollissima* reference sequence indicated that this 40 bp region is monomorphic in the accessions studied, thus lending confidence to my determinations.

The proportion of mutational events (% variability), *sensu* Shaw et al. (2007), was calculated for the six noncoding cpDNA regions studied. The formula is as follows:

$$\% \text{ variability} = (\text{NS} + \text{ID} + \text{IV}) / \text{L} \times 100$$

where NS = number of nucleotide substitutions, ID = number of indels, IV = number of inversions, and L = aligned sequence length in bp. The proportion of variable characters contributed by the following types of polymorphism was also calculated: single nucleotide polymorphisms (SNPs), substitutions longer than 1 bp, insertions/deletions (indels), and chloroplast microsatellites. To facilitate future marker development, I differentiated between chloroplast microsatellites—which are usually mononucleotide repeats less than 15 bp in length (Provan et al. 2001)—and other indels.

The model of nucleotide substitution GTR + G was selected as the best fitting maximum likelihood model, Akaike Information Criterion = 13,742, using jModelTest v.2.2.1 (Guindon and Gascuel 2003; Darriba et al. 2012). The alignment was formatted for jModelTest using the ALignment Transformation EnviRonment (ALTER) (Glez-Peña et al. 2010).

Bayesian Markov Chain Monte Carlo (MCMC) analyses were performed using the MrBayes 3.2.6 (Huelsenbeck and Ronquist 2001) plugin for Geneious. Insertions/deletions (i.e., indels) were coded as binary characters (presence/absence). Two independent runs of Bayesian MCMC were implemented with the following settings: four heated chains were used, trees were sampled every 1,000 generations for 1,000,000 generations, the first 100,000 trees were discarded as burn-in, and the GTR + G model of substitution was used. At the end of the runs, the average standard deviation of split frequencies was <0.01, indicating convergence.

A haplotype network based on statistical parsimony (Templeton et al. 1992) was constructed with TCS v.1.21 (Clement et al. 2000). A 95% connection limit was used, gaps were

treated as missing data, and indels were coded as binary characters. In addition, a haplotype network based on the median-joining method was constructed with Network v.4.6.1.4 (Bandelt et al. 1999).

Geographic Distribution of Haplotypes and Allele Frequencies

Maps of the geographic distribution of cpDNA haplotypes were created using ArcGIS Desktop 10.3 and ArcGIS Online (ESRI, Redlands, California). Layers showing the ranges of *C. dentata*, *C. pumila*, and *C. ozarkensis* were created from shapefiles of Little's (1977) range maps (Figure 1). Allele (i.e., haplotype) frequencies were calculated and used in mapping.

III. RESULTS AND DISCUSSION

Results

Description of Observed cpDNA Haplotypes

Sequence data were obtained for six noncoding cpDNA loci in a total of 91 *Castanea* accessions, resulting in approximately 4,600 bp per individual. Two putative *C. dentata* accessions sent from Arkansas only differed from the *C. mollissima* reference sequence at a 1 bp indel, so these plants were removed from the dataset. Thus, the final dataset comprised 88 North American *Castanea* accessions plus one accession of *C. mollissima* as outgroup. Of these, 13 accessions were those previously genotyped at six loci by Shaw et al. (2012), while the remaining 75 accessions were from my field collections or were recently sent by collaborators.

TCS analysis revealed 33 unique haplotypes in the six-locus dataset comprised of *C. dentata*, *C. pumila*, and *C. ozarkensis* accessions. The median-joining network analysis also found 33 unique haplotypes in the dataset (Figure 7). Twenty-one of these haplotypes are newly documented, while 12 were previously found by Binkley (2008) and Shaw et al. (2012). I have followed Binkley's (2008) method of haplotype nomenclature; that is, haplotypes were named according to their clade and the order in which they were documented. For example, the first haplotype found in the present study grouped with the 'P' clade of Binkley (2008) and Shaw et al. (2012); because 12 haplotypes were previously found by those authors, my first new 'P' haplotype, from an American Chestnut called Old NC10, was given the haplotype designation 'P13'.

For the six noncoding cpDNA regions, a total of 77 polymorphic sites were observed among the three North American *Castanea* spp. (Table 3). 59.7% of the polymorphisms were SNPs, 3.9% were substitutions greater than one nucleotide in length, 15.6% were chloroplast microsatellites (most of which were mononucleotide repeats), and 20.8% were indels. No inversions were observed. Among all cpDNA regions surveyed, *trnV-ndhC* was the most polymorphic region. The percent variability for this region was 3.8%—no other region had a value greater than 1.85% for this measure.

Four broad haplotypic groups were recovered: D haplotypes, P haplotypes, M haplotypes, and O haplotypes. This finding was consistent with that of Binkley (2008) and Shaw et al. (2012). D haplotypes are distinct from all other North American haplotypes due to their 49 bp deletion at *trnV-ndhC*. Of the 77 plants sequenced for the present study, only four *C. dentata*, from the southern Blue Ridge, possessed this haplotype (Figure 8). They were identical to the D2 haplotype of Shaw et al. (2012) at all six loci.

The D2 haplotype differs from the most derived P haplotype, P33 of *C. pumila*, by the 49 bp indel at *trnV-ndhC* and a microsatellite at *atpI-atpH*. Other P haplotypes were found in *C. pumila* and *C. dentata*, but no identical haplotypes were shared among the two species. Three new P haplotypes were documented: P33 in *C. pumila*, P13 in *C. dentata*, and P23 in *C. pumila*. These *C. pumila* were designated as P33 and P23 because a number of haplotypes in the O and M clades were documented after I observed P13, but before I observed P23 and P33. P13 of *C. dentata* differed from P33 of *C. pumila* by one microsatellite at *atpI-atpH*, one microsatellite at *rpL32-trnL*, and one SNP at *rpL32-trnL*. Two of the new haplotypes were found in the Blue Ridge of North Carolina, while haplotype P33 was found in northern Florida.

The M haplotype group is distinguished from the P, D, and O groups by a SNP at *trnV-ndhC* (Figure 9). At this locus, M haplotypes are fixed for guanine, while all other accessions are fixed for adenine. This is the *only* site where all M haplotypes are fixed for one allele and all other haplotypes are fixed for a different allele. Nevertheless, statistical support for the M clade is high in this study, as it has been in previous studies. The following eleven new M haplotypes were observed: M18 in *C. pumila*, M29 in *C. pumila*, M21 in *C. dentata*, M30 in *C. pumila*, M15 in *C. dentata* and *C. pumila*, M16 in *C. pumila*, M14 in *C. pumila*, M17 in *C. pumila*, M22 in *C. dentata*, M19 in *C. dentata* and *C. pumila*, and M20 in *C. dentata*. M6, which was previously documented by Shaw et al. (2012), was observed in *C. pumila* here. Interestingly, M15 was found in both *C. dentata* and *C. pumila* from a sympatric site in the Blue Ridge of South Carolina and in one *C. pumila* from northern Alabama (Figure 10). Similarly, M19 was found in *C. dentata* and *C. pumila*, but the specimens were collected in two different sites, *C. pumila* in the Blue Ridge of South Carolina and *C. dentata* in the Blue Ridge of North Carolina.

The O haplotypes differ from all other haplotype groups at several polymorphisms. For example, at three SNPs in *trnV-ndhC*, all O haplotypes are fixed for one base, while all other haplotypes are fixed for a different base. Such patterns were also seen at substitutions and indels in *rpL32-trnL*, *atpI-atpH*, and *psbA-trnH*. The following seven new O haplotypes were observed: O31 in *C. pumila* accessions that also keyed to *C. floridana* and *C. alnifolia*, O32 in *C. pumila* that also keyed to *C. alnifolia*, O25 in *C. ozarkensis*, O26 in *C. ozarkensis*, O27 in *C. ozarkensis*, O24 in *C. pumila*, and O28 in *C. pumila*. Haplotype O3, which was previously documented in *C. pumila* in southwestern Virginia (Binkley 2008; Shaw et al. 2012), was documented here in *C. pumila* in the Blue Ridge of South Carolina.

Phylogenetic Analyses

Bayesian analyses revealed four strongly supported clades (Figure 11): the clade containing the D haplotypes, the clade containing both the P and D haplotypes, the clade containing the M haplotypes, and clade containing the O haplotypes. The O haplotypes were recovered as monophyletic, the M haplotypes were recovered as monophyletic, and the D haplotypes were recovered as monophyletic. The P haplotypes, however, were paraphyletic with respect to the D haplotypes. In other words, the D haplotypes formed a clade that was nested within the P haplotypes. Therefore, the P group includes some, but not all, of the descendants of their most recent common ancestor. In contrast, Shaw et al. (2012) found that the P haplotypes and D haplotypes formed two clades, which were sister to each other.

Geographic Distribution of cpDNA Polymorphisms

Consistent with the findings of previous studies (Figure 12), the D2 haplotype was restricted to *C. dentata*. All of these accessions were from higher elevations in the Blue Ridge (Table 2).

As in previous studies, P haplotypes were found in *C. dentata* and *C. pumila* in the Southern Appalachians. Unlike previous studies, however, a P haplotype, P33, was found the Coastal Plain (Figure 13, indicated by orange arrow). This haplotype was found in one individual in a population of “trailing” chinquapins that were almost entirely fixed for O haplotypes.

Similar to previous findings, M haplotypes were found in *C. dentata* (Figure 8) and *C. pumila* (Figure 13) in the Southern Appalachians, the Piedmont, and the southern Coastal Plain. Two haplotypes, M15 and M19 were shared among two species (Figure 11). However, M15 and M19 differed in whether they were shared among species that were co-occurring or among

species that were geographically separated. Haplotype M15 was observed in *C. pumila* and *C. dentata* growing at the same locality, sample site CR. But M19 was found in three *C. pumila* at site CR and in one *C. dentata* approximately 200 km to the northeast, at sample site GF.

Interestingly, the O haplotypes were the most geographically widespread of the four groups. As in Binkley (2008) and Shaw et al. (2012), O haplotypes were found in *C. ozarkensis* in the Ozarks and Ouachitas (Figure 14, Figure 15). They were also found in *C. pumila* in the Coastal Plains of Florida, Georgia, and South Carolina, and in the Blue Ridge of South Carolina (Figure 13). Similar to Binkley (2008) and Shaw et al.'s (2012) finding of haplotype O3 in *C. pumila* from southwestern Virginia (Figure 16), this haplotype was also observed in *C. pumila* from the South Carolina mountains (Figure 13).

Haplotypic frequencies are shown for all populations analyzed in Binkley (2008) and in the present study in Figure 17. The distribution of haplotypes for the present study and studies by Binkley (2008) and Kubisiak and Roberds (2006) is shown in Figure 18.

Discussion

For the present study, I was concerned with three topics: the hypothesis of interspecific hybridization between *C. dentata* and *C. pumila*, the evolutionary relationships of previously recognized chinquapin morphotypes, and the broader phylogeography of the North American *Castanea* spp. To address these topics, I sequenced six noncoding chloroplast DNA regions in 77 North American *Castanea* accessions representing the taxonomic and morphological breadth of the group, compared them to sequence data of 13 plants from Shaw et al. (2012), and inferred a cpDNA phylogeny with these data.

Interspecific Hybridization: Molecular and Morphological Evidence

I hypothesized that *C. dentata* and *C. pumila* sometimes hybridize where they grow in sympatry. I therefore predicted that identical chlorotypes would sometimes be found in the two species where they co-occur. Because increasing the number of sequenced loci from one to six would also increase the number of variable sequence characters, I reasoned that such a dataset would enable a more confident test of recent ‘chloroplast capture’. In particular, all 12 of the quickly evolving microsatellites revealed by my sequencing efforts were found in *rpL16*, *trnS-trnG*, *rpL32-trnL*, *atpI-atpH*, and *psbA-trnH* (Table 3); no microsatellites were observed in *trnV-ndhC*, the locus employed in the most extensive sampling of sympatric *Castanea* populations to date (Binkley 2008). These microsatellite loci were often polymorphic *within* populations, and were often the only mutation separating plants that were growing in close proximity. Using this method, I thus expected a lower probability of haplotype sharing due to retention of polymorphisms that predate speciation.

One unambiguous case of haplotype sharing was documented between *C. dentata* and *C. pumila*. However, the *C. pumila* occurred at the CR site in the Blue Ridge of South Carolina, while the *C. dentata* occurred approximately 200 km to the north, in the Blue Ridge of North Carolina. A second possible case of interspecific haplotype sharing was documented among co-occurring *C. pumila* and two plants tentatively identified to *C. dentata*, also from the CR site. Shared haplotypes were not found in my other two localities where *C. dentata* and *C. pumila* were found together, G and BF. In addition, *C. dentata* plants at the CR population had numerous stellate trichomes on their abaxial leaf surfaces—a feature often used to diagnose plants as *C. pumila*. These plants were diagnosed as *C. dentata* with morphological characters on the leaves, buds, and twigs. Unfortunately, plants at the CR site were not producing flowers, so

the number of pistillate flowers per cupule could not be used for species identification. Also, retention of ancestral polymorphism as the cause of the shared alleles in the CR population cannot be entirely rejected, but the fact that these individuals possess identical alleles at all twelve highly polymorphic microsatellites does lend support for the hypothesis of a relatively recent instance of hybridization and introgression—if the *C. dentata* from the site were correctly identified. By ‘recent hybridization,’ I refer to interspecific hybridization that has occurred between *C. dentata* and *C. pumila* where they currently co-occur. Because boreal forests were the dominant vegetation type down to 34°N in the Southern Appalachians during the Late Wisconsin glacial maximum (19,000 to 16,300 years before present) (Delcourt 1979), we can assume that *C. dentata* and *C. pumila* have only occupied their present region of sympatry in the Southern and Central Appalachians after the most recent glacial maximum. While it is possible that the this M15 haplotype predates speciation, I think the geographic context of this shared haplotype—that it was found in two species just a few hundred meters apart—combined with the fact that the six-locus dataset includes 12 microsatellites that often differed within populations provides evidence for gene flow between *C. dentata* and *C. pumila* at or near this site after the last glacial maximum. However, further work will be needed to gather evidence from floral morphology and then statistically test the two competing hypotheses of gene flow and deep coalescence.

Systematic Questions in the North American Chinquapin Complex

I genotyped accessions representing some of the chinquapin morphotypes previously recognized as distinct species by taxonomic authors: *C. alabamensis* Ashe, *C. floridana* (Sargent) Ashe, and *C. alnifolia* Nuttall. For *C. alnifolia* (the Trailing Chinquapins) and *C.*

floridana, I tested two competing hypotheses: (1) that *C. alnifolia* cpDNAs would form a monophyletic group that excluded cpDNAs from the non-trailing *C. floridana* trees growing approximately 25 km to the north, and (2) that the Trailing Chinquapins would not form a monophyletic group, but that they would form a cpDNA clade with the *C. floridana* trees growing approximately 25 km to the north. Despite their unique morphology (forming dense genets comprised of shoots often less than two meters tall at maturity) and severe habitat (frequently burned Longleaf Pine savannahs), haplotypes of nearly all of the *C. alnifolia* (sites B and P) were identical to those of the tree forming *C. floridana* (site S) growing a few miles away. Specifically, all representatives of *C. floridana* and 7 of the 9 representatives of *C. alnifolia* had the O31 haplotype. The eighth accession of *C. alnifolia* only differed from the former plants at a microsatellite, and was given the haplotype O32. The ninth accession of *C. alnifolia* had the haplotype P33.

There are a few reasons this result is not surprising. First, the morphotype previously recognized as *C. alnifolia* may be highly phenotypically plastic. In the absence of frequent fire, these plants may eventually grow into large trees. Second, numerous studies have documented the predominance of cytoplasmic gene flow over nuclear gene flow in plants (reviewed by Rieseberg and Soltis 1991). Under these conditions, geographically proximal taxa are generally most closely related in terms of cpDNA, despite being highly differentiated in morphology and at nuclear loci. Because cpDNA was not different for the two distinct morphotypes, nDNA datasets will be necessary to thoroughly test these taxonomic hypotheses.

My *C. pumila* collections from northern and central Alabama were perhaps the most morphologically intriguing of the study (Figure 19). In the field, these plants appeared to be *C. dentata* due to their arborescent habit, narrow-lanceolate blades, and glabrous abaxial laminar

surfaces. Upon closer inspection, however, I found that these plants had ciliate leaf margins and produced one nut per bur. The latter features are typically thought of as being unique to the chinquapins. Unlike other *C. dentata* and *C. pumila* sampled in my study, the chinquapins from northern Alabama had a glaucous abaxial surface.

After comparing these collections with W.W. Ashe's accessions of *C. alabamensis* Ashe (Figure 2), I found that my plants from sites CH and AG were identical to the historical *C. alabamensis* specimens. As noted in the literature review section, Johnson (1988) and W.H. Duncan (unpublished annotations) determined that Ashe's collections were from a disjunct population of Ozark Chinquapin in Alabama. After field work near Ashe's collection sites, Johnson (1988) concluded that Ozark Chinquapin had been extirpated from Alabama by chestnut blight. Morphological study of *C. ozarkensis* from Arkansas, Ashe's *C. alabamensis*, and my own collections does not seem to support the hypothesis that the plants from northern Alabama are Ozark Chinquapins. The Alabama plants' complete lack of stellate trichomes is a feature I have not observed in *C. ozarkensis*. Sequence data for the northern Alabama collections revealed a diversity of haplotypes in the two localities: M18 in sample AG7, M15 in sample CH22, M16 in sample AG8, M6 in samples AG1-AG6, and M17 in samples CH15, CH18, and CH21. These results are consistent with those of Binkley (2008), who found that a population of taxonomically confounding plants from northwestern Georgia only contained M haplotypes. In another recent phylogeography study, Li and Dane (2013) discussed the morphological "intermediacy" of their samples from central Alabama, but they did not note the similarity to Ashe's *C. alabamensis* or Johnson's disjunct population of Ozark Chinquapins. Li and Dane (2013) posited that these plants might be the result of hybridization between *C. dentata* and *C. pumila*. Interestingly, earlier taxonomists (Elias 1971; Little 1979) have treated these chinquapins from northern and

central Alabama as a hybrid taxon, *C. × alabamensis*. While further molecular and morphological study will be needed to infer the origin of this morphological ‘mosaic’, it seems that the morphological coherence seen in this plant where it co-occurs with *C. dentata* may hint at the biological reality of this taxon. For these reasons, these morphologically unique chinquapin populations should be the focus of rigorous taxonomic and population genetic study in the future.

IV. CONCLUSIONS AND FUTURE DIRECTIONS

Conclusions

The most recent studies of genetic variation in the North American *Castanea* have made the following conclusions:

- Single-locus cpDNA haplotypes are sometimes shared between *C. dentata* and *C. pumila*, and between *C. pumila* and *C. ozarkensis* (Binkley 2008).
- cpDNA haplotypes are largely found in separate geographic ranges, but haplotype distributions overlap in the Southern Appalachians (Shaw et al. 2012).
- A northwest Georgia population thought to represent the hybrid taxon, *C. × neglecta*, is nearly fixed for the M5 haplotype, which was, however, not found in either putative parent species (Shaw et al. 2012).

Conclusions of my study are the following:

- In one case, identical six-locus haplotypes were shared among plants confidently identified as *C. dentata* and *C. pumila*, and in a second case, identical six-locus haplotypes were shared among plants confidently identified as *C. pumila* and plants tentatively identified as *C. dentata*. Pending a greenhouse study of floral morphology in the latter plants, this would provide the first concurrent molecular and morphological evidence for interspecific hybridization in North American *Castanea* species.
- The four broad haplotype groups are not geographically restricted. However, unique six-locus haplotypes do appear to be geographically restricted.

- Morphology and chloroplast DNA sequence data do not support a disjunction of Ozark Chinquapin in northern Alabama.
- While Binkley (2008) and Shaw et al. (2012) found highest haplotypic diversity in the Southern Appalachians, the present work documented high haplotypic diversity in both the Southern Appalachians and the Gulf Coastal Plain.
- Two morphologically divergent, yet geographically proximal, populations of chinquapins in the Coastal Plain are nearly fixed for an identical cpDNA haplotype. Phenotypic plasticity as a result of fire regime or chloroplast capture between two distinct taxa are proposed as possible explanations.
- I observed 21 new cpDNA haplotypes in the North American *Castanea* using the same loci as Shaw et al. (2012). Combined with the results of the earlier study, 33 haplotypes have been documented at these cpDNA loci.

I recommend that future work strive to address the following:

- Test the hypotheses of interspecific hybridization between *C. pumila* and *C. dentata* and between *C. pumila* and *C. ozarkensis* by employing molecular markers with different patterns of inheritance. Bi-parentally inherited nuclear DNA loci should be used in conjunction with maternally inherited cpDNA loci to help determine which shared haplotypes are due to deep coalescence and which are due to interspecific hybridization. Topological conflict in the cpDNA and nDNA trees would be amenable to statistical tests designed to test hypotheses of hybridization (e.g., Joly et al. (2009)).
- Use a standardized sampling and genotyping scheme to infer the phylogeography of the North American *Castanea*. Sampling only a few accessions from geographic regions

risks underestimating the genetic diversity of certain populations and making false-positive conclusions about the genetic structure of the North American *Castanea*.

- Use both nuclear and chloroplast DNA sequence data to infer species-level phylogenies for the group. Previous work in other plant taxa has shown that chloroplast DNA-based phylogenies of closely related species must be viewed with skepticism if they are not corroborated by evidence from nuclear DNA-based datasets (Rieseberg 1991).
- Perform controlled crosses and genetic mapping to understand the genetic factors underlying morphological characters used to distinguish the North American *Castanea* species.

Experimental Approaches for Future Studies

Chloroplast DNA phylogenetics revealed one unambiguous case and one ambiguous case of haplotype sharing among *C. dentata* and *C. pumila*. The latter case involved *C. dentata* and *C. pumila* that were growing in close proximity, at the CR sample site. However, the absence of flowers or fruit on plants at the CR site means that future work, using floral morphology, will be needed to confirm my species diagnoses. This uncertainty can be addressed during the next winter, by grafting scionwood from these CR plants onto Chinese Chestnut rootstock and inducing flowering in the greenhouse or nursery.

Accurate species identification would be aided through a more thorough understanding of the genetic determinants underlying diagnostic traits. The CR accessions identified as *C. dentata* had numerous bud, twig, and leaf characters typical of *C. dentata*, but also possessed the stellate trichomes typical of *C. pumila*. These morphological features are consistent with Johnson's (1988) description of the progeny of controlled *C. dentata* × *C. pumila* crosses performed by

R.A. Jaynes. However, since mapping studies of these diagnostic traits have not been performed, the genetic factors controlling these phenotypes are not understood. A future study might cross well-characterized accessions of *C. dentata* and *C. pumila* and use quantitative trait locus (QTL) mapping to delineate the genomic intervals correlated with these traits. However, since chestnuts typically have a generation time of five years under orchard conditions (J.H. Craddock, pers. comm.), at least five years would be required to produce F₂ or BC₁ mapping populations. An alternative approach, association mapping, would not require crosses to be made, and would therefore be a quicker process; however, association mapping would require more genomic resources and computational power (Allendorf et al. 2013).

The possible hybrid origin of morphologically “intermediate” *Castanea* in the southern and central Appalachians, i.e., *C. × neglecta*, has yet to be thoroughly investigated. To address the hypothesized hybrid origin of a species in the genus *Helianthus*, Bock et al. (2014b) employed a method known as “genome skimming.” With this method, the high-copy genomic fraction—comprising the plastid genome, the mitochondrial genome, and nuclear ribosomal DNA (rDNA)—are assembled and analyzed. Because markers from rDNA are biparentally inherited, mtDNA and cpDNA are maternally inherited, and a large amount of data is generated for a relatively low cost, it seems that this may be an appropriate technique to help resolve the hypothesis of natural hybridity in the North American *Castanea* (L.H. Rieseberg, U. of British Columbia, pers. comm.). The use of whole plastome data to infer evolution in the North American *Castanea* is particularly appealing. With such a dataset in hand, statistical methods (e.g., McDonald-Kreitman tests (1991)) could be used to test for evidence of selection in extra-nuclear genes of interest. Tests for selection may be useful in parsing out the roles of founder effects and natural selection on the geographic patterns of genetic variation in the group.

A different approach, however, may be needed to resolve a problem that has plagued this and previous studies: whether the shared haplotypes are due to interspecific hybridization or “deep coalescence” of alleles that predate speciation. Recent empirical work has shown that sampling large numbers of nuclear loci may be needed to ensure enough informative loci to accurately identify and validate shallow-scale divergences (Hime et al. 2016). Shallow-scale divergence between the North American *Castanea* species may indeed be the case here, and sampling large numbers of nuclear loci with a technique such as genotyping-by-sequencing (Elshire et al. 2011) or RAD-seq (Davey et al. 2011) may help to detect such shallow scale divergences among the North American chestnuts and chinquapins.

An important unresolved question concerns the direction of hybridization and introgression in the North American *Castanea*, if it has occurred recently. In other words, is hybridization between *C. dentata* and *C. pumila* asymmetric? Extensive studies in the sympatric European oak species, *Quercus petraea* and *Q. rubra*, have shown that the post-glacial spread of *Q. petraea* relied on pollination of *Q. robur* (an early successional species and a better disperser of seed) and repeated pollination of the F₁ progeny, BC₁ progeny, and later generation hybrids (Petit et al. 2004). The resulting plant represents a ‘resurrection’ of the *Q. petraea* nuclear genome, with a chloroplast genome of *Q. robur*. Petit et al. (2004) argued that *Q. petraea* populations emerging from such a process would be better adapted to local conditions than those arising from regular seed dispersal if QTLs involved in local adaptation were transferred between species during the process. Similar ecological models may help us better understand the dynamics of introgression in the North American *Castanea* species.

To better understand the direction of gene flow, a potential future study might employ 12 nuclear microsatellite markers—one per linkage group—in conjunction with the cpDNA markers

used here. In fact, this was my plan at the outset of this project. However, time and resources did not permit this. Similar methods have been used recently to understand hybridization between walnut (*Juglans*) species (Hoban et al. 2012). This approach was used by Hoban et al. (2012) to determine the identity of the seed parents and to classify hybrids as F₁, F₂, BC₁, or the products of more complex crosses. The influence of landscape type on directionality of gene flow was also determined by the authors. Implementing such an experimental design for chestnut, however, would require further marker development, since no species-specific microsatellite alleles have been identified in *C. dentata*, *C. pumila*, or *C. ozarkensis*, to my knowledge.

Another equally important question regards the fact that *C. dentata* and *C. pumila* have both been documented with M and P chlorotypes, but no *C. pumila* have been documented with the D chlorotype. Dane (2009) reported two *C. pumila* accessions, from Rabun Co., GA, that possessed deletions thought to be restricted to *C. dentata*, however, it is not clear if this haplotype, called H_{P7}, is the same as the D haplotypes defined here, since the two studies employed different cpDNA loci. Nevertheless, it is interesting that no D haplotypes have been found in the 164 *C. pumila* accessions screened in this and previous studies to use the same loci (Binkley 2008; Shaw et al. 2012).

A recently discovered feature of *C. dentata*'s reproductive biology may provide a mechanism for this pattern. Sisco et al. (2014) found that in crosses between *C. dentata* as female × *C. mollissima* as male, a *C. dentata* mother tree with the D haplotype always yielded progeny with the male-sterile phenotype (i.e., the inability to produce pollen). Since extra-nuclear genomes are inherited maternally in chestnut, all of the male-sterile F₁ progeny had inherited the D haplotype from the female parent. The reciprocal cross, with *C. mollissima* as female and D haplotype *C. dentata* as male, resulted in male-fertile F₁ progeny. In contrast, the

cross of a P or M haplotype *C. dentata* mother tree \times *C. mollissima* father yielded male-fertile F₁ progeny. It was hypothesized that the male-sterile phenotype in these progeny was caused by the interaction of nuclear genes inherited from *C. mollissima* with mitochondrial genes inherited from *C. dentata*. While the crosses analyzed in this study were between *C. dentata* and its Asian congeners, it is possible that the sterilizing cytoplasm associated with the *C. dentata*'s D haplotype may also result in male-sterile F₁ progeny in crosses with its sympatric congener *C. pumila*. Since cytoplasmic male sterility (CMS) is thought to be under frequency-dependent selection in populations of hermaphroditic plants (Rieseberg and Willis 2007), *C. pumila* with the CMS phenotype may occur rarely in natural populations.

To my knowledge, R.A. Jaynes has conducted the only phenotypic studies on progeny of controlled crosses between *C. dentata* and *C. pumila* (Jaynes 1961). Unfortunately, at the time of data collection, F₁ hybrids of *C. dentata* \times *C. pumila* were not phenotyped for male-sterility. The aforementioned work involving controlled crosses could be used in an attempt to answer the question of male-sterility in *C. dentata* \times *C. pumila* hybrids.

In conclusion, a more thorough knowledge of gene flow between *C. dentata* and *C. pumila* is likely to benefit taxonomic work and conservation efforts. Recent theoretical and empirical studies have suggested an important adaptive role for variation in the cytoplasmic genomes of plants (Bock et al. 2014a). In particular, such studies have found that certain chloroplast and mitochondrial genotypes may be more fit in harsh environments. If this is indeed the case, it provides further impetus to conserve a diversity of extra-nuclear genotypes in the North American *Castanea* species, as has been previously recommended (Dane and Sisco 2014).

REFERENCES

- Alexander L. 2016. Rapid, effective DNA isolation from *Osmanthus* via modified alkaline lysis. *Journal of Biomolecular Techniques* **27**: 000-000.
- Allendorf FW, Luikart G, Aitken SN. 2013. *Conservation and the genetics of populations*. Wiley-Blackwell, West Sussex, UK.
- Anagnostakis SL. 2001. The effect of multiple importations of pests and pathogens on a native tree. *Biological Invasions* **3**: 245-254.
- Ashe WW. 1922. Notes on trees and shrubs of the southeastern United States. *Bulletin of the Torrey Botanical Club* **49**: 43-48.
- . 1923. Further notes on trees and shrubs of the southeastern United States. *Ibid* **50**: 359-363
- . 1924. Notes on woody plants. *Journal of the Elisha Mitchell Scientific Society* **40**: 43-48.
- . 1925. Notes on woody plants. *Charleston Museum Quarterly* **1**: 28-32.
- . 1927. *Magnolia cordata* and other woody plants. *Bulletin of the Torrey Botanical Club* **54**: 579-582.
- Bandelt H-J, Forster P, Röhl A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* **16**: 37-48.
- Binkley MA. 2008. The Phylogeography of North American Chestnuts and Chinquapins (*Castanea* Mill., Fagaceae). M.S. thesis. University of Tennessee at Chattanooga.
- Bock DG, Andrew RL, Rieseberg LH. 2014a. On the adaptive value of cytoplasmic genomes in plants. *Molecular Ecology* **23**: 4899-4911.
- Bock DG, Kane NC, Ebert DP, Rieseberg LH. 2014b. Genome skimming reveals the origin of the Jerusalem Artichoke tuber crop species: neither from Jerusalem nor an artichoke. *New Phytologist* **201**: 1021-1030.
- Burnham CR. 1988. The restoration of the American chestnut: Mendelian genetics may solve a problem that has resisted other approaches. *American Scientist* **76**: 478-487.

- Camus A. 1929. Les chataigniers. Monographie des genres *Castanea* et *Castanopsis*. in *Encyclopedie economique de sylviculture*. Paul Lechevalier, Paris.
- Clement M, Posada D, Crandall KA. 2000. TCS: a computer program to estimate gene genealogies. *Molecular Ecology* **9**: 1657-1659.
- Craddock JH. 2014. A tree crop archetype. *Acta Horticulturae* **1019**: 45-48.
- Dane F. 2009. Comparative phylogeography of *Castanea* species. *Acta Horticulturae* **844**: 211-221.
- Dane F, Hawkins LK, Huang H. 1999. Genetic variation and population structure of *Castanea pumila* var. *ozarkensis*. *Journal of the American Society for Horticultural Science* **124**: 666-670.
- Dane F, Sisco PH. 2014. Genetic diversity of American chestnut is highest in the Southern US: the evidence from nuclear and chloroplast DNA studies. *Journal of the American Chestnut Foundation* **28**: 9-13.
- Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* **9**: 772-772.
- Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML. 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics* **12**: 499-510.
- Delcourt HR. 1979. Late Quaternary vegetation history of the eastern Highland Rim and adjacent Cumberland Plateau of Tennessee. *Ecological Monographs* **49**: 255-280.
- Dode L-A. 1908. Notes dendrologiques. VI. Sur les chataigniers. *Bulletin de la Societe Dendrologique de France* **8**: 140-159.
- Elias TS. 1971. The genera of Fagaceae in south-eastern United States. *Journal of the Arnold Arboretum* **52**: 159-195.
- Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS One* **6**: e19379.
- Faison EK, Foster DR. 2014. Did American Chestnut really dominate the eastern forest? *Arnoldia* **72**: 18-32.
- Fu Y, Dane F. 2003. Allozyme variation in endangered *Castanea pumila* var. *pumila*. *Annals of Botany* **92**: 223-230.
- Gielly L, Taberlet P. 1994. The use of chloroplast DNA to resolve plant phylogenies: noncoding versus *rbcL* sequences. *Molecular Biology and Evolution* **11**: 769-777.

- Glez-Peña D, Gómez-Blanco D, Reboiro-Jato M, Fdez-Riverola F, Posada D. 2010. ALTER: program-oriented conversion of DNA and protein alignments. *Nucleic Acids Research* **38**: W14-W18.
- Gronovius J. 1739. *Flora Virginica*. Cornelius Haak, Leiden.
- Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology* **52**: 696-704.
- Hardin JW, Johnson GP. 1985. Atlas of foliar surface features in woody plants, VIII. *Fagus* and *Castanea* (Fagaceae) of eastern North America. *Bulletin of the Torrey Botanical Club* **112**: 11-20.
- Hebard FV, Georgi L, Donahue J, Bevins D, Coalson E. 2013. Meadowview Notes 2011-2012. *Journal of the American Chestnut Foundation* **26**: 19-25.
- Hime PM, Hotaling S, Grewelle RE, O'Neill EM, Voss SR, Shaffer HB, Weisrock DW. 2016. The influence of locus number and information content on species delimitation: an empirical test case in an endangered Mexican salamander. *Molecular Ecology* **In press**.
- Hoban SM, McCleary TS, Schlarbaum SE, Anagnostakis SL, Romero-Severson J. 2012. Human-impacted landscapes facilitate hybridization between a native and an introduced tree. *Evolutionary Applications* **5**: 720-731.
- Huang H, Dane F, Kubisiak TL. 1998. Allozyme and RAPD analysis of the genetic diversity and geographic variation in wild populations of the American chestnut (Fagaceae). *American Journal of Botany* **85**: 1013-1021.
- Huelsenbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**: 754-755.
- Jacobs DF, Dalglish HJ, Nelson CD. 2013. A conceptual framework for restoration of threatened plants: the effective model of American chestnut (*Castanea dentata*) reintroduction. *New Phytologist* **197**: 378-393.
- Jansen RK, Saski C, Lee S-B, Hansen AK, Daniell H. 2011. Complete plastid genome sequences of three rosids (*Castanea*, *Prunus*, *Theobroma*): evidence for at least two independent transfers of rpl22 to the nucleus. *Molecular Biology and Evolution* **28**: 835-847.
- Jaynes RA. 1961. Genetic and cytological studies in the genus *Castanea*. PhD dissertation. Yale University.
- . 1975. Chestnuts. in *Advances in Fruit Breeding* (eds. J Janick, JN Moore), pp. 490-503. Purdue University Press, West Lafayette, IN

- Johnson GP. 1988. Revision of *Castanea* sect. *Balanocastanon* (Fagaceae). *Journal of the Arnold Arboretum* **69**: 25-49.
- Joly S, McLenachan PA, Lockhart PJ. 2009. A statistical approach for distinguishing hybridization and incomplete lineage sorting. *American Naturalist* **174**: E54-E70.
- Kremer A, Abbott AG, Carlson JE, Manos PS, Plomion C, Sisco P, Staton ME, Ueno S, Vendramin GG. 2012. Genomics of Fagaceae. *Tree Genetics & Genomes* **8**: 583-610.
- Kubisiak T, Nelson C, Staton M, Zhebentyayeva T, Smith C, Olukolu B, Fang G-C, Hebard F, Anagnostakis S, Wheeler N et al. 2013. A transcriptome-based genetic map of Chinese chestnut (*Castanea mollissima*) and identification of regions of segmental homology with peach (*Prunus persica*). *Tree Genetics & Genomes* **9**: 557-571.
- Kubisiak TL, Roberds JH. 2003. Genetic variation in natural populations of American Chestnut. *Journal of the American Chestnut Foundation* **16**: 43-48.
- . 2006. Genetic structure of American chestnut populations based on neutral DNA markers. in *Restoration of American Chestnut to Forest Lands: Proceedings of a Conference and Workshop at the North Carolina Arboretum* (eds. K Steiner, J Carlson), pp. 109-122. North Carolina Arboretum, Asheville, NC.
- Lang P, Dane F, Kubisiak TL, Huang H. 2007. Molecular evidence for an Asian origin and a unique westward migration of species in the genus *Castanea* via Europe to North America. *Molecular Phylogenetics and Evolution* **43**: 49-59.
- Li X, Dane F. 2013. Comparative chloroplast and nuclear DNA analysis of *Castanea* species in the southern region of the USA. *Tree Genetics & Genomes* **9**: 107-116.
- Linnaeus C. 1753. *Species plantarum*. Laurentii Salvii, Stockholm.
- Little EL. 1977. *Castanea*. in *Atlas of United States trees-v 4: Minor eastern hardwoods*. US Government Printing Office, Washington, DC.
- . 1979. *Castanea*. in *Checklist of United States trees (native and naturalized)*. US Government Printing Office, Washington, DC.
- McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* **351**: 652-654.
- Milgroom MG, Cortesi P. 2004. Biological control of chestnut blight with hypovirulence: a critical analysis. *Annual Review of Phytopathology* **42**: 311-338.
- Miller P. 1768. *The Gardeners Dictionary*. John and Francis Rivington, London.
- Nei M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* **89**: 583-590.

- Nixon KC. 1997. Fagaceae Dumortier, Beech Family. in *Flora of North America North of Mexico* (ed. Flora of North America Editorial Committee), pp. 436-437. Oxford University Press, New York, New York.
- Paillet FL. 1993. Growth form and life histories of American chestnut and Allegheny and Ozark chinquapin at various North American sites. *Bulletin of the Torrey Botanical Club* **120**: 257-268.
- . 2002. Chestnut: history and ecology of a transformed species. *Journal of Biogeography* **29**: 1517-1530.
- Payne JA, Miller G, Johnson GP, Senter SD. 1994. *Castanea pumila* (L.) Mill.: an underused native nut tree. *HortScience* **29**: 62, 130-131.
- Petit RJ, Bodénès C, Ducouso A, Roussel G, Kremer A. 2004. Hybridization as a mechanism of invasion in oaks. *New Phytologist* **161**: 151-164.
- Provan J, Powell W, Hollingsworth PM. 2001. Chloroplast microsatellites: new tools for studies in plant ecology and evolution. *Trends in Ecology & Evolution* **16**: 142-147.
- Rieseberg LH. 1991. Homoploid reticulate evolution in *Helianthus* (Asteraceae): evidence from ribosomal genes. *American Journal of Botany* **78**: 1218-1237.
- Rieseberg LH, Beckstrom-Sternberg SM, Liston A, Arias DM. 1991. Phylogenetic and systematic inferences from chloroplast DNA and isozyme variation in *Helianthus* sect. *Helianthus* (Asteraceae). *Systematic Botany*: 50-76.
- Rieseberg LH, Soltis DE. 1991. Phylogenetic consequences of cytoplasmic gene flow in plants. *Evolutionary Trends in Plants* **5**: 65-84.
- Rieseberg LH, Willis JH. 2007. Plant speciation. *Science* **317**: 910-914.
- Sargent CS. 1917. *Plantae Wilsonianae*. The University Press, Cambridge, Massachusetts.
- . 1919. Notes on North American trees. IV. *Botanical Gazette* **67**: 208-242.
- Shaw J, Craddock JH, Binkley MA. 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.
- Shaw J, Lickey EB, Beck JT, Farmer SB, Liu W, Miller J, Siripun KC, Winder CT, Schilling EE, Small RL. 2005. The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *American Journal of Botany* **92**: 142-166.
- Shaw J, Lickey EB, Schilling EE, Small RL. 2007. Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. *American Journal of Botany* **94**: 275-288.

- Shaw J, Shafer HL, Leonard OR, Kovach MJ, Schorr M, Morris AB. 2014. Chloroplast DNA sequence utility for the lowest phylogenetic and phylogeographic inferences in angiosperms: The tortoise and the hare IV. *American Journal of Botany* **101**: 1987-2004.
- Sisco P, Neel T, Hebard F, Craddock J, Shaw J. 2014. Cytoplasmic male sterility in interspecific hybrids between American and Asian *Castanea* species is correlated with the American *D* chloroplast haplotype. *Acta Horticulturae* **1019**: 215-222.
- Small JK. 1933. *Manual of the Southeastern Flora*. By the author, New York.
- Taberlet P, Gielly L, Pautou G, Bouvet J. 1991. Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology* **17**: 1105-1109.
- Templeton AR, Crandall KA, Sing CF. 1992. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics* **132**: 619-633.
- Tucker G. 1975. *Castanea pumila* var. *ozarkensis* (Ashe) Tucker, comb. nov. *Proceedings of the Arkansas Academy of Sciences* **29**: 67-69.
- Weakley AS. 2015. *Castanea* P. Miller 1754 (Chestnut, Chinquapin). in *Flora of the Southern and Mid-Atlantic States*. University of North Carolina Herbarium, Chapel Hill, North Carolina.
- Wen J. 1999. Evolution of eastern Asian and eastern North American disjunct distributions in flowering plants. *Annual Review of Ecology and Systematics* **30**: 421-455.
- Zhang B, Oakes AD, Newhouse AE, Baier KM, Maynard CA, Powell WA. 2013. A threshold level of oxalate oxidase transgene expression reduces *Cryphonectria parasitica*-induced necrosis in a transgenic American chestnut (*Castanea dentata*) leaf bioassay. *Transgenic Research* **22**: 973-982.

APPENDIX

A. TABLES

Table 1 Information for Ashe's collections annotated as Ozark Chinquapin
Annotations by Johnson (1988) and W.H. Duncan (unpublished). W.W. Ashe accessions were received on loan from the University of North Carolina Chapel Hill Herbarium (NCU).

NCU barcode #	Locality	Collection date	Ashe's determination
00053427	Lawrence Co., AL	8/2/28	<i>C. alabamensis</i>
00053426	Lawrence Co., AL	8/2/28	<i>C. alabamensis</i>
00050202	Lawrence Co., AL	7/31/28	<i>C. alabamensis</i>
00053434	Lawrence Co., AL	7/30/28	<i>C. alabamensis</i>
00054320	Bibb Co., AL	8/12/26	<i>C. alabamensis</i>
00053425	Winston Co., AL	5/27/25	<i>Castanea</i> sp.
00053431	Lawrence Co., AL	5/24/25	<i>C. alabamensis</i>
00053429	Lawrence Co., AL	5/24/25	<i>C. alabamensis</i>
00053428	Lawrence Co., AL	5/24/25	<i>C. alabamensis</i>
00053433	Lawrence Co., AL	5/24/25	<i>C. alabamensis</i>
00050092	Lawrence Co., AL	5/24/25	<i>C. alabamensis</i>
00050093	Lawrence Co., AL	10/25	<i>C. alabamensis</i>
00050094	Lawrence Co., AL	7/30/28	<i>C. alabamensis</i>
00050095	Lawrence Co., AL	5/24/25	<i>C. alabamensis</i>
00050096	Lawrence Co., AL	5/24/25	<i>C. alabamensis</i>
00053432	Lawrence Co., AL	5/24/25	<i>C. alabamensis</i>
00050097	Lawrence Co., AL	5/23/25	<i>C. alabamensis</i>
00050098	Lawrence Co., AL	5/23/25	<i>C. alabamensis</i>
00000049	Lawrence Co., AL	5/20/24	<i>C. alabamensis</i>

Table 2 North American *Castanea* accessions genotyped for the present study

All accessions will be deposited at the University of Tennessee at Chattanooga Herbarium (UCHT). Sequencing failures occurred at a few loci for some plants; therefore, haplotypes for these plants are to be determined (TBD).

Haplotype	Species	Morphotype (if applicable)	County, state	Site code	Sample #	Date collected	Elevation (m)	Collectors
P13	<i>C. dentata</i>		McDowell, NC	Old NC10	N/A	9/15/14	426	P.H. Sisco
M14	<i>C. pumila</i>		Charleston, SC	A	1	5/11/15	16	T. Perkins, M. Klinghard
M14	<i>C. pumila</i>		Charleston, SC	A	2	5/11/15	16	T. Perkins, M. Klinghard
M14	<i>C. pumila</i>		Charleston, SC	A	3	5/11/15	16	T. Perkins, M. Klinghard
M14	<i>C. pumila</i>		Charleston, SC	A	4	5/11/15	16	T. Perkins, M. Klinghard
M14	<i>C. pumila</i>		Charleston, SC	A	5	5/11/15	16	T. Perkins, M. Klinghard
M14	<i>C. pumila</i>		Charleston, SC	A	6	5/11/15	16	T. Perkins, M. Klinghard
M14	<i>C. pumila</i>		Charleston, SC	A	7	5/11/15	16	T. Perkins, M. Klinghard
M14	<i>C. pumila</i>		Charleston, SC	A	8	5/11/15	16	T. Perkins, M. Klinghard
M14	<i>C. pumila</i>		Charleston, SC	A	9	5/11/15	16	T. Perkins, M. Klinghard
M14	<i>C. pumila</i>		Charleston, SC	A	10	5/11/15	16	T. Perkins, M. Klinghard
O24	<i>C. pumila</i>		Sumter, SC	WB	1	5/11/15	44	T. Perkins, M. Klinghard
O24	<i>C. pumila</i>		Sumter, SC	WB	2	5/11/15	44	T. Perkins, M. Klinghard
M22	<i>C. dentata</i>		Pickens, SC	CEF	1	5/12/15	214	T. Perkins
M22	<i>C. dentata</i>		Pickens, SC	CEF	2	5/12/15	214	T. Perkins
M22	<i>C. dentata</i>		Pickens, SC	CEF	3	5/12/15	216	T. Perkins
M22	<i>C. dentata</i>		Pickens, SC	CEF	4	5/12/15	216	T. Perkins
D2	<i>C. dentata</i>		Transylvania, NC	G	1	5/15/15	807	T. Perkins, J. Tzeng
P23	<i>C. pumila</i>		Transylvania, NC	G	2	5/15/15	892	T. Perkins, J. Tzeng
P23	<i>C. pumila</i>		Transylvania, NC	G	3	5/15/15	892	T. Perkins, J. Tzeng
O3	<i>C. pumila</i>		Oconee, SC	BF	1	5/22/15	336	T. Perkins, J. Tzeng
O3	<i>C. pumila</i>		Oconee, SC	BF	2	5/22/15	333	T. Perkins, J. Tzeng
O3	<i>C. pumila</i>		Oconee, SC	BF	3	5/22/15	322	T. Perkins, J. Tzeng

Haplotype	Species	Morphotype (if applicable)	County, state	Site code	Sample #	Date collected	Elevation (m)	Collectors
O3	<i>C. pumila</i>		Oconee, SC	BF	4	5/22/15	313	T. Perkins, J. Tzeng
M20	<i>C. dentata</i>		Oconee, SC	BF	5	5/22/15	336	T. Perkins, J. Tzeng
M21	<i>C. dentata</i>		Oconee, SC	BF	6	5/22/15	338	T. Perkins, J. Tzeng
M15	<i>C. dentata</i>		Oconee, SC	CR	1	5/17/15	431	T. Perkins
M15	<i>C. dentata</i>		Oconee, SC	CR	2	5/17/15	426	T. Perkins
M15	<i>C. pumila</i>		Oconee, SC	CR	3	5/17/15	433	T. Perkins
M19	<i>C. pumila</i>		Oconee, SC	CR	4	5/17/15	436	T. Perkins
M15	<i>C. pumila</i>		Oconee, SC	CR	5	5/17/15	440	T. Perkins
M19	<i>C. pumila</i>		Oconee, SC	CR	6	5/17/15	438	T. Perkins
M19	<i>C. pumila</i>		Oconee, SC	CR	7	5/17/15	435	T. Perkins
D2	<i>C. dentata</i>		Cocke, TN	CC	1	6/10/15	1276	T. Perkins, J.H. Craddock
D2	<i>C. dentata</i>		Madison, NC	MP	1	6/10/15	1258	T. Perkins, J.H. Craddock
TBD	<i>C. dentata</i>		Avery, NC	GF	1	6/13/15	1418	T. Perkins, M. Klinghard
M19	<i>C. dentata</i>		Avery, NC	GF	2	6/13/15	1430	T. Perkins, M. Klinghard
O28	<i>C. pumila</i>		Wheeler, GA	L	1	7/30/15	58	T. Perkins, J.H. Craddock
O28	<i>C. pumila</i>		Wheeler, GA	L	2	7/30/15	52	T. Perkins, J.H. Craddock
O31	<i>C. pumila</i>	<i>C. floridana</i>	Suwannee, FL	S	1	7/31/15	41	R. Simons, T. Perkins, J.H. Craddock
O31	<i>C. pumila</i>	<i>C. floridana</i>	Suwannee, FL	S	2	7/31/15	41	R. Simons, T. Perkins, J.H. Craddock
O31	<i>C. pumila</i>	<i>C. floridana</i>	Suwannee, FL	S	3	7/31/15	41	R. Simons, T. Perkins, J.H. Craddock
O31	<i>C. pumila</i>	<i>C. alnifolia</i>	Suwannee, FL	B	1	7/31/15	20	R. Simons, T. Perkins, J.H. Craddock
O31	<i>C. pumila</i>	<i>C. alnifolia</i>	Suwannee, FL	B	10	7/31/15	19	R. Simons, T. Perkins, J.H. Craddock
O31	<i>C. pumila</i>	<i>C. alnifolia</i>	Suwannee, FL	B	14	7/31/15	18	R. Simons, T. Perkins, J.H. Craddock
O31	<i>C. pumila</i>	<i>C. alnifolia</i>	Suwannee, FL	B	15	7/31/15	18	R. Simons, T. Perkins, J.H. Craddock
O31	<i>C. pumila</i>	<i>C. alnifolia</i>	Suwannee, FL	B	19	7/31/15	19	R. Simons, T. Perkins, Craddock

Haplotype	Species	Morphotype (if applicable)	County, state	Site code	Sample #	Date collected	Elevation (m)	Collectors
O31	<i>C. pumila</i>	<i>C. alnifolia</i>	Suwannee, FL	P	2	7/31/15	30	R. Simons, T. Perkins, J.H. Craddock
O31	<i>C. pumila</i>	<i>C. alnifolia</i>	Suwannee, FL	P	4	7/31/15	31	R. Simons, T. Perkins, J.H. Craddock
O31	<i>C. pumila</i>	<i>C. alnifolia</i>	Suwannee, FL	P	9	7/31/15	31	R. Simons, T. Perkins, J.H. Craddock
O31	<i>C. pumila</i>	<i>C. alnifolia</i>	Suwannee, FL	P	12	7/31/15	31	R. Simons, T. Perkins, J.H. Craddock
P33	<i>C. pumila</i>	<i>C. alnifolia</i>	Suwannee, FL	P	13	7/31/15	31	R. Simons, T. Perkins, J.H. Craddock
O32	<i>C. pumila</i>	<i>C. alnifolia</i>	Suwannee, FL	P	15	7/31/15	32	R. Simons, T. Perkins, J.H. Craddock
M29	<i>C. pumila</i>	<i>C. floridana</i>	Okaloosa, FL	F	1	8/1/15	4	T. Perkins, J.H. Craddock
M29	<i>C. pumila</i>	<i>C. floridana</i>	Okaloosa, FL	F	2	8/1/15	2	T. Perkins, J.H. Craddock
TBD	<i>C. pumila</i>		Walton, FL	N	1	8/8/15	24	F. Cuchens, T. Perkins
M29	<i>C. pumila</i>		Walton, FL	N	2	8/8/15	18	F. Cuchens, T. Perkins
M30	<i>C. pumila</i>		Walton, FL	N	3	8/8/15	23	F. Cuchens, T. Perkins
M30	<i>C. pumila</i>		Walton, FL	N	4	8/8/15	25	F. Cuchens, T. Perkins
M30	<i>C. pumila</i>		Walton, FL	N	5	8/8/15	28	F. Cuchens, T. Perkins
M17	<i>C. pumila</i>	<i>C. alabamensis</i>	Calhoun Co., AL	CH	15	10/29/15	368	J. Agricola, D. Morris, T. Perkins
M17	<i>C. pumila</i>	<i>C. alabamensis</i>	Calhoun Co., AL	CH	18	10/29/15	415	J. Agricola, D. Morris, T. Perkins
M17	<i>C. pumila</i>	<i>C. alabamensis</i>	Calhoun Co., AL	CH	21	10/29/15	394	J. Agricola, D. Morris, T. Perkins
M15	<i>C. pumila</i>	<i>C. alabamensis</i>	Calhoun Co., AL	CH	22	10/29/15	385	J. Agricola, D. Morris, T. Perkins
M6	<i>C. pumila</i>	<i>C. alabamensis</i>	Clay, AL	AG	1	11/5/15	448	D. Morris, L. Brasher, T. Perkins
M6	<i>C. pumila</i>	<i>C. alabamensis</i>	Clay, AL	AG	2	11/5/15	454	D. Morris, L. Brasher, T. Perkins
M6	<i>C. pumila</i>	<i>C. alabamensis</i>	Clay, AL	AG	3	11/5/15	466	D. Morris, L. Brasher, T. Perkins
M6	<i>C. pumila</i>	<i>C. alabamensis</i>	Clay, AL	AG	4	11/5/15	468	D. Morris, L. Brasher, T. Perkins
M6	<i>C. pumila</i>	<i>C. alabamensis</i>	Clay, AL	AG	6	11/5/15	466	D. Morris, L. Brasher, T. Perkins
M18	<i>C. pumila</i>	<i>C. alabamensis</i>	Clay, AL	AG	7	11/5/15	449	D. Morris, L. Brasher, T. Perkins
M16	<i>C. pumila</i>	<i>C. alabamensis</i>	Clay, AL	AG	8	11/5/15	450	D. Morris, L. Brasher, T. Perkins

Haplotype	Species	Morphotype (if applicable)	County, state	Site code	Sample #	Date collected	Elevation (m)	Collector
TBD	<i>C. pumila</i>	<i>C. alabamensis</i>	Clay, AL	AG	9	11/5/15	449	D. Morris, L. Brasher, T. Perkins
O25	<i>C. ozarkensis</i>		Madison, AR	Q	116	N/A	453	T. Witsell
O25	<i>C. ozarkensis</i>		Madison, AR	Q	117	N/A	489	T. Witsell
O26	<i>C. ozarkensis</i>		Madison, AR	Q	118	N/A	414	T. Witsell
<i>C. mollissima</i>	<i>C. mollissima</i>	putative <i>C. dentata</i>	Madison, AR	Q	119	N/A	414	T. Witsell
<i>C. mollissima</i>	<i>C. mollissima</i>	putative <i>C. dentata</i>	Madison, AR	Q	120	N/A	414	T. Witsell
O27	<i>C. ozarkensis</i>		Logan, AR	Q	121	N/A	713	T. Witsell

Table 3 Summary of DNA sequence polymorphisms observed in six noncoding cpDNA regions in *C. dentata*, *C. pumila*, and *C. ozarkensis*
 Percent variability was calculated using the method described by Shaw et al. (2007).

noncoding cpDNA region	SNPs	substitutions > 1 nt	microsatellites	indels	total polymorphic sites	aligned sequence length (bp)	% variability
<i>trnV-ndhC</i>	10	2	0	2	14	397	3.5
<i>rpL16</i>	5	0	1	1	7	866	0.8
<i>trnS-trnG</i>	4	0	3	4	11	600	1.8
<i>rpL32-trnL</i>	16	1	3	4	24	1296	1.85
<i>atpI-atpH</i>	5	0	4	4	13	926	1.4
<i>psbA-trnH</i>	6	0	1	1	8	524	1.5
Total (% of total polymorphic sites)	46 (59.7 %)	3 (3.9%)	12 (15.6%)	16 (20.8%)	77	4609	

APPENDIX

B. FIGURES

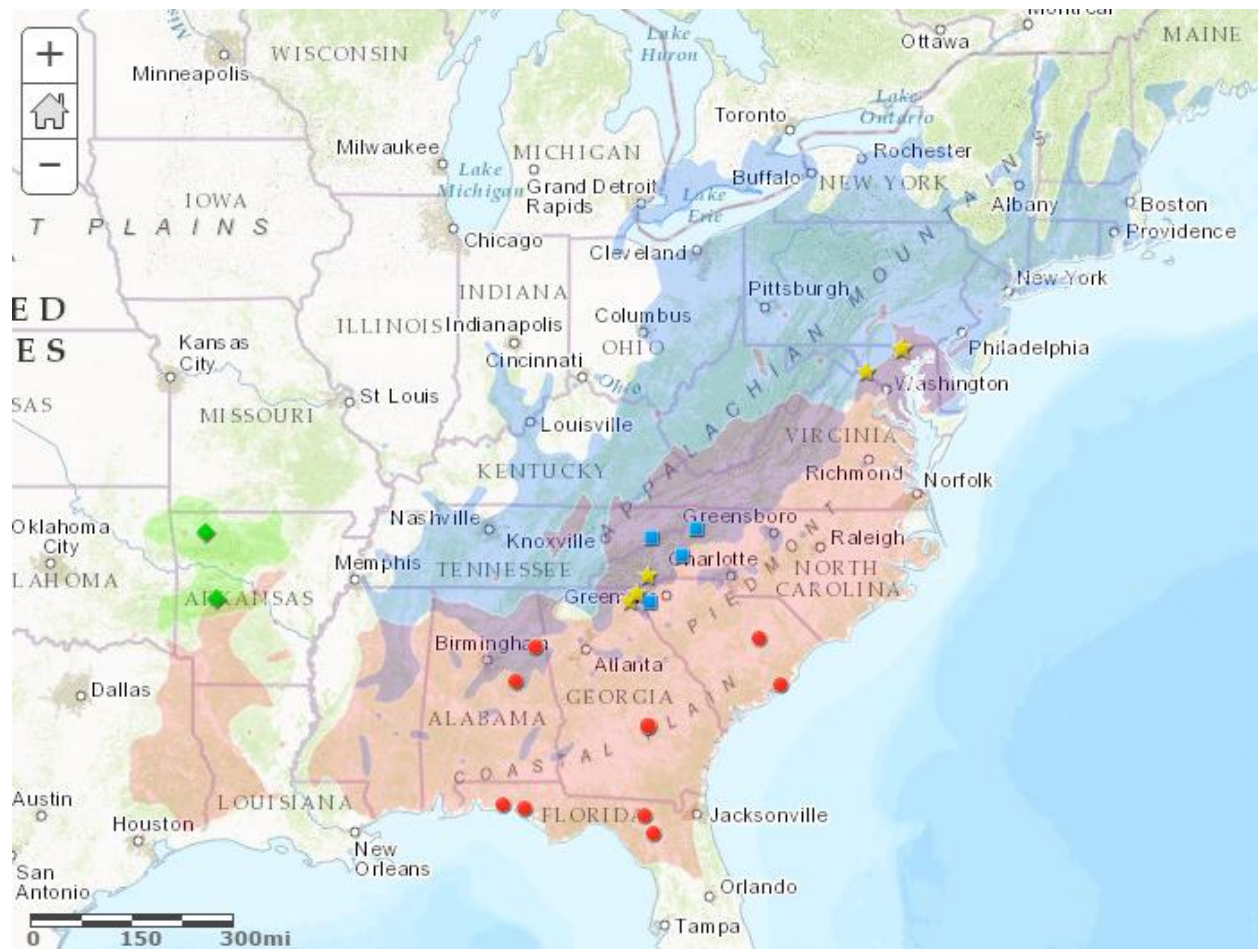


Figure 1 Distribution of three North American *Castanea* species and samples sites of the present study

Green shading indicates the distribution of *C. ozarkensis*; blue shading indicates the distribution of *C. dentata*; and red shading indicates the distribution of *C. pumila*. Note the large area of sympatry for *C. dentata* and *C. pumila*, indicated by the purple shading. The distribution data were taken from Little (1977). Symbols indicate the species sampled at the different sites. Green diamonds indicate *C. ozarkensis*; blue squares indicate *C. dentata*; red circles indicate *C. pumila*; and yellow stars indicate sites where both *C. dentata* and *C. pumila* were sampled. Only sample sites for which DNAs were extracted are shown.



Figure 2 Lectotype of *Castanea alabamensis* Ashe
 G.P. Johnson's *C. pumila* var. *ozarkensis* annotation is in the lower left-hand corner of the herbarium sheet. Specimen image courtesy of the University of North Carolina Chapel Hill Herbarium.

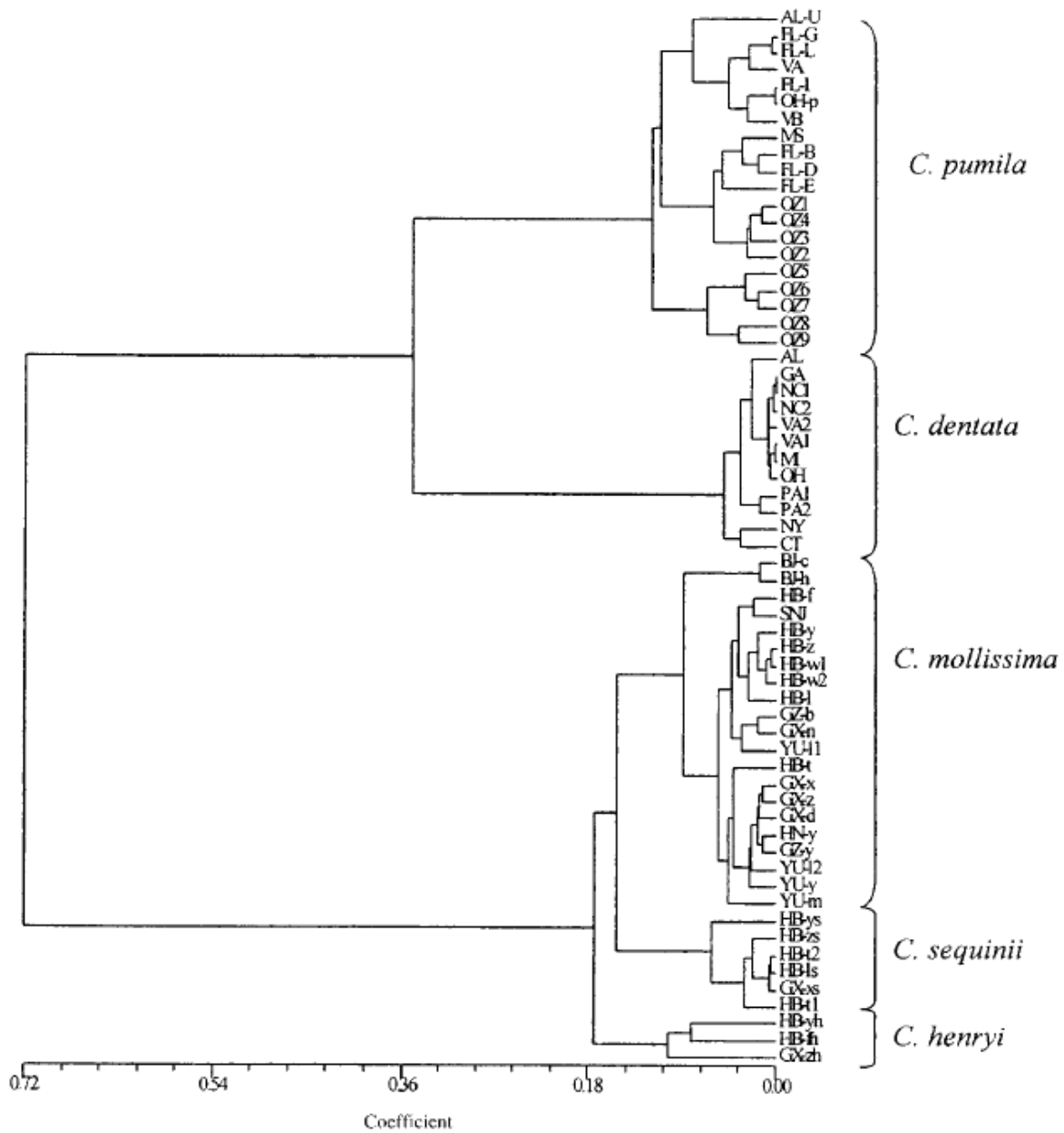


Figure 3 Dendrogram of allozyme divergence among five *Castanea* species from Dane et al. (2003)

C. pumila accessions with the identifier “OZ#” correspond to *C. pumila* var. *ozarkensis*, while all other *C. pumila* analyzed were *C. pumila* var. *pumila*.

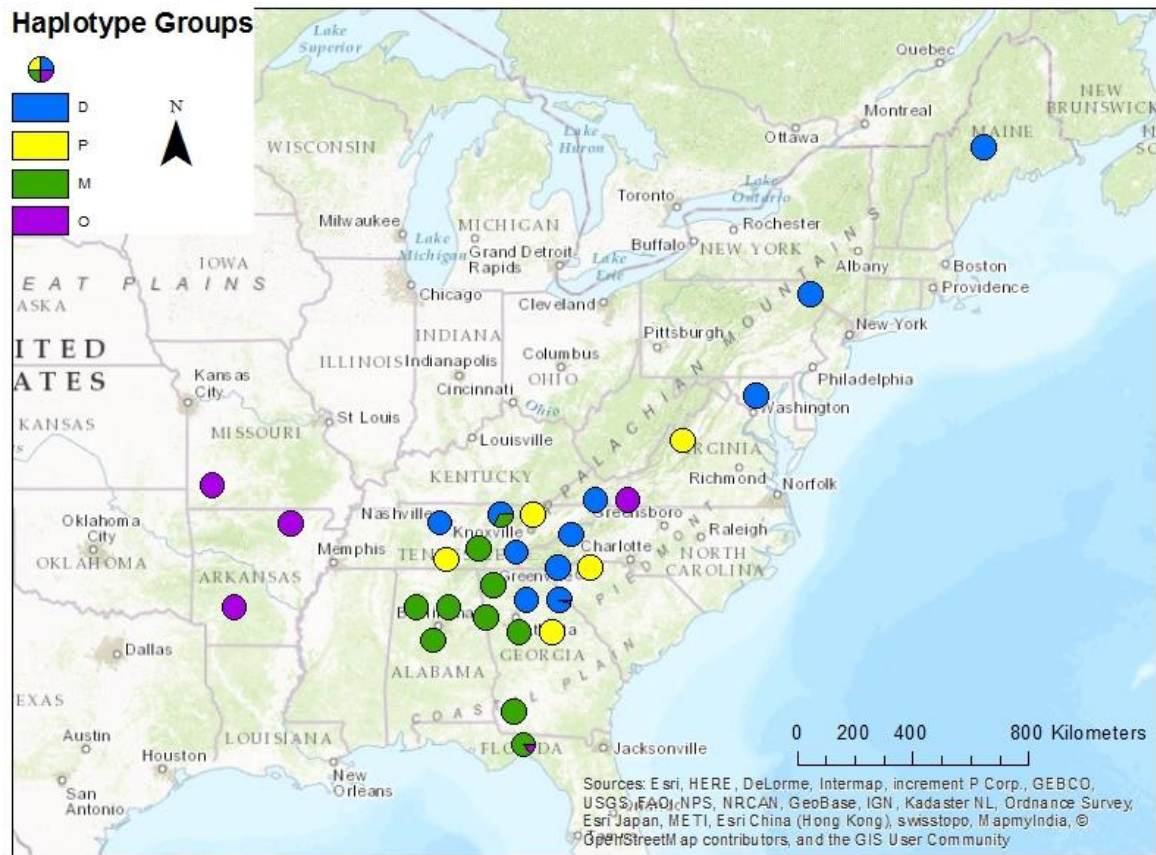


Figure 4 Allele frequencies in *C. dentata*, *C. pumila*, and *C. ozarkensis* populations sampled by Binkley (2008)
 Pie charts indicate populations sampled. Genotypic data were generated from the cpDNA locus *trnV-ndhC*.

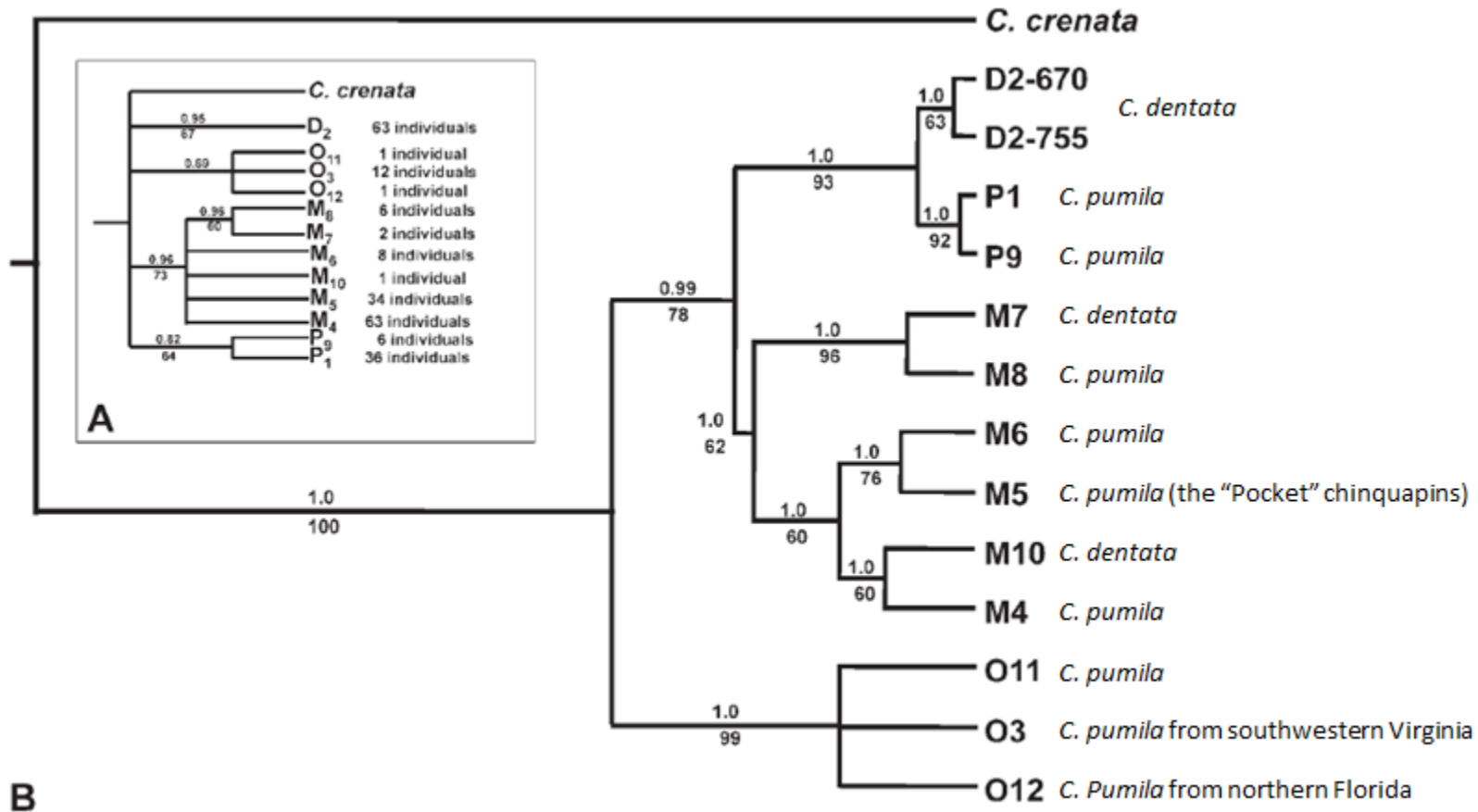


Figure 5 cpDNA phylogenies recovered by Binkley (2008)(inset A) and by Shaw et al. (2012)(inset B)

To show the haplotypes observed in different species, I have added species identifiers to the right of their respective branch tips on the tree of Shaw et al. (2012). It should be noted that only one accession for each of Binkley's (2008) twelve haplotypes (except D2) was sequenced at six loci by Shaw et al. (2012).

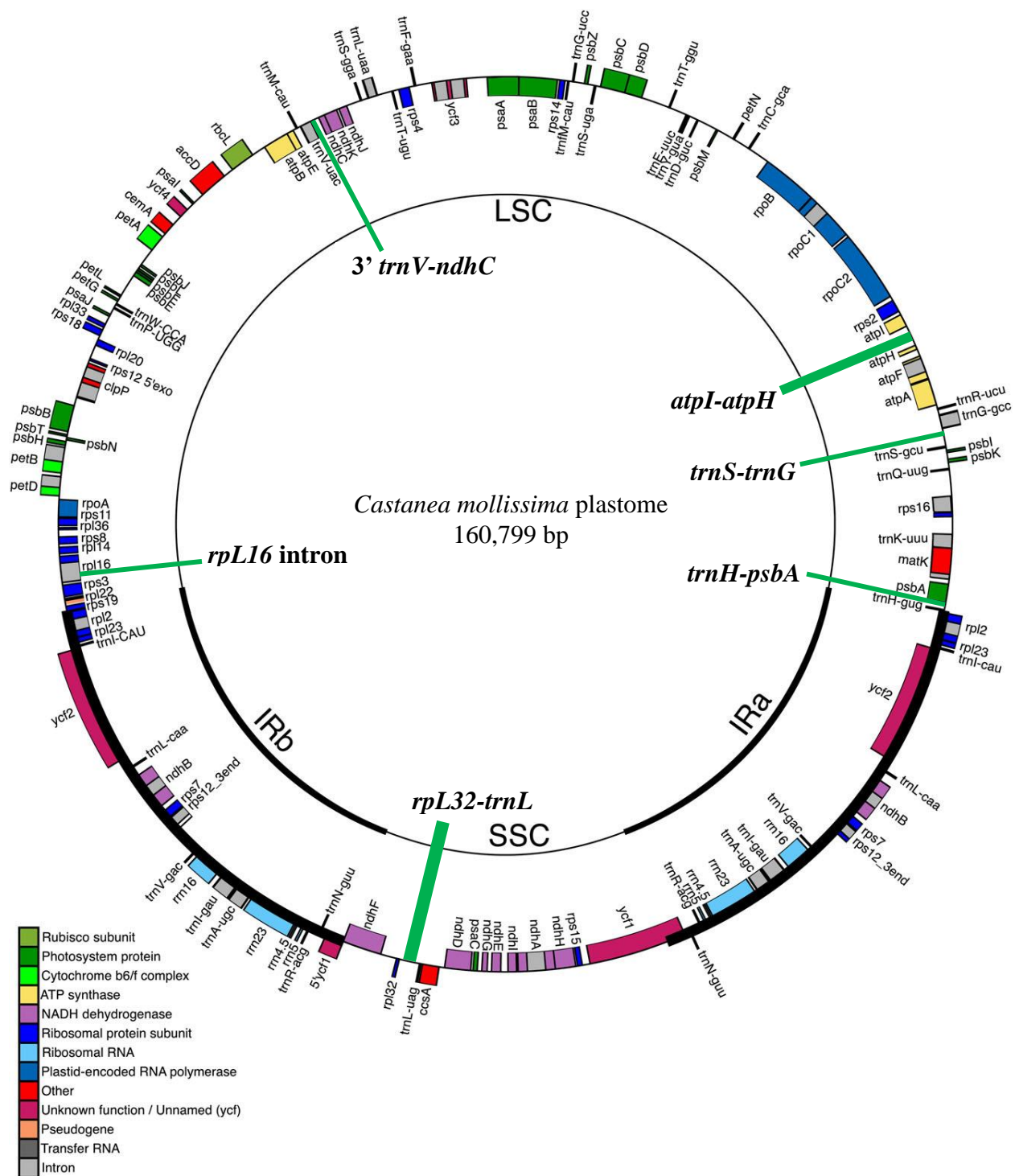


Figure 6 Circularized gene map of the plastid genome of *Castanea mollissima* (from Jansen et al. 2011) with the six loci sequenced for the present study. The six loci analyzed here are indicated by green lines. Colors of gene regions correspond to their function as indicated by the key above.

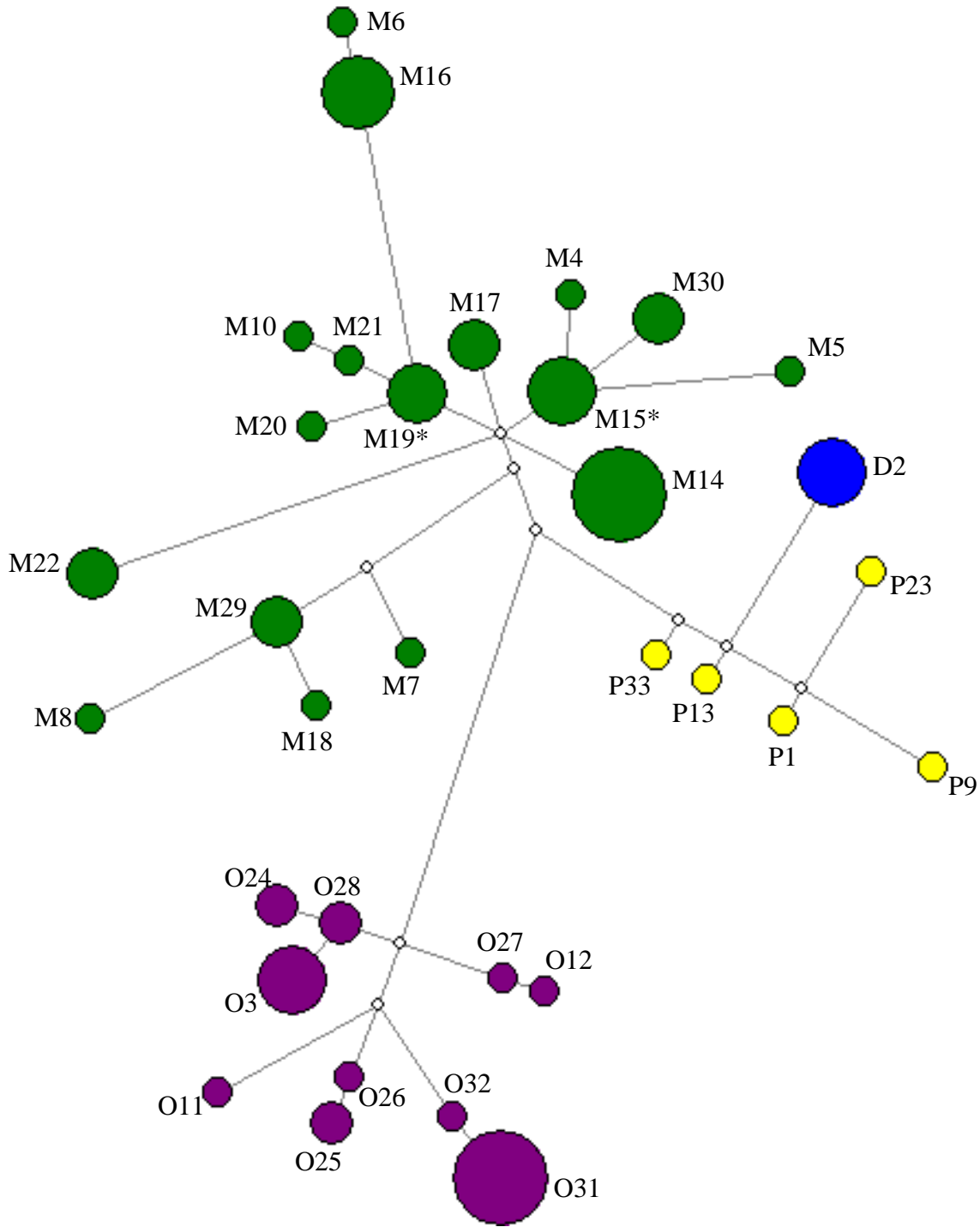


Figure 7 Median-joining network of 82 North American *Castanea* accessions
 33 unique cpDNA haplotypes—indicated by colored circles—were documented. Circle size is proportional to frequency of each haplotype. Colors correspond to the different clades revealed by Bayesian analysis: purple = O, yellow = P, blue = D, green = M. Branch lengths are approximately proportional to the number of mutations separating haplotypes. Haplotypes that were shared among two species are noted with an asterisk.

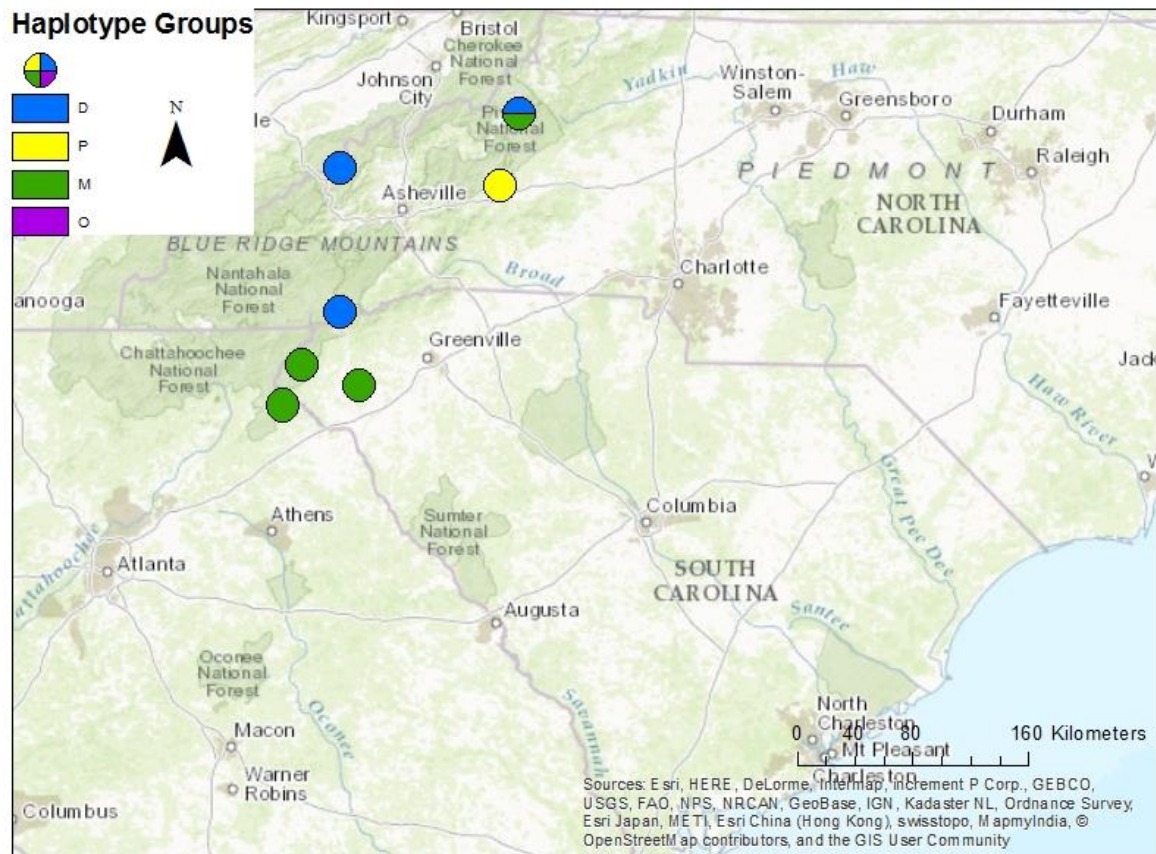


Figure 8 Allele frequencies in *C. dentata* populations sampled in the present study. Pie charts indicate the *C. dentata* populations sampled. Genotypic data were generated from the noncoding cpDNA loci *trnV-ndhC*, *rpL16*, *trnS-trnG*, *rpL32-trnL*, *atpI-atpH*, *psbA-trnH*.

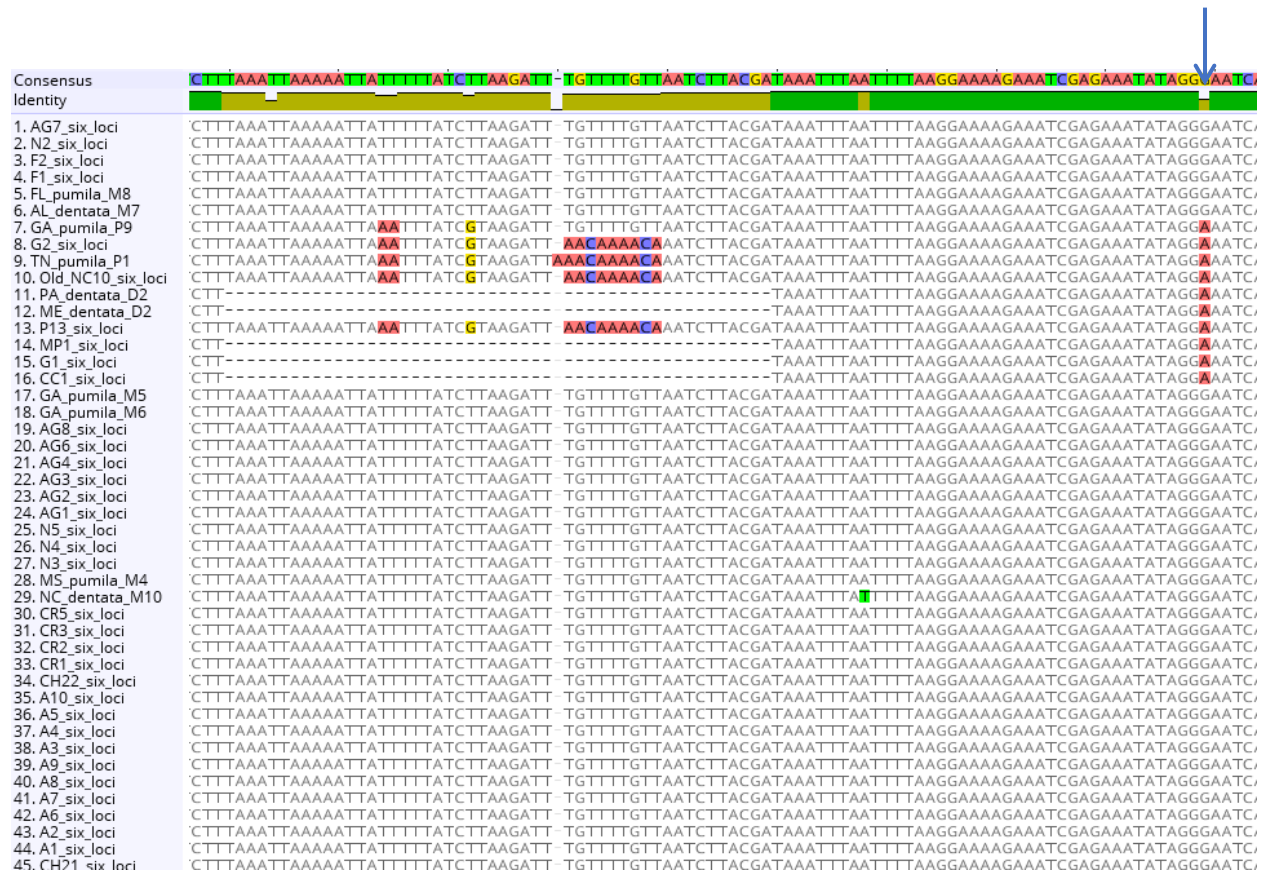


Figure 9 The distinguishing SNP for the M haplotypes occurs within the *trnV-ndhC* intergenic spacer. The distinguishing SNP is indicated by the blue arrow. This polymorphic site occurs 39 nt downstream of the indel that distinguishes D from non-D haplotypes. At this diagnostic SNP, plants with M haplotypes are fixed for guanine, while plants with O, P, and D haplotypes are fixed for adenine. Only plants with M, P, and D haplotypes are shown here.

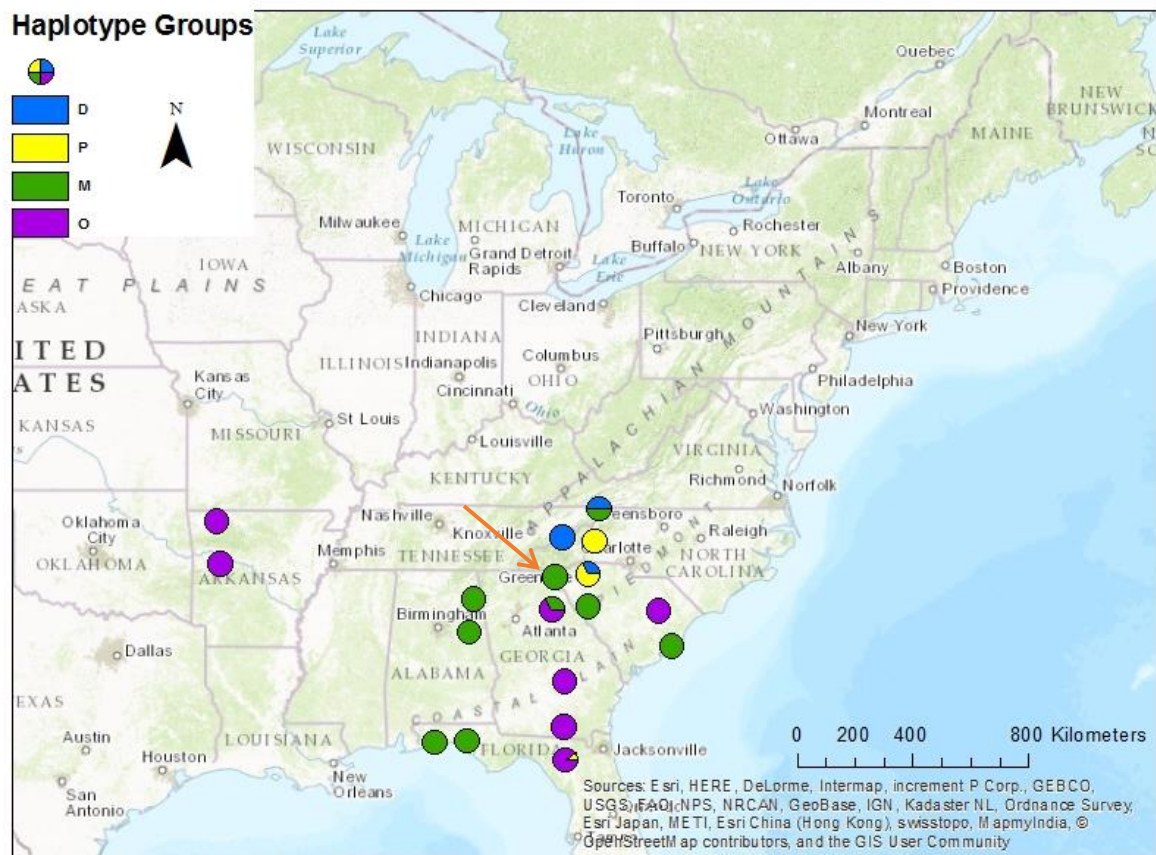


Figure 10 Allele frequencies in *C. dentata*, *C. pumila*, and *C. ozarkensis* populations sampled in the present study

Pie charts indicate the populations sampled. Genotypic data are for the noncoding cpDNA loci *trnV-ndhC*, *rpL16*, *trnS-trnG*, *rpL32-trnL*, *atpI-atpH*, *psbA-trnH*.

Twenty-one unique six-locus haplotypes were found in this dataset, however, for simplicity, only the broader haplotypic groups are shown. The sample site where identical haplotypes were found in *C. pumila* and plants tentatively identified as *C. dentata* is indicated by an orange arrow. At this sample site, CR, the M15 haplotype was observed.

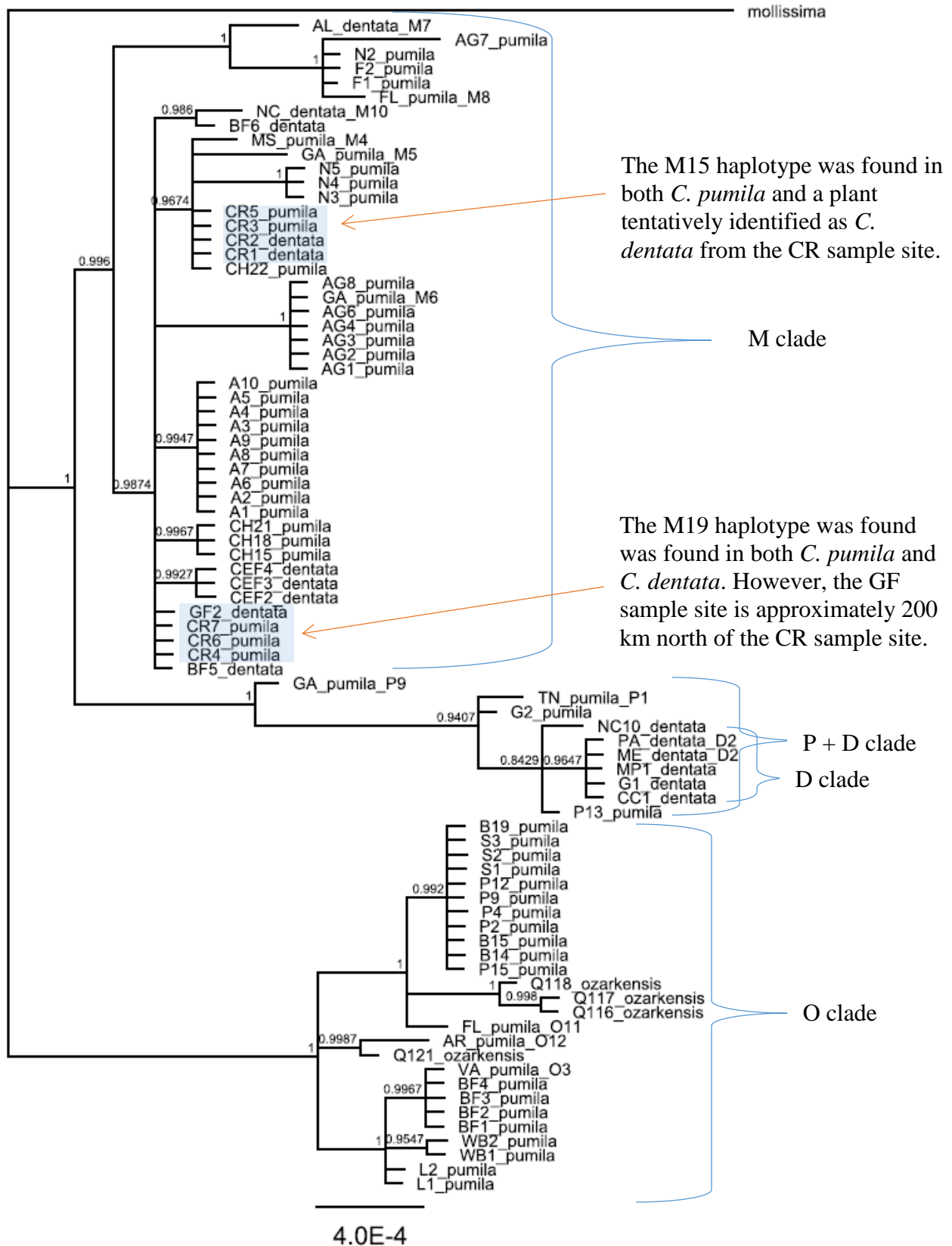


Figure 11 Results of Bayesian MCMC analysis for 83 accessions
cpDNA phylogeny was generated using the noncoding regions *trnV-ndhC*, *rpL16*,
trnS-trnG, *rpL32-trnL*, *atpI-atpH*, *psbA-trnH*
Format for names of my samples is as follows: sample site and sample
number_species (e.g., L1_pumila). The format of Shaw et al.'s (2012) sample names
is as follows: state where collected_species_haplotype (e.g., VA_pumila_O3). Two
haplotypes were observed in multiple species. These plants are indicated by orange
arrows. The branch lengths are drawn proportional to the expected number of
mutations per site (indicated by the scale bar). Posterior probabilities are shown
beside each node.

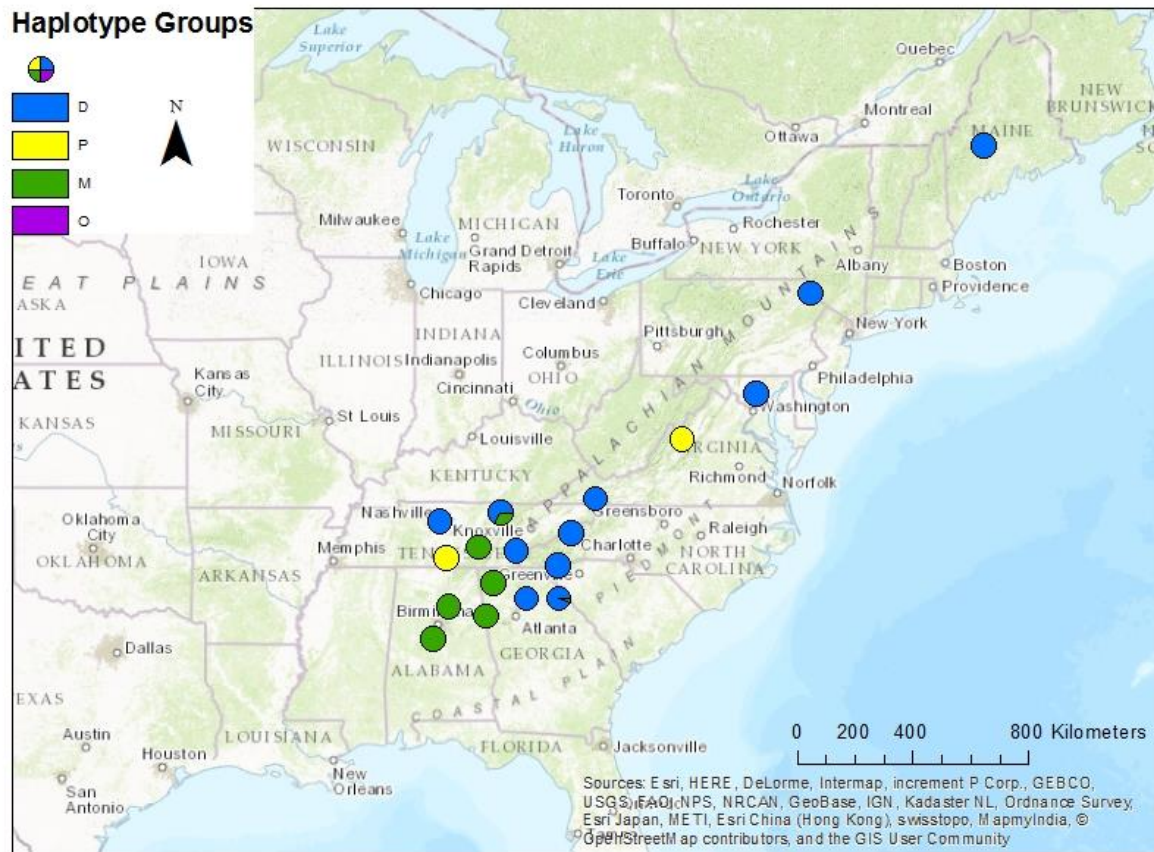


Figure 12 Allele frequencies in *C. dentata* populations sampled by Binkley (2008). Pie charts indicate the populations sampled. Genotypic data were generated from the noncoding cpDNA locus *trnV-ndhC*. Note that in Binkley's (2008) study, the D haplotype was the predominate haplotype in the northern, central, and Southern Appalachian portions of the American Chestnut's range.

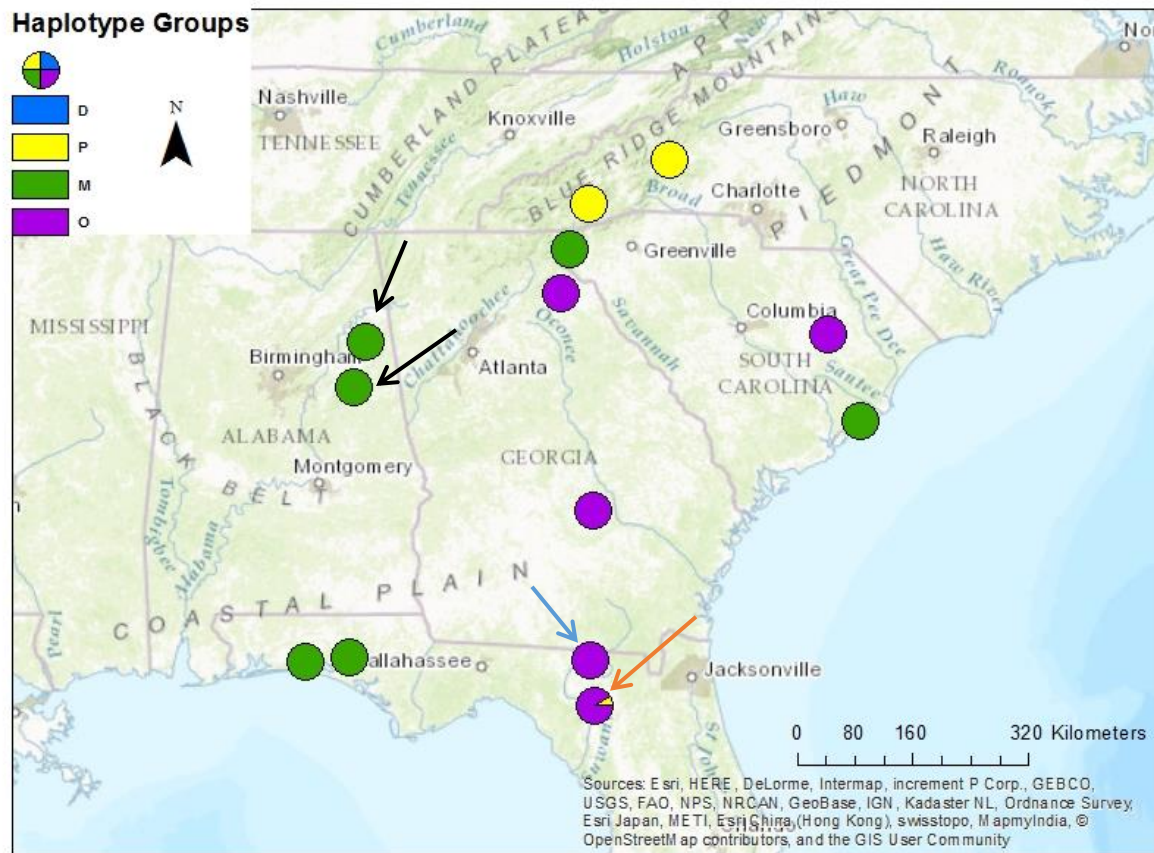


Figure 13 Allele frequencies in *C. pumila* populations sampled in the present study. Pie charts indicate populations sampled. Genotypic data were generated from the cpDNA loci *trnV-ndhC*, *rpL16*, *trnS-trnG*, *rpL32-trnL*, *atpI-atpH*, *psbA-trnH*. Two populations due west of Jacksonville, FL, showed morphology consistent with the chinquapin synonyms *C. alnifolia* (Trailing Chinquapin), which is restricted to Longleaf Pine savannahs, and *C. floridana*, a large tree native to less frequently burned sites in the Coastal Plain. *C. floridana* specimens were collected at the more northerly site, indicated by a blue arrow. *C. alnifolia* specimens were collected at the more southerly site, indicated by an orange arrow. The haplotype O31 was fixed in the *C. floridana* population, while O31 was nearly fixed in the *C. alnifolia* population. Haplotype P33 was found in one individual in the *C. alnifolia* population. The two populations of *C. alabamensis* (or the hypothesized disjunct of *C. ozarkensis*) are indicated by black arrows. These chinquapins from northern Alabama did not possess haplotypes found in *C. ozarkensis* from Arkansas. Also, despite clear morphological differentiation from other chinquapins, the *C. alabamensis* plants possessed haplotypes that were found in morphologically “typical” *C. pumila* in other localities.

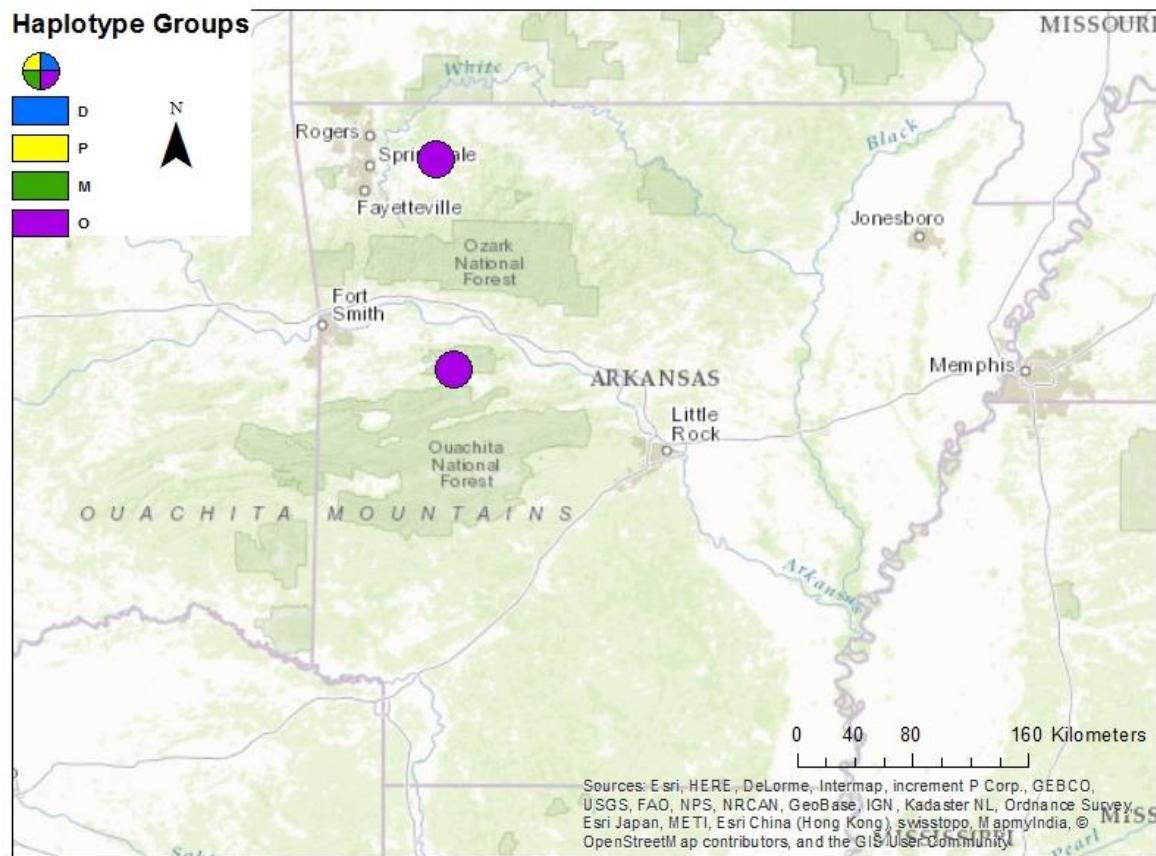


Figure 14 Allele frequencies in *C. ozarkensis* populations sampled in the present study. Pie charts indicate populations sampled. Genotypic data are for the noncoding cpDNA loci *trnV-ndhC*, *rpL16*, *trnS-trnG*, *rpL32-trnL*, *atpI-atpH*, *psbA-trnH*. Three haplotypes in the O clade were observed here. Haplotypes O25 and O26 were present in the more northerly site, while O27 was present in the only plant sampled at the more southerly site.

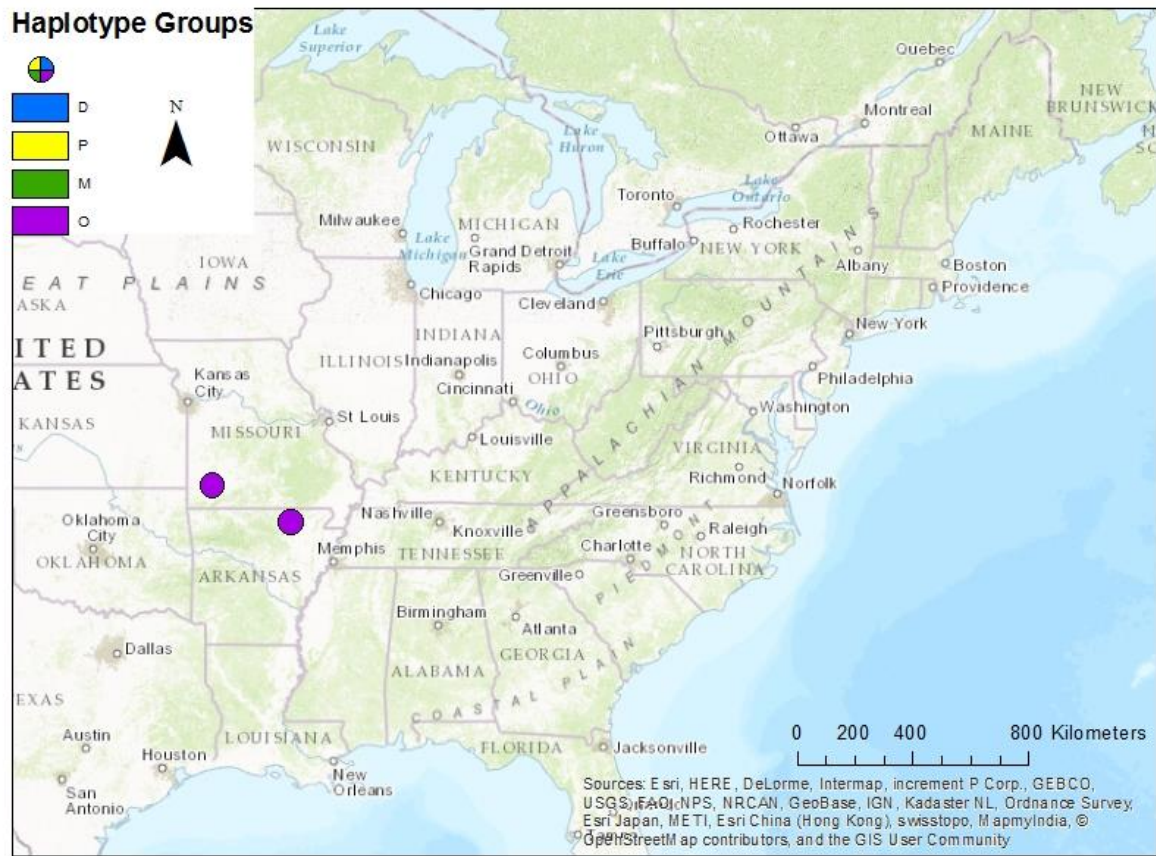


Figure 15 Allele frequencies in *C. ozarkensis* populations sampled by Binkley (2008)
 Pie charts indicate populations sampled. Genotypic data were generated from the cpDNA locus *trnV-ndhC*.

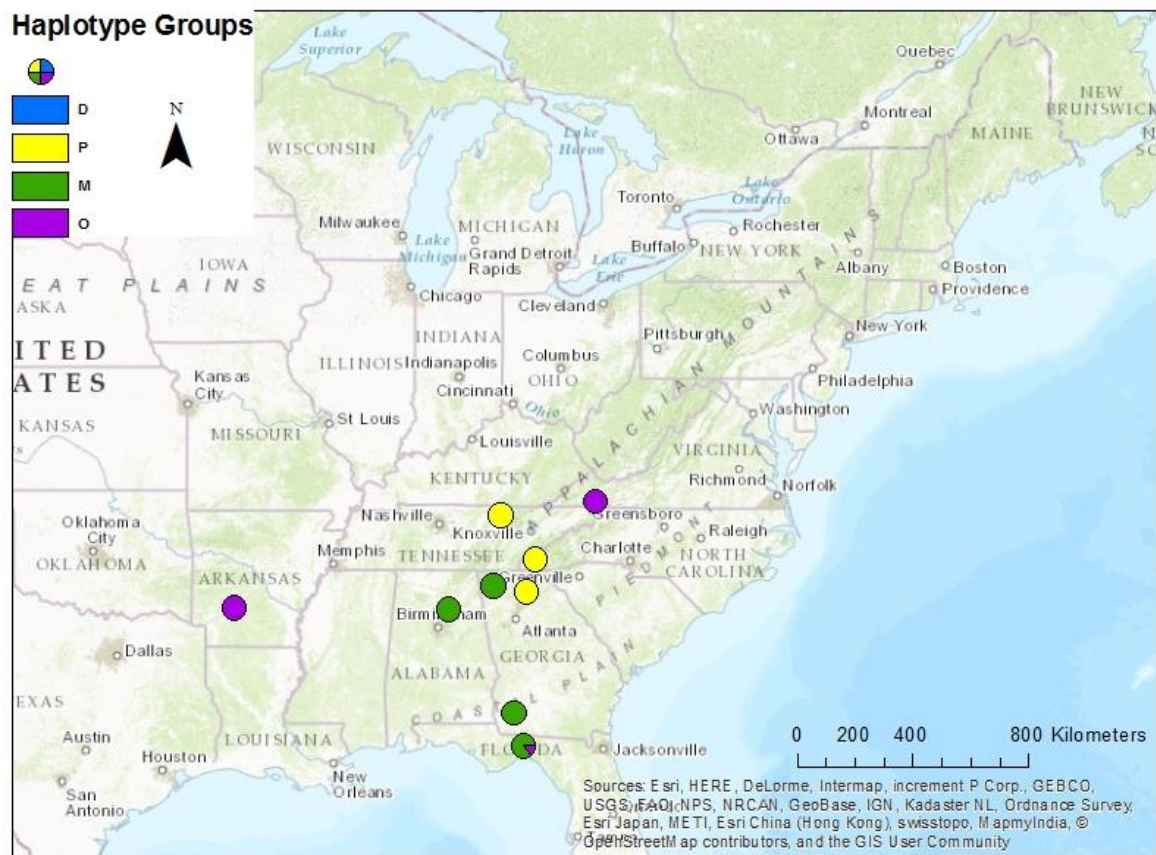


Figure 16 Allele frequencies in *C. pumila* populations sampled by Binkley (2008)
Pie charts indicate populations sampled. Genotypic data were generated from the cpDNA locus *trnV-ndhC*.

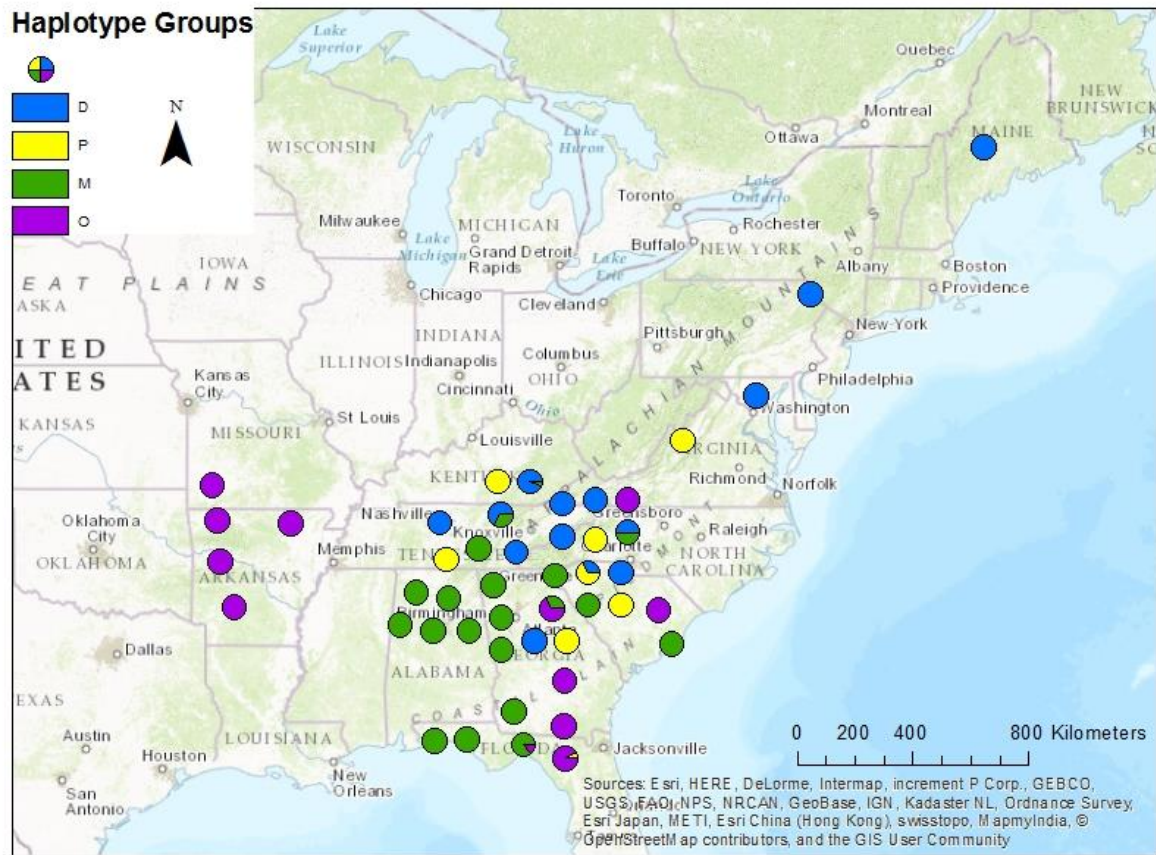


Figure 17 Allele frequencies in *C. dentata*, *C. pumila*, and *C. ozarkensis* populations sampled by Binkley (2008) and for the present study
Pie charts indicate populations sampled. To allow comparisons of my results with Binkley's data, genotypic data are only for the locus *trnV-ndhC*.

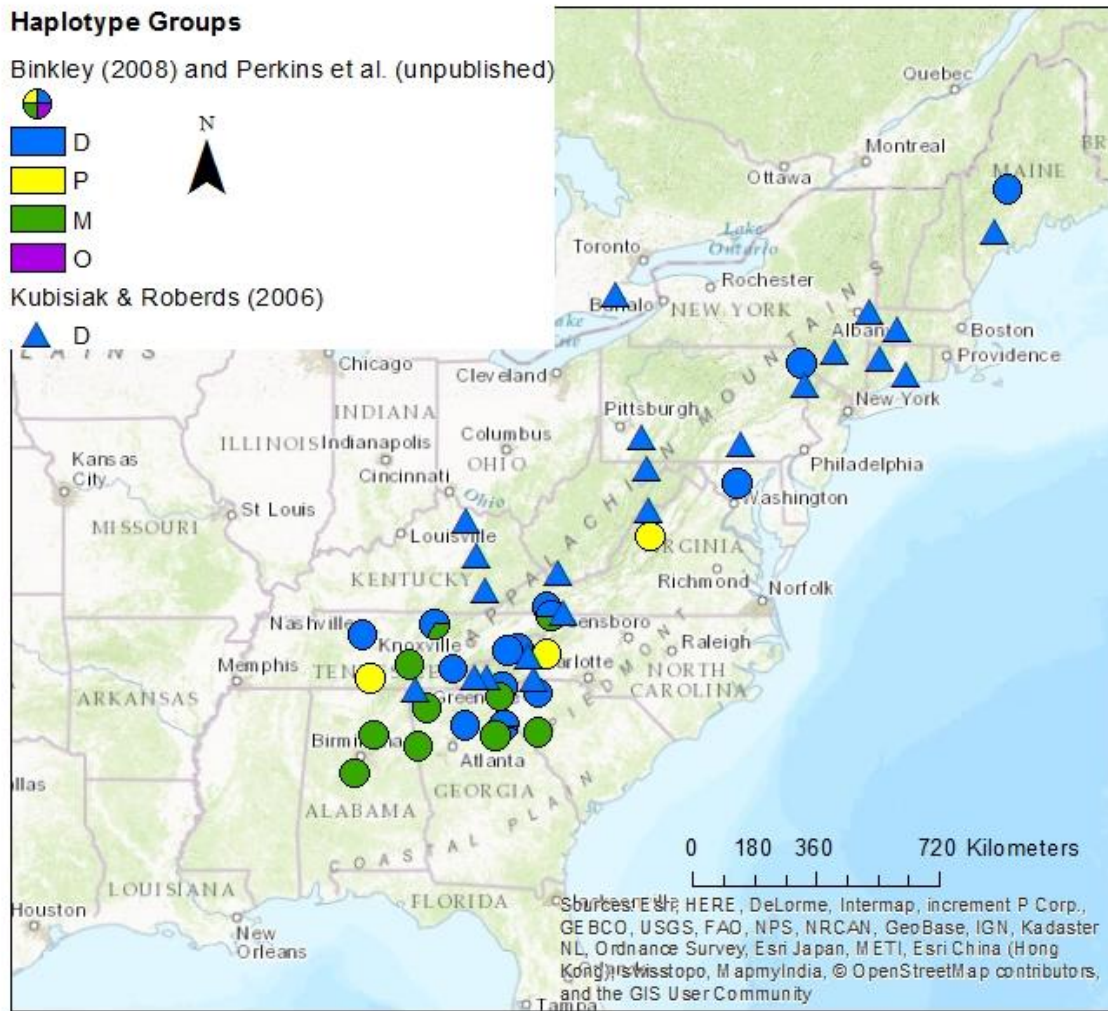


Figure 18 Allele frequencies in *C. dentata* populations sampled by Kubisiak and Roberds (2006), by Binkley (2008), and for the present study
 Pie charts indicate *C. dentata* populations sampled here and by Binkley (2008).
 Triangles indicate *C. dentata* populations sampled by Kubisiak and Roberds (2006).



Figure 19 A specimen from 2015 field work near W.W. Ashe's collection sites of *C. alabamensis*. These taxonomically difficult collections are identical to a morphotype that has been variously identified as *C. alabamensis*, *C. × alabamensis*, and *C. pumila* var. *ozarkensis*.

VITA

Taylor Perkins was born to Teresa Thornton Perkins and Kenneth L. Perkins on October 25, 1988 in Chattanooga, TN. He was followed by a brother, Cameron Perkins, in 1994. Taylor's fascination with nature began at an early age, as he was often outside, using a dip-net to catch fish, crawdads, and snapping turtles from small creeks near home. After graduating from Hixson High School, he entered the University Honors Program at the University of Tennessee at Chattanooga. He graduated in May 2012 with a B.S in Biology, minor in Chemistry (University Honors, Departmental Honors, *cum laude*). His Departmental Honors project, *The effect of phosphite on mycorrhiza formation in American Chestnut*, under Dr. J. Hill Craddock, began his interest in plant breeding and conservation of plant germplasm resources. In 2014, he was accepted into the Environmental Science Graduate Program at UTC. During the fall semester of 2014, he participated in an individual studies course with Dr. Joey Shaw, titled "Molecular Methods for Systematic Questions," and a population genetics course with Dr. Eric M. O'Neill that sparked interests in plant evolution and conservation genetics, respectively. During the spring semester of 2015, he was a visiting scholar at the Clemson University Genomics Institute. While at Clemson, he assisted Dr. Paul H. Sisco and Dr. Tatyana Zhebentyayeva in their study of the quantitative genetics and genomics of ink disease resistance in chestnut. He completed his M.S. in Environmental Science in December 2016 and plans to embark upon a career in the plant sciences.