MOLECULAR PHYLOGENTICS OF *AMORPHA* L. (FABACEAE) AND CONSERVATION GENETICS OF RARE *AMORPHA* SPECIES

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Shannon Christine Kennedy Straub

May 2010

© 2010 Shannon Christine Kennedy Straub

MOLECULAR PHYLOGENTICS OF *AMORPHA* L. (FABACEAE) AND CONSERVATION GENETICS OF RARE *AMORPHA* SPECIES

Shannon Christine Kennedy Straub, Ph.D.
Cornell University 2010

Amorpha L. (Fabaceae) is a distinctive North American genus of 16 species in which the corolla is reduced to a single petal. Continuous morphological variation and phenotypic plasticity have made delimiting species difficult, especially in the highly variable, tetraploid *A. fruticosa* complex. Most species have very limited distributions except *A. fruticosa,* whose broad range overlaps with those of all other species. A molecular phylogenetic study of *Amorpha* was undertaken to assess monophyly, ascertain species relationships, and explore polyploid origins. Results from chloroplast data indicated that *Amorpha* is monophyletic; however, nuclear gene analyses indicated that *Amorpha* is paraphyletic due to the placement of *Errazurizia rotundata* and *Parryella filifolia*. Other key relationships recovered included *A. californica* plus *A. apiculata* as the earliest diverging lineage in *Amorpha* and a close relationship between *A. georgiana* and *A. nana*. *Amorpha fruticosa* and *A. herbacea* were identified as putative progenitors of two tetraploid species: *A. confusa* and *A. crenulata.* The origin of *A. fruticosa* remains unclear, but continued gene flow from sympatric species through hybridization and introgression is a likely cause of its pattern of tremendous morphological variation. Genetic diversity, population structure, and recent population bottlenecks in *A. georgiana*, an endangered species, were explored to inform conservation management. Levels of genetic diversity were high

and the single remaining Georgia population was well differentiated from populations of the Carolinas, which had weaker structure among them. Only a geographically disjunct population, where hybridization with *A. herbacea* was also detected, showed strong evidence of a recent population bottleneck, perhaps due to a recent founder event. The success of restoration genetics efforts for *A. crenulata*, a federally endangered Florida endemic, were evaluated by assessing genetic diversity and population structure in wild, *ex situ*, and restored populations. Genetic diversity was high in wild populations and was partitioned into three geographically correlated genetic clusters. High genetic diversity was captured in *ex situ* and restored populations, with the notable exception of one large restored population. Genetic replication of wild gene pools was successful for overall diversity and preservation of genetic clusters, but accuracy was low because many rare alleles were absent

BIOGRAPHICAL SKETCH

Shannon Christine Kennedy Straub was born to Donna Lynn (Posner) Kennedy and Craig Arrowsmith Kennedy on June 8, 1979 in Denver, Colorado, USA. Two sisters, Kimberly Ann in 1981 and Melissa Lynn in 1986, followed her. Shannon attended St. Vincent de Paul Catholic School in Denver, Colorado, where she participated in her first science fair by building a model of the human eye. In 1990, she moved on to West Middle School in Greenwood Village, Colorado when her family relocated from Denver to Englewood (now Centennial), Colorado. As a ninth grader at Cherry Creek High School in Englewood, Colorado, Shannon's interest in science as a career began to crystallize when she began participating in the Science Research program with Mr. Steve Lantz. That year she was awarded a gold medal from the Colorado-Wyoming Junior Academy of Science for her research entitled: *Are polyacrylamides in the environment harmful to sensitive organisms, such as Daphnia?* She graduated from Cherry Creek High School in 1997.

Following high school, Shannon attended the University of Colorado at Boulder. While at the University of Colorado, she explored many of her interests by majoring in Journalism; Molecular, Cellular, and Developmental Biology; and English with an emphasis on Creative Writing at various times before eventually settling on Environmental, Population, and Organismic Biology. Shannon minored in biochemistry. As an undergraduate she developed an interest in conservation and conducted an independent research project exploring the competitive abilities of *Lasthenia conjugens* (Asteraceae), an endangered vernal pool annual, in the lab of Dr. Sharon Collinge. Shannon found her niche in science after taking Dr. Tom Ranker's

Plant Systematics course. The diversity of plants fascinated her, as did the use of DNA data to explore their evolution. In 2000, she got a summer student assistantship in Dr. Ranker's lab, which extended into an honors thesis project investigating the molecular systematics of the fern family Grammitidaceae over the following two semesters. Shannon graduated *summa cum laude* from the University of Colorado in 2001.

Following her graduation from college, Shannon spent a month in Costa Rica volunteering on a green turtle (*Chelonia mydas*) conservation project and traveling with her partner, Frank. The incredible plant diversity she was able to experience on this trip, especially in the Reserva Biológica Bosque Nuboso Monteverde, further cemented her interest in pursuing plant systematics as a career. Upon her return, Shannon worked for the Colorado Public Research Interest Group raising money for environmental causes before taking a job as a professional research assistant in Dr. Sherry Leonard's lab in the Schizophrenia Research Center at the University of Colorado Health Sciences Center in Denver. In 2003, Shannon enrolled at Cornell University in order to pursue her Ph.D. in Plant Systematics under the direction of Dr. Jeff Doyle.

Shannon is married to her amazing husband, Francis Karol Straub, whom she met at the University of Colorado when they lived on the same floor of Hallett Hall during their freshman year. They were married in 2002. Frank is a middle school science teacher, naturalist, and artist. Shannon and Frank have a wonderful son, Oliver Kennedy Straub, born in 2009, who is a constant source of joy for them both.

For my beloved husband, Francis Karol.
You never cease to amaze, inspire, and challenge me.
Your unwavering love, support, and encouragement have made all the difference in my life.

ACKNOWLEDGMENTS

First, I would like to thank my advisor, Dr. Jeff Doyle, for his guidance and mentorship. Jeff's door was always open whenever I needed advice or assistance. His comments on my written work were always thorough and near instantaneous. I appreciated his encouragement, constructive criticism, and patience over the years. As I move on in my career, I hope I can carry Jeff's enthusiasm for science forward with me. I also thank the other members of my special committee Dr. Melissa Luckow, Dr. Richard Harrison, and Dr. William Crepet for their guidance over the years.

Next I would like to acknowledge all of the people at Cornell, especially in the L.H. Bailey Hortorium, Department of Plant Biology, and the Doyle Lab, who helped me in so many different ways throughout my years as a graduate student. I learned so much from my fellow students and I will value them as friends and colleagues in the years to come. I especially want to recognize Cindy Skema, Marcela Martínez Millán, Jim Cohen, Janelle Burke, Mariana Yazbek, Jeremy Coate, Liz Hermsen, and Fronny Plume. Other faculty members who were not on my committee, especially Jim Reveal and Kevin Nixon, were instrumental in my development as a plant systematist. Bernard Pfeil, Iván Maureira-Butler, Sue Sherman-Broyles, and Ashley Egan, the postdocs in the Doyle Lab during my tenure there, were wonderful mentors and friends over the years as they helped me to broaden my expertise and think about problems in new ways. I could fill another volume writing about the contributions of each person at Cornell that helped me in some way, so I will have to thank the rest of them by listing their names: Charlotte Acharya, Sheri Addy, Mac Alford, Véronique Amiard, Steve Bogdanowicz, Kathryn Bushley, Kuei-Chiu Chen, Tom Clausen, Ed Cobb, Denise Costich, Lin Davidson, Jerry

Davis, Bob Dirig, Jane Doyle, Dick Ecklund, Peter Frassinet, Alejandra Gandolfo, Natalie Gengel, Rod Getchell, Tee Havananda, Laurel Hester, Dan Ilut, John Jantz, Karin Jantz, Caroline Kellogg, Damon Little, Tara Nihill, Karl Niklas, Jeremy Royal, Jackie Salazar, Tom Silva, Jennifer Svitko, Anna Stalter, Sherry Vance, Pat Welch, and Wendy Worth.

Many people outside of Cornell gave me valuable assistance by providing plant material, assistance in the field, access to critical references or information, help obtaining permits, valuable discussions, and insight into the role of *Amorpha* in their local areas. Thank you to: Nell Allen, Daron Barnes, Dan Bell, Greg Chandler, Dave Crumb, Robert Dana, Dwayne Estes, Holly Forbes, Misty Franklin, Laura Gadd, Joie Goodman, Janet Gray, Craig Kennedy, Richard LeBlond, Carl Lewis, John Lohde, Marie Long, Ginger Morgan, Katrina Morris, Cindy Osborne, Tom Patrick, Daniela Roth, Bruce Sorrie, Frank Straub, Tim Thibault, Gin Wall, Michael Wall, Alan Weakley, Kristie Wendelberger, and Theo Whitsell.

Thank you to my undergraduate mentors at the University of Colorado, who encouraged me to go on to graduate school: Tom Ranker, Sharon Collinge, Jennifer Ott Geiger, and Jennifer Ramp Neale.

I am grateful to the instructors and students of ConGen 2008 in Porto, Portugal. I learned so much from them all.

I thank the curators of ARIZ, MIN, MO, NCU, OKL, and TEX-LL for loans of *Amorpha* specimens or access to specimens.

I thank the Arkansas Natural Heritage Commission, Florence County South Carolina Parks and Recreation, Florida Department of Environmental Conservation, Fort Bragg Military Reservation, Georgia Division of Natural Resources, The Nature Conservancy, North Carolina Department of Agriculture and Consumer Services Plant Protection Program, and North

Carolina Wildlife Resources Commission for permission to collect plant material.

I acknowledge my sources of funding: the Cornell Graduate School, the Department of Plant Biology and the Department of Neurobiology and Behavior at Cornell University, a U.S. National Science Foundation (NSF) Doctoral Dissertation Improvement Grant (DEB-0709960), a Catherine H. Beatty Fellowship from the Garden Club of America, the Harold E. Moore, Jr. Endowment Fund, the American Society of Plant Taxonomists, the Botanical Society of America, and research assistantships funded through grants to Jeff Doyle (NSF DEB-0516673; NSF DBI-0822258; United States – Israel Binational Agricultural Research and Development Fund grant IS-4017-07). Part of this work was carried out by using the resources of the Computational Biology Service Unit from Cornell University, which is partially funded by Microsoft Corporation.

On a personal note, I would like to thank my family for their constant love and support. First, I thank my husband Frank. He agreed to uproot our life in Colorado and move to Ithaca so that I could attend Cornell. He was always ready for an adventure and was a great help to me collecting plants in the field. He was always there to give me a pep talk at the end of a hard day or to make sure that I had hot meal when I was swamped with work. He has always been my biggest supporter and cheerleader and I could not have done it without him. I am grateful to my parents, Craig and Donna Kennedy, for teaching me the value of education, to always take pride in my work and to endeavor to perform to my potential, and for allowing me to pursue my many interests over the years. I would also like to thank my sisters, Kimberly Cox and Melissa Kennedy, grandparents, Clyde and Marie Kennedy, grandparents, Harry and Marge Posner, aunt, Karen Kennedy, and great-uncle, Glen

Kennedy for believing in me and always supporting me, even when they were not sure what I was actually trying to do by putting in all of those long hours in the lab. My sister Kimberly's faithful weekly phone calls were a cherished link to home.

I also thank the rest of my family, parents-in-law, Wes and Fran Straub; brothers and sisters-in-law, Ted and Crissy Straub, Wesley and Kris McGinn-Straub, Christine and Geoff Delisio; nieces and nephews, Maggy, Jimmy, Teddy, Saydee, Solia, Finnegan, and Stella for their support and encouragement over the years. I thank my brother-in-law, Wesley Straub, for his mentorship as a scientist one step ahead of me in the process, as he finished his PhD just as I was starting mine. His advice proved valuable in choosing a graduate program and in helping me focus while finishing my dissertation. I am especially grateful for the help of my mother-in-law, Fran, who kept our household running, including providing day care for our son, while I was coming down the home stretch with my dissertation.

Last, but certainly not least, I would like to thank my son, Oliver, for helping me to keep everything in perspective and for never failing to light up my day with his smile.

TABLE OF CONTENTS

Biographical sketch iii

Dedication v

Acknowledgments vi

Table of contents x

List of Figures xiv

List of Tables xv

Chapter One - Introduction to *Amorpha* L. (Fabaceae: Amorpheae) 1

Chapter Two - Molecular phylogenetics of *Amorpha* (Fabaceae: Amorpheae): An evaluation of monophyly, species relationships, and polyploid origins 17

Abstract 17

Introduction 18

Materials and Methods 22

DNA Extraction, Amplification, and Sequencing 22

Alignment and Phylogenetic Analyses 32

Network Analyses 35

Results 36

Chloroplast Data 36

Low Copy Nuclear Gene Data 40

Discussion 60

Evaluation of the Monophyly of Amorpha and the Placements of Errazurizia rotundata and Parryella filifolia 60
Species Relationships in Amorpha 61
Polyploid Origins and Evolution in Amorpha 67
Conclusion 71
References 73

Chapter Three - Characterization of 12 polymorphic microsatellite markers for Georgia false indigo (*Amorpha georgiana* Wilbur var. *georgiana*), an endangered species, and their utility in other dwarf *Amorpha* L. species 79
Primer Note 79
References 88

Chapter Four - Conservation genetics of *Amorpha georgiana* (Fabaceae), an endangered legume of the southeastern United States 90
Abstract 90
Introduction 91
Materials and Methods 94
Sampling and Genotyping 94
Data Quality Control 95
Genetic Diversity 98
Population Structure 101
Recent Demographic Events 103
Assessment of hybridization with A. herbacea 104

Results 105
Genetic Diversity 105
Population Structure 106
Recent Demographic Events 112
Assessment of Hybridization with A. herbacea 117
Discussion 121
Genetic Diversity and Population Structure 121
Recent Demographic Events 125
Assessment of Hybridization with A. herbacea 127
Implications for Conservation Management of A. georgiana 128
References 133

Chapter Five - Applications of genetics to endangered species restoration: A case study evaluating natural genetic diversity and the genetic success of *ex situ* conservation and reintroduction efforts for *Amorpha crenulata* (Fabaceae) 143
Abstract 143
Introduction 144
Materials and Methods 151
Sample Collection and Preparation 151
Genotyping 151
Data Quality Control 153
Analyses of Genetic Diversity and Population Structure 155
Results 158
Genotyping Success and Ploidy of Amorpha crenulata 158
Overall Genetic Diversity in Amorpha crenulata 159

Genetic Diversity and Population Structure in Wild Amorpha crenulata Populations 159
Genetic Diversity in and Distinctiveness of Extirpated Amorpha crenulata Populations 162
Genetic Diversity of Restored Amorpha crenulata Populations 166
Discussion 171
Genetic Diversity and Population Structure in Wild Populations of Amorpha crenulata 171
The Genetic Success of Conservation Efforts for Amorpha crenulata . 174
Implications for Future Conservation Management of Amorpha crenulata 180
References 183

Appendix A: A survey of genome size in *Amorpha fruticosa* using flow cytometry 191
References 193

LIST OF FIGURES

Figure 2.1 Phylogenetic relationships of chloroplast haplotypes 37

Figure 2.2 Phylogram of relationships of *Amorpha* chloroplast 42

Figure 2.3 Phylogenetic relationships of *CNGC5* alleles 45

Figure 2.4 Phylogram of relationships of *CNGC5* alleles 48

Figure 2.5 Network relationships of *CNGC5* alleles 50

Figure 2.6 Phylogenetic relationships of *MinD* alleles 53

Figure 2.7 Phylogram of relationships of *MinD* alleles 55

Figure 2.8 Network of relationships of *MinD* alleles 58

Figure 4.1 Map of the southeastern United States showing collecting . . . 96

Figure 4.2 Population structure among *Amorpha georgiana* populations . . 107

Figure 4.3 Factorial correspondence analysis of multilocus 119

Figure 5.1 Map of the current and historical distribution 147

Figure 5.2 Population structure among the remaining 163

Figure 5.3 Population structure among the remaining 168

LIST OF TABLES

Table 2.1 List of voucher specimens 23

Table 2.2 Allele-specific sequencing primers 31

Table 2.3 Sequence length and variability for DNA regions 39

Table 2.4 Results of separate parsimony analyses for each incomplete . . .41

Table 3.1 Characterization of twelve novel microsatellite loci 83

Table 3.2 Cross-species testing of *A. georgiana* 86

Table 4.1 Genetic variation in *A. georgiana* populations 99

Table 4.2 *Amorpha georgiana* selfing rates 110

Table 4.3 Pairwise F_{ST} values for *A. georgiana* collecting localities 111

Table 4.4 Pairwise F_{ST} values for *A. georgiana* populations 112

Table 4.5 Results of AMOVA analysis for *A. georgiana* populations 112

Table 4.6 Table of *P*-values for one-tailed Wilcoxon 114

Table 4.7 Genetic variation found in *Amorpha herbacea* 118

Table 5.1 Summary of the extent of sampling for wild 152

Table 5.2 Success of the use of *Amorpha georgiana* microsatellite 154

Table 5.3 Genetic differentiation among populations 160

Table 5.4 Genetic diversity observed in the four remaining wild 161

Table 5.5 Differentiation among wild populations 162

Table 5.6 Genetic diversity observed in the *ex situ* conservation 167

Table 5.7 Genetic diversity observed in restored 170

Table A1 Estimates of genome size for *Amorpha fruticosa* 192

CHAPTER ONE

INTRODUCTION TO *AMORPHA* L. (FABACEAE: AMORPHEAE)

Amorpha L. (Fabaceae Juss.: Amorpheae Boriss.) is notable among papilionoid legumes for having a non-papilionaceous corolla consisting solely of a petaloid banner, although deviation from the normal papilionoid floral form is common throughout Amorpheae. The name of the genus is derived from Latin meaning without (a-) shape (morphos) and is undoubtedly a reference to its unique flowers. Even though the recognition of this strictly North American genus has never been in question due to its easily distinguished floral characters, delimitation of species within the genus has long caused consternation among taxonomists. The degree and intergradation of morphological variation have resulted in a prodigious list of recognized species, varieties, and forms, most often associated with the *A. fruticosa* L. complex. Early taxonomic treatments failed to result in a satisfactory circumscription of the genus (e.g., Schneider 1907; Rydberg 1919; Palmer 1931), but the insightful and thorough work of Wilbur (1975) was a marked improvement and was closely followed by Isely (1998), and also, with the exception of recognition of *A. confusa* (Wilbur) S.C.K. Straub, Sorrie & Weakley at the rank of species (Straub *et al.* 2009b) in this dissertation.

Taxonomic uncertainty is unfortunate because *Amorpha* has been a group of interest for many years in the search for biologically active compounds, and much of the focus has been on the most plastic and polymorphic species, *A. fruticosa*. In the area of medicinal biochemistry, anti-

tumor and anti-inflammatory compounds have been isolated from *A. fruticosa* (e.g., Konoshima *et al.* 1993; Li *et al.* 1993; Cho *et al.* 2000). Antimicrobial agents have been identified in both *A. fruticosa* and *A. nana* Nutt. (Mitscher *et al.* 1981, 1985). *Amorpha fruticosa* has also been investigated for its insecticidal and insect repellant properties (e.g., Roark 1947). The glands on the fruits of this species have been shown to contain compounds that poison numerous types of insects through ingestion or contact (e.g., chinch bug, cotton aphid, pea aphid, spotted cucumber beetle, mosquito larvae) and also have insect repellant properties (e.g., striped cucumber beetle, flour beetles, dog fleas, and houseflies; fide Brett 1946a, 1946b).

Native Americans of the Great Plains employed several of the more common *Amorpha* species for a variety of uses. *Amorpha fruticosa* was used for bedding, horse feed, and arrow shafts, and stems were arranged on the ground to create a clean surface on which to put butchered meat (Gilmore 1919; Vestal & Schultes 1939; Rogers 1980). *Amorpha canescens* Pursh was used to treat stomach pain, intestinal worms, eczema, neuralgia, and rheumatism, and its powdered leaves were applied to wounds (Hoffman 1891; Gilmore 1913, 1919; Smith 1928). The leaves were also mixed with buffalo fat and smoked or used to make tea (Gilmore 1919). The dried leaves of *A. nana* were used to treat catarrh (Elmore 1944).

A few *Amorpha* species are found regularly in cultivation, the most common being *A. fruticosa* and *A. canescens*, while *A. nana*, *A. californica* Nutt. ex Torr. & A. Gray, and *A. herbacea* Walter are less commonly part of the horticultural trade. *Amorpha fruticosa* has also been used in the United States and abroad for soil stabilization, erosion control, and in windbreaks, and has been investigated as a potential forage and biomass crop (e.g., Huh & Huh

1997; Wang *et al.* 2002; DeHaan *et al.* 2006). Cultivation of *A. fruticosa* has led to its escape and naturalization in many parts of Europe and Asia (e.g., Huh & Huh 1997; Wang *et al.* 2002; Karrenberg *et al.* 2003; Török *et al.* 2003).

Some species of *Amorpha* are commonly studied from an ecological perspective. *Amorpha canescens* is an important component of upland prairie communities of the central United States. In this role it has been studied in ecological contexts ranging from responses to grazing (e.g., Hickman & Hartnett 2002) and habitat fragmentation (e.g., Slagle & Hendrix 2009), to estimation of the effects of invasive plants on native communities (e.g., Mattingly *et al.* 2007), and to physiology (e.g., Nipert & Knapp 2007; Nippert *et al.* 2007) and nutrient cycling (e.g., Norris & Reich 2009). Ecological factors affecting recruitment and propagation techniques for endangered species, such as *A. crenulata* Rydb., have also been examined (e.g., Roncal *et al.* 2006; Wendelberger *et al.* 2008; Wendelberger & Maschinski 2009). *Amorpha* has also been used as a model for exploring whether shade tolerance is an important factor in species rarity in fire-maintained communities now subject to fire suppression (Marchin *et al.* 2009).

From an evolutionary perspective, the "amorphoid" clade of Amorpheae, to which *Amorpha* belongs, has been strongly supported as monophyletic in molecular analyses (McMahon & Hufford 2004), although most relationships among genera in the clade remain unclear. Analyses of chloroplast, nuclear ribosomal DNA (nrDNA), and low-copy nuclear gene sequence data have indicated that *Amorpha* may not be monophyletic because either *Parryella filifolia* Torr. & A. Gray or *Errazurizia rotundata* (Wooton) Barneby, or both, are nested within it (McMahon & Hufford 2004, 2005; McMahon 2005). Some of these analyses have recovered *P.*

filifolia and *E. rotundata* as sister species (combined analyses of chloroplast *trnK, matK, ITS/5.8S* nrDNA, nuclear *CNGC4*), which may or may not be nested in *Amorpha*; other analyses indicated that *E. rotundata* is nested among *Amorpha* species whereas *P. filifolia* may or may not be (*trnK*, *matK* analyzed alone, *CNGC4* analyzed alone). These conflicting outcomes have left unclear the relationship of these two species to each other and to *Amorpha*, as well as whether *Amorpha*, as traditionally circumscribed, is a natural group.

Morphological evidence also supports the close association of *P. filifolia*, *E. rotundata* (originally described as *P. rotundata* Wooton) and *Amorpha*. Both Wilbur (1975) and Barneby (1977) suggested this association: although Barneby was undecided about the true affinities of *E. rotundata*, which he had recently transferred to *Errazurizia* Phil. from *Parryella* Torr. & A. Gray ex A. Gray (Barneby 1962). *Parryella filifolia* lacks a corolla, and the corolla of *E. rotundata* is either absent or consists of only a single petal, which suggests a closer association in terms of floral evolution to *Amorpha* than to the other 5-petaled genera of Amorpheae, an observation supported by developmental studies (McMahon & Hufford 2005). Mahler (1965) hypothesized a close relationship of *Amorpha*, *P. filifolia*, and *E. rotundata* based on shared pollen characteristics. Further testing of the monophyly of *Amorpha* is presented in Chapter 2; however, the monophyly of *Amorpha* remains an open question.

Tribal level analyses in Amorpheae included several exemplars from the genus, but the evolutionary relationships among *Amorpha* species as a group have never been studied. As in any systematic study, the first goal of the research presented here was to achieve a better understanding of a poorly understood genus. Because the underlying pattern of evolution that

led to the current morphologies, ecologies, and distributions of these species is unknown, the working out of evolutionary relationships in *Amorpha* could allow these aspects to be better appreciated and a classification based on them to be produced. Toward that end, the first phylogenetic analyses of the genus, based on chloroplast and nuclear DNA sequences, are presented here (Chapter 2). Determination of relationships among species in the genus was predicted to be difficult from the outset due to taxonomic issues, phenotypic plasticity, and assumed close relationships among species; this made incomplete lineage sorting, hybridization and introgression, as well as lack of DNA sequence variation, very real problems (e.g., Wendel & Doyle 1998). The work presented here based on phylogenetic analysis of chloroplast haplotype and low-copy nuclear gene data has resulted in a better understanding of relationships, but many affinities remain unclear due to the aforementioned problems. Key findings include the *A. californica/A. apiculata* Wiggins lineage as the earliest diverging in the genus, a close relationship of the morphologically similar *A. georgiana* Wilbur and *A. nana*, as well as complicated patterns of variation involving polyploid *A. fruticosa*, and identification of putative progenitors for other polyploid species in the genus, *A. confusa* and *A. crenulata*. The overall patterns observed in the phylogenetic analysis were indicative of a rapid radiation in the genus and/or of continued gene flow and partial genetic homogenization through hybridization and introgression, possibly mediated by widespread species. Further work will be needed to explore the complicated relationships in *Amorpha* and to understand the patterns of morphological and genetic variation.

Polyploidy is increasingly appreciated as an important evolutionary mechanism, and not only in plants (e.g., Mable 2004). In plants polyploidy is now understood to be even more prevalent than previously thought (Soltis *et al.* 2004), and is a major source of evolutionary novelty (Osborn *et al.* 2003). Chromosome counts exist for only five of the species of *Amorpha*, and often it is not clear from the literature which species' chromosomes were actually counted due to the taxonomic difficulty of the genus and lack of voucher specimens. This dearth of information additionally means that the ploidy of most *Amorpha* species is unknown. Ongoing genetic work has suggested that polyploidy is more common in *Amorpha* than previously recognized. *Amorpha fruticosa* has long been known to be a tetraploid ($2n = 4x = 40$; e.g., Kreuter 1930), and new information from nuclear gene DNA sequences and microsatellites has indicated that *A. confusa*, *A. crenulata*, and *A. roemeriana* Scheele may also be tetraploids, although this is yet to be confirmed by chromosome counts (Straub *et al.* 2009a; Chapters 2, 3, and 5).

Polyploidy contributes to the major source of disagreement among taxonomists, the circumscription of *A. fruticosa* and allies. The list of synonymy for *A. fruticosa* in Wilbur's (1975) revision includes 39 names, due to numerous regional segregates and copious amounts of morphological variation across its range, which encompasses that of all other *Amorpha* species and consists of much of North America. This species complex shows "spectacular" environmental plasticity (Wilbur 1975, p. 338). Even with all of the readily apparent morphological variation, little variation in genome size has been observed in *A. fruticosa* (Appendix A). Whether the one or more polyploidy events that led to the formation of this complex were fundamentally allopolyploid or autopolyploid is unknown.

Understanding the taxonomy and evolution of the complex may lead to a better understanding of its components and their ability to successfully invade and colonize new habitats. Identification of the most noxious genotypes of *A. fruticosa* may aid in focusing of control efforts in areas where it is an invasive weed, as has been true for *Phragmites australis* (Cav.) Steud. (Saltonstall 2002; Saltonstall 2003). As a polyploid member of a group of generally narrowly distributed diploids, *A. fruticosa* is like many weedy and invasive plants (e.g., Amsellem *et al*. 2001; Pandit *et al*. 2006) and could join such taxa as *Spartina anglica* C.E. Hubb. (Ainouche *et al*. 2004) as a model for studying the relationship between ploidy and invasiveness.

A first step to understanding *A. fruticosa* is presented in Chapter 2 and Appendix A. Phylogenetic analyses indicated that its relationship with diploid, as well as other polyploid, species in the genus is complicated. It shares chloroplast haplotypes with geographically proximal species throughout its range and there is putative recombination of nuclear alleles derived from diploid species in the polyploid indicating introgression, chloroplast or nuclear, from multiple different species and contributing to overall variation. There is also an indication that the other *A. fruticosa*-like species of *Amorpha* are regional variants and may not comprise distinct evolutionary lineages.

A current and comprehensive exploration of evolutionary relationships and genetic diversity in *Amorpha* is needed to aid biodiversity conservation efforts because more than half of all species of *Amorpha* have very limited distributions and many of these are of conservation concern. Consideration of genetics is essential to the management of rare species to avoid inbreeding depression and to preserve evolutionary potential (i.e.

genetic diversity) for response to future environmental challenges (e.g., Frankham *et al*. 2002; Spielman *et al*. 2004). As researchers of biological diversity, systematists can and should make an increased contribution to the field of conservation biology by providing updated taxonomy based on molecular data and identification of distinct lineages for consideration as conservation units (Soltis & Gitzendanner 1999), as well as by exploring molecular diversity and population structure within these units.

Among rare *Amorpha* species, *A. crenulata*, *A. georgiana*, *A. confusa*, *A. herbacea* Walter var. *floridana* (Rydb.) Wilbur, *A. paniculata* Torr. & A. Gray, *A. schwerinii* C.K. Schneid., *A. glabra* (Pers.) Poir., and *A. ouachitensis* Wilbur are recognized with some designation of conservation concern (e.g., endangered, threatened, species of concern) at the federal or state levels. More common species that also have conservation designations at some level, usually at the edges of their ranges, include *A. nitens* F.E. Boynton, *A. canescens*, and *A. nana*. Rare or locally endemic species or varieties with small ranges that probably should have some sort of conservation designation, but thus far do not, are: *A. californica* Nutt. var. *napensis* Jeps., a localized endemic in Marin, Napa, and Sonoma counties in northern CA; *A. roemeriana*, an endemic of the Edwards Plateau in TX, with a disjunct distribution to Coahuila, Mexico; and *A. apiculata*, an endemic of the Sierra San Pedro Martir in Baja CA, Mexico. Another species, *A. laevigata* Nutt., is a special case because it is an undercollected, ill-defined species with characters overlapping with those of several other species. Further study may show that it does not warrant recognition or, alternatively, that it is an extremely rare endemic in need of conservation concern.

In-depth study of genetic diversity and population structure in a conservation context was undertaken for two endangered species of *Amorpha*. *Amorpha georgiana* is an endangered species of the southeastern United States with only about 900 individuals remaining in scattered populations, due to habitat fragmentation in longleaf pine savannas (Sorrie 1995; Miller 2004). Microsatellite markers were developed (Chapter 3) in order to survey genetic variation and population structure for this species (Chapter 4). Even though this species is rare and has experienced severe habitat degradation and fragmentation, genetic diversity remains high and population structure corresponding to geography was detected. Each of the several geographical units can be used in future management and *ex situ* conservation planning. Another important observation was the detection of hybridization with the more widespread and abundant *A. herbacea* at one locality; this could lead to genetic swamping of *A. georgiana,* which could be of conservation concern (e.g., Rhymer & Simberloff 1996).

The conservation genetics of *A. crenulata* in the context of genetic evaluation of restoration efforts for this federally endangered species was explored (Chapter 5) using the markers developed for *A. georgiana.* For the most part, current *ex situ* and restoration efforts for this species have done a good job of capturing genetic diversity, which remains high in the remnant wild populations, probably in part due to the tetraploid nature of the species. Distinct genetic clusters were identified that will be critical in representing in future *ex situ* and restoration conservation efforts in order to preserve the evolutionary potential for this species.

The research presented here provides a solid foundation for further work with *Amorpha* in the arenas of phylogenetics, polyploidy research,

population biology, and conservation. Clearly, additional data are needed to resolve species relationships within *Amorpha* and understand its evolution, to update species concepts within the genus, and to determine whether or not *Amorpha* is monophyletic as circumscribed or if *P. filifolia* and *E. rotundata* should be added to *Amorpha*. Obtaining chromosome counts to assess ploidy for the majority of the species in the genus will also be an essential next step in understanding the evolution of these plants, especially in light of the previously unsuspected tetraploids recently detected amongst *Amorpha* species. One of the other areas most in need of additional work is continued exploration of the incredible variation present in the *A. fruticosa* complex. In-depth population-level studies spanning the range of the species may be the best way to begin to parse morphological, chemical, and ecologically important variation, based on a thorough sampling of genetic variation and an understanding of underlying population structure. Understanding the contribution of environmental factors to phenotypic plasticity in the complex will also be important. Although the work reported here has made a considerable contribution to adding a genetic component to conservation and restoration of two species, *A. georgiana* and *A. crenulata*, with so many other rare and threatened species in the genus, more work is urgently needed to prevent loss of diversity and to preserve evolutionary potential in *Amorpha*.

REFERENCES

Ainouche ML, Baumel A, Salmon A (2004) *Spartina anglica* C. E. Hubbard: A natural model system for analysing early evolutionary changes that affect allopolyploid genomes. *Biological Journal of the Linnean Society*, **82**, 475-484.

Amsellem L, Chevallier M, Hossaert-McKey M (2001) Ploidy level of the invasive weed *Rubus alceifolius* (Rosaceae) in its native range and in areas of introduction. *Plant Systematics and Evolution*, **228**, 171-179.

Barneby RC (1962) A synopsis of *Errazurizia* [Leguminosae; Psoraleae]. *Leaflets of Western Botany*, **9**, 209-214.

Barneby RC (1977) *Dalea Imagines*: An illustrated revision of *Errazurizia, Psorothamnus, Marina* and *Dalea* including all Species of Leguminosae tribe Amorpheae ever referred to *Dalea*. *Memoirs of the New York Botanical Garden*, **27**, 1-891.

Brett CH (1946a) Insecticidal properties of the indigobush (*Amorpha fruticosa*). *Journal of Agricultural Research*, **73**, 81-96.

Brett CH (1946b) Repellent properties of extract of *Amorpha fruticosa*. *Journal of Econmic Entomology*, **39**, 810.

Cho JY, Kim PS, Park J, *et al.* (2000) Inhibitor of tumor necrosis factor-alpha production in lipopolysaccharide-stimulated RAW264.7 cells from *Amorpha fruticosa*. *Journal of Ethnopharmacology*, **70**, 127-133.

DeHaan LR, Ehlke NJ, Sheaffer CC, Wyse DL, DeHaan RL (2006) Evaluation of diversity among North American accessions of false indigo (*Amorpha fruticosa* L.) for forage and biomass. *Genetic Resources and Crop Evolution*, **53**, 1463-1476.

Elmore FH (1944) *Ethnobotany of the Navajo*. University of New Mexico Press, Albuquerque, New Mexico.

Frankham R, Ballou DJ, Briscoe DA (2002) *Introduction to Conservation Genetics*. Cambridge University Press, Cambridge.

Gilmore MR (1913) A study in the ethnobotany of the Omaha Indians. *Nebraska State Historical Society Collections*, **17**, 314-357.

Gilmore MR (1919) Uses of plants by the Indians of the Missouri River Region. *Annual Report of the Bureau of American Ethnology*, **33**, 1-126.

Hickman KR, Hartnett DC (2002) Effects of grazing intensity on growth, reproduction, and abundance of three palatable forbs in Kansas tallgrass prairie. *Plant Ecology*, **159**, 23-33.

Hoffman WJ (1891) The Midewiwin or "Grand Medicine Society" of the Ojibwa. *Annual Report of the Bureau of American Ethnology*, **7**, 143-300.

Huh MK, Huh HW (1997) Allozyme variation and genetic structure of *Amorpha fructicosa* L. population in Korea. *Korean Journal of Genetics*, **19**, 39-47.

Isely D (1998) *Amorpha*. In: *Native and Naturalized Leguminosae (Fabaceae) of the United States (Exclusive of Alaska and Hawaii)*, pp. 132-144, Brigham Young University, Provo, Utah.

Karrenberg S, Kollmann J, Edwards PJ, Gurnell AM, Petts GE (2003) Patterns in woody vegetation along the active zone of a near-natural Alpine river. *Basic and Applied Ecology*, **4**, 157-166.

Konoshima T, Terada H, Kokumai M, *et al.* (1993) Studies on inhibitors of skin tumor promotion: XII. Rotenoids from *Amorpha fruticosa*. *Journal of Natural Products*, **56**, 843-848.

Kreuter E (1930) Beitrag zur karyologisch-systematischen studien an Galegeen. *Planta*, **11**, 1-44.

Li L, Wang H, Chang J, *et al.* (1993) Antitumor agents, 138: Rotenoids and isoflavones as cytotoxic constituents from *Amorpha fruticosa*. *Journal of Natural Products*, **56**, 690-698.

Mable BK (2004) 'Why polyploidy is rarer in animals than in plants': myths and mechanisms. *Biological Journal of the Linnean Society*, **82**, 453-466.

Mahler WF (1965) *The pollen morphology of the tribe Psoraleae (Leguminosae).* National Science Foundation Research Participation Report, 1-24.[1]

Marchin RM, Bhandari RK, Wall WA, Hohmann MG, Gray JB, Hoffmann WA (2009) Are rare species less shade tolerant than common species in fire-prone environments? A test with seven *Amorpha* (Fabaceae) species. *Plant Ecology*, **205**, 249-260.

Mattingly WB, Hewlate R, Reynolds HL (2007) Species evenness and invasion resistance of experimental grassland communities. *Oikos*, **116**, 1164-1170.

McMahon MM (2005) Phylogenetic relationships and floral evolution in the papilionoid legume clade Amorpheae. *Brittonia*, **57**, 397-411.

McMahon M, Hufford L (2004) Phylogeny of Amorpheae (Fabaceae: Papilionoideae). *American Journal of Botany*, **91**, 1219-1230.

McMahon M, Hufford L (2005) Evolution and development in the amorphoid clade (Amorpheae : Papilionoideae : Leguminosae): Petal loss and dedifferentiation. *International Journal of Plant Sciences*, **166**, 383-396.

Miller VM (2004) *Habitat characterization of Amorpha georgiana var. georgiana groups at Fort Bragg, North Carolina*. Masters Thesis, North Carolina State University, Raleigh, North Carolina.

Mitscher LA, Park YH (1981) Amorfrutin A and B, bibenzyl antimicrobial agents from *Amorpha fruticosa*. *Phytochemistry*, **20**, 781-786.

Mitscher LA, Gollapudi SR, Drake S, Oburn DS (1985) Amorphastilbol an antimicrobial agent from *Amorpha nana*. *Phytochemistry*, **24**, 1481-1484.

Nippert JB, Fay PA, Knapp AK (2007) Photosynthetic traits in C-3 and C-4 grassland species in mesocosm and field environments. *Environmental and Experimental Botany*, **60**, 412-420.

[1] A mimeograph copy of this unpublished report was obtained from the Mertz Library of the New York Botanical Garden.

Nippert JB, Knapp AK (2007) Linking water uptake with rooting patterns in grassland species. *Oecologia*, **153**, 261-272.

Norris MD, Reich PB (2009) Modest enhancement of nitrogen conservation via retranslocation in response to gradients in N supply and leaf N status. *Plant and Soil*, **316**, 193-204.

Osborn TC, Pires JC, Birchler JA, *et al.* (2003) Understanding mechanisms of novel gene expression in polyploids. *Trends in Genetics*, **19**, 141-147.

Palmer EJ (1931) Conspectus of the genus *Amorpha*. *Journal of the Arnold Arboretum*, **12**, 157-220.

Pandit MK, Tan HTW, Bisht MS (2006) Polyploidy in invasive plant species of Singapore. *Botanical Journal of the Linnean Society*, **151**, 395-403.

Rhymer JM, Simberloff D (1996) Extinction by hybridization and introgression. *Annual Review of Ecology and Systematics*, **27**, 83-109.

Roark RC (1947) Some promising insecticidal plants. *Economic Botany*, **1**, 437-445.

Rogers DJ (1980) *Lakota Names and Traditional Uses of Native Plants by Sicangu (Brule) People in the Rosebud Area, South Dakota: A study based on Father Eugene Buechel's collection of plants of Rosebud around 1920*. Rosebud Educational Society, St. Francis, South Dakota.

Roncal J, Fisher JB, Fellows MQN, Wendelberger KS, Maschinski J, Fidelibus MW (2006) Propagation protocol for the endangered Crenulate Lead Plant *Amorpha herbacea* var. *crenulata*. *Native Plants Journal*, **7**, 89-93.

Rydberg PA (1919) *Amorpha*. *North American Flora*, **24**, 26-34.

Saltonstall K (2002) Cryptic invasion by a non-native genotype of the common reed, *Phragmites australis*, into North America. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 2445-2449.

Saltonstall K (2003) Microsatellite variation within and among North American lineages of *Phragmites australis*. *Molecular Ecology*, **12**, 1689-1702.

Schneider CK (1907) Conspectus generis Amorphae. *Botanical Gazette*, **43**, 297-307.

Slagle MW, Hendrix SD (2009) Reproduction of *Amorpha canescens* (Fabaceae) and diversity of its bee community in a fragmented landscape. *Oecologia*, **161**, 813-823.

Smith HH (1928) Ethnobotany of the Meskwaki Indians. *Bulletin of the Public Museum of the City of Milwaukee*, **4**, 175-326.

Soltis DE, Soltis PS, Tate JA (2004) Advances in the study of polyploidy since Plant Speciation. *New Phytologist*, **161**, 173-191.

Soltis PS, Gitzendanner MA (1999) Molecular systematics and the conservation of rare species. *Conservation Biology*, **13**, 471-483.

Sorrie BA (1995) *Status Survey of Amorpha georgiana var. georgiana*. US Fish and Wildlife Service, Office of Endangered Species, Asheville, North Carolina, and North Carolina Natural Heritage Program, Division of Parks and Recreation, Raleigh, North Carolina.

Spielman D, Brook BW, Briscoe DA, Frankham R (2004) Does inbreeding and loss of genetic diversity decrease disease resistance? *Conservation Genetics*, **5**, 439-448.

Straub SCK, Bogdanowicz SM, Doyle JJ (2009a) Characterization of twelve polymorphic microsatellite markers for Georgia false indigo (*Amorpha georgiana* Wilbur var. *georgiana*), an endangered species, and their utility in other dwarf *Amorpha* L. species. *Molecular Ecology Resources*, **9**, 225-228.

Straub SCK, Sorrie BA, Weakley AS (2009b) *Amorpha confusa*, a New Name for an Old *Amorpha* (Fabaceae: Amorpheae). *Journal of the Botanical Research Institute of Texas*, **3**, 151-155.

Török K, Botta-Dukát Z, Dancza I, *et al.* (2003) Invasion gateways and corridors in the Carpathian Basin: Biological invasions in Hungary. *Biological Invasions*, **5**, 349-356.

Vestal PA, Schultes RE (1939) *The Economic Botany of the Kiowa Indians as it Relates to the History of the Tribe.* Botanical Museum, Harvard University, Cambridge, Massachusetts.

Wang ET, Rogel MA, Sui XH, Chen WX, Martinez-Romero E, van Berkum P (2002) *Mesorhizobium amorphae*, a rhizobial species that nodulates *Amorpha fruticosa*, is native to American soils. *Archives of Microbiology*, **178**, 301-305.

Wendel JF, Doyle JJ (1998) *Phylogenetic incongruence: Window into genome history and molecular evolution*. In: *Molecular Systematics of Plants II: DNA Sequencing* (eds. Soltis, DE, Soltis PS, Doyle JJ), pp 265-296, Kluwer Academic Publishers, Boston, Massachusetts.

Wendelberger KS, Fellows MQN, Maschinski J (2008) Rescue and restoration: Experimental translocation of *Amorpha herbacea* Walter var. *crenulata* (Rydb.) Isely into a novel urban habitat. *Restoration Ecology*, **16**, 542-552.

Wendelberger KS, Maschinski J (2009) Linking geographical information systems and observational and experimental studies to determine optimal seedling microsites of an endangered plant in a subtropical urban fire-adapted ecosystem. *Restoration Ecology*, **17**, 845-853.

Wilbur RL (1975) A revision of the North American genus *Amorpha* (Leguminosae-Psoraleae). *Rhodora*, **77**, 337-409.

CHAPTER TWO

MOLECULAR PHYLOGENETICS OF *AMORPHA* (FABACEAE: AMORPHEAE): AN EVALUATION OF MONOPHYLY, SPECIES RELATIONSHIPS, AND POLYPLOID ORIGINS

ABSTRACT

Amorpha L. is a North American legume genus of 16 species of shrubs, which is most diverse in the southeastern United States and distinctive due to the reduction of the corolla to a single petal. Most of the species have very limited distributions, but the tetraploid *A. fruticosa* species complex is widely distributed and its range overlaps with those of all of the other species. Morphological variation in the genus is characterized by gradation of characters among species and it has been the subject of repeated taxonomic study due to the difficulty in delimiting species, especially among *A. fruticosa* and allies. This study presents the first phylogenetic and network analyses for evaluation of relationships amongst *Amorpha* species based on three non-coding chloroplast regions (*trnD-trnT*, *trnH-psbA*, *ycf6-psbM*) and two low-copy nuclear genes (*CNGC5*-like, *MinD*-like). The monophyly of the genus was also evaluated to determine if *Errazurizia rotundata* and *Parryella filifolia* were nested in *Amorpha*, rendering it paraphyletic, as has been suggested by work at the tribal level. Chloroplast analyses supported a monophyletic *Amorpha* with *P. filifolia* and *E. rotundata* as successive sister lineages; however, nuclear gene analyses supported the nesting of these two species and thus a paraphyletic *Amorpha*. Relationships among species of *Amorpha* were best resolved in the chloroplast phylogeny and in most cases were concordant with expectations based on morphology.

Relationships based on the nuclear gene phylogenies were less clear due to lack of informative variation (*CNGC5*-like) or conflict in the data set (*MinD*-like). The origins of *A. fruticosa* were unclear, but the chloroplast phylogeny revealed that this species shares the same or similar chloroplast haplotype as other species in a geographic region. Putative recombination of diploid species' alleles was evident in the *MinD*-like network. Phenotypic plasticity in combination with gene flow into this species from different diploids, or even tetraploids, across its range may account for the incredible morphological diversity of the *A. fruticosa* species complex. Putative progenitors for two other suspected allotetraploid species, *A. confusa* and *A. crenulata*, were identified as *A. fruticosa* and *A. herbacea*.

INTRODUCTION

Amorpha L. (Fabaceae Juss.: Amorpheae Boriss.) is a genus of 16 species native to North America with a center of diversity in the southeastern United States (Wilbur 1975; Isely 1998; Straub *et al.* 2009b; Straub *et al.* in press). These plants can be shrubs, subshrubs or herbaceous perennials and are found in a variety of habitats, including riparian corridors, savannas, and prairies. Departure from the canonical papilionaceous floral form is common in Amorpheae and in the flowers of *Amorpha*, only the banner petal is present in the corolla. Over half of all species and varieties of *Amorpha* have very limited distributions and several are considered threatened or endangered (e.g., *A. crenulata* Rydb., *A. confusa* (Wilbur) S.C.K. Straub, Sorrie & Weakley, and *A. georgiana* Wilbur; Federal Register 1985; North Carolina Department of Agriculture & Consumer Services 2008).

There have been numerous, often conflicting, taxonomic treatments of *Amorpha* (e.g., Schneider 1907; Rydberg 1919; Palmer 1931; Wilbur 1975, Isely 1998) largely due to the high degree of environmental plasticity within species and the gradation of character variation between species, most notably involving leaf, floral, and gland characteristics (Palmer 1931; Wilbur 1975; Isely 1998). Disagreement exists as to the delimitation of several species, but the major source of confusion is *Amorpha fruticosa* L., whose list of synonymy includes at least 39 names (Palmer 1931; Wilbur 1975). This sometimes invasive tetraploid species complex ($2n = 4x = 40$; e.g., Kreuter 1930; Turner 1956; Löve 1982) shows "spectacular" environmental plasticity (Wilbur 1975, p. 338) over its large geographic range, which overlaps that of all of the other species and spans most of the United States. Most other taxonomic disagreements involve other currently recognized species that are similar in morphology to *A. fruticosa* and may be weakly differentiated or regional variants (*A. glabra* (Pers.) Poir., *A. nitens* F.E. Boynton, *A. laevigata* Nutt., *A. ouachitensis* Wilbur; Isely 1998). One such example is *A. nitens* whose main differentiating characteristics are that pressed specimens blacken upon drying and that the adaxial sides of the leaflets are shiny; however, some *A. fruticosa* also blacken upon drying. In addition to the *A. fruticosa*-like species, another distinctive group within the genus is the "dwarf amorphas," species of which are characterized as being subshrubs or herbaceous perennials of less than one meter in height that generally also have petioles shorter than the length of the lowermost leaflet and revolute leaflet margins (Wilbur 1975, Isely 1998).

Amorpha itself has never been the subject of a phylogenetic analysis, but recent phylogenetic work on higher-level relationships in tribe

Amorpheae has provided a good foundation for molecular phylogenetic study of the genus. Earlier studies included only a small subset of *Amorpha* species and so do not reveal details of relationships among them, but had enough sampling to bring the monophyly of the genus into question. Analyses of nuclear (*ITS*, *5.8S*, *CNGC4*) and chloroplast (*trnK*, *matK*) markers firmly place *Amorpha* in the amorphoid subclade of Amorpheae, which also includes *Apoplanesia* C. Presl, *Errazurizia* Phil., *Eysenhardtia* Kunth, and *Parryella* Torr. & A. Gray ex A. Gray (McMahon & Hufford 2004, 2005). The amorphoid clade as a whole is strongly supported as monophyletic (McMahon & Hufford 2004), but the relationships among genera remain unclear. The monophyly of *Amorpha* is questionable due to the placement of *Errazurizia rotundata* (Wooton) Barneby and *Parryella filifolia* Torr. & A. Gray, one or both of which may be nested in *Amorpha* (McMahon & Hufford 2004; McMahon & Hufford 2005; McMahon 2005). Total evidence analyses of chloroplast, nuclear ribosomal DNA (nrDNA), and low-copy nuclear gene data indicated that *E. rotundata* and *P. filifolia* are sister taxa with high bootstrap support and could be either the sister group of, or nested within *Amorpha* (McMahon 2005; McMahon & Hufford 2005). However, when the chloroplast data set and *CNGC4* data sets were analyzed individually, only *E. rotundata* was nested in *Amorpha* with support, while *P. filifolia* was either recovered as sister to the rest of the amorphoid clade, potentially sister to *Amorpha* plus *E. rotundata*, or as possibly nested in that group (McMahon & Hufford 2004; McMahon 2005).

Close relationships among *Amorpha*, *E. rotundata*, and *P. filifolia* based on molecular data are largely consistent with relationships suggested by morphology. Wilbur (1975) hypothesized that *Parryella*, which was then

recognized as containing both *P. filifolia* and *E. rotundata* (*Parryella rotundata* Wooton), was most closely related to *Amorpha* among the genera of Amorpheae. Barneby (1977) suggested a close relationship of *Amorpha*, *P. filifolia*, and *Eysenhardtia*. He excluded *E. rotundata* from this group, pending further evidence to place it either with the other species of *Errazurizia* or in *Parryella*, the genus of its first description. Palynological evidence also suggests a close relationship of *Amorpha*, *E. rotundata*, and *P. filifolia*. The pollen observed for this group is tricolporate, has type A os, and is operculate, a combination of characters which is not observed in the other genera of the amorphoid clade of Amorpheae (Mahler 1965). Floral characteristics further support this relationship: *Amorpha* has a single banner petal, *E. rotundata* occasionally has a vestigial banner petal, and *P. filifolia* has no petals, as opposed to the other genera of Amorpheae, which all have five petals (Barneby 1977; McMahon 2005; McMahon & Hufford 2005).

In addition to *A. fruticosa*, several other tetraploid species have been identified in the genus. *Amorpha confusa* and *A. crenulata* are thought to be tetraploids based on patterns observed for microsatellite loci (Chapter 5; Straub *et al.* 2009a). The origins of these species and whether they are autopolyploid or allopolyploid are unknown. Chromosome counts confirming diploidy ($2n = 2x = 20$) have only been done for a handful of other species (*A. californica* Nutt.; Kreuter 1930; *A. canescens* Pursh; Löve 1982; *A. nana* Nutt.; Kreuter 1930; Löve 1982) leaving open the possibility that other *Amorpha* species are also polyploid.

The goals of this study were to use DNA sequence data obtained from three non-coding chloroplast regions and two low-copy nuclear loci to: test the monophyly of *Amorpha*; resolve the relationships of *E. rotundata* and

P. filifolia to *Amorpha* and each other; elucidate the relationships among *Amorpha* species; determine if groups suggested by morphology (e.g., *A. fruticosa*-like amorphas, dwarf amorphas) are monophyletic; and explore the polyploid origins of the tetraploids *A. fruticosa*, *A. confusa*, and *A. crenulata*.

MATERIALS AND METHODS

Sampling Strategy

All sixteen currently recognized species of *Amorpha* (Wilbur 1975; Straub *et al.* 2009b; Straub *et al.* in press) were sampled for this study. When possible, multiple individuals per species, including representatives of named varieties or individuals from different areas of large species ranges, were sampled for a total of 57 individuals. Collection was especially intense across the native and introduced range of the widespread and morphologically variable tetraploid, *A. fruticosa*, in order to capture any underlying genetic variation. In order to test the monophyly of *Amorpha*, samples were obtained from representatives of the other genera of the "amorphoid" clade of Amorpheae (McMahon & Hufford 2004): monotypic *Apoplanesia*, two species of *Errazurizia* including *E. rotundata*, two species of *Eysenhardtia*, and monotypic *Parryella*. An additional outgroup species, *Psorothamnus kingii* (S. Watson) Barneby, was sampled from the "daleoid" clade of Amorpheae. Voucher information for specimens is given in Table 2.1.

DNA Extraction, Amplification, and Sequencing

DNA was extracted from silica dried tissue or fragments of herbarium specimens using a standard cetyl trimethylammonium bromide (CTAB) extraction protocol (Doyle & Doyle 1987) modified by adding 2% PVP-40

Table 2.1. List of voucher specimens.

Species	Voucher Specimen	Herbarium Location	Additional Information
Amorpha apiculata	Fishbein 3745	ARIZ	
Amorpha californica var. *californica*	Kennedy s.n.	BH	
Amorpha californica var. *californica*	Wall (RSA 8060) – A specimen will be made by Michael Wall in spring of 2010.	RSA	Munz & Balls 8060 seed collection from the Santa Rosa Mountains, Riverside Co., CA cultivated at Rancho Santa Ana Botanic Garden
Amorpha californica var. *napensis*	Straub 7	BH	
Amorpha canescens	Bussmann 15941 (MoBot961600-1)	MO	Cultivated at Missouri Botanical Garden – Accession 961600-1
Amorpha canescens	Cohen 24	BH	
Amorpha canescens	Cohen & Straub 18	BH	
Amorpha canescens	Straub 5	BH	Cultivated at Cornell Plantations
Amorpha canescens	Wheeler 16290	MIN	
Amorpha confusa	Straub 3-1	BH	
Amorpha confusa	Straub 37-8	BH	
Amorpha confusa	Weakley #1*	NCU/WNC	
Amorpha crenulata	Lewis s.n.	BH	
Amorpha crenulata	FTG T557 – no specimen made	-	Fairchild Tropical Botanic Garden wild plant #T557
Amorpha fruticosa	Cohen 36	BH	
Amorpha fruticosa	Cohen & Straub 10	BH	
Amorpha fruticosa	Cohen & Straub 28	BH	
Amorpha fruticosa	Cohen & Straub 34	BH	
Amorpha fruticosa	Cohen & Straub 45	BH	
Amorpha fruticosa	Cohen & Straub 54	BH	
Amorpha fruticosa	Cohen & Straub 73	BH	
Amorpha fruticosa	Cohen & Straub 91	BH	
Amorpha fruticosa	Doyle 1591	BH	
Amorpha fruticosa	Smith 27167	MIN	
Amorpha fruticosa	Smith 29307	MIN	
Amorpha fruticosa	Straub 2	BH	

Table 2.1 (Continued)

Species	Voucher Specimen	Herbarium Location	Additional Information
Amorpha fruticosa	Straub 11 – A specimen will be made by Holly Forbes in spring of 2010.	JEPS	Cultivated at Berkeley Botanic Garden Accession 65.1040 collected by W & M Roderick s.n. San Diego Co., CA
Amorpha fruticosa	Straub 12 – A specimen will be made by Holly Forbes in spring of 2010.	JEPS	Cultivated at Berkeley Botanic Garden – Accession 82.0562 obtained from University of Guelph Arboretum, collected from Pelee Island, Ontario, Canada
Amorpha fruticosa	Straub 20	BH	
Amorpha fruticosa	Straub 27	BH	
Amorpha fruticosa	Straub 34	BH	
Amorpha fruticosa	Straub 40	BH	
Amorpha fruticosa	Straub 48	BH	
Amorpha fruticosa	Straub 50	BH	
Amorpha fruticosa	Weakley #1*	NCU/WNC	
Amorpha georgiana	Straub 1-2	BH	
Amorpha georgiana	Straub 15-18	BH	
Amorpha georgiana	Sorrie #2*	NCU/WNC	
Amorpha glabra	Cohen & Straub 87	BH	
Amorpha glabra	Straub 32	BH	
Amorpha herbacea var. *herbacea*	Straub 63	BH	
Amorpha herbacea var. *herbacea*	Weakley #1*	NCU/WNC	
Amorpha herbacea var. *herbacea*	Weakley #2*	NCU/WNC	
Amorpha herbacea var. *floridana*	Straub s.n.	BH	Individual #2 of population sample from Manatee Springs State Park
Amorpha laevigata	Taylor 29265	OKL	
Amorpha nana	Dana s.n.	BH	
Amorpha nana	Smith 28049	MIN	

Table 2.1 (Continued)

Species	Voucher Specimen	Herbarium Location	Additional Information
Amorpha nana	Straub 51	BH	Cultivated, Dryden, NY
Amorpha nitens	Straub 46	BH	
Amorpha nitens	Walker #1*	NCU/WNC	
Amorpha nitens	Whitsell 04-745	BH	
Amorpha ouachitensis	Straub 41	BH	
Amorpha ouachitensis	Zaencker 182 (MoBot 88-1955)	MO	Plant cultivated at Missouri Botanical Garden – Accession number 88-1955
Amorpha paniculata	Cohen & Straub 72	BH	
Amorpha paniculata	Cohen & Straub 76	BH	
Amorpha roemeriana	Cohen 133	BH	
Amorpha roemeriana	Cohen & Straub 59	BH	
Amorpha schwerinii	Diamond 696	NC Zoo	
Amorpha schwerinii	Straub 75	BH	
Apoplanesia paniculata	Burke 84	BH,CICY	
Errazurizia benthamii	Moran 10691	BH	
Errazurizia rotundata	Roth 1857	BH	
Eysenhardtia orthocarpa	Straub s.n.	BH	Plant cultivated at Cornell University from USDA seed accesstion XDL 90-0208
Eysenhardtia texana	Cohen & Straub 55	BH	
Parryella filifolia	Roth 1859	BH	
Psorothamnus kingii	Lavin 6190	BH	

*Tissue for these samples was obtained from Dr. Gregory Chandler at the University of North Carolina, Wilmington. The voucher specimens are housed at WNC, NCU, or both, but the collector numbers were unavailable. It will be requested that labels be attached to the specimens indicating which sheets are the vouchers for this study.

to the extraction buffer. The DNA extraction buffer was further modified for for herbarium specimens and species expected to contain higher amounts of polyphenolics (e.g., *A. nitens*) by doubling the amounts of polyvinylpyrrolidone (PVP-40) and 2-mercaptanol used. CTAB incubation and ethanol precipitation times were also doubled for herbarium specimens.

DNA sequence data were collected for three non-coding chloroplast regions shown to be variable in other angiosperms (Kress *et al.* 2005; Shaw *et al.* 2005; Kress & Erickson 2007). Following Shaw *et al.* (2005), for polymerase chain reaction (PCR) amplification of the *trnH-psbA* region, the trnH^{GUG} primer of Tate and Simpson (2003) and psbA primer of Sang *et al.* (1997) were employed. Amplification of the *trnD-trnT* region was accomplished using the trnD^{GUC} and trnT^{GUU} primers of Demesure *et al.* (1995). Amplification of the *ycf6-psbM* region was achieved using the ycf6F and psbMR primers of Shaw *et al.* (2005). Each PCR reaction contained 1 μL template DNA and 0.5 U of *Taq* DNA polymerase (New England BioLabs, Inc.) per 12.5 μL of final reaction volume and final concentrations for other reagents were 1x Standard *Taq* Buffer (New England BioLabs, Inc.), 0.2 μM of each primer, and 0.2 mM deoxyribonucleotide triphosphate (dNTP) equimolar mix of dATP, dCTP, dGTP, and dTTP. This standard PCR cocktail was modified for the *ycf6-psbM* primer combination by increasing the final concentration of $MgCl_2$ by 2 mM. PCR cycling conditions for *trnH-psbA* were 94 °C for 3 min followed by 40 cycles of 94 °C for 30 s, 58 °C for 30 s, 72 °C for 1 min and a final extension at 72 °C of 5 min. PCR cycling conditions for *trnD-trnT* were 94 °C for 3 min followed by 30 cycles of 94 °C for 30 s, 52 °C for 1 min 15 s, 72 °C for 1 min and a final extension at 72 °C of 5 min. PCR cycling conditions for *ycf6-psbM* were 94 °C for 2 min

followed by 30 cycles of 94 °C for 15 s, 60 °C for 15 s, 72 °C for 1 min and a final extension at 72 °C of 5 min.

DNA sequence data were also collected for two low-copy nuclear genes: *CNGC5*-like and *MinD*-like (hereafter referred to as *CNGC5* and *MinD*). The nuclear genes were amplified in a smaller group of samples than were included in the chloroplast data set due to difficulty in amplification in some samples and of determining phase of alleles for heterozygous individuals. *CNGC5* was amplified using the primers developed for use in *Medicago* L., *Trifolium* L., and *Trigonella* L. by Maureira-Butler *et al.* (2008). PCR primers for *MinD* (Forward 5' - CCGCAACTCGCCGGCGAAACCCCGCG - 3' and Reverse 5' - CATGCTATCTTGCTCCACGAGCCTCCA - 3'; I.J. Maureira-Butler unpublished data) were developed using the same strategy described by Maureira-Butler *et al.* (2008) for development of the *CNGC5* primers. Each PCR reaction contained 1 μL template DNA and 0.5 U of *Taq* DNA polymerase (New England BioLabs, Inc.) per 12.5 μL of final reaction volume and final concentrations for other reagents were 67 mM Tris, 2 mM $MgCl_2$, 2% dimethyl sulfoxide (DMSO), 0.4 μM of each primer, and 250 μM of each dNTP. PCR reactions for *CNGC5* also contained 0.002% bovine serum albumin (BSA). Touchdown PCR cycling conditions for *CNGC5* were 94 °C for 3 min followed by 12 cycles of 94 °C for 30 s, 62 °C for 30 s, 72 °C for 1 min with a decrease in annealing temperature of 0.5 °C per cycle, followed by 28 cycles with a 56 °C annealing temperature and a final extension at 72 °C of 5 min. Touchdown PCR cycling conditions for *MinD* were 94 °C for 2 min followed by 8 cycles of 94 °C for 30 s, 64 °C for 30 s, 72 °C for 1 min with a decrease in annealing temperature of 0.5 °C per cycle,

followed by 27 cycles with a 60 °C annealing temperature and a final extension at 72 °C of 5 min.

Following PCR cycling, the success of all reactions was checked by agarose gel electrophoresis using 2 -5 μL of PCR product, 1% agarose gels, and ethidium bromide visualization. If small volume PCRs were successful for nuclear loci, large volume reactions of 50 – 100 μL were performed and run out on 1.25% agarose gels at 80-90 V for 3 – 4 hrs in order to separate any minor PCR products from the predominant one. The major band was then excised from the gel and DNA extracted using the QIAquick Gel Extraction Kit (Qiagen).

Cycle sequencing was accomplished using BigDye Terminator v3.1 chemistry (Applied Biosystems, Inc.) followed by ethanol precipitation. DNA sequencing was done at the Cornell University Life Sciences Core Laboratories Center (CLC) using 3730 DNA Analyzers (Applied Biosystems, Inc.). Both forward and reverse primers were used for sequencing except the psbMR primer, which was not used due to a homopolymer repeat close to the priming site that prevented readable sequence from being obtained. Due to several homopolymer runs and the large size of the amplicon, sequencing of the *trnD-trnT* amplicon required the use of the internal sequencing primers $trnE^{UUC}$ and $trnY^{GUA}$ described by Shaw *et al.* (2005). All sequences were checked for accuracy and edited using Sequencher 4.8 (GeneCodes Corp.).

Following direct sequencing of nuclear loci, in cases where individuals were heterozygous for more than one position in a particular amplicon for nuclear loci, making phase determination difficult, allele-specific primers designed using the method for mismatch amplification mutation assay

(MAMA) strategy were used to obtain separate allele sequences (Table 2.2; Cha *et al.* 1992; Rauscher *et al.* 2002). MAMA allele-specific sequencing primers were substituted for the original PCR primers, but otherwise sequencing was as described above for products amplified using the original PCR primers.

In some cases where the use of allele-specific primers did not resolve the phase issue for particular individuals, especially in the case of tetraploid individuals, single strand conformation polymorphism (SSCP; Orita *et al.* 1989) analysis was one attempted strategy for their separation. SSCP was accomplished using PCR products produced as described above separated on 0.7x MDE™ (Cambrex, Inc.) gels run at 100V and 5-10 mA for 52 -96 hrs. Bands were visualized using SYBR Gold stain (Invitrogen, Inc.), excised from the gel, mashed in 50 μL of TE buffer, and incubated at 60 °C for 5 min. Following centrifugation, 1.25 μL of supernatant was used as template for PCR re-amplifications using the reaction conditions for PCR of *CNGC5* and *MinD* described above. PCR cycling conditions were 40 cycles of 94 °C for 15 s, 58 °C for 30 s, 72 °C for 45 s and a final incubation at 72 °C of 5 min. Reactions were gel purified and submitted for DNA sequencing as described above.

A third strategy employed for phase determination of alleles, and in a few cases the sole method for determination of alleles when direct sequence was not available, was cloning of PCR products using TOPO® TA Cloning® Kits for Sequencing with TOP10 One Shot® Competent Cells (Invitrogen, Inc.) following manufacturer protocols modified by doing ½ volume reactions. Bacterial colonies were picked into 50 μL Buffer AE (10 mM Tris·Cl; 0.5 mM EDTA, pH 9.0; Qiagen) and lysed at 97 °C. PCR of

Table 2.2. Allele-specific sequencing primers designed using the mismatch amplification mutation assay method (MAMA; Cha *et al.* 1992) in order to determine allele phase for individuals heterozygous for *CNGC5* and *MinD*. All sequences are given in the 5' – 3' orientation.

Primer Name	Direct Sequence	MAMA Primers
CNGC5209F	GTGACGACATTATGTTATCATTGTY	GTGACGACATTATGTTATCATTGGT
		GTGACGACATTATGTTATCATTGGC
CNGC5230F	GTGGTTATATGGTCATTTTGCGY	GTGGTTATATGGTCATTTTGCCC
		GTGGTTATATGGTCATTTTGCCT
CNGC5444R	GCATCAAGCAATCTCTCATCCR	GCATCAAGCAATCTCTCATCAA
		GCATCAAGCAATCTCTCATCAG
CNGC5640R	CCAGGTTAAAAGCTCCTCCCCR	CCAGGTTAAAAGCTCCTCCCAA
		CCAGGTTAAAAGCTCCTCCCAG
MinD123F	GCCCGCCTCGGCTTCTCTGTCGTM	GCCTCGGCTTCTCTGTCGCA
		GCCTCGGCTTCTCTGTCGCC
MinD129F	CGGCTTCTCTGTCGTCGCCGTY	GCTTCTCTGTCGTCGCCGGT
		GCTTCTCTGTCGTCGCCGGC
MinD171F	CGCAACCTCGACCTCCTCCTCGGY	AACCTCGACCTCCTCCTCGTC
		AACCTCGACCTCCTCCTCGTT
MinD360F	CGTCGATGCCCTCAAATCTCGY	GTCGATGCCCTCAAATCTCCT
		GTCGATGCCCTCAAATCTCCC
MinD360R	GAACTGAGGGGAACCCTCGGGR	AACTGAGGGGAACCCTCGGTA
		AACTGAGGGGAACCCTCGGTG
MinD554R	TAATCATATCTGTCCTCACACGR	TAATCATATCTGTCCTCACACCA
		TAATCATATCTGTCCTCACACCG
MinD638R	TAACCTCAGAATCTTCAGGAATCR	TAACCTCAGAATCTTCAGGAATGG
		TAACCTCAGAATCTTCAGGAATGA
MinD642R	CTAATAACCTCAGAATCTTCAGGR	CTAATAACCTCAGAATCTTCAGTA
		CTAATAACCTCAGAATCTTCAGTG
MinD664R	GTACCCTCTGTTGGTACTCCK	GGTACCCTCTGTTGGTACTCAG
		GGTACCCTCTGTTGGTACTCAT
MinD684R	GAGGCTTATTCAAAACCAAAGGR	GAGGCTTATTCAAAACCAAAGGAA
		GAGGCTTATTCAAAACCAAAGGAG
MinD691R	GATCGTTGGAGGCTTATTCAAAAY	GATCGTTGGAGGCTTATTCAAATC
		GATCGTTGGAGGCTTATTCAAATT

plasmids was accomplished by using 1 μL colony lysate as template, 0.05 μL each of kit-provided M13F and M13R primers (Invitrogen, Inc.) and 0.5 U of *Taq* DNA polymerase (New England BioLabs, Inc.) in 10 μL total volume reactions with final ingredient concentrations of 1x Standard *Taq* Buffer (New England BioLabs, Inc.) and 0.2 mM dNTP equimolar mix. PCR cycling conditions were 94 °C for 3 min followed by 25 cycles of 94 °C for 30 s, 56 °C for 30s, 72 °C for 1 min and a final extension of 72 °C for 15 min. The presence of plasmid inserts of the correct size was confirmed by agarose gel electrophoresis as described above. Insert-positive reactions were cleaned by adding a mixture of 5 U of exonuclease I (New England BioLabs, Inc.), 1.25 U of Antarctic phosphatase (New England BioLabs, Inc.), and 1.5 μL of 10X Standard *Taq* Buffer (New England BioLabs, Inc.) and incubating the reactions at 37 °C for 45 min followed by 10 min at 90 °C. DNA sequencing was as described above using the original *CNGC5* or *MinD* PCR primers.

If efforts to determine phase experimentally failed, in some simple cases, the sequences of a few alleles used in the subsequent phylogenetic analyses were hypothesized if a common allele in other individuals could be determined by subtraction from the pattern observed for those heterozygous individuals (Clark 1990).

Alignment and Phylogenetic Analyses

The DNA sequences obtained for each region were aligned using the MUSCLE web server (http://www.ebi.ac.uk/Tools/muscle/index.html; Edgar 2004). Alignments were inspected and adjusted by eye using BIOEDIT v7.0.9.0 (Hall 1999). Informative gaps were coded using the simple indel

coding method of Simmons and Ochoterena (2000). Gaps in the alignments due to homopolymer runs or microsatellites were treated as missing data, not coded as informative gaps. For chloroplast regions, only one copy of each unique sequence obtained for a region was included for alignment. Sequences were still considered unique if they only differed due to homopolymer runs, excluding ambiguous regions of alignment due to the extreme length of some homopolymer runs leading to deteriorated sequence quality (positions 381-415 for *trnD-trnT* and positions 123-180 *trnH-psbA*). A matrix consisting of unique chloroplast haplotypes was constructed by merging matrices for each individual region using WINCLADA 1.7 (Nixon 1999). Two types of matrices were then constructed for the chloroplast haplotypes. The first matrix included haplotypes for which there was not a significant amount of missing data. Seven other matrices were constructed in order to test the placement of haplotypes that did have missing data and consisted of the "no missing data" matrix plus a single haplotype missing one or a significant part of one of the three regions. For the two nuclear regions, two types of data matrices were also constructed. The first consisted of sequences for as many individuals as possible. If the phase of alleles had been successfully determined for an individual, those sequences were included, but if determination of phase was unsuccessful, IUPAC ambiguity codes were used for those heterozygous individuals. A second matrix consisted of only unambiguous sequences for which phase had been determined, except for some outgroup species for which ambiguity codes were retained.

Maximum parsimony (MP) analyses were conducted using TNT v1.1 (Goloboff *et al.* 2008). Tree searches were performed with uninformative

characters deactivated and maximum trees to hold in memory set to 20,001. The parsimony search strategy consisted of 2000 heuristic tree bisection and reconnection (TBR) search replicates with 20 starting trees per replicate and a random starting seed followed by 5000 iterations of the parsimony ratchet with the probabilities of up weighting or down weighting characters set to 5, followed by 50 iterations of tree drift, followed by tree fusion. The number of trees allowed to be held in memory was then increased to 1,000,000 and all trees were swapped to completion using TBR. All most parsimonious trees from these searches were then used to make a strict consensus with all unsupported nodes collapsed. Ten thousand bootstrap replicates with 10,001 trees allowed in memory were conducted with 20 TBR searches with 20 starting trees per search followed by 200 parsimony ratchet iterations with the probabilities of up weighting or down weighting characters set to 5 per bootstrap replicate. Bootstrap support values were calculated on the strict consensus tree to assess support for clades in optimal trees using WINCLADA.

Models of molecular evolution for the chloroplast haplotype data set and two nuclear gene data sets with gap coding removed and with and without ambiguity codes were determined using jMODELTEST 1.0.1 (Posada 2008) set to evaluate 88 different models using an ML optimized base tree. The best model for each data set was selected using the Akaike Information Criterion (AIC; Akaike 1974). Maximum likelihood (ML) analyses were conducted in GARLI v0.96 (Zwickl 2006). A total of 200 ML search replicates were conducted. Fifty search replicates were conducted using stepwise addition starting trees and fifty search replicates were conducted using random starting trees and default search settings. Search settings were

then adjusted for a more thorough search by decreasing the selection intensity to 0.1, decreasing the optimization precision (startoptprec) to 0.2 and increasing the maximum distance a branch could be moved in subtree pruning and regrafting (SPR; limsprrange) to 7. An additional 50 search replicates apiece were conducted with these settings using either stepwise addition or random starting trees. The trees with the highest –ln likelihood from each of the four searches were then compared for differences in topology. To assess support for clades in the most likely trees, 10,000 bootstrap replicates were performed using default search settings and one search replicate per bootstrap replicate. Bootstrap support values were calculated on the most likely tree out of the four search strategy sets using WINCLADA.

Parsimony tree figures were created using WINCLADA and maximum likelihood tree figures were created using FIGTREE 1.3.1 (Rambaut 2006).

Network Analyses

Network analyses were conducted for each nuclear data set. Matrices for these analyses were edited to remove as much missing data as possible as follows. After the removal of outgroup sequences from the matrices, the ends of all sequences were trimmed to the shortest length observed for any individual, missing data due to homopolymer runs were removed, and missing data for other sequence gaps were reduced to a single one base pair gap. Statistical parsimony (Templeton *et al.* 1992) implemented in TCS 1.21 (Clement *et al.* 2000) was used to visualize genetic distances between haplotypes or alleles, while allowing for non-hierarchical relationships or ancestor-descendant relationships to be present (Posada & Crandall 2001).

The confidence interval for the sequence connection limit for these analyses was 95% and gaps were treated as a fifth state.

RESULTS

Chloroplast Data

For the chloroplast data set, DNA sequence data were obtained for representatives of 15 of the 16 currently recognized species of *Amorpha*, *E. rotundata*, *P. filifolia*, and five Amorpheae outgroup species. No fresh material of *A. laevigata* could be obtained and although DNA extraction from a herbarium specimen was successful, only the *trnH-psbA* region of the chloroplast could be amplified. Due to lack of sufficient data, this sequence (type "C"; see Figure 2.1) was excluded from the phylogenetic analyses. An additional seven individuals had missing data for either *trnH-psbA* or *ycf6-psbM*. Thirty-five distinct chloroplast haplotypes were observed for the 50 individuals sequenced for all three regions. Alignment length, sequence variability, and gap coding information for each individual region and the combined haplotype data set are given in Table 2.3.

Maximum parsimony analyses resulted in recovery of six most parsimonious trees with a length of 245 steps, consistency index (Ci) of 91, and a retention index (Ri) of 89. A strict consensus of the six most parsimonious trees with unsupported nodes collapsed is shown in Figure 2.1. Placement of collection locality information on the tree revealed a roughly geographic pattern of haplotype relationships, especially with reference to the *A. fruticosa* haplotypes. Analyses of each chloroplast region individually were consistent with overall haplotype relationships in the combined analysis for *trnD-T* and *trnH-psbA*. However, the individual analysis of *ycf6-psbM*

Figure 2.1. Phylogenetic relationships of chloroplast haplotypes based on maximum parsimony analysis of three non-coding chloroplast regions (*trnD-trnT*, *trnH-psbA*, and *ycf6-psbM*). The strict consensus of six most parsimonious trees is shown. Numbers above the branches indicate bootstrap support calculated on the strict consensus for 10,000 bootstrap replicates. Samples are identified by collector numbers. The geographic origin of each sample, represented by U.S. state abbreviations, is given. A lowercase "c" indicates a cultivated individual. The combination of numbers and letters given for each terminal in parentheses indicates the sequence variants that make up each haplotype ("H" in the combination). The number represents the *trnD-trnT* type followed by a letter for the *trnH-psbA* type*,* followed by a second letter for *ycf6-psbM* type. Thus, for example, *E. rotundata* has allele 16 for *trnD-trnT*, allele J for *trnH-psbA*, and allele C for *ycf6-psbM*, combining to produce haplotype H16JC. The clade containing all dwarf amorphas, except *A. confusa*, is marked "D" and the clades containing *A. fruticosa* and the other *A. fruticosa*-like species are marked "F1" and "F2."

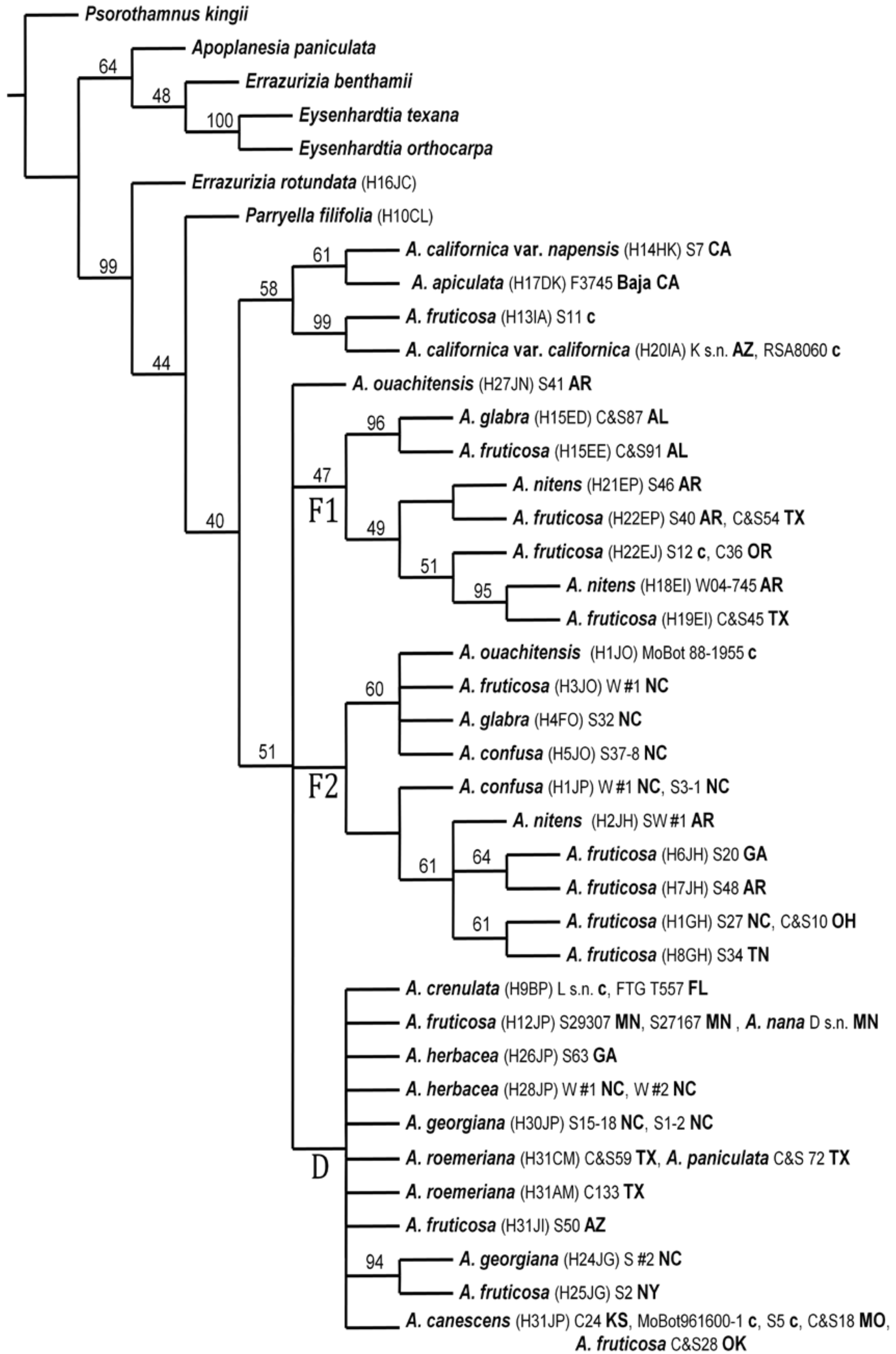

Psorothamnus kingii
Apoplanesia paniculata
Errazurizia benthamii
Eysenhardtia texana
Eysenhardtia orthocarpa
Errazurizia rotundata (H16JC)
Parryella filifolia (H10CL)
A. californica var. napensis (H14HK) S7 CA
A. apiculata (H17DK) F3745 Baja CA
A. fruticosa (H13IA) S11 c
A. californica var. californica (H20IA) K s.n. AZ, RSA8060 c
A. ouachitensis (H27JN) S41 AR
A. glabra (H15ED) C&S87 AL
A. fruticosa (H15EE) C&S91 AL
A. nitens (H21EP) S46 AR
A. fruticosa (H22EP) S40 AR, C&S54 TX
A. fruticosa (H22EJ) S12 c, C36 OR
A. nitens (H18EI) W04-745 AR
A. fruticosa (H19EI) C&S45 TX
A. ouachitensis (H1JO) MoBot 88-1955 c
A. fruticosa (H3JO) W #1 NC
A. glabra (H4FO) S32 NC
A. confusa (H5JO) S37-8 NC
A. confusa (H1JP) W #1 NC, S3-1 NC
A. nitens (H2JH) SW #1 AR
A. fruticosa (H6JH) S20 GA
A. fruticosa (H7JH) S48 AR
A. fruticosa (H1GH) S27 NC, C&S10 OH
A. fruticosa (H8GH) S34 TN
A. crenulata (H9BP) L s.n. c, FTG T557 FL
A. fruticosa (H12JP) S29307 MN, S27167 MN , A. nana D s.n. MN
A. herbacea (H26JP) S63 GA
A. herbacea (H28JP) W #1 NC, W #2 NC
A. georgiana (H30JP) S15-18 NC, S1-2 NC
A. roemeriana (H31CM) C&S59 TX, A. paniculata C&S 72 TX
A. roemeriana (H31AM) C133 TX
A. fruticosa (H31JI) S50 AZ
A. georgiana (H24JG) S #2 NC
A. fruticosa (H25JG) S2 NY
A. canescens (H31JP) C24 KS, MoBot961600-1 c, S5 c, C&S18 MO,
A. fruticosa C&S28 OK
64
48
100
99
44
40
58
61
99
51
47
96
49
51
95
60
61
64
61
94
F1
F2
D

Table 2.3. Sequence length and variability for DNA regions sequenced for phylogenetic analysis of *Amorpha*. The "All" category encompasses all samples sequenced including outgroups and the "Ingroup" category includes all *Amorpha* samples plus *E. rotundata* and *P. filifolia*.

DNA Region	Genome	Aligned Length (bp)	All # Variable/ Informative Characters	Ingroup # Variable/ Informative Characters	# Gaps Coded
cp combined	chloroplast	2372	208/119	61/34	8
trnD-trnT	chloroplast	1550	137/78	42/20	2
trnH-psbA	chloroplast	308	23/12	8/0	0
ycf6-psbM	chloroplast	514	48/24	11/4	6
CNGC5	nuclear	771	83/20	55/15	0
MinD	nuclear	708	129/84	72/50	0

resulted in a sister group relationship for the "H" and "P" sequence types with bootstrap support of 61 in contrast to the relationship of haplotypes including these sequence types in the overall analysis (Figure 2.1). Results for parsimony analyses to determine the phylogenetic placement of partial chloroplast haplotype data are given in Table 2.4.

For the maximum likelihood analyses of the chloroplast data set the TPM1uf+G model (Kimura 1981) was used. The final model parameters estimated in GARLI for one of the most likely replicates were a nucleotide substitution relative rate matrix of AC = 1.000, AG = 1.249, AT = 0.317, CG = 0.317, CT = 1.249, GT = 1.000 and estimated equilibrium state frequencies of A = 0.3419, C = 0.1405, G = 0.1513, T = 0.3663. The rate heterogeneity model included an alpha shape parameter of 0.3031 and rate categories of 0.0056, 0.1069, 0.6010, and 3.2865 with a proportion of 0.25 each. For each of the four search strategies, the best –ln likelihoods ranged from -4635.1831 to -4635.1830 and the –ln likelihoods for 191 of 200 replicates were identical to or fell within 0.1 of these best values. Relationships apparent in the phylograms for each of the best search replicates from the four search strategies were identical to each other and consistent with the strict consensus of most parsimonious trees (Figure 2.2).

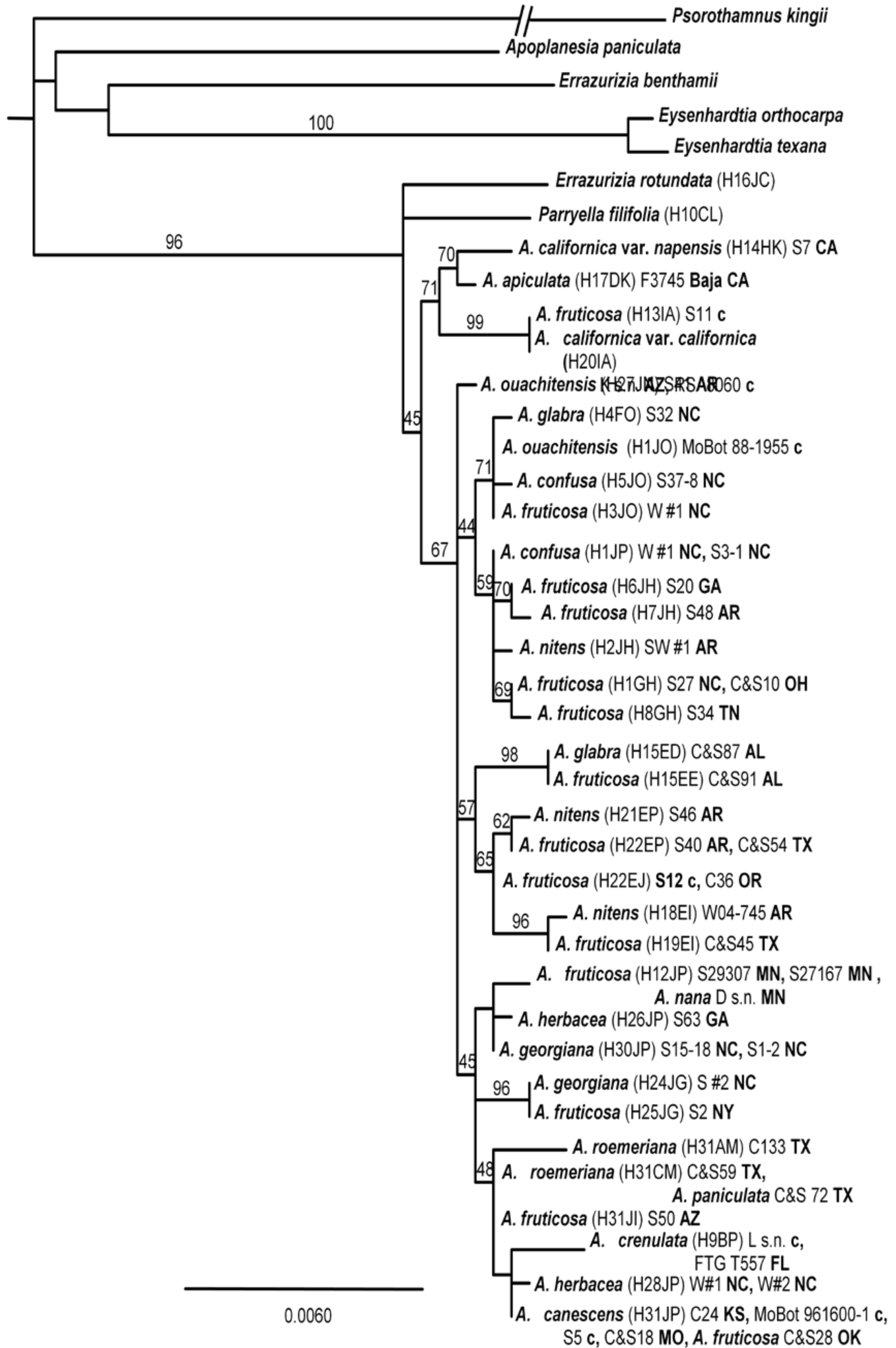
Low-copy Nuclear Gene Data

For the *CNGC5* data set, DNA sequence data were obtained for 28 individuals that were either homozygous or heterozygous and for which allele phase could be determined. These individuals represented 13 of 16 *Amorpha* species, *E. rotundata*, and *P. filifolia.* Amongst *Amorpha* species 27 alleles were observed. Data were collected for an additional 3 heterozygous

Table 2.4. Results of separate parsimony analyses for each incomplete chloroplast haplotype.

Species	Sample ID	Missing Region	Length of Missing Region	Haplo-type	Number of MPT	Length	Ci	Ri	Placement	Bootstrap Support
A. canescens	W16290	*ycf6-psbM*	514 bp	H31J_	14	245	91	90	polytomy with members of the clade including H9BP and H12JP	37
A. herbacea var. *floridana*	S s.n.	3' end of *ycf6-psbM*	177 bp	H25CB	6	245	91	90	polytomy with H24JG, H25JG	94
A. fruticosa	C&S73	5' end of *trnH-psbA*	130 bp	H23J?P	125	247	90	89	polytomy with all *Amorpha* haplotypes except *A. californica* clade, tree structure collapses	45
A. fruticosa	D1591	*trnH-psbA*	308 bp	H22_J	6	245	91	90	polytomy with H22EJ, (H18EI,H19EI)	50
A. nana	S51	*trnH-psbA*	308 bp	H12_P	6	245	91	90	sister to H12JP	68
A. nana	S28049	*ycf6-psbM*	514 bp	H11J_	6	246	91	90	sister to H12JP	67
A. schwerinii	S75	ycf6-*psbM*	514 bp	H29J_	14	245	91	90	polytomy with members of the clade including H9BP and H12JP	37

Figure 2.2. Phylogram of relationships of *Amorpha* chloroplast haplotypes determined using maximum likelihood. Numbers given above the branches indicate bootstrap support values calculated based on 10,000 bootstrap replicates. The scale bar represents the number of substitutions per site. Haplotype designations are as in Figure 2.1.



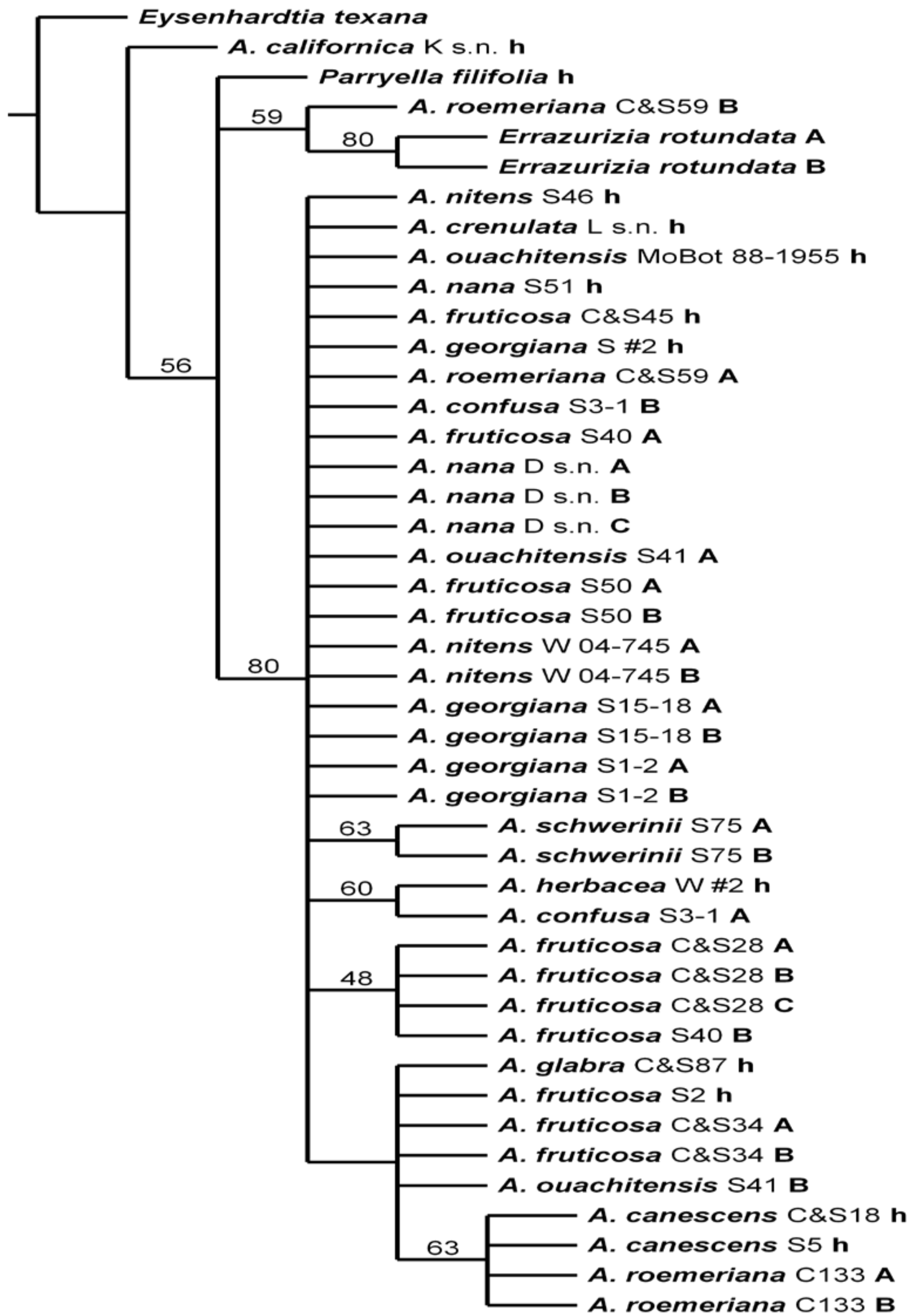
Psorothamnus kingii
Apoplanesia paniculata
Errazurizia benthamii
100
Eysenhardtia orthocarpa
Eysenhardtia texana
Errazurizia rotundata (H16JC)
Parryella filifolia (H10CL)
96
70
A. californica var. napensis (H14HK) S7 CA
A. apiculata (H17DK) F3745 Baja CA
71
A. fruticosa (H13IA) S11 c
99
A. californica var. californica
(H20IA)
A. ouachitensis
45
A. glabra (H4FO) S32 NC
A. ouachitensis (H1JO) MoBot 88-1955 c
71
A. confusa (H5JO) S37-8 NC
A. fruticosa (H3JO) W #1 NC
44
67
A. confusa (H1JP) W #1 NC, S3-1 NC
59
70
A. fruticosa (H6JH) S20 GA
A. fruticosa (H7JH) S48 AR
A. nitens (H2JH) SW #1 AR
69
A. fruticosa (H1GH) S27 NC, C&S10 OH
A. fruticosa (H8GH) S34 TN
98
A. glabra (H15ED) C&S87 AL
A. fruticosa (H15EE) C&S91 AL
57
62
A. nitens (H21EP) S46 AR
A. fruticosa (H22EP) S40 AR, C&S54 TX
65
A. fruticosa (H22EJ) S12 c, C36 OR
96
A. nitens (H18EI) W04-745 AR
A. fruticosa (H19EI) C&S45 TX
A. fruticosa (H12JP) S29307 MN, S27167 MN ,
A. nana D s.n. MN
A. herbacea (H26JP) S63 GA
A. georgiana (H30JP) S15-18 NC, S1-2 NC
45
96
A. georgiana (H24JG) S #2 NC
A. fruticosa (H25JG) S2 NY
A. roemeriana (H31AM) C133 TX
48
A. roemeriana (H31CM) C&S59 TX,
A. paniculata C&S 72 TX
A. fruticosa (H31JI) S50 AZ
A. crenulata (H9BP) L s.n. c,
FTG T557 FL
A. herbacea (H28JP) W#1 NC, W#2 NC
A. canescens (H31JP) C24 KS, MoBot 961600-1 c,
S5 c, C&S18 MO, A. fruticosa C&S28 OK
0.0060

Amorpha individuals and *Ey. texana* Scheele, for which allele phase could not be determined. For the *MinD* data set, DNA sequence data were collected for 26 individuals that were either homozygotes or heterozygotes for which allele phase was successfully determined and represented 13 of 16 *Amorpha* species, *E. rotundata*, *P. filifolia*, and *Ps. kingii.* Amongst *Amorpha* species 30 alleles were observed. No two alleles observed were identical, save those in homozygous individuals, even within the same morphological species. Data were collected for an additional 11 heterozygous individuals for which allele phase could not be determined and two additional Amorpheae outgroup species (*Ey. texana* and *E. benthamii* I.M. Johnst.). Information on alignment length and sequence variability for both data sets is given in Table 2.3.

Maximum parsimony analysis of the *CNGC5* data without ambiguity of allele phase resulted in twelve most parsimonious trees of 89 steps, Ci of 89, and Ri of 91. A strict consensus of the most parsimonious trees with unsupported nodes collapsed is shown in Figure 2.3. Analyses conducted on the expanded data set including ambiguity codes resulted in 380 most parsimonious trees of 89 steps, Ci of 94, and Ri of 91. The strict consensus of these trees resulted in lower resolution than was observed in the previous analysis and did not reveal any additional information about the species relationships in *Amorpha* (data not shown).

For the maximum likelihood analyses of both *CNGC5* data sets, with and without ambiguity of allele phase, the TrN+G (Tamura & Nei 1993) model was used. The final model parameters estimated in GARLI for one of the most likely replicates were a nucleotide substitution relative rate matrix of AC = 1.000, AG = 2.783, AT = 1.000, CG = 1.000, CT = 4.835, GT = 1.000 and estimated equilibrium state frequencies of A = 0.2616, C =

Figure 2.3. Phylogenetic relationships of *CNGC5* alleles from *Amorpha*. The strict consensus of twelve most parsimonious trees is shown. Numbers above the branches indicate bootstrap support values calculated on the strict consensus based on 10,000 bootstrap replicates. Alleles from heterozygous individuals are lettered A-C and alleles from homozygous individuals are denoted by h.

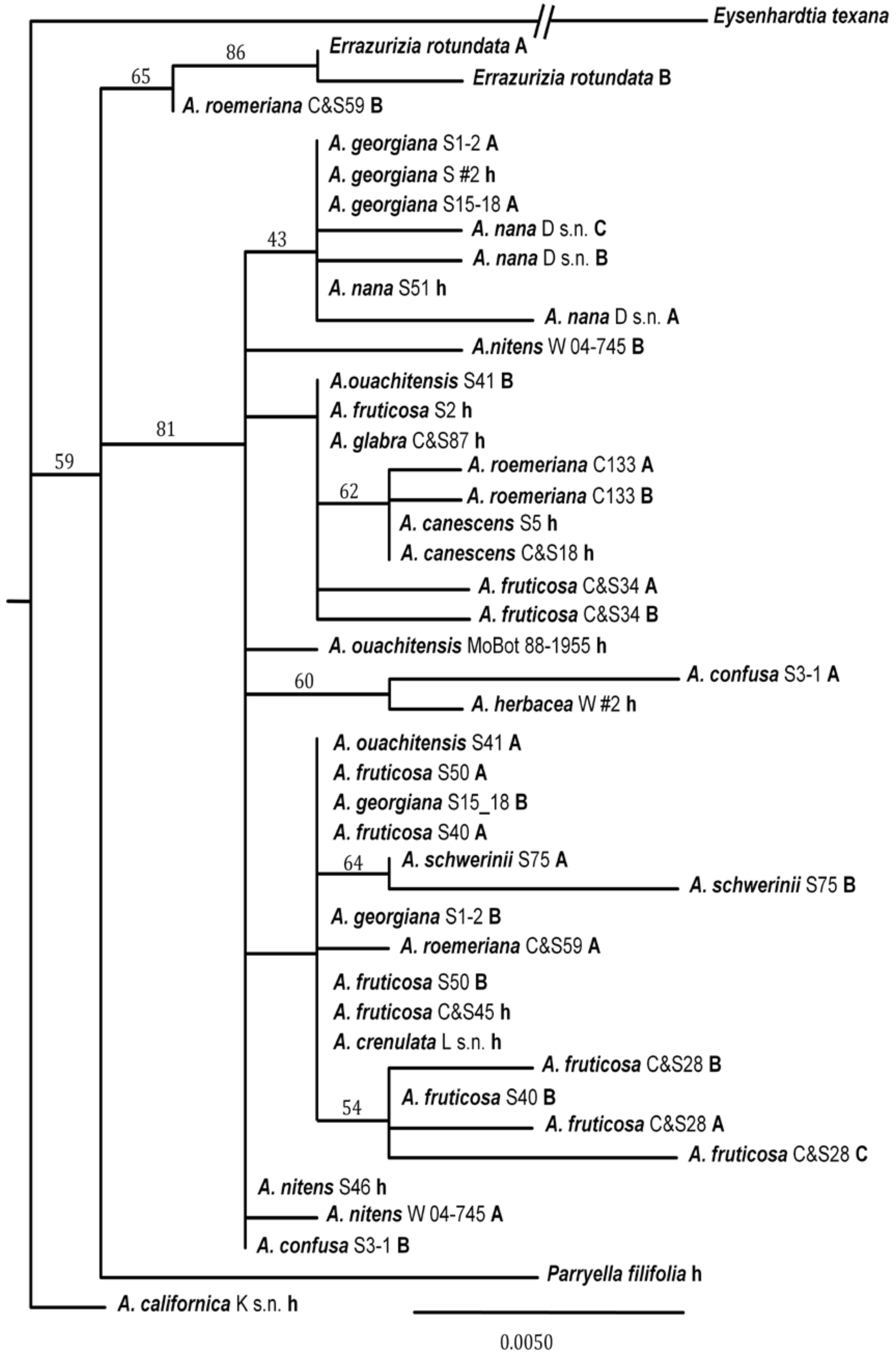

Eysenhardtia texana
A. californica K s.n. h
Parryella filifolia h
59
A. roemeriana C&S59 B
80
Errazurizia rotundata A
Errazurizia rotundata B
A. nitens S46 h
A. crenulata L s.n. h
A. ouachitensis MoBot 88-1955 h
A. nana S51 h
A. fruticosa C&S45 h
A. georgiana S #2 h
56
A. roemeriana C&S59 A
A. confusa S3-1 B
A. fruticosa S40 A
A. nana D s.n. A
A. nana D s.n. B
A. nana D s.n. C
A. ouachitensis S41 A
A. fruticosa S50 A
A. fruticosa S50 B
A. nitens W 04-745 A
80
A. nitens W 04-745 B
A. georgiana S15-18 A
A. georgiana S15-18 B
A. georgiana S1-2 A
A. georgiana S1-2 B
63
A. schwerinii S75 A
A. schwerinii S75 B
60
A. herbacea W #2 h
A. confusa S3-1 A
A. fruticosa C&S28 A
48
A. fruticosa C&S28 B
A. fruticosa C&S28 C
A. fruticosa S40 B
A. glabra C&S87 h
A. fruticosa S2 h
A. fruticosa C&S34 A
A. fruticosa C&S34 B
A. ouachitensis S41 B
A. canescens C&S18 h
63
A. canescens S5 h
A. roemeriana C133 A
A. roemeriana C133 B

0.1866, G = 0.2158, T = 0.3360. The rate heterogeneity model included an alpha shape parameter of 1.4839 and rate categories of 0.2228, 0.5859, 1.0495, and 2.1418 with a proportion of 0.25 each. A best –ln likelihood of -1652.5208 was recovered using three of the four search strategies. Searches using stepwise addition were more likely to produce values the same as or within 0.0002 of the highest –ln likelihood recovered. Searches with random starting trees had more variability in the –ln likelihoods recovered with the lowest being -1655.6920 and only two of 100 replicates hitting the best –ln likelihood of -1652.5208. Even so, apparent relationships in phylograms for each of the best search replicates from the four search strategies were identical to each other and consistent with the strict consensus of most parsimonious trees (Figure 2.4). When ambiguity codes were used for sequences for which allele phase could not be determined, similar results were obtained in terms of the topology apparent in the phylograms, and the same highest –ln likelihood was recovered (data not shown). The only topological difference among the best trees from each search strategy when ambiguous data were analyzed was that there were two alternative placements of the *A. fruticosa* C&S73 ambiguous sequence.

Statistical parsimony network analysis resulted in a network with all haplotypes connected by 11 steps or less, which was the maximum distance allowed for a 95% confidence interval for statistical parsimony (Figure 2.5). This network suggested that the lack of supported resolution in the phylogenetic analyses was likely due to a lack of information because many alleles were identical or differed by only one to a few mutational steps.

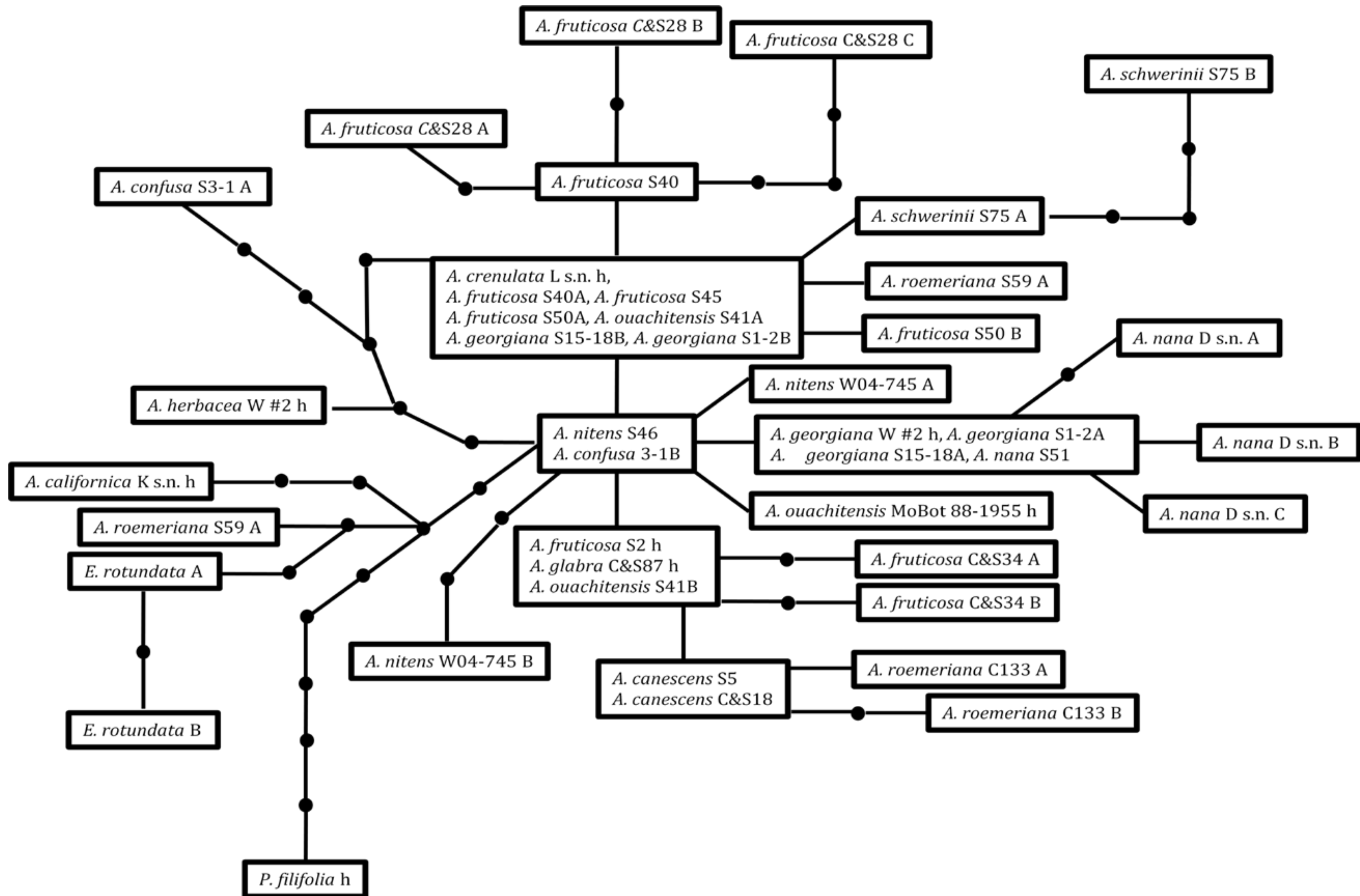
Maximum parsimony analysis of the *MinD* data without ambiguity of allele phase resulted in 157 most parsimonious trees of 250 steps, Ci of 63,

Figure 2.4. Phylogram of relationships of *CNGC5* alleles from *Amorpha* determined using maximum likelihood. Numbers given above the branches indicate bootstrap support values calculated based on 10,000 bootstrap replicates. The scale bar represents the number of substitutions per site. Alleles from heterozygous individuals are lettered A-C and alleles from homozygous individuals are denoted by h.



Eysenhardtia texana
86
Errazurizia rotundata A
65
Errazurizia rotundata B
A. roemeriana C&S59 B
A. georgiana S1-2 A
A. georgiana S #2 h
A. georgiana S15-18 A
A. nana D s.n. C
43
A. nana D s.n. B
A. nana S51 h
A. nana D s.n. A
A.nitens W 04-745 B
A.ouachitensis S41 B
A. fruticosa S2 h
81
A. glabra C&S87 h
59
A. roemeriana C133 A
62
A. roemeriana C133 B
A. canescens S5 h
A. canescens C&S18 h
A. fruticosa C&S34 A
A. fruticosa C&S34 B
A. ouachitensis MoBot 88-1955 h
A. confusa S3-1 A
60
A. herbacea W #2 h
A. ouachitensis S41 A
A. fruticosa S50 A
A. georgiana S15_18 B
A. fruticosa S40 A
64
A. schwerinii S75 A
A. schwerinii S75 B
A. georgiana S1-2 B
A. roemeriana C&S59 A
A. fruticosa S50 B
A. fruticosa C&S45 h
A. crenulata L s.n. h
A. fruticosa C&S28 B
A. fruticosa S40 B
54
A. fruticosa C&S28 A
A. fruticosa C&S28 C
A. nitens S46 h
A. nitens W 04-745 A
A. confusa S3-1 B
Parryella filifolia h
A. californica K s.n. h
0.0050

Figure 2.5. Network of relationships of *CNGC5* alleles determined using statistical parsimony in TCS. The presence of multiple sample names in a box indicates that they possessed the same allele. Each line segment separated by dots between boxes represents one mutational change between alleles. Small circles represent unsampled alleles inferred by TCS. Alleles from heterozygous individuals are lettered A-C and alleles from homozygous individuals are denoted by h.


A. fruticosa C&S28 B
A. fruticosa C&S28 C
A. schwerinii S75 B
A. fruticosa C&S28 A
A. confusa S3-1 A
A. fruticosa S40
A. schwerinii S75 A
A. crenulata L s.n. h,
A. fruticosa S40A, A. fruticosa S45
A. fruticosa S50A, A. ouachitensis S41A
A. georgiana S15-18B, A. georgiana S1-2B
A. roemeriana S59 A
A. fruticosa S50 B
A. nana D s.n. A
A. nitens W04-745 A
A. herbacea W #2 h
A. nitens S46
A. confusa 3-1B
A. georgiana W #2 h, A. georgiana S1-2A
A. georgiana S15-18A, A. nana S51
A. nana D s.n. B
A. californica K s.n. h
A. ouachitensis MoBot 88-1955 h
A. nana D s.n. C
A. roemeriana S59 A
E. rotundata A
A. fruticosa S2 h
A. glabra C&S87 h
A. ouachitensis S41B
A. fruticosa C&S34 A
A. fruticosa C&S34 B
A. nitens W04-745 B
A. canescens S5
A. canescens C&S18
A. roemeriana C133 A
A. roemeriana C133 B
E. rotundata B
P. filifolia h

and Ri of 66. A strict consensus of the most parsimonious trees with unsupported nodes collapsed is shown in Figure 2.6. Analyses conducted on the expanded data set including ambiguity codes resulted in 916,600 most parsimonious trees of 279 steps, Ci of 58, and Ri of 69. Making a strict consensus of these trees led to the collapse of most of the tree structure observed for the non-ambiguous data set, but did reveal a sister group relationship for the allele observed for *A. nana* D s.n. and the ambiguous allele sequence obtained for *A. nana* S51 with bootstrap support of 97 (data not shown). Alleles from *A. fruticosa* C36 for which phase had partially been determined were resolved as sister to the *A. fruticosa* S40 A and *A. fruticosa* S34 alleles respectively.

For the maximum likelihood analysis of the *MinD* data set without ambiguity of allele phase, the TIM2ef+I+G (Posada 2003) model was used. The final model parameters estimated in GARLI for the most likely replicate were a nucleotide substitution relative rate matrix of AC = 1.425, AG = 5.065, AT = 1.425, CG = 1.000, CT = 9.258, GT = 1.000 and equal equilibrium state frequencies of 0.25. The rate heterogeneity model included an alpha shape parameter of 0.5189 with the proportion of invariant sites estimated to be 0.5986 and four additional rate categories of 0.0369, 0.2641, 0.8335, and 2.8655 with a proportion of 0.1003 each. A best –ln likelihood of -2381.0752 was recovered in only two search replicates using stepwise addition starting trees and default search parameters, but values within 0.01 of this value were frequently recovered using all four search strategies. Apparent relationships in phylograms for each of the best search replicates from the four search strategies were identical to each other (Figure 2.7), but

Figure 2.6. Phylogenetic relationships of *MinD* alleles from *Amorpha*. The strict consensus of 157 most parsimonious trees is shown. Numbers above the branches indicate bootstrap support values calculated on the strict consensus based on 10,000 bootstrap replicates. Alleles from heterozygous individuals are lettered A-C and alleles from homozygous individuals are denoted by h.

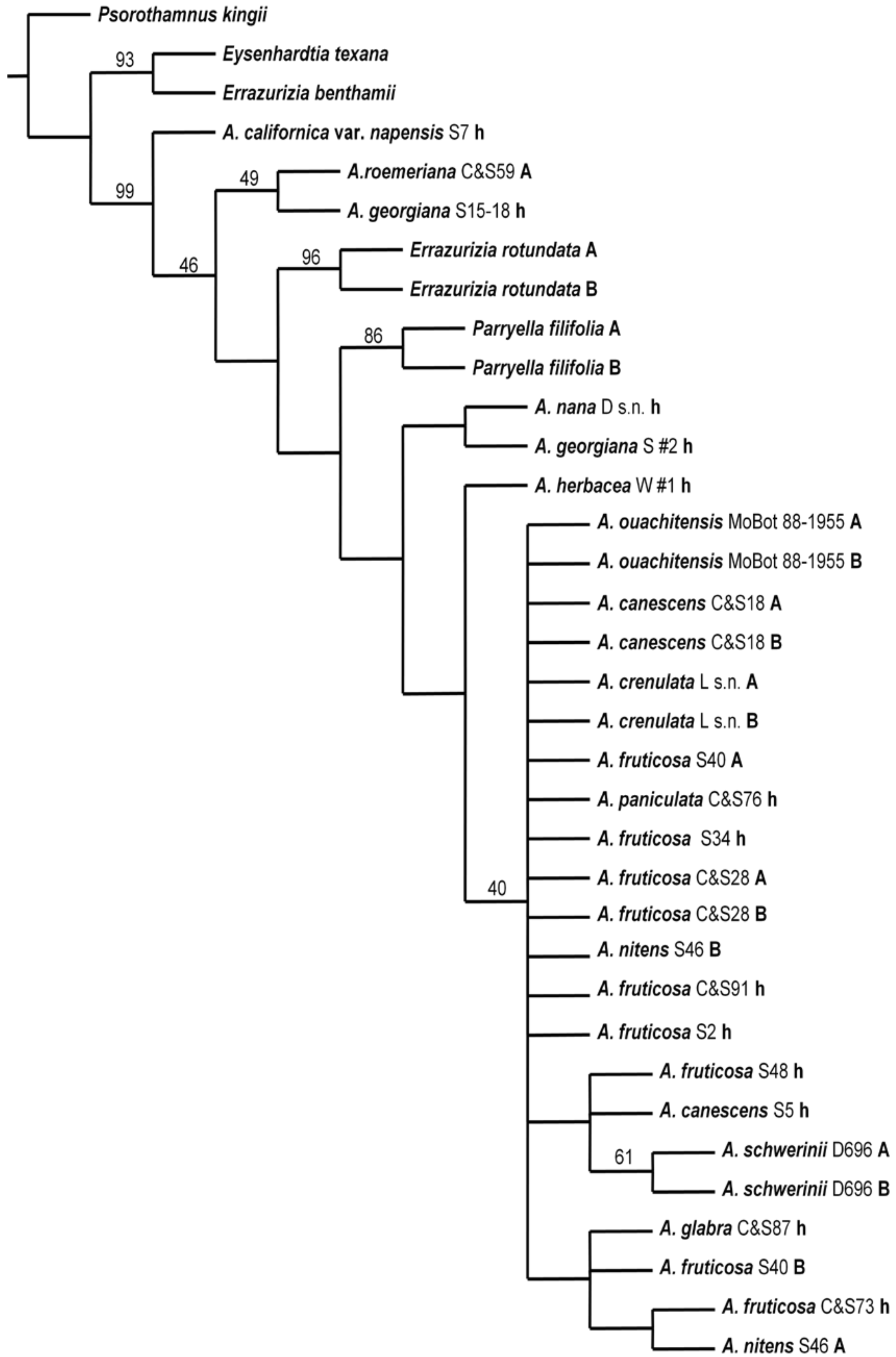
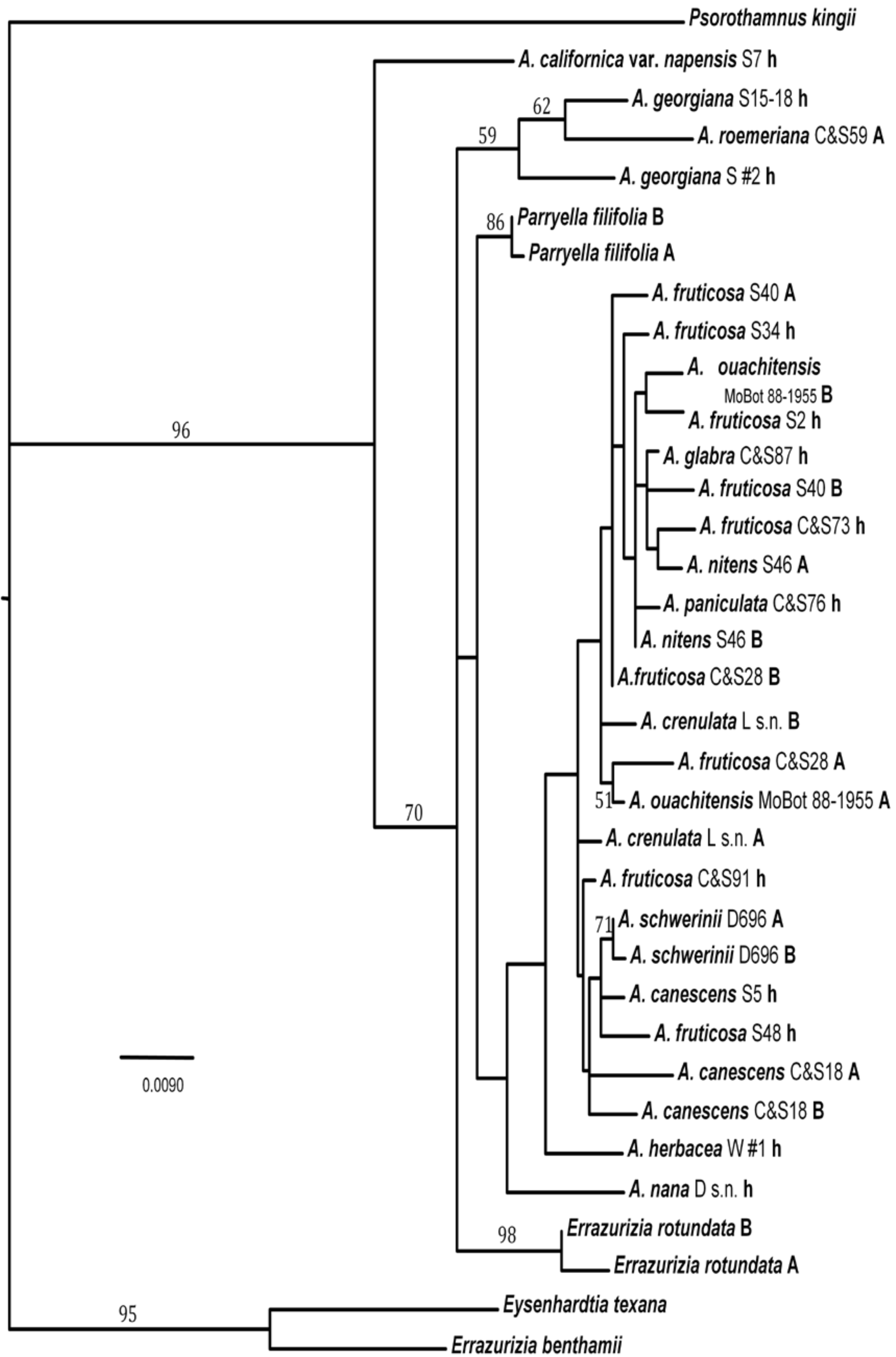

Psorothamnus kingii
93
Eysenhardtia texana
Errazurizia benthamii
A. californica var. napensis S7 h
99
49
A.roemeriana C&S59 A
A. georgiana S15-18 h
46
96
Errazurizia rotundata A
Errazurizia rotundata B
86
Parryella filifolia A
Parryella filifolia B
A. nana D s.n. h
A. georgiana S #2 h
A. herbacea W #1 h
A. ouachitensis MoBot 88-1955 A
A. ouachitensis MoBot 88-1955 B
A. canescens C&S18 A
A. canescens C&S18 B
A. crenulata L s.n. A
A. crenulata L s.n. B
A. fruticosa S40 A
A. paniculata C&S76 h
A. fruticosa S34 h
40
A. fruticosa C&S28 A
A. fruticosa C&S28 B
A. nitens S46 B
A. fruticosa C&S91 h
A. fruticosa S2 h
A. fruticosa S48 h
A. canescens S5 h
61
A. schwerinii D696 A
A. schwerinii D696 B
A. glabra C&S87 h
A. fruticosa S40 B
A. fruticosa C&S73 h
A. nitens S46 A

Figure 2.7. Phylogram of relationships of *MinD* alleles from *Amorpha* determined using maximum likelihood. Numbers given above the branches indicate bootstrap support values calculated based on 10,000 bootstrap replicates. The scale bar represents the number of substitutions per site. Alleles from heterozygous individuals are lettered A-C and alleles from homozygous individuals are denoted by h.

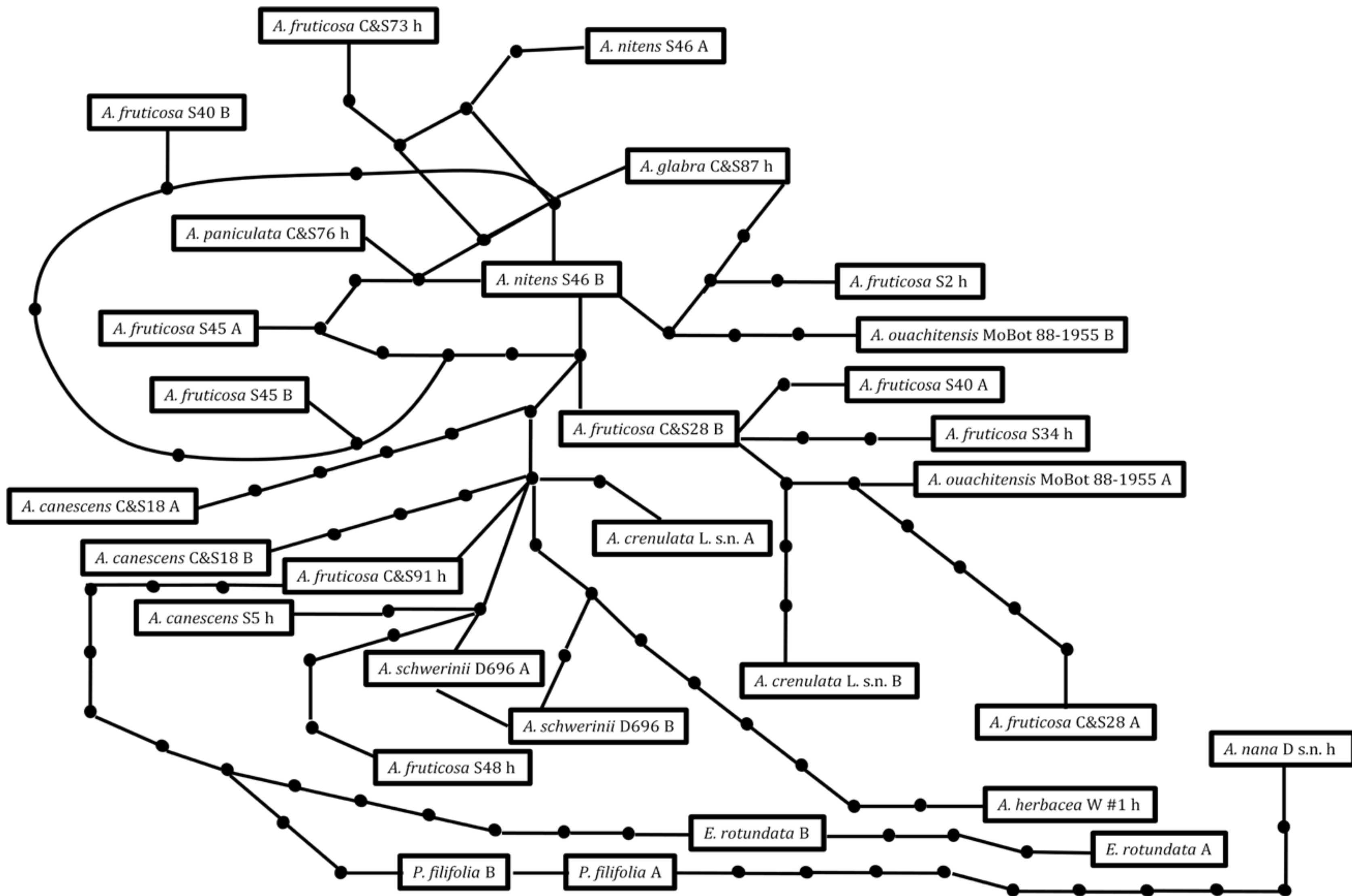

Psorothamnus kingii
A. californica var. napensis S7 h
62
A. georgiana S15-18 h
A. roemeriana C&S59 A
59
A. georgiana S #2 h
86
Parryella filifolia B
Parryella filifolia A
A. fruticosa S40 A
A. fruticosa S34 h
A. ouachitensis
MoBot 88-1955 B
A. fruticosa S2 h
96
A. glabra C&S87 h
A. fruticosa S40 B
A. fruticosa C&S73 h
A. nitens S46 A
A. paniculata C&S76 h
A. nitens S46 B
A.fruticosa C&S28 B
A. crenulata L s.n. B
A. fruticosa C&S28 A
51
A. ouachitensis MoBot 88-1955 A
70
A. crenulata L s.n. A
A. fruticosa C&S91 h
71
A. schwerinii D696 A
A. schwerinii D696 B
A. canescens S5 h
A. fruticosa S48 h
0.0090
A. canescens C&S18 A
A. canescens C&S18 B
A. herbacea W #1 h
A. nana D s.n. h
98
Errazurizia rotundata B
Errazurizia rotundata A
95
Eysenhardtia texana
Errazurizia benthamii

in this case there was incongruence between the ML tree and the strict consensus of most parsimonious trees.

For the *MinD* data set using ambiguity codes for sequences for which allele phase could not be determined the TPM2uf+I+G model (Kimura 1981) was used. The final model parameters estimated in GARLI for one of the most likely replicates were a nucleotide substitution relative rate matrix of AC = 1.895, AG = 9.861, AT = 1.895, CG = 1.000, CT = 9.861, GT = 1.000 and estimated equilibrium state frequencies of A = 0.2057, C = 0.2914, G = 0.2483, T = 0.2546. The rate heterogeneity model included an alpha shape parameter of 0.5109 with the proportion of invariant sites estimated to be 0.6328 and four additional rate categories of 0.0354, 0.2590, 0.8280, and 2.8776 with a proportion of 0.0918 each. A highest –ln likelihood of -2561.8601 was recovered using either stepwise addition or random starting trees and either default or adjusted search parameters. Apparent relationships in phylograms for each of the best search replicates from the four search strategies were identical to each other, but conflicted with the relationships recovered without inclusion of ambiguous sequence (data not shown).

For a 95% confidence interval for statistical parsimony, the distance connection limit was set to eleven steps for *MinD* alleles. *Amorpha californica*, *A. roemeriana*, and *A. georgiana* alleles could not be attached to the network because they differed by more than eleven steps. The network connecting the rest of the observed alleles is shown in Figure 2.8. The reticulate relationships of alleles shown by the six loops in the network indicated that the reason for the lack of resolution in the phylogenetic analyses could be conflict in signal, possibly caused by recombination. Removal of the *A.*

Figure 2.8. Network of relationships of *MinD* alleles that could be connected within the 95% confidence interval for statistical parsimony in TCS. The presence of multiple sample names in a box indicates that they possessed the same allele. Each line segment separated by dots between boxes represents one mutational change between alleles. Small circles represent unsampled alleles inferred by TCS. Alleles from heterozygous individuals are lettered A-C and alleles from homozygous individuals are denoted by h.


A. fruticosa C&S73 h
A. nitens S46 A
A. fruticosa S40 B
A. glabra C&S87 h
A. paniculata C&S76 h
A. nitens S46 B
A. fruticosa S2 h
A. fruticosa S45 A
A. ouachitensis MoBot 88-1955 B
A. fruticosa S40 A
A. fruticosa S45 B
A. fruticosa C&S28 B
A. fruticosa S34 h
A. canescens C&S18 A
A. ouachitensis MoBot 88-1955 A
A. crenulata L. s.n. A
A. canescens C&S18 B
A. fruticosa C&S91 h
A. canescens S5 h
A. schwerinii D696 A
A. crenulata L. s.n. B
A. schwerinii D696 B
A. fruticosa C&S28 A
A. fruticosa S48 h
A. nana D s.n. h
A. herbacea W #1 h
E. rotundata B
E. rotundata A
P. filifolia B
P. filifolia A

fruticosa alleles or removal of all alleles from known polyploids from the data set both resulted in two unconnected networks, one connecting the *E. rotundata*, *P. filifolia*, and *A. nana* alleles and one connecting the remaining *Amorpha* alleles (data not shown). In the *Amorpha* allele network of the six loops observed in the original network only one, involving the two *A. schwerinii* alleles, remained.

DISCUSSION

Evaluation of the Monophyly of *Amorpha* and the Placements of *Errazurizia rotundata* and *Parryella filifolia*

The monophyly of *Amorpha* is still in question due to conflicting results from chloroplast and low-copy nuclear gene data sets. Phylogenetic analyses of chloroplast haplotype data indicated that *Amorpha* chloroplast genomes are monophyletic with moderate bootstrap support (MP 40; ML 45). The clade consisting of *Amorpha* plus *E. rotundata* and *P. filifolia* chloroplast haplotypes was very strongly supported as monophyletic (MP 99; ML 96). Parsimony analyses of haplotypes suggested that the *P. filifolia* chloroplast genome is sister to the *Amorpha* chloroplast genomes and the *E. rotundata* chloroplast genome is sister to this clade. This finding is partially in conflict with relationships recovered by McMahon and Hufford (2004) based on chloroplast data (5’ end of the *trnK* intron and *matK* gene). In that study the relationship of the *E. rotundata* chloroplast genome to the *Amorpha* chloroplast genomes was unresolved, potentially allowing for it to be nested in or sister to the *Amorpha* sequences, but the *P. filifolia* chloroplast genome was strongly supported as being sister to a clade containing the chloroplast genomes of *Amorpha*, *E. rotundata*, *Apoplanesia*, two other *Errazurizia* species,

and two species of *Eysenhardtia*. The findings of this study are consistent with predictions based on morphology (Mahler 1965; Wilbur 1975).

Phylogenetic analyses of *CNGC5* and *MinD* both suggested that *Amorpha* is paraphyletic. The nested placements of alleles of both *E. rotundata* and *P. filifolia* were supported in both maximum parsimony and maximum likelihood analyses and there was no support for a sister group relationship between the alleles of these two species. These results are consistent with results from another low-copy nuclear gene, *CNGC4* (McMahon 2005). However, these results are in conflict with results obtained from another nuclear genome data source, ribosomal DNA. Previous parsimony analyses of *ITS/5.8S* nrDNA indicated that the *Amorpha* alleles formed a monophyletic group and the *E. rotundata* and *P. filifolia* alleles were successive sisters to *Amorpha* (McMahon & Hufford 2004), but maximum likelihood analysis strongly supported the relationship of the *E. rotundata* and *P. filifolia* alleles, suggesting that they are sister species and belong to a clade which also included *Amorpha* species (McMahon & Hufford 2005).

Species Relationships in *Amorpha*

Many of the relationships suggested by the phylogenetic analysis of chloroplast data are consistent with those that might be expected based on morphology. There was a well-supported clade of *A. californica* and *A. apiculata* in both parsimony and likelihood trees. A morphological synapomorphy of these species is the spine-like glands that are present on the stems and leaves. This finding conflicted with the results of previous studies, which did not recover the alleles from these two species as sister to one another using *ITS/5.8S* rDNA sequences (McMahon & Hufford 2004;

McMahon & Hufford 2005). Of the other major clades that were recovered, one contained all of the chloroplast haplotypes of the species of dwarf amorphas (Clade D in Figure 2.1) save *A. confusa*, a tetraploid (see discussion of the origin of *A. confusa* below). Surprisingly, also included in this group were haplotypes from *A. roemeriana* and *A. paniculata*, large shrubs of the south central United States and the partial haplotype from *A. schwerinii*, a rare plant found in the southeastern U.S. There is no clear morphological synapomorphy for this larger set of species.

The remaining two major clades recovered (Clades F1 and F2 in Figure 2.1) contained haplotypes from *A. fruticosa* and the other *A. fruticosa*-like amorphas: *A. glabra*, *A. nitens*, and *A. ouachitensis*. However, in these clades haplotypes from the morphological species did not form monophyletic groups, but rather grouped with haplotypes of individuals of other *A. fruticosa*-like species from the same region. This pattern could be produced by continued gene flow among these species in more local geographic areas and will be discussed further in the context of *A. fruticosa* evolution below. Alternatively, this pattern could be an artifact of the difficult taxonomy of this genus. The differentiation of these species from the morphologically highly polymorphic *A. fruticosa* is difficult to characterize and there is much overlap among distinguishing morphological characters. As noted by Isely (1998), *A. fruticosa* specimens can be found that have the distinguishing characteristics of *A. glabra*, *A. nitens*, and *A. ouachitensis*. Wilbur (1975, p.338) lamented "the number of specimens that clearly possess the characteristics of one 'taxon' in the growth produced early in the season and that of another 'species' or 'variety' in the later growth." More work will

need to be done to determine if these *A. fruticosa*-like isolates should continue to receive taxonomic recognition.

The chloroplast data also provided some information regarding the more recent colonization of habitat by *A. fruticosa*, which can be weedy and is sometimes considered invasive (e.g., Karrenberg *et al.* 2003; Török *et al.* 2003). Historically, *A. fruticosa* did not occur in the Pacific Northwest region of the United States (Wilbur 1975), and has only been observed in that area since the mid 1980's (Glad & Halse 1993). The chloroplast haplotype of an individual collected in Oregon, *A. fruticosa* C36, was identical to that of a cultivated individual whose origin is known to be Ontario, where *A. fruticosa* is also considered introduced (Straub *et al.* in press). This haplotype was firmly placed in a clade of haplotypes from the south central United States, indicating that the invaders originated from genetic stock from that part of the species range. Introductions of *A. fruticosa* are often anthropogenic, as this species is often cultivated or used for erosion control (Glad & Halse 1993). Once escaped from cultivation *A. fruticosa* easily becomes naturalized and occurs in Africa, Asia, and Europe (Straub *et al.* in press). The northeastern United States is another area that was probably more recently colonized by *A. fruticosa* and chloroplast haplotypes collected from New York and Ohio are identical or nearly identical to haplotypes observed in the southeastern United States, indicating a separate source of invasion into these areas than into Ontario and the Pacific Northwest.

Phylogenetic analyses of *CNGC5*, provided some supported relationships among alleles from the same species in both the parsimony and likelihood analyses, such as for *A. schwerinii* and *A. fruticosa.* There was also a supported grouping of alleles from two different *A. canescens* individuals and

one of the sampled *A. roemeriana* individuals, suggesting a possible close relationship between these species whose chloroplast haplotypes also belonged to a supported clade. In the ML tree, but not the MP tree, a supported clade of alleles from *A. nana* and *A. georgiana* was recovered suggesting a close relationship between these two species, which was also supported by the relationship of their chloroplast haplotypes. These two species are very similar morphologically, mainly differing in leaf length, anther color and connation, and ovary pubescence.

The network analysis of *CNGC5* alleles clarified that the lack of resolution in the phylogenetic analysis is likely due to lack of variation among alleles, many of which differed by only a few mutational steps or were identical across species. The network illustrated possible hybrid origins for two species: *A. ouachitensis* and *A. roemeriana*. One of the *A. ouachitensis* S41 alleles was identical to that found in *A. glabra*, while the other allele was the same as those found in *A. fruticosa* and *A. crenulata*. Sampling of another *A. ouachitensis* individual (MoBot 88-1955) indicated that this could be an isolated case of hybridization, rather than a hybrid origin of the species, because it failed to show the same pattern. Alternatively, this may also have been an artifact of the taxonomic difficulty surrounding the *A. fruticosa* species complex and *A. fruticosa*-like plants. The case of a possible hybrid origin of *A. roemeriana* is discussed further below.

Despite having the highest amount of variation and number of phylogenetically informative characters, phylogenetic analyses of the *MinD* alleles failed to result in resolved relationships. These analyses did provide some supported resolution at deeper nodes, but did not reveal much in the way of relationships among most *Amorpha* species. Network analysis of the

MinD data set revealed multiple loops among the alleles in the large polytomy of the MP consensus and clades without bootstrap support in the ML tree. This result suggested either recombination among alleles or parallel substitution and reversal; in either case, the resultant conflict in the data caused collapse of that part of the tree in the strict consensus. Based on *MinD* it may be that there is some phylogenetic structure among species in the genus, while others of the species segregated based on morphology are tokogenetically related due to continued gene flow. The relationships suggested in the phylogenetic part of the tree supported some of the same relationships observed for *CNGC5*. The allele of one of the *A. georgiana* individuals and *A. nana* formed a clade in the MP consensus, but this relationship was absent in the ML tree. The resolved, but not supported by bootstrap values, grouping of alleles from *A. glabra* C&S87, *A. fruticosa* S40B, *A. fruticosa* C&S73, and *A. nitens* S46A was congruent with relationships based on chloroplast data, but incongruent with the relationships suggested by *CNGC5*.

Two interesting patterns that were replicated across analyses deserve comment. The first relationship that was consistently supported in all analyses was *A. californica* as sister to the rest of the ingroup *Amorpha* taxa plus, in the nuclear gene analyses, *E. rotundata* and *P. filifolia*. This pattern is consistent with previous studies, which recovered *A. californica* as sister to the rest of *Amorpha* based on *ITS/5.8S* nrDNA or as unplaced within the *Amorpha* plus *E. rotundata* and *P. filifolia* clade (McMahon & Hufford 2004; McMahon 2005; McMahon & Hufford 2005). The second interesting pattern was the placement of *A. roemeriana* alleles for individual S59 in the nuclear gene analyses. When looking at the analyses individually, the placement of

the *A. roemeriana* alleles might have appeared to be evidence of an orthology/paralogy problem or incomplete lineage sorting. However, the pattern was consistent for both *CNGC5* and *MinD*. In the case of *CNGC5*, one allele was one mutational step away from that found in several *Amorpha* species, while the other allele was two mutational steps away from the *E. rotundata* allele. From sequence data obtained for *MinD* using both allele-specific sequencing primers and SSCP, it was clear that there were at least three alleles present at this locus. This *A. roemeriana* individual is likely of hybrid origin, either a diploid hybrid with a simple gene duplication of *MinD* or an allopolypoid. This is probably not a characteristic pattern for the whole species, as the alleles for a second *A. roemeriana* individual (C133) did not follow the same pattern.

Overall phylogenetic analyses of the nuclear genes provided less information than the chloroplast data set concerning relationships amongst *Amorpha* species due to lack of supported resolution in the best trees. Incongruence of relationships suggested by each of the nuclear genes and non-monophyly of alleles from morphological species made it difficult to translate the information contained in the individual gene trees into a hypothesis of species relationships and is likely caused by processes such as incomplete lineage sorting and hybridization and introgression (Doyle 1992; Doyle 1997; Maddison 1997; Wendel & Doyle 1998). The possibility of extensive hybridization and introgression, especially involving *A. fruticosa* as was evident in the chloroplast phylogeny, further complicates inference of an overall picture of most species relationships to be inferred from the gene trees. Perhaps because of these processes, there is not a single history to be inferred. Indeed, such complicated gene tree relationships among *Amorpha*

species probably could have been predicted by the complex pattern of morphological variation, which makes species delimitation in this genus so difficult, leading previous taxonomic experts to consider the genus "not especially tractable" (Wilbur 1975, p. 337) and "notoriously difficult" (Barneby 1977, p.8).

Polyploid Origins and Evolution in *Amorpha*

The analyses presented have allowed a better understanding of several of the known tetraploid species of *Amorpha*. When the placement of chloroplast haplotypes from the numerous *A. fruticosa* individuals sampled was considered, there was an indication of trans-specific geographic patterns of haplotype variation, as has been observed for chloroplast haplotype variation in other plant genera (e.g., *Quercus*; Whittemore & Schaal 1991; Petit *et al.* 2004) and mitochondrial haplotype variation in animals (e.g., *Laupala*; Shaw 2002) where hybridization is common. *Amorpha fruticosa* haplotypes appeared at various places in the phylogeny and were associated with haplotypes of different species of similar geographic origin. Nuclear gene alleles from *A. fruticosa* were scattered throughout the allele networks and indicated a complex relationship between the polyploid and the other species. For *CNGC5*, *A. fruticosa* individuals actually shared identical alleles with *A. ouachitensis*, *A. glabra*, and *A. georgiana* individuals. Reduction of the number of loops to one following the removal of the *A. fruticosa* alleles from the *MinD* network suggested influx of alleles from multiple, typically narrowly distributed species, into sympatric populations of the widespread polyploid, where gene flow has allowed them to recombine.

One interpretation of these results is that there were multiple origins of the tetraploid, *A. fruticosa*. Multiple origins are a common phenomenon among polyploids (Soltis & Soltis 1999; Wendel & Doyle 2005). Multiple origins of the polyploid could account for morphological diversity in this species complex, especially if different members of this morphologically polymorphic complex actually had different progenitor species. Alternatively, there may have been a single origin of the polyploid followed by gene flow from sympatric species, leading to shared nuclear alleles and chloroplast haplotypes. Perhaps more importantly, regardless of the origins of *A. fruticosa*, there is likely continued gene flow from sympatric species into the polyploid complex, increasing its genetic diversity either from its progenitors or perhaps from its progenitors and other species as well. Gene flow among the tetraploids is probably also occurring. Gene flow and increased genetic diversity may be maintained either through unreduced pollen in the case of diploid contributors, normal pollen from other tetraploids, or through triploid bridges (Ramsey & Schemske 1998; Wendel & Doyle 2005). Any of these scenarios could also help explain the observed introgression of regional chloroplasts into *A. fruticosa.* In support of this hypothesis, putative hybrids between *A. fruticosa* and diploids in the genus have been reported. A named hybrid between *A. fruticosa* and diploid *A. canescens* (Löve 1982), *Amorpha* x *notha*, is known from Missouri (Palmer 1953; Wilbur 1975), although recent flow cytometry work has indicated that there are tetraploid *A. canescens* (E. Baack, Luther College, pers. comm.), so it is possible that the gene flow between these two species is within a ploidy level. A putative hybrid between *A. fruticosa* and diploid *A. nana* (Kreuter 1930; Löve 1982) has also been observed in Minnesota (R. Dana, MN-DNR, pers. comm.).

Hybridization among other *Amorpha* species, not involving *A. fruticosa*, has also been documented (e.g., *A. georgiana* and *A. herbacea*; Straub & Doyle 2009), further complicating possibilites for gene flow in the genus.

Allopolyploidy and multiple origins were suggested by the nuclear and chloroplast data for *A. confusa*. Two different chloroplast haplotypes were observed for this species, each of which belonged to a different clade, indicating multiple origins of the tetraploid. Although it was unclear which species is the maternal progenitor of the polyploid, *A. fruticosa* emerged as the strongest candidate species because it was present in each of the clades containing an *A. confusa* haplotype. Additionally, its range overlaps with the extremely limited distribution of *A. confusa* in southeastern North Carolina, whereas the ranges of *A. ouachitensis*, *A. glabra*, and *A. nitens* do not. In both MP and ML analyses for *CNGC5* there was a supported relationship between one of the *A. confusa* alleles and the allele observed for *A. herbacea*. There was no clear relationship between the other *A. confusa* allele and the alleles of another species; however, if *A. fruticosa* was the other progenitor, relevant variation would not have been detected because none of the *A. fruticosa* individuals collected in geographic proximity to the range *A. confusa* were sequenced for *CNGC5*. Phase could not be determined for the alleles that were obtained for *A. confusa* for *MinD*, but inclusion of ambiguous sequence for this tetraploid resulted in a clade in the ML analysis containing *A. confusa*, *A. herbacea*, and *A. fruticosa* (and *A. crenulata* see below) alleles, further supporting *A. herbacea* and *A. fruticosa* as progenitors of *A. confusa*. Formation of this tetraploid could have occurred directly via unreduced pollen from *A. herbacea* or through a triploid bridge involving normal haploid *A. herbacea* pollen. That *A. confusa* could have arisen through hybridization

between *A. herbacea* and *A. fruticosa* is also plausible based on the relatively intermediate morphology of the tetraploid. This result is a bit surprising because *A. confusa* had long been considered a variety of *A. georgiana* (Wilbur 1964; Straub *et al.* 2009b), which was not implicated as a possible progenitor of the polyploid.

Origins of the third putative tetraploid included in the analysis, *A. crenulata*, were less clear. In the ML tree for the chloroplast data, the chloroplast haplotype recovered for two *A. crenulata* individuals was in a clade with the chloroplast haplotype observed for two of the *A. herbacea* individuals sampled. Based on morphology, *A. crenulata* has sometimes been treated as a variety of *A. herbacea* (Isely 1986; Isely 1998), so the implication of *A. herbacea* as a possible progenitor is not surprising. For *CNGC5*, *A. crenulata* possessed an allele identical to the allele found in several *A. fruticosa*, *A. ouachitensis*, and two *A. georgiana* individuals. If range overlap is used to select a potential progenitor from among these species, only *A. fruticosa* is found in close proximity to the extremely narrow range of *A. crenulata* in Miami-Dade County, Florida. Neither the phylogenetic analysis of *A. crenulata MinD* alleles, which showed that they belonged to a large clade containing alleles of numerous *A. fruticosa* and multiple other *Amorpha* species, nor the network analysis helped clarify clear candidates for a progenitor of *A. crenulata.* However, if ambiguous sequences were considered, the ML analysis resulted in a close relationship of the *A. crenulata* alleles with *A. fruticosa* alleles in a clade that also containined *A. herbacea* and *A. confusa* alleles. The morphology of *A. crenulata* is also suggestive of a hybrid origin for this species, as it has many characteristics commonly observed in dwarf amorphas, but can grow to three meters in height and

have long petioles, characteristics that are common in *A. fruticosa* (Wilbur 1975; Isely 1998). Further sampling of individuals and genetic markers will be required to test if *A. crenulata* is in fact an allopolyploid having *A. herbacea* and *A. fruticosa* as progenitors.

Conclusion

Exploration of species relationships and polyploid origins in *Amorpha* illustrates some of the most difficult problems facing systematists seeking to elucidate of relationships among closely related species, especially the products of recent radiations. In such cases a variety of confounding factors, such as lineage sorting and various levels of gene flow through hybridization and introgression, can make inference of a species tree difficult, if not impossible. Species *Amorpha* are morphologically variable and environmentally plastic enough to make their taxonomy difficult, but have low levels of genetic divergence in both non-coding chloroplast regions and low-copy nuclear genes. So, despite morphological differentiation among *Amorpha* species, there is not much corresponding DNA sequence variation and even less phylogenetically informative variation. The monophyly of the genus is still in question and the observed phylogenetic and tokogenetic patterns suggest that one or more of the following may be true: 1) the lineages in *Amorpha* are quite young and not enough time has passed since speciation for informative variation to accumulate, and observed polymorphism may be ancestral and subject to incomplete lineage sorting; 2) there may not be as many separate lineages as suggested by morphology and taxonomy due to phenotypic plasticity; 3) there may be gene flow among taxa that were distinct but have experienced secondary contact, especially

mediated by *A. fruticosa*. Increased sampling of individuals and nuclear genes may shed additional light on patterns of relationship in the genus, but it is as likely that more data will merely reveal an even more complicated picture of evolution in *Amorpha* as reveal clear taxonomic groupings.

REFERENCES

Akaike H (1974) A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19**, 716-723.

Barneby RC (1977) *Dalea Imagines*: An illustrated revision of *Errazurizia, Psorothamnus, Marina* and *Dalea* including all species of Leguminosae tribe Amorpheae ever referred to *Dalea*. *Memoirs of the New York Botanical Garden*, **27**, 1-891.

Cha RS, Zarbl H, Keohavong P, Thilly WG (1992) Mismatch amplification mutation assay (MAMA): Application to the c-H-*ras* gene. *PCR Methods and Applications*, **2**, 14-20.

Clark AG (1990) Inference of haplotypes from PCR-amplified samples of diploid populations. *Molecular Biology and Evolution*, **7**, 111-122.

Clement M, Posada D, Crandall KA (2000) TCS: A computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1657-1659.

Demesure B, Sodzi N, Petit RJ (1995) A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants. *Molecular Ecology*, **4**, 129-131.

Doyle JJ (1992) Gene trees and species trees: Molecular systematics as one-character taxonomy. *Systematic Botany*, **17**, 144-163.

Doyle JJ (1997) Trees within trees: Genes and species, molecules and morphology. *Systematic Biology*, **46**, 537-553.

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*, **19**, 11-15.

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, **32**, 1792-1797.

Federal Register (1985) Endangered and threatened wildlife and plants: Endangered and threatened status for five Florida pine rockland plants. *Federal Register*, **50**, 138, 29345-29349.

Glad JB, Halse RR (1993) Invasion of *Amorpha fruticosa* L. (Leguminosae) along the Columbia and Snake Rivers in Oregon and Washington. *Madroño*, **40**, 62-63.

Goloboff PA, Farris JS, Nixon KC (2008) TNT, a free program for phylogenetic analysis. *Cladistics*, **24**, 774-786.

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, **41**, 95-98.

Isely D (1986) Notes about *Psoralea* sensu Auct. *Amorpha, Baptisia, Sesbania* and *Chamaecrista* Leguminosae in the Southeastern USA. *Sida Contributions to Botany*, **11**, 429-440.

Isely D (1998) *Amorpha*. In: *Native and Naturalized Leguminosae (Fabaceae) of the United States (Exclusive of Alaska and Hawaii)*, pp. 132-144, Brigham Young University, Provo, Utah.

Karrenberg S, Kollmann J, Edwards PJ, Gurnell AM, Petts GE (2003) Patterns in woody vegetation along the active zone of a near-natural Alpine river. *Basic and Applied Ecology*, **4**, 157-166.

Kimura M (1981) Estimation of evolutionary distances between homologous nucleotide sequences. *Proceedings of the National Academy of Sciences of the United States of America*, **78**, 454-458.

Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, Janzen DH (2005) Use of DNA barcodes to identify flowering plants. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 8369-8374.

Kress WJ, Erickson DL (2007) A two-locus global DNA barcode for land plants: The coding *rbcL* gene complements the non-coding *trnH-psbA* spacer region. *PLoS One*, **2**, 6, ArteNo.:e508.

Kreuter E (1930) Beitrag zur karyologisch-systematischen studien an Galegeen. *Planta*, **11**, 1-44.

Löve A (1982) International Organization of Plant Bio Systematists Chromosome Number Reports 75. *Taxon*, **31**, 342-368.

Maddison WP (1997) Gene trees in species trees. *Systematic Biology*, **46**, 523-536.

Mahler WF (1965) *The pollen morphology of the tribe Psoraleae (Leguminosae).* National Science Foundation Research Participation Report, 1-24.[2]

Maureira-Butler IJ, Pfeil BE, Muangprom A, Osborn TC, Doyle JJ (2008) The reticulate history of *Medicago* (Fabaceae). *Systematic Biology*, **57**, 466-482.

McMahon MM (2005) Phylogenetic relationships and floral evolution in the papilionoid legume clade Amorpheae. *Brittonia*, **57**, 397-411.

McMahon M, Hufford L (2004) Phylogeny of Amorpheae (Fabaceae: Papilionoideae). *American Journal of Botany*, **91**, 1219-1230.

McMahon M, Hufford L (2005) Evolution and development in the amorphoid clade (Amorpheae : Papilionoideae : Leguminosae): Petal loss and dedifferentiation. *International Journal of Plant Sciences*, **166**, 383-396.

Nixon KC (1999-2004) WinClada. Ver. 1.7. Published by the author, Ithaca, New York.

North Carolina Department of Agriculture & Consumer Services (2008) *Plant Industry Division — Plant Protection Section — Plant Conservation Program: Protected Plant List* [Online Data], NCDA & CS, Raleigh, North Carolina, Available at: http://www.ncagr.com/plantindustry/plant/plantconserve/plist.htm. Accessed 3 June 2009.

Orita M, Suzuki Y, Sekiya T, Hayashi K (1989) Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. *Genomics*, **5**, 874-879.

Palmer EJ (1931) Conspectus of the genus *Amorpha*. *Journal of the Arnold Arboretum*, **12**, 157-220.

[2] A mimeograph copy of this unpublished report was obtained from the Mertz Library of the New York Botanical Garden.

Palmer EJ (1953) A hybrid *Amorpha* and new forms and records from Missouri. *Rhodora*, **55**, 157-160.

Petit RJ, Bodénès C, Ducousso A, Roussel G, Kremer A (2004) Hybridization as a mechanism of invasion in oaks. *New Phytologist*, **161**, 151-164.

Posada D (2003) Using Modeltest and PAUP* to select a model of nucleotide substitution. In: *Current Protocols in Bioinformatics* (eds. Baxevanis AD, Davison DB, Page RDM, Petsko GA, Stein LD, Stormo GD), pp. 6.5.1-6.5.14, John Wiley & Sons, New York, New York.

Posada D (2008) jModelTest: Phylogenetic model averaging. *Molecular Biology and Evolution*, **25**, 1253-1256.

Posada D, Crandall KA (2001) Intraspecific gene genealogies: Trees grafting into networks. *Trends in Ecology and Evolution*, **16**, 37-38.

Rambaut, A (2006 – 2009) FigTree v.1.3.1. Published by the author, Edinburgh. Available at: http://tree.bio.ed.ac.uk. Accessed 21 February, 2010.

Ramsey J, Schemske DW (1998) Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annual Review of Ecology and Systematics*, 29, 467-501.

Rauscher JT, Doyle JJ, Brown AHD (2002) Internal transcribed spacer repeat-specific primers and the analysis of hybridization in the *Glycine tomentella* (Leguminosae) polyploid complex. *Molecular Ecology*, **11**, 2691-2702.

Rydberg PA (1919) *Amorpha*. *North American Flora*, **24**, 26-34.

Sang TD, Crawford DJ, Stuessy TF (1997) Chloroplast DNA phylogeny , reticulate evolution, and biogeography of *Paeonia* (Paeoniaceae). *American Journal of Botany*, **84**, 1120-1136.

Schneider CK (1907) Conspectus generis Amorphae. *Botanical Gazette*, **43**, 297-307.

Shaw J, Lickey EB, Beck JT, *et al.* (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 K (2002) Conflict between nuclear and mitochondrial DNA phylogenies of a recent species radiation: What mtDNA reveals and conceals about modes of speciation in Hawaiian crickets. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 16122-16127.

Simmons MP, Ochoterena H (2000) Gaps as characters in sequence-based phylogenetic analyses. *Systematic Biology*, **49**, 369-381.

Soltis DE, Soltis PS (1999) Polyploidy: Recurrent formation and genome evolution. *Trends in Ecology and Evolution*, **14**, 348-352.

Straub SCK, Bogdanowicz SM, Doyle JJ (2009a) Characterization of twelve polymorphic microsatellite markers for Georgia false indigo (*Amorpha georgiana* Wilbur var. *georgiana*), an endangered species, and their utility in other dwarf *Amorpha* L. species. *Molecular Ecology Resources*, **9**, 225-228.

Straub SCK, Doyle JJ (2009) Conservation genetics of *Amorpha georgiana* (Fabaceae), an endangered legume of the Southeastern United States. *Molecular Ecology*, **18**, 4349-4365.

Straub SCK, Sorrie BA, Weakley AS (2009b) *Amorpha confusa*, a New Name for an Old *Amorpha* (Fabaceae: Amorpheae). *Journal of the Botanical Research Institute of Texas*, **3**, 151-155.

Straub SCK, Reveal JL, Weakley AS (in press) *Amorpha*. In: *Flora of North America North of Mexico* (eds. Flora of North America Editorial Committee) 12+ vols. New York and Oxford.

Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution*, **10**, 512-526.

Tate JA, Simpson BB (2003) Paraphyly of *Tarasa* (Malvaceae) and diverse origins of the polyploid species. *Systematic Botany*, **28**, 723-737.

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.

Török K, Botta-Dukát Z, Dancza I, *et al.* (2003) Invasion gateways and corridors in the Carpathian Basin: Biological invasions in Hungary. *Biological Invasions*, **5**, 349-356.

Turner BL (1956) Chromosome numbers in the Leguminosae. *American Journal of Botany*, **43**, 577-581.

Wendel JF, Doyle JJ (1998) *Phylogenetic incongruence: Window into genome history and molecular evolution*. In: *Molecular Systematics of Plants II: DNA Sequencing* (eds. Soltis, DE, Soltis PS, Doyle JJ), pp 265-296, Kluwer Academic Publishers, Boston, Massachusetts.

Wendel J, Doyle JJ (2005) Polyploidy and evolution in plants. In: *Plant Diversity and Evolution: Genotypic and Phenotypic Variation in Higher Plants* (ed. Henry RJ), pp. 97-117, CAB International, Cambridge, Massachusetts.

Whittemore AT, Schaal BA (1991) Interspecific gene flow in sympatric oaks. *Proceedings of the National Academy of Sciences of the United States of America*, **88**, 2540-2544.

Wilbur RL (1964) A revision of the dwarf species of *Amorpha* (Leguminosae). *The Journal of the Elisha Mitchell Scientific Society*, **80**, 51-65.

Wilbur RL (1975) A revision of the North American genus *Amorpha* (Leguminosae-Psoraleae). *Rhodora*, **77**, 337-409.

Zwickl DJ (2006) *Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion*. Ph.D. Dissertation, The University of Texas at Austin, Austin, Texas.

CHAPTER 3

CHARACTERIZATION OF 12 POLYMORPHIC MICROSATELLITE MARKERS FOR GEORGIA FALSE INDIGO (*AMORPHA GEORGIANA* WILBUR VAR. *GEORGIANA*), AN ENDANGERED SPECIES, AND THEIR UTILITY IN OTHER DWARF *AMORPHA* L. SPECIES[3]

ABSTRACT

In order to facilitate the addition of a genetic component to conservation management plans for Georgia false indigo (*Amorpha georgiana* var. *georgiana*), a rare legume of the southeastern USA, 12 polymorphic microsatellite markers were developed. No gametic disequilibrium was detected among locus pairs, but observations for five of the loci significantly deviated from expected Hardy–Weinberg proportions. Cross-species testing was successful and demonstrated the utility of the majority of the markers in congeners *Amorpha georgiana* var. *confusa* and *Amorpha herbacea*. The results also suggested that *A*. *georgiana* var. *confusa* is tetraploid rather than diploid.

PRIMER NOTE

Georgia false indigo (*Amorpha georgiana* Wilbur var. *georgiana*) is a rare legume native to the longleaf pine savannas of the southeastern USA (Wilbur 1975). It is endangered in North Carolina and Georgia and is a species of special concern in South Carolina (South Carolina Department of Natural

---

[3] Straub SCK, Bogdanowicz SM, Doyle JJ (2009) Characterization of twelve polymorphic microsatellite markers for Georgia false indigo (*Amorpha georgiana* Wilbur var. *georgiana*), an endangered species, and their utility in other dwarf *Amorpha* L. species. *Molecular Ecology Resources*, **9**, 225-228. Reproduced with permission from John Wiley & Sons.

Resources 2006; Georgia Department of Natural Resources 2007; North Carolina Department of Agriculture & Consumer Services 2008). The major threats to the continued existence of this plant are habitat degradation and loss, largely due to fire suppression and agriculture (Sorrie 1995). Nothing is currently known about the population genetics of this plant and the development of genetic markers to explore genetic variation and population structure is the essential first step for including a genetic component in conservation management plans.

In this study, we characterized 12 polymorphic microsatellite loci as a tool for genetic exploration of *A. georgiana* var. *georgiana* populations. Additionally, these loci were tested for cross-amplification and utility in *Amorpha georgiana* var. *confusa* Wilbur[4], a threatened North Carolina endemic (North Carolina Department of Agriculture & Consumer Services 2008) soon to be recognized as a distinct species by A. Weakley, B. Sorrie and S. Straub (unpublished data), and *Amorpha herbacea* Walter, a closely related and more widespread species (Wilbur 1975).

Tissue was collected from 40 *A. georgiana* var. *georgiana* individuals at Fort Bragg Military Reservation, North Carolina, 40 *A. georgiana* var. *confusa* individuals at Green Swamp Preserve, North Carolina, and 40 *A. herbacea* individuals from Sampson and Brunswick Counties, North Carolina. DNA was isolated using the cetyltrimethyl ammonium bromide method of Doyle & Doyle (1987) modified by adding 2% PVP-40 to the extraction buffer. Genomic DNA from one *A. georgiana* var. *georgiana* individual was used to construct a DNA library enriched for various di-, tri- and tetranucleotide

[4] This variety has been recognized at the rank of species since the publication of this paper, and is referred to as *Amorpha confusa* throughout the rest of this dissertation.

repeats following the procedure of Hamilton *et al*. (1999) as modified by Dopman *et al*. (2004). Cycle sequencing reactions produced using standard BigDye Terminator version 3.1 chemistry (Applied Biosystems, Inc.) were cleaned by ethanol precipitation and sequenced using 3730 DNA Analysers (Applied Biosystems, Inc.) at the Cornell University Life Sciences Core Laboratories Center (CLC). Sequences were visualized and unique microsatellitecontaining sequences identified using Sequencher 4.7 (Gene Codes Corp.).

Marker amplification primers were designed using PrimerQuest and Primer 3 (Rozen & Skaletsky 2000). A 5′ universal linker (5′-CGAGTTTTCCCAGTCACGAC; modified from Waldbieser *et al*. 2003) was added to the forward primer to facilitate nested polymerase chain reactions (PCR) containing both the marker-specific primers and the universal linker labelled with either 6-FAM, NED, or VIC (Schuelke 2000; Guipponi *et al*. 2005). Function and optimal annealing temperature for primer sets were determined using genomic DNA from the individual used to create the library. Primer sets that produced one clear band visualized in 1% agarose gel stained with ethidium bromide were then tested for amplification on a subset of eight genomic DNA samples for each of the three taxa. Optimized 10 μL PCR for genotyping most loci contained 1 μL template DNA, 0.5 U of *Taq* DNA Polymerase (New England BioLabs, Inc.), and a final concentration of 1x Standard *Taq* Buffer (New England BioLabs, Inc.), 2 mM additional $MgCl_2$, 0.2 mM of each dNTP, 0.05 μM (linker+) forward primer, 0.15 μM labelled universal linker, and 0.2 μM reverse primer. Improved results for loci that proved difficult to cleanly amplify were obtained using 10 μL PCRs that contained 1 μL template

DNA, 0.25 U of Platinum *Taq* DNA Polymerase (Invitrogen), and a final concentration of 1x PCR buffer (Invitrogen), 1.5 mM $MgCl_2$, 0.2 mM each dNTP, 0.025 μM (linker+)forward primer, 0.075 μM labelled universal linker, 0.1 μM reverse primer, and 0.25 M betaine. Optimized PCR cycling conditions in Techne 412, Techne 512, or MJ Research PTC-100 thermal cyclers were as follows: 94 ºC (2 min); 35–40 cycles of 94 ºC (25–30 s), 53–61 ºC (30– 45 s), 72 ºC (45–60 s); 72 ºC (5 min). See Tables 3.1 and 3.2 for primer set specific PCR mixtures and annealing temperatures. Age25 required the following touchdown PCR protocol to eliminate amplification of paralogues: 94 ºC (2 min); 8 cycles of 94 ºC (30 s), 65 ºC minus 0.5 ºC per cycle (30 s), 72 ºC (45 s); 25 cycles of 94 ºC (30 s), 61 ºC (30 s), 72 ºC (45 s); 72 ºC (5 min).

PCR products were diluted with sterile H_2O by a locus-specific empirically determined amount in order to obtain fluorescence values that were on scale and above 200. Dilutions of PCR products with differing fluorescent labels were pooled and then 1–3 μL mixed with 0.2 μL GeneScan 500 LIZ size standard (Applied Biosystems, Inc.) and brought up to a final volume of 20 μL with Hi-Di Formamide (Applied Biosystems, Inc.). Samples were denatured at 94 ºC (3 min), placed immediately on ice, and run on ABI 3730 DNA analysers at the CLC. Genotypes were determined using GeneMapper 4.0 (Applied Biosystems, Inc.). Loci for which test sets indicated clean results were genotyped for the remaining 32 individuals. PCR was repeated two to three times for samples producing unclear genotypes or PCR failures before being marked as missing data. Standard population genetics metrics were calculated using ARLEQUIN 3.11 (Excoffier *et al.* 2005).

Table 3.1. Characterization of twelve novel microsatellite loci in *Amorpha georgiana* var. *georgiana*.

Locus	GenBank Accession No.	Primer Sequence (5'-3')	Clone Repeat Motif	Dye Label Used	PCR Mix	T_a (°C)	n	A	Size Range (bp)†	H_O	H_E	HW *P* value	Percentage of individuals with >2 alleles
Age01	EU678953	F: AATTCTTGCGCCGGTCCAAATGAC R: TGGCTTCGCATAATCCATGTAGCC	$(AAC)_3CAC(CAA)_7$	VIC	NEB	59	40	13	236-324	0.447	0.680	<0.001*	3
Age02	EU678954	F: ATCATGATCCATGTAGCCGACCC R: CAGTGATGGAGTCCTGTTGCATCT	$(GTT)_{12}$	NED	NEB	59	40	10	207-244	0.744	0.790	0.257	0
Age06	EU678955	F: CGTTTGTCTTTACCCTTGTT R: TTCAAAGTTGAACCACTCTTC	$(CTT)_{21}$	VIC	NEB	59	40	18	281-337	0.487	0.914	<0.001*	0
Age07	EU678956	F: ACTCTTCTCTGTCCCAAAAG R: TCATGGCTATCCACAAACTA	$(AC)_{14}$	6-FAM	Inv	57	40	10	431-455	0.538	0.532	0.585	0
Age10	EU678957	F: ATCTTTGTAAGTCCAAGGCCAA R: TAGCTTGCAGTCTAATGTCAAC	$(GT)_4(GAT)_{16}$	VIC	Inv	57	38	14	128-173	0.529	0.881	<0.001*	0
Age13	EU678958	F: AGAGTACCATGTTGCCCA R: GGTAAAGGCAGAAGCGAAAC	$(AAC)_{11}$	6-FAM	NEB	59	40	11	387-422	0.795	0.886	0.044	8
Age15	EU678959	F: CACACAAAGCCTCGCTATCT R: CAACCCAAACTTGGTTGTGAC	$(AC)_5CC(CA)_3TA(TACA)_{18}$	NED	NEB	60	39	13	147-202	0.711	0.905	<0.001*	0
Age18	EU678960	F: TAGCAACTTTGGTTTGGGTGGG R: TGGTAACTTGGCCCGTCTTGAT	$(TCA)_8$	NED	NEB	59	40	3	133-142	0.359	0.475	0.201	0
Age24	EU678961	F: TTCTGCTACCACAAGTCCACTG R: CCGGATAAGCTCTTGGTAATGG	$(TC)_{20}$	NED	Inv	55	40	11	79-113	0.821	0.852	0.134	0
Age25	EU678962	F: TAAATTAACCGGGCCGACCCAA R: TGGTGGCTACTAGGGTTTCTCA	$(GT)_{17}$	VIC	Inv	61	40	13	297-337	0.282	0.902	<0.001*	0
Age29	EU678963	F: GCACATGCCTTCACAAACAACC R: GCTAGATCTCGCTCAGATTACT	$(ATG)_{10}$	6-FAM	NEB	60	40	13	238-281	0.919	0.884	0.181	5
Age30	EU678964	F: ACCAAGGCAGGAGGATTAAGCA R: CACTGCTGCTTCTGCTGTCAAA	$(AAG)_{23}$	NED	NEB	60	40	15	264-309	0.974	0.907	0.749	0

NEB: PCR optimized using New England BioLabs *Taq*; Inv: PCR optimized using Invitrogen *Taq*; T_a: annealing temperature; n: number of individuals successfully genotyped out of 40; A: number of alleles observed; H_O: observed heterozygosity; H_E: gene diversity
† Sizes given reflect the subtraction of the 20bp universal linker sequence.
*Significant deviation from expected Hardy-Weinberg proportions (HW) following Bonferroni correction.

Overall, 93 unique microsatellite-containing sequences were identified through sequencing the inserts of 253 positive clones. Primers were designed and tested for a total of 32 loci with various repeat types: 13 complex, 5 di-, 12 tri- and 2 tetranucleotide repeats. Of these, 12 amplified reliably, gave clean signal and were polymorphic in *A. georgiana* var. *georgiana* (Table 3.1). No gametic disequilibrium was detected in pairwise comparisons of loci, but observed allele frequencies significantly deviated from Hardy–Weinberg proportions for five loci. Further analysis indicated the possible presence of null alleles ($P < 0.05$) for these five loci due to the excess of observed homozygotes (Micro-Checker 2.2; van Oosterhout *et al.* 2004). In addition to the possibility of null alleles, selfing could also be contributing to increased homozygosity. Not much is known about the breeding system of *A. georgiana*, but a closely related species is known to be able to self-pollinate (Possley *et al.* 2004).

Cross-species test amplifications and genotyping for *A. georgiana* var. *confusa* and *A. herbacea* were highly successful for all loci except Age29 (Table 3.2). Interestingly, primer sets generally produced two to four peaks, indicating the presence of more than two alleles, in many or most *A. georgiana* var. *confusa* individuals. The repetition of this pattern across loci indicates the likely presence of multiple copies of each locus and suggests that this variety is tetraploid rather than diploid. Multiple copies of a single locus, Age13, are amplified in *A. herbacea*, as most individuals had more than two alleles. The presence of more than two alleles for a low percentage of *A. georgiana* var. *georgiana* and *A. herbacea* individuals for a few of the loci (Tables 3.1 and 3.2) suggests that the primers are amplifying paralogues in some

Table 3.2. Cross-species testing of *A. georgiana* var. *georgiana* microsatellite loci in *A. georgiana* var. *confusa* and *A. herbacea.*

	A. georgiana var. *confusa*							*A. herbacea*						
Locus	PCR Mix	T_a (°C)	Ampli-fied?	n	A	Size Range (bp)*	Percentage of individuals with >2 alleles	PCR Mix	T_a (°C)	Ampli-fied?	n	A	Size Range (bp)*	Percentage of individuals with >2 alleles
*Age*01	NEB	59	Y	40	18	276-325	75	NEB	59	Y	40	5	279-286	0
*Age*02	NEB	59	Y	39	16	213-260	62	NEB	59	Y	40	9	218-249	10
*Age*06	NEB	55	Y	40	29	226-372	40	NEB	55	Y	40	11	245-377	8
*Age*07	Inv	57	Y	40	19	426-462	38	Inv	57	Y	40	7	434-455	0
*Age*10	Inv	55	Y	39	22	129-175	64	Inv	55	Y	37	13	128-154	22
*Age*13	NEB	59	Y	40	11	356-405	83	Inv	53	Y	40	9	356-397	83
*Age*16	NEB	60	Y	40	21	122-210	45	Inv	60	Y	40	4	142-158	0
*Age*18	NEB	59	Y	40	3	139-145	0	NEB	53	Y	40	2	142-147	0
*Age*24	Inv	55	Y	38	22	79-129	84	Inv	55	Y	40	4	235-283	10
*Age*25	Inv	61	Y	39	14	254-339	41	Inv	61	Y	40	7	303-336	0
*Age*29	NEB	58	Y/N	28	10	189-281	18	NEB	53	Y/N	8	1	233	0
*Age*30	NEB	58	Y	39	32	259-331	92	NEB	55	Y	40	14	268-297	0

NEB: PCR optimized using New England BioLabs *Taq*; Inv: PCR optimized using Invitrogen *Taq*; T_a: annealing temperature; n: number of individuals successfully genotyped out of 40; A: number of alleles observed
*Sizes given reflect the subtraction of the 20bp universal linker sequence.

individuals or that the populations may be of mixed ploidy with a low proportion of polyploids existing among a majority of diploids.

The development of these 12 microsatellite markers is the foundation for adding a genetic component to conservation management plans for *A. georgiana* var. *georgiana* and will facilitate study of genetic diversity and population structure across the remaining populations. Cross-species testing was successful in two other *Amorpha* species and has opened an avenue for population genetics studies of these plants or for comparative studies.

REFERENCES

Dopman EB, Bogdanowicz SM, Harrison RG (2004) Genetic mapping of sexual isolation between E and Z pheromone strains of the European corn borer (*Ostrinia nubilalis*). *Genetics*, **167**, 301–309.

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*, **19**, 11–15.

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

Georgia Department of Natural Resources (2007) *Special Concern Plant Species in Georgia* [Online Data]. GADNR, Atlanta, Georgia. Available at: http://georgiawildlife.dnr.state.ga.us/content/specialconcernplants.asp. Accessed 3 January 2008.

Guipponi M, Herbert S, Toh MY, Poetter K, Forrest S, Scott HS (2005) Universal fluorescent labeling of PCR products for DHPLC analysis: reducing cost and increasing sample throughput. *BioTechniques*, **39**, 34–40.

Hamilton MB, Pincus EL, Di Fiore A, Fleischer RC (1999) Universal linker and ligation procedures for construction of genomic DNA libraries enriched for microsatellites. *BioTechniques*, **27**, 500–507.

North Carolina Department of Agriculture & Consumer Services (2008) *Plant Industry Division — Plant Protection Section — Plant Conservation Program: Protected Plant List* [Online Data]. NCDA & CS, Raleigh, North Carolina. Available at: http://www.ncagr.com/plantindustry/plant/plantconserve/plist.htm. Accessed 3 January 2008.

van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) Micro-Checker: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, **4**, 535–538.

Possley J, Fellows M, Wendelberger K (2004) *Conservation Action Plan — Public Version Amorpha herbacea var. crenulata* [online data]. Fairchild Tropical Botanic Garden, Coral Gables, Florida. Available at: http://www.fairchildgarden.org/index.cfm?section=centerfortropicalplantconservation&subsection=speciesandhabitatconservation&page=speciesweworkwith. Accessed 4 August 2008.

Rozen S, Skaletsky H (2000) Primer 3 on the WWW for general users and for biologist programmers. In: *Bioinformatics Methods and Protocols: Methods in Molecular Biology* (eds Krawetz S, Misener S), pp. 365–386, Humana Press, Totowa, New Jersey.

Schuelke M (2000) An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology*, **18**, 233–234.

Sorrie BA (1995) *Status Survey of Amorpha georgiana var. georgiana*. US Fish and Wildlife Service, Office of Endangered Species, Asheville, North Carolina, and North Carolina Natural Heritage Program, Division of Parks and Recreation, Raleigh, North Carolina.

South Carolina Department of Natural Resources (2006) *South Carolina Rare, Threatened, And Endangered Species Inventory* [online data]. SCDNR, Columbia, South Carolina. Available at: https://www.dnr.sc.gov/pls/heritage/county_species.list?pcounty=all. Accessed 3 January 2008.

Waldbieser GC, Quiniou SMA, Karsi A (2003) Rapid development of gene-tagged microsatellite markers from bacterial artificial chromosome clones using anchored TAA repeat primers. *BioTechniques*, **35**, 976–979.

Wilbur RL (1975) A revision of the North American genus *Amorpha* (Leguminosae-Psoraleae). *Rhodora*, **77**, 337–409.

CHAPTER FOUR

CONSERVATION GENETICS OF *AMORPHA GEORGIANA* (FABACEAE), AN ENDANGERED LEGUME OF THE SOUTHEASTERN UNITED STATES[5]

ABSTRACT

Amorpha georgiana (Fabaceae) is an endangered legume species found in longleaf pine savannas in the Southeastern United States. Approximately 900 individuals and 14 populations remain, most of which are concentrated in North Carolina. Eleven microsatellite loci were used to explore genetic diversity, population structure and recent population bottlenecks using genotypic data from 132 individuals collected at ten different localities. Although *A. georgiana* is quite rare, it exhibited high levels of genetic diversity (17.7 alleles / locus; $H_O = 0.65$, $H_E = 0.75$). Most of the genetic variation was found within rather than between populations of this species. The single remaining Georgia population was well differentiated from populations of the Carolinas ($F_{ST} > 0.1$), which had weaker structure among them ($F_{ST} < 0.1$). Only a geographically disjunct population showed strong evidence of a recent population bottleneck, perhaps due to a recent founder event. Hybridization with *A. herbacea* was also detected. For conservation

---

[5] Straub SCK, Doyle JJ (2009) Conservation genetics of *Amorpha georgiana* (Fabaceae), an endangered legume of the Southeastern United States. *Molecular Ecology,* **18**, 4349-4365. Reproduced with permission from John Wiley & Sons.

management plans, *A. georgiana* populations in each geographic region (North Carolina, South Carolina and Georgia) plus a disjunct population in North Carolina (Holly Shelter) should be treated as separate management units for which *in situ* conservation, including habitat restoration and use of prescribed burns, should ensure persistence of this species and preservation of its evolutionary potential.

INTRODUCTION

Longleaf pine (*Pinus palustris* Mill.) savanna, once the dominant habitat type in the southeastern United States and home to numerous animal species and one of the world's most diverse temperate floras, has largely been destroyed and degraded since European settlement; only approximately 2% of the original habitat remains (Noss 1989; Ware *et al.* 1993; Outcalt 2000; Frost 2006; Jose *et al.* 2006; Means 2006). Currently, only 0.2% of the remaining habitat is healthy enough to sustain its historical complement of species (Frost 2006). Further complicating the patterns of species endangerment in this habitat type is the fact that 26.7% of the plant taxa found in the region are endemic, and 54% of these are narrow endemics (Sorrie & Weakley 2006). Anthropogenic factors, including the historical naval stores industry, logging, agriculture, development and fire suppression, have all contributed to the decline of the longleaf pine savannas (Ware *et al.* 1993; Outcalt & Sheffield 1996; Outcalt 2000; Frost 2006; Jose *et al.* 2006). The subject of this conservation genetics study, *Amorpha georgiana* Wilbur (Fabaceae: Amorpheae), is one of the exceedingly rare, narrow endemic inhabitants of this devastated ecosystem.

Amorpha georgiana is a suffrutescent subshrub that mainly occurs near the high water mark above blackwater rivers or occasionally at the edges of swampy areas in longleaf pine savanna (Wilbur 1975; Sorrie 1995; Isely1998; Straub *et al.* 2009b). Although the longleaf pine savanna once extended from southern Virginia to eastern Texas, the range of this species currently and historically only included riparian corridors from North Carolina southward to Georgia in the Mid-Atlantic and South Atlantic coastal plains (Sorrie & Weakley 2006). Herbarium records indicate that this species was historically represented by at least one population in nine North Carolina counties, two South Carolina counties, and three Georgia counties, with reports of populations in two additional counties; however, *A. georgiana* has since been extirpated from five counties, three in North Carolina and two in Georgia (Sorrie 1995; Straub *et al.* 2009b). It is possible that this species was never common, judging from the scarcity of herbarium collections (Wilbur 1954, 1964) and narrowness of its habitat requirements.

In 1995 the U.S. Fish and Wildlife Service conducted a baseline status survey for *A. georgiana* that indicated that approximately 855 individuals in eleven populations remained (Sorrie 1995). Since 1995, additional populations have been discovered or historical populations located, bringing the current census to about 900 and number of populations to approximately 14, if none of the populations assessed during the status survey have been lost or suffered reductions in size in the intervening years. The largest concentration of individuals (400–500) is found along the Little River in and around Fort Bragg Military Reservation in North Carolina in approximately 20 subpopulations (Sorrie 1995; Miller 2004). Outside of North Carolina, there are three or four small populations known from South

Carolina and a single population, corresponding to the type locality for the species (Wilbur 1954), extant in Georgia. Based on its rarity, this species has been recognized as endangered at the state level in North Carolina and Georgia, and as a species of concern in South Carolina (South Carolina Department of Natural Resources 2006; Georgia Department of Natural Resources 2007; North Carolina Department of Agriculture & Consumer Services 2008).

Given the severe fragmentation of the habitat in which *A. georgiana* occurs and its low number of populations and small census size, this species is likely to have experienced recent population bottlenecks and may be suffering genetic consequences that often befall endangered species, including decreased genetic diversity and gene flow and increased inbreeding (e.g., Barrett & Kohn 1991; Ellstrand & Elam 1993; Fenster & Dudash 1994; Young *et al.* 1996; Booy *et al.* 2000; Amos & Balmford 2001; Hedrick 2001; Frankham 2003). The negative impacts of these factors affect rare species before ecological or demographic factors cause extinction (Frankham & Ralls 1998; Saccheri *et al.* 1998; Spielman *et al.* 2004a) and have been detected in multiple endangered plant species (e.g., Travis *et al.* 1996; Godt & Hamrick 1998; Helenurm 2003; Peakall *et al.* 2003; Spielman *et al.* 2004a; Lázaro & Traveset 2006). Adding a genetic component to conservation planning has become more critical as more is learned about the role of genetics in extinction (Young *et al.* 1996; Frankham & Ralls 1998; Hedrick 2001; Frankham *et al.* 2002; Frankham 2003; Allendorf & Luikart 2007).

Hybridization and introgression are also factors of concern in the conservation of *A. georgiana*. It co-occurs with another *Amorpha* species, *A. herbacea* Walter, in a single locality: Holly Shelter Game Land in North

Carolina. Morphologically intermediate individuals have been observed at this site, suggesting hybridization and possibly genetic introgression from *A. herbacea*. Introgressive hybridization is thought to be a danger to many rare species, due to the potential for outbreeding depression and the possibility that their genetic uniqueness and adaptation to local environments might be swamped out when they come into contact with closely related species (Harrison 1990; Ellstrand & Elam 1993; Rieseberg & Wendel 1993; Levin *et al.* 1996; Rhymer & Simberloff 1996; Allendorf *et al.* 2001; Wolf *et al.* 2001; Levin 2002). Even with strong reproductive barriers, small amounts of hybridization between two species can lead to widespread genetic introgression (Yatabe *et al.* 2007), a phenomenon that has been recently observed in studies of hybridizing plant (e.g., Burgess *et al.* 2005; Carlson & Meinke 2008) and animal (e.g., Barilani *et al.* 2007; Muñoz-Fuentes *et al.* 2007) species (Rhymer & Simberloff 1996) where one of the pair is rare or endangered.

The aims of this study included comprehensive population-level sampling of *A. georgiana* in order to survey genetic diversity, detect population structure, determine if recent habitat loss and fragmentation have caused genetic bottlenecks, assess the extent of introgressive hybridization with *A. herbacea*, and consider how genetic information can inform conservation management of this species.

MATERIALS AND METHODS

Sampling and Genotyping

Tissue was collected and placed in silica gel from 145 individuals at ten different localities from seven populations throughout the species range

identified from herbarium records, locality data provided by the North Carolina Natural Heritage Program, and Sorrie's (1995) status survey (Figure 4.1; map created using M. Weinelt's Online Map Creation tool found at http://www.aquarius.ifm-geomar.de/make_map.html). The Georgia population was not found during the status survey, but was located by S. Straub and T. Patrick (GADNR) for this study. Permit restrictions, due to the endangered status of *A. georgiana* in North Carolina, limited the collection of individuals at any one locality to 20. At sites where fewer than 20 individuals occurred, tissue was collected from every individual located. The same collecting protocol was applied in South Carolina and Georgia. Due to the rhizomatous habit of these plants, care was taken not to remove tissue samples from stems growing close to one another, so as to sample genets rather than ramets (Miller 2004). Voucher specimens for each population were deposited at BH.

DNA was obtained using a standard CTAB DNA extraction protocol (Doyle & Doyle 1987) modified by adding 2% PVP-40 to the buffer. Individuals were genotyped for 12 unlinked microsatellite loci (Age01, Age02, Age06, Age07, Age10, Age13, Age15, Age18, Age24, Age25, Age29 and Age30) developed specifically for *A. georgiana* using the primers polymerase chain reaction (PCR) conditions, and general genotyping procedures described by Straub *et al.* (2009a).

Data Quality Control

PCR was attempted at least three times if the initial amplification failed. Samples exhibiting unclear banding patterns, rare alleles, or mismatched genotypes between two runs were re-genotyped at least once. In

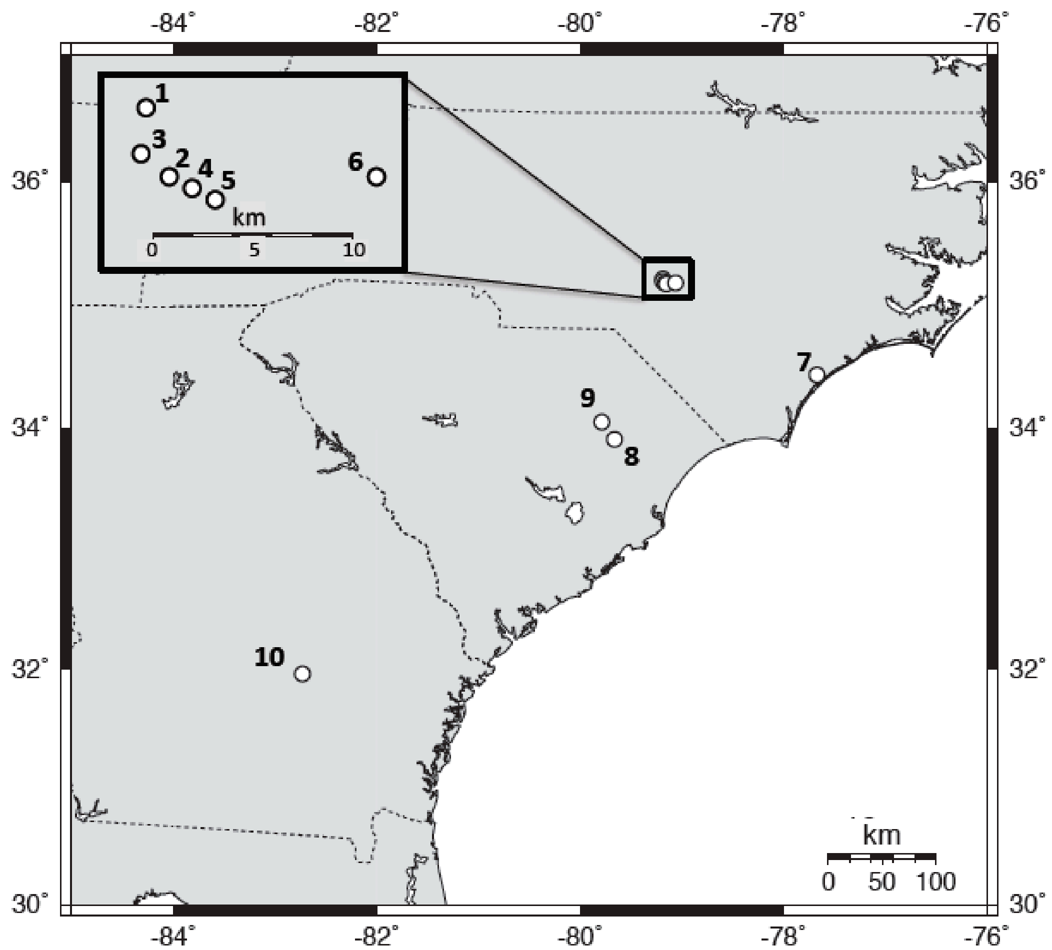


Figure 4.1. Map of the southeastern United States showing collecting localities for *Amorpha georgiana* in North Carolina (1 – Crain's Creek, 2 – Creek Bend, 3 – Fort Bragg 1, 4 – Fort Bragg 2, 5 – Fort Bragg 3, 6 – Fort Bragg 4, 7 – Holly Shelter), South Carolina (8 – Lynches River, 9 – County Park) and Georgia (10 – Georgia).

a few cases, multiple PCR and genotyping runs for a sample gave a heterozygous pattern in some instances and a homozygous pattern in others. These were regarded as examples of allelic dropout (nonamplification of one of the two alleles in a heterozygote) if the intermittently observed allele was found in other samples and these individuals were coded as heterozygotes; otherwise the allele was considered a false allele and the genotype was coded as homozygous. If a clear banding pattern could not be obtained or PCR repeatedly failed, the genotype was coded as missing data for that locus. Overall data quality was assessed by re-genotyping 10%of the individuals for every locus. Averaged across all loci the success rate was 96%. All instances of mismatches between the first and second runs occurred in two loci, Age18 and Age25, and were cases of allelic dropout.

Some samples gave clear and repeatable genotypes consisting of more than the two alleles expected for diploids. When this pattern was repeated over multiple loci, it is likely that these individuals are polyploids and that some *A. georgiana* populations are of mixed ploidy. Single putative polyploids were observed at Creek Bend and Holly Shelter; both were removed from the data set prior to analysis. For individuals that exhibited more than two alleles for only one locus, perhaps due to amplification of paralogues, the genotype for that locus alone was coded as missing data. All data for Age10 were eliminated because 14% of individuals had more than the two expected alleles, likely due consistent paralogue amplification, leaving 11 loci in the final data set.

DROPOUT 1.3 (McKelvey & Schwartz 2005) was used to identify loci for which mistakes may have been presenting the data set due to genotyping or human error. Each of the 11 loci was carefully rechecked and problems

(e.g., allele scoring or data entry) were corrected. DROPOUT 1.3 was also used to identify identical multilocus genotypes and pairs of genotypes differing at fewer than three loci in the data set. Identical multilocus genotypes were regarded as accidental sampling of ramets rather than genets. Pairs of individuals with highly similar multilocus genotypes were evaluated to determine if ramets had been sampled and one of the samples suffered from allelic dropout at a particular locus. In these cases, the sample with the homozygous genotype was assumed to have allelic dropout and was removed from the data set. Accidental sampling of ramets was detected in the Fort Bragg 2, Holly Shelter, Lynches River and Georgia sites, lowering the number of individuals included in the analyses for these localities (Table 4.1).

Genetic Diversity

HP-RARE 1.1 (Kalinowski 2005) was used to calculate the numbers of alleles observed and allelic richness corrected for sample size to allow comparisons between populations. The number of private alleles and private allelic richness were calculated for each locus in each sampling locality, and in hierarchically nested groups of sampling localities (Kalinowski 2004). ARLEQUIN 3.1(Excoffier *et al.* 2005) was used to calculate observed heterozygosity and gene diversity (H_O, H_E; Nei 1987). A one-way ANOVA, in combination with a Tukey-Kramer honestly significant difference (HSD) test ($\alpha = 0.05$), was conducted in JMP 7.0 (SAS Institute, Inc.) and used to compare sample means across loci for the basic population statistics in Table 4.1. ARLEQUIN 3.1 was used to test for significant departures from expected Hardy–Weinberg equilibrium proportions in each sampling locality and each

Table 4.1. Genetic variation in *A. georgiana* populations based on multilocus genotype data from eleven microsatellite loci. Variation observed for both individual collecting localities and groups of collecting localities is reported. The "All FB" group includes all North Carolina (NC) populations except Holly Shelter. Statistical significance of the differences among groups was tested among the individual localities and the larger groups separately. Superscript letters indicate significant differences among localities not linked by the same letter. Significance for differences in the number of private alleles is based on the mean number of private alleles per locus.

Localities	N_c	N_i	A	A_R	P	P_R	H_O	H_E	Loci Departing from HW
NC - Crain's Creek	8	8	6.5^{BC}	4.60^{A}	3^{BC}	1.92^{AB}	0.727^{A}	0.784^{A}	None
NC - Creek Bend	12	11	7.9^{AB}	4.85^{A}	3^{BC}	2.40^{A}	0.668^{A}	0.794^{A}	Age15, Age25
NC - Fort Bragg 1	10	10	5.9^{BC}	4.12^{AB}	4^{ABC}	1.84^{AB}	0.660^{A}	0.747^{A}	None
NC - Fort Bragg 2	20	19	8.6^{AB}	4.58^{A}	4^{ABC}	2.16^{A}	0.641^{A}	0.748^{A}	Age15
NC - Fort Bragg 3	20	20	9.9^{A}	4.79^{A}	4^{ABC}	2.37^{A}	0.697^{A}	0.803^{A}	Age06, Age25
NC - Fort Bragg 4	10	10	6.5^{BC}	4.29^{A}	4^{ABC}	1.82^{AB}	0.585^{A}	0.749^{A}	Age25
NC - Holly Shelter	20	16	7.1^{AB}	4.55^{A}	11^{AB}	2.32^{A}	0.691^{A}	0.797^{A}	None
SC - Lynches River	20	18	6.5^{BC}	4.02^{AB}	6^{ABC}	1.66^{AB}	0.565^{A}	0.747^{A}	Age01, Age15
SC - County Park	4	4	3.4^{C}	3.36^{B}	0^{C}	1.17^{B}	0.682^{A}	0.646^{A}	None
GA - Georgia	20	16	6.1^{BC}	3.87^{AB}	14^{A}	2.12^{AB}	0.571^{A}	0.678^{A}	None
All FB	80	78	14.7^{A}	4.54^{A}	48^{A}	2.08^{A}	0.665^{A}	0.804^{A}	Age01, Age06, Age15, Age25
All NC	100	94	15.7^{A}	4.54^{A}	79^{A}	2.20^{A}	0.669^{A}	0.814^{A}	Age01, Age06, Age15, Age25
All SC	24	22	7.0^{B}	3.69^{B}	6^{B}	1.42^{B}	0.586^{A}	0.749^{A}	Age01, Age15
All GA	20	16	6.1^{B}	3.87^{AB}	14^{B}	2.12^{AB}	0.571^{A}	0.678^{A}	None
All	144	132	17.7	4.30	N/A	N/A	0.649	0.749	N/A

N_c = number of individuals collected; N_i = number of individuals included in analyses; A = avg. number of alleles per locus; A_R = avg. allelic richness; P = number of private alleles; P_R = private allelic richness; H_O = avg. observed heterozygosity; H_E = avg. gene diversity; HW = expected Hardy-Weinberg proportions

geographic region using a test analogous to Fisher's exact test with a Markov chain of 1,000,000 iterations and 100,000 dememorization steps (Guo & Thompson 1992). A sequential Bonferroni correction was applied to determine significance following multiple tests ($\alpha = 0.05$; Rice 1989).

Population Structure

INSTRUCT (Gao *et al.* 2007) was used to determine the extent of population structure among *A. georgiana* populations. The underlying algorithms of INSTRUCT are based on those of the program STRUCTURE (Pritchard *et al.* 2000). Unlike STRUCTURE, INSTRUCT's estimation of population structure allows for partial selfing and less severe forms of inbreeding to be a source of allele correlation due to nonrandom mating, in addition to population substructure, by calculating the expected genotype frequencies using selfing rates rather than assuming Hardy–Weinberg equilibrium for each subpopulation (Gao *et al.* 2007). Use of STRUCTURE in cases where inbreeding occurs can lead to detection of false population structure due to violation of the assumptions of the method (Falush *et al.* 2003; Gao *et al.* 2007). Because mixed mating systems are common among plants (e.g., Goodwillie *et al.* 2005), INSTRUCT is the most appropriate choice of analysis program when the details of the mating system of the study species are unknown, as is the case with *A. georgiana*, and therefore could violate the total outcrossing assumption of STRUCTURE.

Population selfing rates and population structure were determined simultaneously in INSTRUCT implemented on computer clusters at Cornell University's Computational Biology Service Unit (CUCBSU; http://cbsuapps.tc.cornell.edu/InStruct.aspx). To determine a reasonable

range of K values for more detailed analyses, preliminary analyses were conducted for K = 1–10 using the following settings: Chain Number = 5; MCMC Iterations Number = 500,000; Burn-in = 200,000; Thinning = 20; Significance level for Posterior Credible Interval = 0.95. More thorough analyses were then conducted for K = 1–7 using the following settings: Chain Number = 10; MCMC Iterations Number =1,200,000; Burn-in = 500,000; Thinning = 30; Significance level for Posterior Credible Interval = 0.95. Convergence of chains for population structure and selfing rates was verified using the Gelman-Rubin statistics (Gelman & Rubin 1992) calculated by INSTRUCT.

Several different approaches for choosing K were employed and the results compared. These methods included the ΔK method (Evanno *et al.* 2005), the deviance information criterion (DIC; implemented in INSTRUCT), and the ad hoc ln Pr($X|K$) method (Pritchard *et al.* 2000). The output generated by INSTRUCT was visualized using the CUCBSU web interface for DISTRUCT (Rosenberg 2004; http://cbsusrv02.tc.cornell.edu/Distruct/distruct.aspx).

To explore further the magnitude of population structure among *A. georgiana* populations, a permutation analysis (10,000 permutations) was conducted in SPAGeDi 1.2 (Hardy & Vekemans 2002) to assess whether allele size was a significant factor in genetic differentiation (Hardy *et al.* 2003) and thereby to determine whether F_{ST} or R_{ST} is a more appropriate descriptor for the *A. georgiana* data set. Pairwise F_{ST} values were then calculated using ARLEQUIN 3.1 for all pairs of sampling localities and each geographic region. A permutation test (50,000 permutations) followed by a sequential Bonferroni correction for multiple tests was performed to

determine the significance of each value ($\alpha = 0.05$; Rice 1989). A locus-by-locus analysis of molecular variance (AMOVA; Excoffier *et al.* 1992) was conducted and a permutation test (15,000 permutations) done to determine the significance of each component in this hierarchical analysis of variance using ARLEQUIN 3.1.

Recent Demographic Events

BOTTLENECK version 1.2.02 (Piry *et al.* 1999) was used to test for the genetic signature of bottlenecks in the recent histories of *A. georgiana* populations. Three models of mutation were used: the stepwise mutation model (SMM; Ohta & Kimura 1973), the infinite allele model (IAM; Kimura & Crow 1964), and the two-phase model (TPM; Di Rienzo *et al.* 1994). Microsatellite loci with repeat motifs of 3–5 units are thought to follow a mutational model closer to the SMM, whereas dinucleotide repeats or loci with compound motifs are thought to follow a mutational model deviating from the SMM towards the IAM (Shriver *et al.* 1993; Estoup *et al.* 1995; Cornuet & Luikart 1996), and because most loci evolve according to a model somewhere between the two extremes, the TPM is most appropriate (Di Rienzo *et al.* 1994; Cornuet & Luikart 1996; Luikart & Cornuet 1998; Piry *et al.* 1999). For this study, approximately 70% of the loci fall into the first category and 30% into the second category (Straub *et al.* 2009a), consequently the TPM with 70% SMM character was employed.

A sensitivity analysis was conducted to determine which outcomes were robust to parameter permutation. Different variances in the number of steps allowed in multistep mutations were tested: 5, 10, 20, 30 and 40. The number of iterations used in the simulations was set at 1000, 2000 or 5000.

The model used was permuted by testing 100% SMM and 100% IAM and then by varying the percentage of SMM in the TPM between 50% and 90% in increments of 10%. The suggested settings of Piry *et al.* (1999) for microsatellite data, a step variance of 12% and 95% SMM in the TPM, were also tested with 2000 iterations. One-tailed Wilcoxon signed rank tests were used to determine if a significant excess or deficiency of heterozygosity existed for each collecting locality and each larger geographical group of localities suggested to be populations by INSTRUCT (Piry *et al.* 1999). Loci exhibiting significant deviations from Hardy–Weinberg proportions in specific populations and populations with sample sizes of fewer than 10 individuals were excluded prior to analysis (Cornuet & Luikart 1996).

Assessment of hybridization with *A. herbacea*

At Holly Shelter, tissue was collected on silica gel from several morphologically-intermediate putative hybrids, as well as 36 *A. herbacea* individuals. Putative hybrids were included with the *A. georgiana* Holly Shelter population for all analyses. Tissue was also collected from 20 individuals at another site approximately 60 km northwest of Holly Shelter where *A. herbacea* occurs in the absence of congeners. DNA was extracted as described above. *Amorpha herbacea* individuals were genotyped for the same loci as *A. georgiana* individuals with the exception of Age10, which was problematic in *A. georgiana*, Age13, which is duplicated in *A. herbacea*, and Age29, which is not readily amplifiable in this species (Straub *et al.* 2009a). Age25 proved difficult to amplify and genotype consistently and was also excluded from the analysis, bringing the total number of loci included for *A. herbacea* to eight. Data quality control and calculation of basic population genetics

measures for *A. herbacea* were as described above. Regenotyping success was 100%. No ramets were sampled. Comparisons of allelic diversity, allelic richness and heterozygosity between *A. herbacea* and *A. georgiana* were made using a t-Test ($\alpha = 0.05$) implemented in JMP 7.0. GENETIX 4.03 (Belkhir *et al.* 1996–2004) was used to perform a factorial correspondence analysis of the multilocus genotypes for *A. herbacea* and *A. georgiana* to assess the extent of hybridization between the two species.

RESULTS

Genetic Diversity

Overall, genetic diversity in *A. georgiana* was high with an average of 17.7 alleles per locus and average gene diversity of 0.75 (Table 4.1). Several populations had significantly higher allelic diversity than others, but after correction for sample size, all of the North Carolina populations, except Crain's Creek, had significantly higher allelic richness than the site with the least allelic diversity and richness, County Park. The remaining nine populations could not be distinguished statistically based on the allelic richness metric (Table 4.1).

Georgia and Holly Shelter were the most differentiated in terms of allele complement, each having over twice as many private alleles as the other localities (Table 4.1). When the average number of private alleles per locus was taken into account Georgia had significantly more private alleles per locus than Crain's Creek, Creek Bend and County Park and Holly Shelter had significantly more private alleles per locus than County Park. With sample size taken into account, several of the North Carolina populations, including Holly Shelter, had significantly more private allelic richness than

County Park; however the private allelic richness of Georgia was similar to all of the other populations.

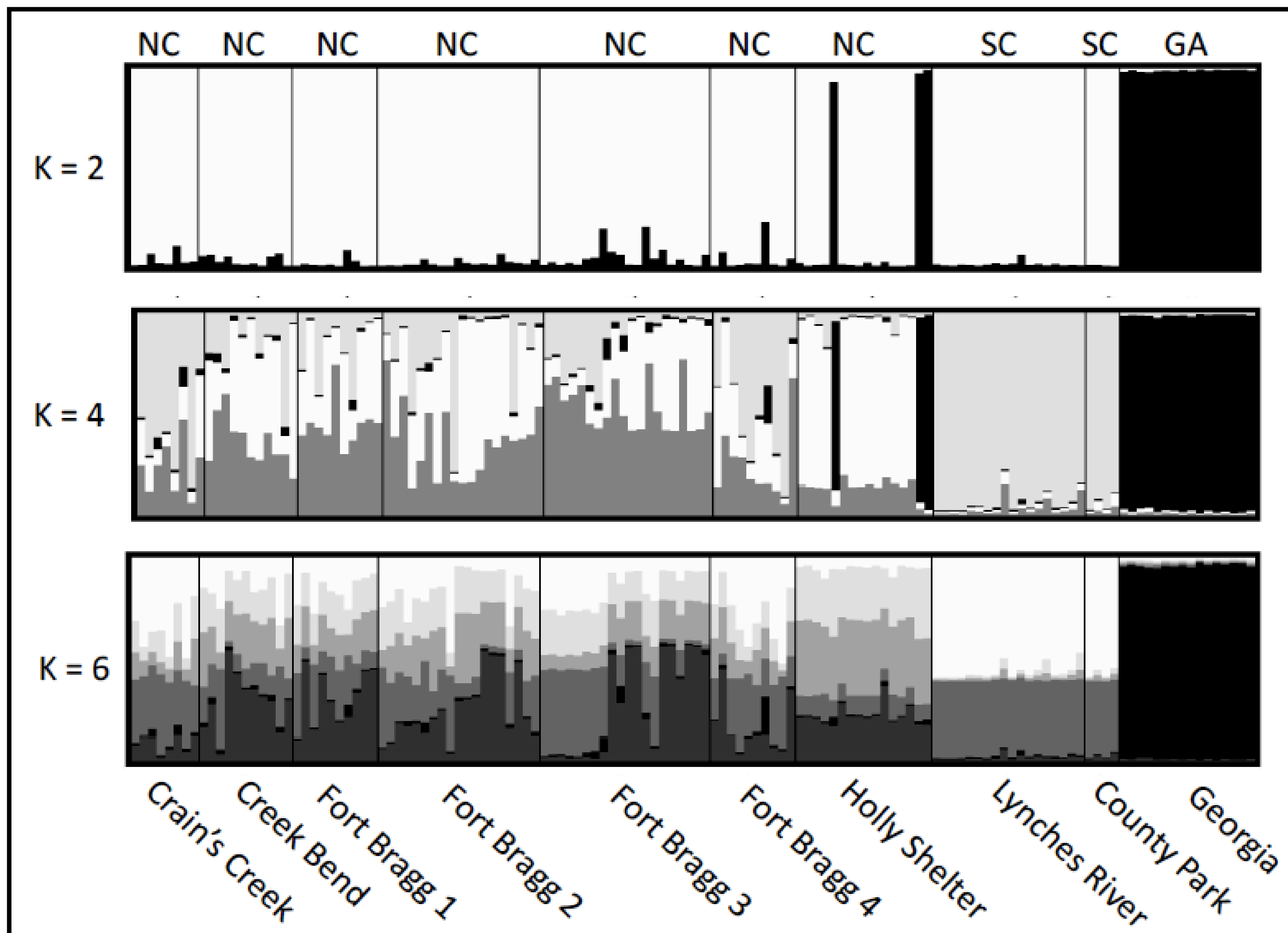
When the sampling localities were grouped by geographic region, North Carolina had significantly higher allelic diversity and average private alleles per locus than South Carolina or Georgia and significantly higher allelic richness and private allelic richness than South Carolina (Table 4.1). In this case, correction for sample size showed that genetic diversity and genetic differentiation, as assessed by private alleles, was similar in North Carolina and Georgia.

No significant differences were observed among populations or geographic regions for levels of observed and expected heterozygosity (Table 4.1). In all cases, except for the extremely small County Park population, the observed heterozygosity was less than the expected heterozygosity, as is commonly observed for microsatellites (Nybom 2004). Only in a few cases did the observed values deviate significantly from expected Hardy–Weinberg proportions (Table 4.1). Age15 and Age25 displayed significant deviations in 30% of the localities, indicating that these loci might have a more generalized underlying problem, such as null alleles or allelic dropout. Deviations for Age01 and Age06 were restricted to single populations and are likely due to processes acting at the local level, such as inbreeding.

Population Structure

All three methods used for determining the number of clusters present in the data set (K) resulted in different estimations of K for the output from INSTRUCT (Figure 4.2). The ΔK method of Evanno *et al.* (2005) indicated that K = 2. At K = 2 the highest level of population structure was

Figure 4.2. Population structure among *Amorpha georgiana* populations for one K = 2 (ΔK), K = 4 [ln Pr(*X*|*K*)], and K = 6 (DIC) chain. Each method of K-choice suggested a different value of K was best as indicated in parentheses after the K-value. Each cluster is shown in a different shade of grey. Each vertical bar represents an individual and the colours represent the proportion of its genotype assigned to each cluster. Broad geographical localities are listed across the top and specific localities are listed across the bottom of the graph.


NC
NC
NC
NC
NC
NC
NC
SC
SC
GA
K = 2
K = 4
K = 6
Crain's Creek
Creek Bend
Fort Bragg 1
Fort Bragg 2
Fort Bragg 3
Fort Bragg 4
Holly Shelter
Lynches River
County Park
Georgia

detected, as might be expected for this method of K choice (Evanno *et al.* 2005). Application of the ad hoc ln Pr($X|K$) method of Pritchard *et al.* (2000) resulted in choosing K = 4 as the best hypothesis. The DIC favoured the most complex model of population structure, K = 6. Recovery of each individual as admixed among most populations is indicative of choosing too large a value of K when the underlying biology does not support such a complex model of population structure (Pritchard *et al.* 2000, 2007). The ln Pr($X|K$) performed the best because it lead to the choice of the least complex model of population structure for *A. georgiana* that was biologically interpretable (unlike K = 6) and that did not lose information due to application of too simplistic a model (unlike K = 2). The general patterns of population structure apparent for K = 4 included a well-differentiated Georgia population and a homogeneous, but less well differentiated, group comprised of the two South Carolina populations. Individuals in the North Carolina populations had a mixture of assignments to multiple clusters, most with no clear majority assignment of the preponderance of individuals to a single cluster. The exception was Holly Shelter, where most individuals had a majority assignment of their genotypes to one cluster (white in Figure 4.2). Within the Holly Shelter population three individuals had majority assignment to the black cluster. These three individuals were identified as having intermediate morphology between *A. georgiana* and *A. herbacea* and their assignment to the most divergent cluster most likely reflects their putative hybrid ancestry. In addition, subsets of the data consisting of groups of the North Carolina individuals alone and the Fort Bragg individuals alone were run in separate INSTRUCT analyses to determine if other levels of population structure were detectable when the more divergent clusters were

removed (data not shown; see Coulon *et al*. 2008; Pritchard *et al*. 2007). These analyses did not allow further insight into the problem of admixture for most individuals or detection of finer patterns within the population structure hierarchy.

Selfing rates for each population detected by INSTRUCT were estimated simultaneously with population structure (Table 4.2). These values

Table 4.2. *Amorpha georgiana* selfing rates estimated for each cluster detected by INSTRUCT for K =4. See Figure 4.2 for corresponding grayscale shade designations.

INSTRUCT Population	Mean	Median	95% Credible Interval
Cluster 1 (dark gray)	0.191	0.186	0.045 - 0.374
Cluster 2 (white)	0.196	0.191	0.053 - 0.361
Cluster 3 (black)	0.263	0.256	0.087 - 0.473
Cluster 4 (light gray)	0.275	0.273	0.131 - 0.434

suggested a greater degree of selfing in the black and light grey clusters corresponding to the smaller and more isolated populations found in Georgia and South Carolina than in the other two, mainly North Carolina, clusters. However, the 95% credible intervals were overlapping and none of the selfing proportions were significantly different. The magnitude of the population structure detected in INSTRUCT was further evaluated using F_{ST} and AMOVA. A permutation analysis to determine whether allele size was a significant factor in population differentiation indicated that it was not because the observed R_{ST} value was not significantly different from the permuted pR_{ST} value (Hardy *et al.* 2003). Consequently, pairwise F_{ST} values, rather than R_{ST} values, were calculated (Tables 4.3 and 4.4). Only pairwise comparisons involving the Georgia population had sizeable F_{ST} values

Table 4.3. Pairwise F_{ST} values for *A. georgiana* collecting localities. Values in bold are significant based on permutation tests following a sequential Bonferroni correction.

	NC - Crain's Creek	**NC - Creek Bend**	**NC – Fort Bragg 1**	**NC – Fort Bragg 2**	**NC – Fort Bragg 3**	**NC – Fort Bragg 4**	**NC - Holly Shelter**	**SC - Lynches River**	**SC - County Park**	**Georgia**
NC - Crain's Creek	-	-	-	-	-	-	-	-	-	-
NC - Creek Bend	0.023	-	-	-	-	-	-	-	-	-
NC - Fort Bragg 1	**0.065**	0.038	-	-	-	-	-	-	-	-
NC - Fort Bragg 2	**0.032**	0.006	**0.073**	-	-	-	-	-	-	-
NC - Fort Bragg 3	**0.045**	0.024	0.034	**0.044**	-	-	-	-	-	-
NC - Fort Bragg 4	**0.054**	0.024	**0.080**	**0.051**	**0.066**	-	-	-	-	-
NC - Holly Shelter	**0.057**	**0.039**	**0.074**	**0.055**	**0.070**	**0.069**	-	-	-	-
NC - Lynches River	**0.050**	**0.061**	**0.087**	**0.073**	**0.070**	**0.081**	**0.092**	-	-	-
NC - County Park	0.074	**0.097**	**0.126**	**0.130**	**0.116**	**0.118**	**0.121**	0.083	-	-
Georgia	**0.138**	**0.132**	**0.185**	**0.139**	**0.128**	**0.165**	**0.156**	**0.173**	**0.263**	-

Table 4.4. Pairwise F_{ST} values for *A. georgiana* populations detected by INSTRUCT. All values are significant based on permutation tests following a sequential Bonferroni correction.

	NC	SC	GA
NC	-	-	-
SC	**0.048**	-	-
GA	**0.115**	**0.176**	-

ranging from 0.13 to 0.26, indicating differentiation from the populations in the Carolinas. Many of the other pairwise F_{ST} values were lower (<0.1), but significantly different from zero, generally supporting the results of the INSTRUCT analysis of stronger differentiation of the Georgia population and weaker differentiation among the populations in the Carolinas. The higher F_{ST} values observed for County Park were most likely artefacts of small sample size. The results of the AMOVA analysis further supported the extent of population structure found by INSTRUCT (Table 4.5).

Table 4.5. Results of AMOVA analysis for *A. georgiana* populations. All collecting localities were treated as populations and each group represents the collecting localities found in a particular geographic region (NC, SC, and GA). P-value indicates percentages of variation significantly different from zero.

Source of Variation	% of the Variation	P-value
Among groups	5.75	<0.00001
Among populations within groups	4.74	<0.00001
Within populations	89.51	<0.00001

Recent Demographic Events

The outcome of the bottleneck permutation analysis was affected by changing the model of molecular evolution and the allowed variance in step

size for multi-step mutations, but not by changing the number of iterations. When groups of individuals were analysed using collecting locality as populations, only Holly Shelter showed strong evidence for an excess of gene diversity, indicating a genetic bottleneck sometime in the last 0.2–4 N generations (Table 4.6). As might be expected under the most conservative scenarios, the only cases in which a significant excess of heterozygosity was not obtained were under the strict SMM and the TPM with 95% SMM character. A recent bottleneck was weakly supported for the Lynches River population under the strict IAM and TPM variants with the most IAM character. The genetic signature of a bottleneck was not detected or very weakly supported in the other localities. Interestingly, Fort Bragg 4 may have experienced population expansion because it displayed a weak signal for population growth by having significantly reduced gene diversity compared to equilibrium expectations under models with a high SMM character or with a low variance (Luikart & Cornuet 1998). Although Georgia had a significant excess and Fort Bragg 3 a significant deficiency of heterozygosity under the most extreme models, these results were not robust to the sensitivity analysis.

When populations were grouped into the clusters indicated by INSTRUCT, there was weak evidence that as a whole the North Carolina Fort Bragg area populations have recently experienced population expansion because this result was only recovered under the strict SMM model and closest TPM variant (Table 4.6). In addition, when both South Carolina populations were combined there was only weak evidence for a recent bottleneck overall in that area based on a significant result under the strict

Table 4.6. Table of *P*-values for one-tailed Wilcoxon signed rank tests for heterozygosity excess and deficiency for each *A. georgiana* population under the infinite allele model (IAM), stepwise mutation model (SMM), and two phase model (TPM) using BOTTLENECK 1.2.02. Significant P-values are shown in bold. Analyses were not valid for the Crain's Creek and County Park populations alone due to low sampling (n < 10), but these populations were included in the combined analyses for Fort Bragg, NC, and SC. The Fort Bragg Combined population included Crain's Creek, Creek Bend, and Fort Bragg 1-4.

Model	% SMM	Variance	Type of Deviation from Equilibrium	NC - Creek Bend	NC - Fort Bragg 1	NC - Fort Bragg 2	NC - Fort Bragg 3	NC - Fort Bragg 4	NC - Holly Shelter	SC - Lynches River	Georgia	Fort Bragg Combined	NC Combined	SC Combined
IAM	0	N/A	heterozygosity excess	0.285	0.062	0.080	0.326	0.652	**<0.001**	**0.002**	**0.011**	0.148	0.148	**0.007**
			heterozygosity deficiency	0.752	0.949	0.935	0.715	0.385	1.000	0.999	0.992	0.945	0.945	0.995
SMM	100	N/A	heterozygosity excess	0.875	0.449	0.812	0.986	0.999	0.260	0.787	0.650	0.988	1.000	0.715
			heterozygosity deficiency	0.150	0.584	0.216	**0.019**	**<0.001**	0.768	0.248	0.382	**0.020**	**0.004**	0.326
TPM	50	30	heterozygosity excess	0.674	0.160	0.216	0.715	0.920	**0.002**	**0.010**	0.062	0.406	0.469	0.125
			heterozygosity deficiency	0.367	0.861	0.813	0.326	0.097	0.999	0.993	0.949	0.656	0.594	0.898
TPM	60	30	heterozygosity excess	0.715	0.207	0.216	0.715	0.920	**0.002**	**0.014**	0.062	0.469	0.469	0.125
			heterozygosity deficiency	0.327	0.817	0.813	0.326	0.097	0.998	0.990	0.949	0.594	0.594	0.898
TPM	70	30	heterozygosity excess	0.715	0.232	0.278	0.752	0.947	**0.003**	0.082	0.062	0.656	0.469	0.213
			heterozygosity deficiency	0.326	0.793	0.754	0.285	0.065	0.998	0.936	0.949	0.406	0.594	0.820
TPM	80	30	heterozygosity excess	0.752	0.232	0.423	0.752	0.947	**0.008**	0.150	0.120	0.656	0.656	0.367
			heterozygosity deficiency	0.285	0.793	0.615	0.285	0.065	0.994	0.875	0.897	0.406	0.406	0.674
TPM	90	30	heterozygosity excess	0.787	0.289	0.423	0.875	0.993	**0.042**	0.326	0.289	0.852	0.945	0.410
			heterozygosity deficiency	0.248	0.740	0.615	0.150	**0.009**	0.966	0.715	0.740	0.188	0.148	0.633
TPM	70	40	heterozygosity excess	0.326	0.207	0.278	0.715	0.935	**0.003**	0.082	0.062	0.469	0.469	0.180
			heterozygosity deficiency	0.715	0.817	0.754	0.326	0.080	0.998	0.936	0.949	0.594	0.594	0.850
TPM	70	30	heterozygosity excess	0.715	0.232	0.278	0.752	0.947	**0.003**	0.082	0.062	0.656	0.469	0.213
			heterozygosity deficiency	0.326	0.793	0.754	0.285	0.065	0.998	0.936	0.949	0.406	0.594	0.820
TPM	70	20	heterozygosity excess	0.752	0.232	0.385	0.752	0.947	**0.003**	0.082	0.074	0.656	0.594	0.326
			heterozygosity deficiency	0.285	0.793	0.652	0.285	0.065	0.998	0.936	0.938	0.406	0.469	0.715
TPM	70	10	heterozygosity excess	0.752	0.232	0.385	0.787	0.947	**0.006**	0.150	0.139	0.656	0.711	0.367
			heterozygosity deficiency	0.285	0.793	0.652	0.248	0.065	0.995	0.875	0.880	0.406	0.344	0.674
TPM	70	5	heterozygosity excess	0.787	0.232	0.423	0.850	0.991	**0.026**	0.213	0.232	0.852	0.852	0.367
			heterozygosity deficiency	0.248	0.793	0.615	0.180	**0.012**	0.989	0.820	0.793	0.188	0.188	0.674

Table 4.6 (Continued)

Model	% SMM	Vari-ance	Type of Deviation from Equilibrium	NC - Creek Bend	NC - Fort Bragg 1	NC - Fort Bragg 2	NC - Fort Bragg 3	NC - Fort Bragg 4	NC - Holly Shelter	SC - Lynches River	Georgia	Fort Bragg Com-bined	NC Com-bined	SC Com-bined
TPM*	95*	12*	heterozygosity excess	0.850	0.350	0.577	0.976	0.995	0.160	0.590	0.449	0.973	0.996	0.500
			heterozygosity deficiency	0.180	0.681	0.461	0.064	**0.007**	0.861	0.455	0.584	**0.039**	**0.008**	0.545

*Suggested settings of Piry et al. 1999.

IAM. These results were not well supported under the range of more realistic models used for the sensitivity analysis.

Assessment of Hybridization with *A. herbacea*

Overall genetic diversity and loci departing from expected Hardy–Weinberg proportions were calculated for *A. herbacea* (Table 4.7). *Amorpha herbacea* had a similar number of alleles per locus and allelic richness as *A. georgiana* ($P > 0.05$). Observed heterozygosity ($P = 0.017$) and gene diversity ($P = 0.001$) were significantly lower in *A. herbacea* than in *A. georgiana*.

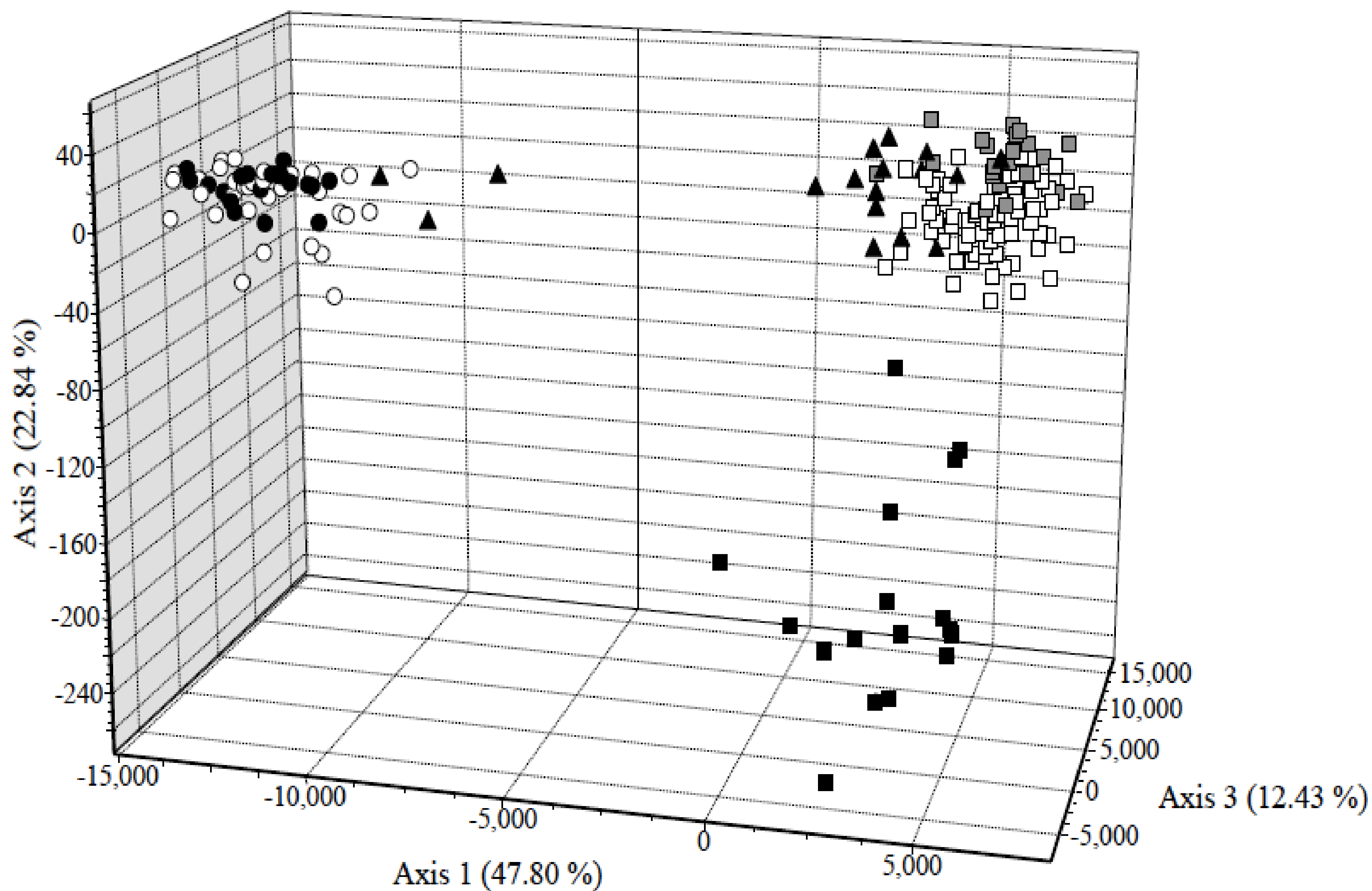
The factorial correspondence analysis showed that the three putative hybrid individuals identified in the INSTRUCT analysis did fall in a somewhat intermediate position between the two species, although much closer to *A. herbacea*, based on their multilocus genotypes (Figure 4.3). It is unlikely that these individuals are pure *A. herbacea* because each was found to possess two to three alleles diagnostic of *A. georgiana* in addition to two to three alleles diagnostic of *A. herbacea*. Age13 was duplicated in these individuals, as is expected for *A. herbacea*, and Age29, which does not amplify well in *A. herbacea*, only amplified in one of the three individuals. The single putative tetraploid individual identified at this site (see Methods) may also be of hybrid origin because it had two alleles diagnostic of *A. herbacea* and eleven alleles diagnostic of *A. georgiana*. Additionally, this individual had many alleles for Age13, suggesting more than two copies of this locus, and had only two alleles rather than the expected three or four for Age29 indicating amplification of only the *A. georgiana* locus. All of this evidence points to genetic contributions from both *Amorpha* species to this polyploid.

Table 4.7. Genetic variation found in *Amorpha herbacea* populations based on multilocus genotype data from eight microsatellite loci. Richness values were calculated to allow comparisons with values for *A. georgiana* in Table 1. Superscript letters indicate significant differences among localities not linked by the same letter. Significance for differences in the number of private alleles is based on the mean number of private alleles per locus.

Localities	N	A	A_R	P	P_R	H_O	H_E	Loci Departing from HW
NC - Holly Shelter	36	7.0^{A}	4.45^{A}	18^{A} (8)	1.56^{A}	0.523^{A}	0.590^{A}	Age01
NC - Other	20	6.4^{A}	4.53^{A}	13^{A} (7)	1.64^{A}	0.565^{A}	0.593^{A}	None
All	56	8.6	4.49	N/A	N/A	0.538	0.596	N/A

N = number of individuals; A = avg. number of alleles per locus; A_R = avg. allelic richness; P = number of private alleles (number of private alleles with respect to *A. georgiana*); P_R = private allelic richness; H_O = avg. observed heterozygosity; H_E = avg. gene diversity; HW = expected Hardy-Weinberg proportions

Figure 4.3. Factorial correspondence analysis of multilocus genotype data for *Amorpha georgiana* and *A. herbacea* for eight microsatellite loci. Each point represents one individual and each colour/shape combination represents a different population [*A. herbacea*: Holly Shelter (white circles), NC – other (black circles); *A. georgiana*: Fort Bragg (white squares), Holly Shelter (black triangles), South Carolina (grey squares), Georgia (black squares)]. The three Holly Shelter individuals nearest the *A. herbacea* cluster are the three putative hybrid individuals identified during the INSTRUCT analysis.


Axis 2 (22.84 %)
40
0
-40
-80
-120
-160
-200
-240
Axis 1 (47.80 %)
-15,000
-10,000
-5,000
0
5,000
Axis 3 (12.43 %)
15,000
10,000
5,000
0
-5,000

DISCUSSION

Species with small populations are at increased risk of extinction due to genetic considerations, such as genetic drift and inbreeding, as well as demographic and stochastic environmental factors. Genetic drift becomes amore powerful force in small populations and they suffer from other problems, such as reduced genetic diversity, increased inbreeding and genetic load, and diminished gene flow among subpopulations or populations (e.g., Barrett & Kohn 1991; Ellstrand & Elam 1993; Fenster & Dudash 1994; Young *et al.* 1996; Booy *et al.* 2000; Amos & Balmford 2001; Hedrick 2001; Frankham 2003). All of these issues decrease the ability of a species to respond to environmental challenges, such as climate change, new pathogens, or invasive species, thereby leading to increased extinction risk (e.g., Booy *et al.* 2000; Hedrick 2001; Frankham 2003; Spielman *et al.* 2004b). Consideration of genetics is essential for conservation planning and the results of this study will contribute to future conservation efforts for *A. georgiana*.

Genetic Diversity and Population Structure

There is no evidence that habitat fragmentation and loss has resulted in generally genetically depauperate populations when microsatellite variation is used as a proxy for overall genetic variation. The genetic diversity observed for *A. georgiana* was high overall and higher than might have been expected based on the scarcity of individuals and populations (Table 4.1). The average values for *A. georgiana* for observed and expected heterozygosity are higher than calculated averages for plant studies utilizing microsatellite data compiled from the literature ($H_O = 0.58 \pm 0.22$, $H_E = 0.61 \pm 0.21$;

Nybom 2004) and fall within one standard deviation of that mean indicating that *A. georgiana* harbours as much genetic diversity as the average plant species. However, this could still be less genetic diversity than was present in this species prior to habitat fragmentation.

When larger geographical units were designated as populations, North Carolina had significantly more allelic diversity and number of private alleles than Georgia or South Carolina. When sample size was taken into account, this pattern somewhat equalized the genetic diversity observed in North Carolina and Georgia, but South Carolina populations could still be considered to have lower genetic diversity than populations in the other geographical regions. This pattern could be said to be an artefact of sampling because, in general, small populations are known to harbour less genetic diversity than large populations (e.g., Honnay & Jacquemyn 2007). However, most populations other than those sampled have been destroyed. Consequently, this is the pattern of genetic diversity that will need to be considered in future conservation planning no matter what the historical distribution of diversity might have been.

The population structure analyses identified populations that were genetically distinct, in the case of the Georgia, or somewhat genetically differentiated, in the case of South Carolina, Fort Bragg area, and Holly Shelter (Figure 4.2, Tables 4.3 and 4.4). These largely separated based on geographic distance from one another. The recovery of multiple clusters as probable genetic source pools for nearly all individuals within the populations in and around Fort Bragg (Figure 4.2) could be due to the distribution of the individuals in small subpopulations along the river. Isolation by distance could limit gene flow among nonadjacent populations,

which could lead to inference of patterns by Bayesian clustering analysis that are difficult to interpret biologically (Figure 4.2; Pritchard *et al.* 2000, 2007).

Differentiation, as measured by the number of private alleles, is most apparent in populations on the edges of the current range of *A. georgiana*. Many of those in the Holly Shelter population are due to hybridization with *A. herbacea*, whereas those in Georgia have probably resulted from geographical isolation. The grouping of the putatively hybrid Holly Shelter individuals in the same cluster as the Georgia (GA) individuals in the INSTRUCT analysis probably partially stems from this phenomenon. These have the most private alleles and different allele frequencies, so they are placed together by the algorithm because they are the most dissimilar from everything else.

The life history traits of species have an effect on both genetic diversity and population structure (e.g., Hamrick & Godt 1996; Nybom 2004). Numerous long-lived perennials with outcrossing or mixed mating systems, like *A. georgiana*, have higher gene diversity and maintain more of their genetic variation within rather than between populations compared to plants that are annuals or short-lived perennials or primarily self-pollinators (e.g., Hamrick & Godt 1996; Nybom & Bartish 2000; Nybom 2004). Relatively recent habitat destruction compared to the longevity of these plants provides an explanation for high genetic variation in *A. georgiana* maintained within populations rather than among them (Tables 4.3, 4.4 and 4.5; Lowe *et al.* 2005). Individual genets of rhizomatous plants with similar life history traits as *A. georgiana* might survive for decades (Ehrlén & Lehtilä 2002). This is presumably an adaptation to frequent fire in healthy longleaf pine savannas, in addition to being one strategy that allows plant populations

to persist even when habitats are fragmented or degraded (Honnay & Bossuyt 2005). The longevity and presumably long generation times of the plants may have thus far prevented genetic drift and inbreeding from more quickly eroding genetic diversity and leading to population differentiation following habitat fragmentation (Young *et al.* 1996).

Taken together, the genetic diversity and population structure results suggest a possible history of the past distribution of *A. georgiana* populations. In light of the extensive habitat loss and degradation experienced by this species, perhaps *A. georgiana* is naturally a narrow endemic with small populations. Under this scenario larger F_{ST} values might be expected because of limited gene flow among scattered small populations, however, the F_{ST} values observed for *A. georgiana* (Tables 4.3 and 4.4) are below the averages observed for plants in general (0.26), long-lived perennials (0.19), and plants with outcrossing (0.22) or mixed mating (0.26) systems (Nybom 2004). Considering population structure, or lack thereof, as an indirect measure of gene flow, this pattern suggests that in the recent past *A. georgiana* was likely more continuously distributed in the middle and inner Coastal Plain from North Carolina south to Georgia. Although individual populations may have been small, gene flow, via pollen or by seeds perhaps dispersed by water, could have been maintained thereby slowing the development of population structure. Longleaf pine savanna has had a long history of being the dominant vegetation type in the southeastern United States, reaching most of its pre-settlement range by about 5,000–8,000 years ago (Delcourt & Delcourt 1981), making this scenario possible for numerous generations. Current habitat destruction will serve as a barrier to gene flow in the future and promote the development of further population structure.

Recent Demographic Events

The possibility of recent bottlenecks in *A. georgiana* was explored because of the reduction in size that populations may have suffered due to habitat loss. For populations that have experienced a recent bottleneck (<4 N_e generations ago), an excess of heterozygosity (H_E) compared to the expected mutation–drift equilibrium heterozygosity (H_{eq}) is expected for neutral loci because allelic diversity decreases more quickly than heterozygosity during a population contraction (Nei *et al.* 1975; Maruyama & Fuerst 1985; Allendorf 1986; Hedrick *et al.*1986; Cornuet & Luikart 1996; Luikart & Cornuet 1998). This study provides strong support for a recent bottleneck at Holly Shelter and weak support for a recent bottleneck at Lynches River based on significant excesses of gene diversity relative to mutation–drift equilibrium expectations (Table 4.6). If there had been bottlenecks in the other populations, the method of Cornuet & Luikart (1996) should have been able to detect it based on the species' demographic characteristics (Williamson-Natesan 2005).

A possible explanation for the weak support for a bottleneck at Lynches River and little evidence of bottlenecks for the other populations is that limited sampling (<30 individuals per population genotyped for 11 loci) did not provide enough statistical power to detect bottlenecks (Cornuet & Luikart 1996; Piry *et al.* 1999), which could be tested by increased sampling of loci. Alternatively, if the natural condition of *A. georgiana* populations is to be small, then it is possible that recent habitat loss has not caused genetic bottlenecks and most of the sampling localities were in fact at mutation–drift equilibrium, thus explaining the lack of excess gene diversity. Many of the remaining populations occur in areas that have been protected from human

development and are managed in an attempt to maintain natural ecosystem functions. This is especially true for the Fort Bragg Military Reservation where periodic controlled burns are common practice (Miller 2004). Maintenance of ecosystem processes and protection from most anthropogenic disturbances could contribute to maintaining equilibrium in most populations and the deficiency of heterozygosity in some populations at Fort Bragg (e.g., Fort Bragg 4) indicating possible recent population expansions (Table 4.6; Luikart & Cornuet 1998). Although, it is important to note that this possible expansion could be due to events that pre-date anthropogenic influence in the area, not necessarily recent management practices. There is also still the possibility that the choice of model may have led to recovery of a false signal of heterozygosity deficiency due to the drawbacks of using strict models or to violation of assumptions of the analysis, including the presence of underlying population substructure when collecting localities were pooled into larger geographic areas and the inclusion of localities with significant deviations from equilibrium expectations in the larger geographic populations (Cornuet & Luikart 1996).

Interestingly, Holly Shelter, the single population displaying the signature of a recent bottleneck event, occurs in a geographically disjunct locality and in atypical habitat for this species. *Amorpha georgiana* is most commonly encountered on sandy river terraces in longleaf pine savanna in the middle and inner Coastal Plain, but Holly Shelter, although in longleaf pine savanna habitat, is not near any rivers or creeks and occurs in the outer Coastal Plain. However, it does occupy a swale where seasonally inundated pocosin habitat meets the drier savanna, and therefore inhabits an area with a moisture regime similar to that of river terrace populations. This

population was not discovered until 1995, so *A. georgiana* has been known from this locality for a short time. It is possible that this population displays the genetic signature of a recent bottleneck due to a founder effect because it was recently established through natural or anthropogenic means. If so, this was most likely from central North Carolina because this is the group with which it is clustered genetically (Figure 4.2). However, hybridization between *A. georgiana* and *A. herbacea* at Holly Shelter followed by introgression of *A. herbacea* alleles should erase the signature of a recent bottleneck by introducing more rare alleles into the gene pool, while leaving gene diversity about the same.

Assessment of Hybridization with *A. herbacea*

This study has provided evidence that *A. georgiana* hybridizes with *A. herbacea* and has produced allopolyploids. The hybrid individuals observed at Holly Shelter are likely not F1 hybrids, but rather backcrosses to the more abundant species, *A. herbacea* (Figure 4.3). The risks posed to the persistence of *A. georgiana* overall by hybridization with and introgression from *A. herbacea* are probably minimal from a conservation perspective for several reasons. The two species only hybridize at this single geographically isolated location where only approximately 40 *A. georgiana* individuals are found. Perhaps more importantly, hybridization appears to be rare at the site and introgression of *A. herbacea* alleles into *A. georgiana* limited (Figure 4.3). However, there is still a risk that this population could eventually be genetically overwhelmed due to the rarity of *A. georgiana* and the frequency of *A. herbacea* at the site if hybridization were to become more extensive. Habitat differentiation is important in establishment of a stable hybrid zone

(Wolf *et al.* 2001) and *A. georgiana* and *A. herbacea* have been observed to occur in different areas of the savanna according to moisture regime, with the hybrids being found in intermediate habitat. This difference in habitat requirements may protect *A. georgiana* from expansion of hybrids into its preferred habitat, though if hybrid seed is frequently produced by *A. georgiana* individuals, the growth potential of the Holly Shelter population may be limited (Levin *et al.* 1996).

Implications for Conservation and Management of *A. georgiana*

When planning conservation efforts for any endangered species, in addition to ecological and demographic issues, genetics plays a key role. Maintenance of allelic diversity, heterozygosity, and the overall evolutionary potential of the species, while avoiding inbreeding depression, founder effects and outbreeding depression is fundamental to well-designed conservation and restoration programs (Allendorf 1986; Fenster & Dudash1994; Hufford & Mazer 2003; Allendorf & Luikart 2007). Of the two most pressing problems for species with low census sizes, inbreeding and loss of genetic diversity, inbreeding leading to loss of heterozygosity is a more immediate concern for *A. georgiana* because it shows effects before loss of genetic diversity (Booy *et al.* 2000; Amos & Balmford 2001), which is not yet apparent in this species. Inbreeding may have already contributed to the excess of homozygotes and departure from expected Hardy–Weinberg proportions observed in several populations (Table 4.1). Most populations have census sizes less than 50 individuals and several have less than 20 individuals, probably rendering mating between close relatives unavoidable. Evidence from this study indicates that approximately 20% of matings may

involve selfing or less severe forms of inbreeding, such as mating between siblings or family groups occurring in close proximity (Table 4.2). If small population sizes have existed in the long term, then genetic load might also be a concern because inbreeding will cause deleterious alleles to occur in the homozygous condition more often, reducing individual and population fitness (Hedrick2001). Another consideration is that the effective population size for *A. georgiana* is certainly lower than the census size, as is the case for most species (e.g., Frankham 1995). Fires can eliminate reproduction in any given year, but plants will generally flower within several years following a burn (Miller 2004). However, some populations do not reproduce in some years, regardless of when the last burn occurred, nor do very small plants often flower.

Population structure has developed among the remaining *A. georgiana* populations, as evidenced by the clear differentiation of Georgia, and weaker differentiation of the South Carolina populations and Holly Shelter, from the Fort Bragg area. Each of these four areas could be considered a separate management unit based on genetic differentiation, different population sizes, presence of private alleles and other factors, including hybridization. Outside of the genetically diverse, relatively abundantly populated, and well managed Fort Bragg area, the Georgia population deserves special conservation consideration because it has numerous private alleles, is geographically isolated, and has historical value as the type locality for the species. Focused effort on conservation of this population is needed because its habitat has been severely degraded through mechanical site preparation for tree farming and the overgrowth of brush due to prolonged fire suppression. Furthermore, the census size of 30 individuals is low in stark contrast to

Wilbur's (1954) description of the plants at this site as abundant. The recent reclassification of this species as endangered in Georgia, in combination with the evidence for genetic distinctiveness presented here, should focus more attention on the conservation of this species and the longleaf pine ecosystem to which it belongs.

In general, ecosystem level conservation and restoration approaches for longleaf pine savanna (e.g., Johnson & Gjerstad 2006; Walker & Silletti 2006) and *in situ* management of the remaining *A. georgiana* will likely be most effective. A key component for restoration and management would be to return regular fires to the ecosystem. Fire is especially important because it may be a vital factor in triggering reproduction in *A. georgiana* (Miller 2004) and ensuring adequate recruitment. Current *in situ* conservation through good management practices, including habitat protection and regular controlled burns, has been achieved at Fort Bragg Military Reservation and Holly Shelter Game Land, but the outlook for most other *A. georgiana* populations is poor due to habitat degradation and their location on private or unmanaged land. At Fort Bragg, burns occur at least once every 10 years, often more frequently (Miller 2004), to simulate the natural fire regime of burns every 1–3 years (Frost 2006). Management at Fort Bragg has allowed the preservation of the largest, most genetically diverse population of *A. georgiana* to persist, preserving an indispensable source that may be the key to future conservation and restoration efforts.

It is essential that an *ex situ* collection of plants and a seed bank be developed for *A. georgiana* following recognized plant conservation strategies (e.g., Brown & Briggs 1991; Center for Plant Conservation 1991; Eberhart *et al.* 1991). Because most of the remaining individuals occur in a relatively

small geographic area, they are vulnerable to stochastic environmental factors, and protected reserves of individuals and seed in other locations would provide insurance for species persistence and provide a source for population augmentation. The majority of the genetic variation and amplitude of ecological variation and adaptation present in the species could be maintained by vegetative propagation and seed banking using individuals from the Fort Bragg area and Georgia (Brown & Briggs 1991; Allendorf & Luikart 2007). Material from the Georgia population for *ex situ* populations and seed banks should be collected over several years and so that removal of propagules would not negatively impact the probability of persistence of that population (Center for Plant Conservation 1991). These measures may be required to ensure successful reproduction in the Georgia population, followed by supplementation of population size through breeding program. Care should be taken not to cross the North Carolina stock and the Georgia stock until it is known whether outbreeding depression could pose a problem (e.g., Barrett & Kohn 1991; Ellstrand & Elam 1993; Fenster & Dudash 1994; Booy *et al.* 2000; Hufford & Mazer 2003).

Future genetic work should focus on further exploring inbreeding and its relationship to population persistence in *A. georgiana*. The effects of increased inbreeding are more easily detected in progeny, rather than through surveys of the established reproductive individuals, as was done in this study, therefore analysis of progeny arrays will be essential for understanding the extent of inbreeding and its relationship to population size and other factors for this species (Lowe *et al.* 2005; Aguilar *et al.* 2008). In combination with further genetic studies, demographic and ecological studies are needed to learn more about recruitment, pollination biology and the

effects of fire on reproduction in *A. georgiana*. In conclusion, if steps are taken to preserve remaining populations and their genetic diversity, implement good management practices for *in situ* conservation, and develop *ex situ* resources, *A. georgiana* should be protected from extinction and maintain evolutionary potential as a species.

REFERENCES

Aguilar R, Quesada M, Ashworth L, Herrerias-Diego Y, Lobo J (2008) Genetic consequences of habitat fragmentation in plant populations: Susceptible signals in plant traits and methodological approaches. *Molecular Ecology*, **17**, 5177-5188.

Allendorf FW (1986) Genetic drift and the loss of alleles versus heterozygosity. *Zoo Biology*, **5**, 181-190.

Allendorf FW, Leary RF, Spruell P, Wenburg JK (2001) The problems with hybrids: Setting conservation guidelines. *Trends in Ecology and Evolution*, **16**, 613-622.

Allendorf FW, Luikart G (2007) *Conservation and the Genetics of Populations.* Blackwell Publishing Ltd., Malden, Massachusetts.

Amos W, Balmford A (2001) When does conservation genetics matter? *Heredity*, **87**, 257-265.

Barilani M, Bernard-Laurent A, Mucci N, *et al.* (2007) Hybridisation with introduced chukars (*Alectoris chukar*) threatens the gene pool integrity of native rock (*A. graeca*) and red-legged (*A. rufa*) partridge populations. *Biological Conservation*, **137**, 57-69.

Barrett SCH, Kohn JR (1991) Genetic and evolutionary consequences of small population size in plants: Implications for conservation. In: *Genetics and Conservation of Rare Plants* (eds. Falk DA, Holsinger KE), pp. 3-30, Oxford University Press, New York, New York.

Belkhir K., Borsa P, Chikhi L, Raufaste N, Bonhomme F (1996–2004) *GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations*. Laboratoire Génome, Populations, Interactions, CNRS UMR 5171, Université de Montpellier II, Montpellier (France). Available at: http://www.genetix.univ-montp2.fr/genetix/constr.htm#download. Accessed 18 February 2009.

Booy G, Hendriks RJJ, Smulders MJM, Van Groenendael JM, Vosman B (2000) Genetic diversity and the survival of populations. *Plant Biology*, **2**, 379-395.

Brown AHD, Briggs JD (1991) Sampling strategies for genetic variation in *ex situ* collections of endangered plant species. In: *Genetics and Conservation of Rare Plants* (eds. Falk DA, Holsinger KE), pp. 99-119, Oxford University Press, New York, New York.

Burgess KS, Morgan M, Deverno L, Husband BC (2005) Asymmetrical introgression between two *Morus* species (*M. alba*, *M. rubra*) that differ in abundance. *Molecular Ecology*, **14**, 3471-3483.

Carlson ML, Meinke RJ (2008) Hybridization between a rare and introduced *Oenothera* along the north Pacific Coast. *Western North American Naturalist*, **68**, 161-172.

Center for Plant Conservation (1991) Genetic sampling guidelines for conservation collections of endangered plants. In: *Genetics and Conservation of Rare Plants* (eds. Falk DA, Holsinger KE), pp. 225-238, Oxford University Press, New York, New York.

Cornuet JM, Luikart G (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics,* **144**, 2001-2014.

Coulon A, Fitzpatrick JW, Bowman R, *et al.* (2008) Congruent population structure inferred from dispersal behaviour and intensive genetic surveys of the threatened Florida scrub-jay (*Aphelocoma coerulescens*). *Molecular Ecology*, **17**, 1685-1701.

Delcourt PA, Delcourt HR (1981) Vegetation maps for Eastern North America: 40,000 Yr. B.P. to the present. In: *Geobotany II* (ed. Romans RC), pp. 123-165, Plenum, New York, New York.

Di Rienzo A, Peterson AC, Garza JC, Valdes AM, Slatkin M, Freimer NB (1994) Mutational processes of simple-sequence repeat loci in human-populations. *Proceedings of the National Academy of Sciences of the United States of America*, **91**, 3166-3170.

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*, **19**, 11-15.

Eberhart SA, Roos EE, Towill LE (1991) Strategies for long-term management of germplasm collections. In: *Genetics and Conservation of*

Rare Plants (eds. Falk DA, Holsinger KE), pp. 135-145, Oxford University Press, New York, New York.

Ehrlén J, Lehtilä K (2002) How perennial are perennial plants? *Oikos*, **98**, 308-322.

Ellstrand NC, Elam DR (1993) Population genetic consequences of small population-size - Implications for plant conservation. *Annual Review of Ecology and Systematics*, **24**, 217-242.

Estoup A, Garnery L, Solignac M, Cornuet J (1995) Microsatellite variation in honey bee (*Apis mellifera* L.) populations: Hierarchical genetic structure and test of the infinite allele and stepwise mutation models. *Genetics*, **140**, 679-695.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611-2620.

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes application to human mitochondrial DNA restriction data. *Genetics*, **131**, 479-491.

Excoffier L, Laval G, Schneider S (2005) ARLEQUIN ver. 3.0: an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, **1**, 47-50.

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics*, **164**, 1567-1587.

Fenster CB, Dudash MR (1994) Genetic considerations for plant population restoration and conservation. In: *Restoration of Endangered Species Conceptual Issues, Planning, and Implementation* (eds. Bowles ML, Whelan CJ), pp. 34-62, Cambridge University Press, Cambridge.

Frankham R (1995) Effective population size/adult population size ratios in wildlife: a review. *Genetical Research*, **66**, 95-107.

Frankham R (2003) Genetics and conservation biology. *Comptes Rendus Biologies*, **326** (Suppl 1), S22-S29.

Frankham R, Ralls K (1998) Conservation biology - Inbreeding leads to extinction. *Nature*, **392**, 441-442.

Frankham R, Ballou DJ, Briscoe DA (2002) *Introduction to Conservation Genetics.* Cambridge University Press, Cambridge.

Frost C (2006) History and future of the longleaf pine ecosystem. In: *The Longleaf Pine Ecosystem Ecology, Silviculture, and Restoration* (eds. Jose S, Jokela EJ, Miller DL), pp. 9-42, Springer Science+Business Media, New York, New York.

Gao H, Williamson S, Bustamante CD (2007) A Markov chain Monte Carlo approach for joint inference of population structure and inbreeding rates from multilocus genotype data. *Genetics*, **176**, 1635-1651.

Gelman A, Rubin DB (1992) Inference from iterative simulation using multiple sequences. *Statistical Science*, **7**, 457-511.

Georgia Department of Natural Resources (2007) *Special concern plant species in Georgia.* [Online Data], GADNR, Atlanta, Georgia, Available at: http://georgiawildlife.dnr.state.ga.us/content/specialconcernplants.asp. Accessed 3 June 2009.

Godt MJW, Hamrick JL (1998) Low allozyme diversity in *Schwalbea americana* (Scrophulariaceae), an endangered plant species. *Journal of Heredity*, **89**, 89-93.

Goodwillie C, Kalisz S, Eckert CG (2005) The evolutionary enigma of mixed mating systems in plants: Occurrence, theoretical explanations, and empirical evidence. *Annual Review of Ecology Evolution and Systematics*, **36**, 47-79.

Guo SW, Thompson EA (1992) Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics*, **48**, 361-372.

Hamrick JL, Godt MJW (1996) Effects of life history traits on genetic diversity in plant species. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, **351**, 1291-1298.

Hardy OJ, Vekemans X (2002) SPAGeDi: A versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes*, **2**, 618-620.

Hardy OJ, Charbonnel N, Freville H, Heuertz M (2003) Microsatellite allele sizes: A simple test to assess their significance on genetic differentiation. *Genetics*, **163**, 1467-1482.

Harrison RG (1990) Hybrid zones windows on evolutionary process. In: *Oxford Surveys in Evolutionary Biology,* Vol.7 (eds. Futuyma D, Antonovics J.), pp. 69-128, Oxford University Press, Oxford.

Hedrick PW (2001) Conservation genetics: where are we now? *Trends in Ecology & Evolution*, **16**, 629-636.

Hedrick PW, Brussard PF, Allendorf FW, Beardmore JA, Orzack S (1986) Protein variation, fitness, and captive propagation. *Zoo Biology*, **5**, 91-99.

Helenurm K (2003) Genetic diversity in the rare, insular endemic *Sibara filifolia* (Brassicaceae). *Madroño*, **50**, 181-186.

Honnay O, Bossuyt B (2005) Prolonged clonal growth: Escape route or route to extinction? *Oikos*, **108**, 427-432.

Honnay O, Jacquemyn H (2007) Susceptibility of common and rare plant species to the genetic consequences of habitat fragmentation. *Conservation Biology*, **21**, 823-831.

Hufford KM, Mazer SJ (2003) Plant ecotypes: Genetic differentiation in the age of ecological restoration. *Trends in Ecology & Evolution*, **18**, 147-155.

Isely D (1998) *Native and Naturalized Leguminosae (Fabaceae) of the United States.* Brigham Young University, Provo, Utah.

Johnson R, Gjerstad D (2006) Restoring the overstory of longleaf pine ecosystems. In: *The Longleaf Pine Ecosystem Ecology, Silviculture, and Restoration* (eds. Jose S, Jokela EJ, Miller DL), pp. 271-295, Springer Science+Business Media, New York, New York.

Jose S, Jokela EJ, Miller DL (2006) The longleaf pine ecosystem. In: *The Longleaf Pine Ecosystem Ecology, Siliviculture, and Restoration* (eds. Jose S,

Jokela EJ, Miller DL), pp. 3-8, Springer Science+Business Media, New York, New York.

Kalinowski ST (2004) Counting alleles with rarefaction: Private alleles and hierarchical sampling designs. *Conservation Genetics*, **5**, 539-543.

Kalinowski ST (2005) HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic richness. *Molecular Ecology Notes*, **5**, 187-189.

Kimura M, Crow JF (1964) The number of alleles that can be maintained in a finite population. *Genetics*, **49**, 725-738.

Lázaro A, Traveset A (2006) Reproductive success of the endangered shrub *Buxus balearica* Lam. (Buxaceae): pollen limitation, and inbreeding and outbreeding depression. *Plant Systematics and Evolution*, **261**, 117-128.

Levin DA (2002) Hybridization and extinction. *American Scientist*, **90**, 254-260.

Levin DA, Francisco-Ortega J, Jansen RK (1996) Hybridization and the extinction of rare plant species. *Conservation Biology*, **10**, 10-16.

Lowe AJ, Boshier D, Ward M, Bacles CFE, Navarro C (2005) Genetic resource impacts of habitat loss and degradation; reconciling empirical evidence and predicted theory for neotropical trees. *Heredity*, **95**, 255-273.

Luikart G, Cornuet JM (1998) Empirical evaluation of a test for identifying recently bottlenecked populations from allele frequency data. *Conservation Biology*, **12**, 228-237.

Maruyama T, Fuerst PA (1985) Population bottlenecks and nonequilibrium models in population genetics. II. Number of alleles in a small population that was formed by a recent bottleneck. *Genetics*, **111**, 675-689.

McKelvey KS, Schwartz MK (2005) DROPOUT: a program to identify problem loci and samples for noninvasive genetic samples in a capture-mark-recapture framework. *Molecular Ecology Notes*, **5**, 716-718.

Means DB (2006) Vertebrate faunal diversity of longleaf pine ecosystems. In: *The Longleaf Pine Ecosystem Ecology, Silviculture, and Restoration* (eds. Jose S, Jokela EJ, Miller DL), pp. 157-213, Springer Science+Business Media, New York, New York.

Miller VM (2004) *Habitat characterization of Amorpha georgiana var. georgiana groups at Fort Bragg, North Carolina*. Masters Thesis, North Carolina State University, Raleigh, North Carolina.

Muñoz-Fuentes V, Vila C, Green AJ, Negro JJ, Sorenson MD (2007) Hybridization between white-headed ducks and introduced ruddy ducks in Spain. *Molecular Ecology*, **16**, 629-638.

Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York, New York.

Nei M, Maruyama T, Chakraborty R (1975) The bottleneck effect and genetic variability in populations. *Evolution*, **29**, 1-10.

North Carolina Department of Agriculture & Consumer Services (2008) *Plant Industry Division — Plant Protection Section — Plant Conservation Program: Protected Plant List* [Online Data], NCDA & CS, Raleigh, North Carolina, Available at: http://www.ncagr.com/plantindustry/plant/plantconserve/plist.htm. Accessed 3 June 2009.

Noss RF (1989) Longleaf pine and wiregrass: Keystone components of an endangered ecosystem. *Natural Areas Journal*, **9**, 211-213.

Nybom H (2004) Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. *Molecular Ecology*, **13**, 1143-1155.

Nybom H, Bartish IV (2000) Effects of life history traits and sampling strategies on genetic diversity estimates obtained with RAPD markers in plants. *Perspectives in Plant Ecology Evolution and Systematics*, **3**, 93-114.

Ohta T, Kimura M (1973) Model of mutation appropriate to estimate number of electrophoretically detectable alleles in a finite population. *Genetical Research*, **22**, 201-204.

Outcalt KW (2000) The longleaf pine ecosystem of the South. *Native Plants Journal*, **1**, 42-51.

Outcalt KW, Sheffield RM (1996) *The longleaf pine forest: Trends and current conditions. Resourse Bulletin SRS-9.* U.S. Department of Agriculture, Forest Service, Southern Research Station, Asheville, North Carolina.

Peakall R, Ebert D, Scott LJ, Meagher PF, Offord CA (2003) Comparative genetic study confirms exceptionally low genetic variation in the ancient and endangered relictual conifer, *Wollemia nobilis* (Araucariaceae). *Molecular Ecology*, **12**, 2331-2343.

Piry S, Luikart G, Cornuet JM (1999) BOTTLENECK: A computer program for detecting recent reductions in the effective population size using allele frequency data. *Journal of Heredity*, **90**, 502-503.

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

Pritchard JK, Wen X, Falush D (2007) *Documentation for structure software: Version 2.2.* [online data], Available at: http://pritch.bsd.uchicago.edu/structure.html. Accessed 3 June 2009.

Rhymer JM, Simberloff D (1996) Extinction by hybridization and introgression. *Annual Review of Ecology and Systematics*, **27**, 83-109.

Rice WR (1989) Analyzing tables of statistical tests. *Evolution*, **43**, 223-225.

Rieseberg LH, Wendel JF (1993) Introgression and its consequences in plants. In: *Hybrid Zones and the Evolutionary Process* (ed. Harrison RG), pp. 70-109, Oxford University Press, New York, New York.

Rosenberg NA (2004) DISTRUCT: A program for the graphical display of population structure. *Molecular Ecology Notes*, **4**, 137-138.

Saccheri I, Kuussaari M, Kankare M, Vikman P, Fortelius W, Hanski I (1998) Inbreeding and extinction in a butterfly metapopulation. *Nature*, **392**, 491-494.

Shriver MD, Jin L, Chakraborty R, Boerwinkle E (1993) VNTR allele frequency distributions under the stepwise mutation model: A computer simulation approach. *Genetics*, **134**, 983-993.

Sorrie BA (1995) *Status Survey of Amorpha georgiana var. georgiana*. US Fish and Wildlife Service, Office of Endangered Species, Asheville, North Carolina, and North Carolina Natural Heritage Program, Division of Parks and Recreation, Raleigh, North Carolina.

Sorrie BA, Weakley AS (2006) Conservation of the endangered *Pinus palustris* ecosystem based on Coastal Plain centres of plant endemism. *Applied Vegetation Science*, **9**, 59-66.

South Carolina Department of Natural Resources (2006) *South Carolina Rare, Threatened, And Endangered Species Inventory*. [online data], SCDNR, Columbia, South Carolina, Available at: https://www.dnr.sc.gov/pls/heritage/county_species.list?pcounty=all. Accessed 3 June 2009.

Spielman D, Brook BW, Frankham R (2004a) Most species are not driven to extinction before genetic factors impact them. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 15261-15264.

Spielman D, Brook BW, Briscoe DA, Frankham R (2004b) Does inbreeding and loss of genetic diversity decrease disease resistance? *Conservation Genetics*, **5**, 439-448.

Straub SCK, Bogdanowicz SM, Doyle JJ (2009a) Characterization of twelve polymorphic microsatellite markers for Georgia false indigo (*Amorpha georgiana* Wilbur var. *georgiana*), an endangered species, and their utility in other dwarf *Amorpha* L. species. *Molecular Ecology Resources*, **9**, 225-228.

Straub SCK, Sorrie BA, Weakley AS (2009b) *Amorpha confusa*, a new name for an old *Amorpha* (Fabaceae: Amorpheae). *Journal of the Botanical Research Institute of Texas*, **3**, 151-155.

Travis SE, Maschinski J, Keim P (1996) An analysis of genetic variation in *Astragalus cremnophylax* var. *cremnophylax*, a critically endangered plant, using AFLP markers. *Molecular Ecology*, **5**, 735-745.

Walker JL, Silletti AM (2006) Restoring the ground layer of longleaf pine ecosystems. In: *The Longleaf Pine Ecosystem Ecology, Silviculture, and Restoration* (eds. Jose S, Jokela EJ, Miller DL), pp. 297-325, Springer Science+Business Media, New York, New York.

Ware S, Frost C, Doerr PD (1993) Southern mixed hardwood forest: The former longleaf pine forest. In: *Biodiversity of the Southeastern United States Lowland Terrestrial Communities* (eds. Martin WH, Boyce SG, Echternacht AC), pp. 447-493, John Wiley & Sons, Inc., New York, New York.

Wilbur RL (1954) An undescribed dwarf species of *Amorpha* from Georgia. *Rhodora*, **56**, 261-265.

Wilbur RL (1964) A revision of the dwarf species of *Amorpha* (Leguminosae). *The Journal of the Elisha Mitchell Scientific Society*, **80**, 51-65.

Wilbur RL (1975) A revision of the North American genus *Amorpha* (Leguminosae-Psoraleae). *Rhodora*, **77**, 337-409.

Williamson-Natesan EG (2005) Comparison of methods for detecting bottlenecks from microsatellite loci. *Conservation Genetics*, **6**, 551-562.

Wolf DE, Takebayashi N, Rieseberg LH (2001) Predicting the risk of extinction through hybridization. *Conservation Biology*, **15**, 1039-1053.

Yatabe Y, Kane NC, Scotti-Saintagne C, Rieseberg LH (2007) Rampant gene exchange across a strong reproductive barrier between the annual sunflowers, *Helianthus annuus* and *H. petiolaris*. *Genetics*, **175**, 1883-1893.

Young A, Boyle T, Brown T (1996) The population genetic consequences of habitat fragmentation for plants. *Trends in Ecology & Evolution*, **11**, 413-418.

CHAPTER FIVE

APPLICATIONS OF GENETICS TO ENDANGERED SPECIES RESTORATION: A CASE STUDY EVALUATING NATURAL GENETIC DIVERSITY AND THE GENETIC SUCCESS OF *EX SITU* CONSERVATION AND REINTRODUCTION EFFORTS FOR *AMORPHA CRENULATA* (FABACEAE)

ABSTRACT

Consideration of genetic factors is important for the design of *ex situ* and restoration conservation efforts for endangered species because these species often suffer detrimental genetic effects of small population size, such as loss of genetic diversity and inbreeding depression. Ensuring that adequate genetic diversity is present in conservation collections and reintroduced populations is essential for the sustainability, long-term viability, and preservation of the evolutionary potential of these species. Choice of source populations is important because high genetic diversity gained through mixing of different sources may have negative consequences, such as outbreeding depression, so knowledge of population structure, in addition to amounts of genetic diversity, may be vital. In order to evaluate the genetic success of conservation efforts for *Amorpha crenulata*, a federally endangered legume endemic to Miami-Dade County in Florida, wild, *ex situ*, and restored populations were assessed for levels of genetic diversity and population structure using eight microsatellite markers. Multilocus genotypes revealed a pattern consistent with tetraploidy, as opposed to the assumed diploid nature of this species. Genetic success was defined as capturing high

genetic diversity, good replication of natural genetic diversity, and preservation of distinct gene pools. Genetic diversity was high in wild populations of *A. crenulata* and was partitioned into three distinct genetic clusters correlating with geography. Overall, genetic diversity captured in *ex situ* and restored populations was also high, with one notable exception. Genetic replication of wild gene pools was good in terms of overall levels of diversity, but examination of complements of private alleles showed that accuracy could be considered low because many rare alleles were not captured. All genetic clusters identified in the wild populations were represented in *ex situ* and restored populations, but the information presented here will provide a guide for augmenting representation of two of the clusters in future conservation efforts. High genetic diversity in wild populations and genetic differentiation among them argue for inclusion of only single genetic sources in new *ex situ* or restored populations in order to best replicate natural diversity and to avoid outbreeding depression.

INTRODUCTION

Habitat loss and degradation due to human activities are major causes of species endangerment through reduction of the number of populations and of population sizes. Endangered species are imperiled due to the detrimental effects of small population size from both a genetic and an ecological perspective. Small populations often suffer from loss of genetic diversity due to decreased gene flow among populations and genetic drift, as well as higher inbreeding leading to inbreeding depression due to the increased probability of individuals carrying deleterious alleles in the homozygous condition or cases where heterozygotes have a selective

advantage (e.g., Barrett & Kohn 1991; Ellstrand & Elam 1993; Fenster & Dudash 1994; Young *et al.* 1996; Hedrick 2001; Frankham 2003). Ecologically, small populations are at increased risk of extinction from stochastic environmental and demographic factors and other effects of habitat loss and degradation, such as edge effects (e.g., Huenneke 1991; Lande 1999). Genetic factors have also been shown to have critical impacts on extinction risk in timeframes that are comparable to or shorter than ecological factors (Frankham & Ralls 1998; Saccheri *et al.* 1998; Spielman *et al.* 2004a).

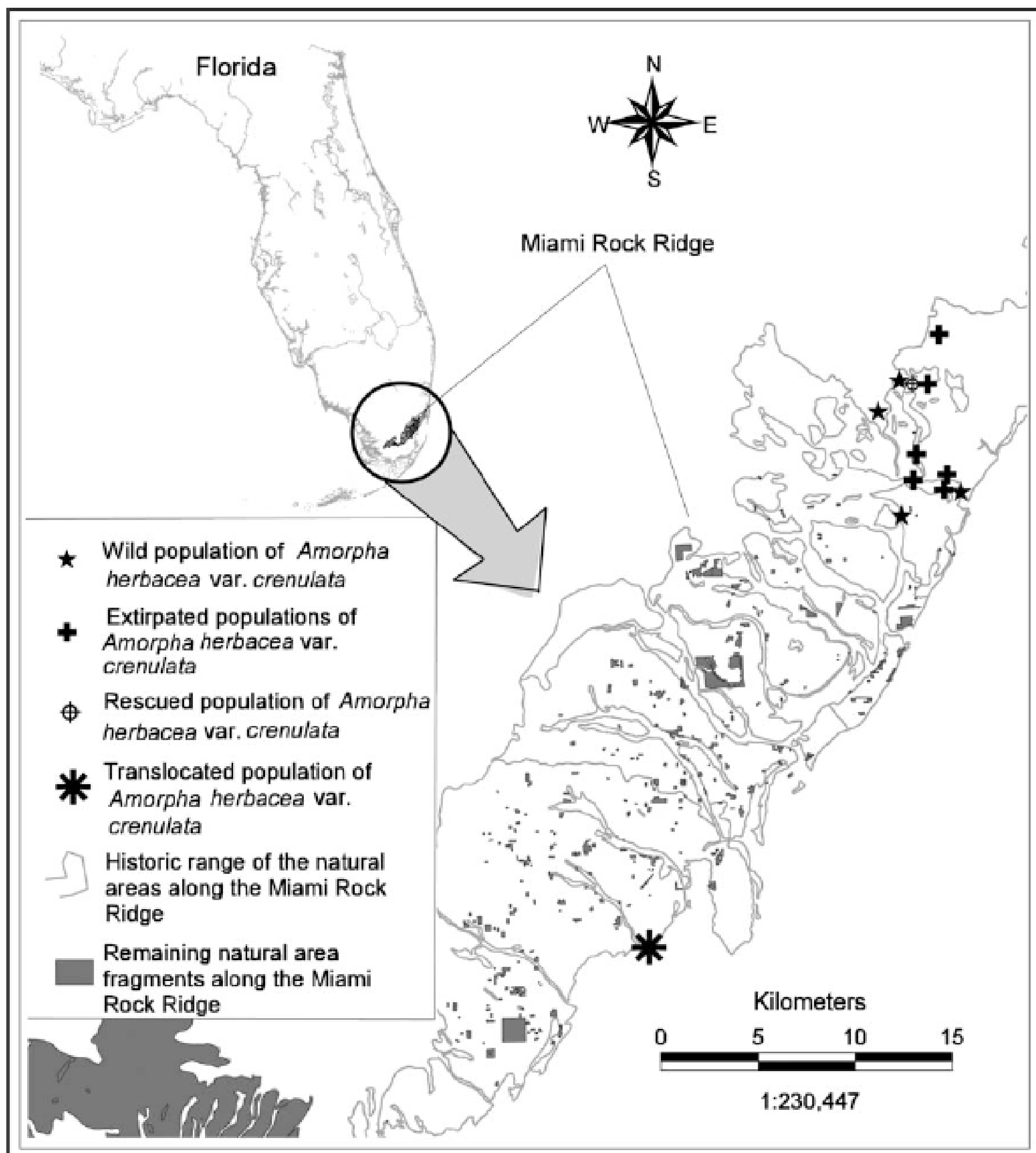
In order to decrease the probability of extinction for endangered plant species and mitigate the negative consequences of small population size, development of *ex situ*, or off site, populations and population restoration efforts are often key components of conservation programs (Brown & Briggs 1991). Consideration of genetics is essential for the design of restoration projects because maintenance of genetic variability is associated with population viability and is essential for preservation of mean population fitness and the evolutionary potential of the endangered species by providing the genetic basis for responses to changing environmental conditions and challenges (Fenster & Dudash 1994; Lesica & Allendorf 1999; Montalvo *et al.* 1997; Hedrick 2001; Frankham 2003; Spielman *et al.* 2004b; Menges 2008). Creating a restoration design that maintains high genetic diversity and avoids inbreeding depression, while preserving local adaptation and co-adapted gene complexes to avoid outbreeding depression is one of the big challenges for species restoration (McKay *et al.* 2005). Populations used as sources for restoration efforts should mirror as closely as possible wild populations in their complements and frequencies of alleles and should have

high genetic diversity in order to avoid founder effects (Lacy 1994; Hufford & Mazer 2003). However, source populations should be carefully chosen because in some cases high genetic diversity in a restored population created by using individuals from genetically distinct source populations, as recommended, for example, by Lesica & Allendorf (1999), can increase genetic diversity at the cost of decreased fitness of the population due to outbreeding depression (Fenster & Dudash 1994, Pavlik 1996; Montalvo & Ellstrand 2001; Hufford & Mazer 2003; Allendorf & Luikart 2007).

Relatively few studies have considered genetic factors in ecological restoration designs (Young 2000; Seddon *et al.* 2007). When genetic evaluation of programs is conducted, it often occurs only after the design and establishment of a restoration program based on ecological or demographic factors, and not when it would be most useful, prior to the establishment of the program. In some cases, *ex post facto* genetic evaluations show that sampling of multiple source populations based on ecological amplitude and potentially adaptive variation has also ensured maintenance of genetic diversity in conservation collections and restored populations of endangered species (e.g., Huenneke 1991; Travis *et al.* 2002; Ramp *et al.* 2006; Le Gouar *et al.* 2008; González-Pérez *et al.* 2009). However, in other cases restored populations have lower genetic diversity compared to their wild counterparts and *a priori* information about the genetics of the species could have improved conservation efforts (e.g., Helenurm & Parsons 1997; Williams 2001; Tani *et al.* 2006; Fant *et al.* 2008).

Amorpha crenulata Rydb. (Crenulate Lead-plant) is a naturally rare, federally endangered legume currently and historically found only in Miami-Dade county in Florida (Federal Register 1985; Figure 5.1). These primarily

Figure 5.1. Map of the current and historical distribution of *Amorpha crenulata*. This figure was originally published by Wendelberger *et al.* (2008) and is reproduced with permission from John Wiley & Sons. Wendelberger *et al.* (2008) followed Isely (1986, 1998) in treating *A. crenulata* as a variety of *A. herbacea*. The stars on the map correspond to wild populations A, B, D, and C respectively when reading from north to south. The rescued population corresponds to the extirpated Railroad population. The asterisk corresponds to restored population Y.


Florida
N
W
E
S
Miami Rock Ridge
Wild population of *Amorpha herbacea* var. *crenulata*
Extirpated populations of *Amorpha herbacea* var. *crenulata*
Rescued population of *Amorpha herbacea* var. *crenulata*
Translocated population of *Amorpha herbacea* var. *crenulata*
Historic range of the natural areas along the Miami Rock Ridge
Remaining natural area fragments along the Miami Rock Ridge
Kilometers
0
5
10
15
1:230,447

outcrossing shrubs are endemics of pine rocklands and ecotones between pinelands and wet prairies of the Miami Rock Ridge (Possley *et al.* 2004). Destruction of this naturally rare habitat by human development and by fire suppression has decreased its occurrence by more than 98% and led to severe population reduction for this and other endangered species (Federal Register 1984, 1985; Wendelberger *et al.* 2008) with 50% of the known wild *A. crenulata* populations having been destroyed since 1995 (Roncal *et al.* 2006). Water diversion and invasive species have also contributed to the decline of *A. crenulata* (Wendelberger *et al.* 2008). Currently the approximately 300 – 350 wild individuals of this species are found at only four localities, two larger populations of approximately 150 individuals apiece, a smaller population with fewer than twenty individuals, and a site where a single individual is found. Although these plants are easily propagated (Roncal *et al.* 2006), natural recruitment appears to be rare, especially in degraded habitat where forest duff depth is not optimal for seedling establishment (Possley *et al.* 2004; Wendelberger *et al.* 2008; Wendelberger & Maschinski 2009).

Conservation efforts for *A. crenulata* have thus far been ecologically and horticulturally based (Roncal *et al.* 2006). A conservation action plan has been developed for this species and is administered by Fairchild Tropical Garden in Coral Gables, Florida (Possley *et al.* 2004). The primary goals of the plan are prevention of extinction and stabilization of this species through protection and monitoring of the remaining wild populations, protection of existing habitat and acquisition of new habitat for population restoration, and maintenance or improvement of habitat quality, such as through introducing controlled burns (Possley *et al.* 2004). In addition to strategies to

stabilize wild populations, *ex situ* conservation measures have also been enacted. An *ex situ* collection of plants was created at Fairchild Tropical Garden consisting of individuals rescued from five extirpated wild populations that were growing in areas slated for development. Individuals from one of the extant wild populations have also been protected by transplantation to the *ex situ* collection. In addition to the *ex situ* collection of plants at the garden, three populations have been restored in habitat suitable for this species (Wendelberger *et al.* 2008). In addition to these conservation efforts, an *ex situ* seed bank of nearly 50,000 seeds has been established at the National Center for Genetic Resources Preservation in Fort Collins, CO (Roncal *et al.* 2006; Wendelberger *et al.* 2008).

Methods of propagation for this species have been evaluated through extensive exploration of germination requirements, rooting techniques for cuttings, and plant translocation (Roncal *et al.* 2006). The restored populations of *A. crenulata* have been established under three different experimental designs utilizing these techniques. The first population was established in 1995 to determine whether translocation of individuals could be successful, but by 2003 only 52% of the initial 190 individuals were still alive (Wendelberger *et al.* 2008). Another restored population was started with several hundred individuals transplanted from a population about to be extirpated through development. A third outplanting was created as a restored population and experiment to evaluate the success of different propagation methods for this species by transplanting whole individuals of different genetic sources and ages, planting rooted cuttings from multiple genetic sources, and planting seeds from both wild populations and the those produced in the *ex situ* garden collection (Wendelberger *et al.* 2008).

The purposes of this study were to use variation at eight microsatellite loci to explore genetic diversity and population structure of remaining wild populations of *Amorpha crenulata*; to evaluate how the genetic diversity of naturally occurring *A. crenulata* was affected by extirpation of populations which were preserved in the *ex situ* conservation collection and restored populations; and to evaluate the success of conservation efforts for *A. crenulata* from a genetic perspective by comparing levels of genetic diversity in wild populations with those found in the *ex situ* and restored populations. How consideration of genetics might be useful in future conservation planning for *A. crenulata* is also discussed.

MATERIALS AND METHODS

Sample Collection and Preparation

Two hundred fifty-five *Amorpha crenulata* leaf tissue samples from the four remaining wild populations, three restored populations, and the *ex situ* conservation collection were collected and dried on silica gel by staff at Fairchild Tropical Garden in Coral Gables, Florida (Table 5.1). DNA was extracted from tissue samples using a modified CTAB extraction protocol (Doyle & Doyle 1987) with 2% PVP-40 added to the extraction buffer.

Genotyping

Twelve microsatellite loci developed for *Amorpha georgiana* Wilbur (Straub *et al.* 2009) were tested for transferability to *A. crenulata* using the polymerase chain reaction (PCR) conditions optimized for *A. herbacea* Walter (Straub *et al.* 2009), a species of which *A. crenulata* is sometimes considered a variety (Isely 1986, 1998). Genotypic data were collected using the

Table 5.1. Summary of the extent of sampling for wild, *ex situ*, and restored populations of *Amorpha crenulata*, including descriptions of the genetic source of each population and estimates of current population sizes.

Population	Type	Genetic Source	Estimated Population Size	Number of Samples Collected	Number of Samples Included in Analyses
Wild A	Wild	N/A	~ 150	23	23
Wild B	Wild	N/A	~ 150	22	22
Wild C	Wild	N/A	< 20	10	7
Wild D	Wild	N/A	1	1	1
ex situ Wild B	*ex situ* from extant wild	Wild B extant wild	18	13	11
Snapper Creek Canal	*ex situ* from extirpated wild	N/A	3	1	1
84th St.	*ex situ* from extirpated wild	N/A	17	7	6
ex situ Railroad	*ex situ* from extirpated wild	N/A	> 200	28	26
Restored Railroad	Restored	Railroad extirpated wild	~ 350	32	32
Restored Y	Restored	Multiple	126	52	50
Restored Z	Restored	unknown	~ 200	66	66
			Totals:	**255**	**245**

genotyping procedure described by Straub *et al.* (2009) for all samples for eight loci (Age01, Age02, Age06, Age07, Age13, Age15, Age24, Age30), which amplified well, had no to minimal stutter, and were variable (Table 5.2). These loci are unlinked in *A. georgiana* (Straub *et al.* 2009) and were assumed to be unlinked in *A. crenulata* for this study.

Data Quality Control

Genotyping success was calculated by re-genotyping approximately 12% of the individuals for each locus to determine if both runs produced identical genotypes. The overall success rate was 91.2% and all instances of mismatches between the first and second genotyping run were suspected cases of allelic dropout (non-amplification of a particular allele) in either the first or second PCR reaction. In these cases, the genotype for that individual was coded to include the allele that was missing from one of the runs. PCR reactions for a few individuals faithfully produced more than the four bands for particular loci. Genotypes for these individuals were coded as missing data for those loci.

GENOTYPE (Miermans & van Tienderen 2004) was used to check for similar multilocus genotypes, which could indicate sampling of ramets rather than genets. The program was run with the threshold genetic distance between genotypes under the infinite alleles model set to 0 or 1. For sample pairs from the same population that shared identical genotypes, one of the two samples was eliminated from the data set. For sample pairs with identical genotypes from the *ex situ* and restored populations, both genotypes remained in the final data set because the plant in the *ex situ* collection was the likely source, through cuttings, of the plant being used in population

Table 5.2. Success of the use of *Amorpha georgiana* microsatellite locus primer sets in *Amorpha crenulata*. The percentage of individuals with greater than two alleles for each locus is reported.

A. georgiana **Locus**	**Amplified and Cleanly Genotyped in *A. crenulata*?**	**Polymorphic in *A. crenulata*?**	**% of individuals with > 2 alleles**
Age01	Yes	Yes	66.5
Age02	Yes	Yes	51.4
Age06	Yes	Yes	57.6
Age07	Yes	Yes	58.0
Age10	No	N/A	N/A
Age13	Yes	Yes	35.1
Age15	Yes	Yes	57.1
Age18	Yes	No	N/A
Age24	Yes	Yes	81.6
Age25	No	N/A	N/A
Age29	No	N/A	N/A
Age30	Yes	Yes	89.4

restoration. Samples that were found to share similar genotypes were then examined to see if the similar, but not identical genotypes could be cases of allelic dropout. If allelic dropout was suspected at one locus for two samples in the same population, the sample with the missing allele was removed from the final data set. Samples with similar genotypes where allelic dropout would have to be invoked for more than one locus were left as two separate individuals in the analysis because they could be closely related individuals rather than ramets.

Analyses of Genetic Diversity and Population Structure

Genotyping for *Amorpha crenulata* individuals indicated that this species is tetraploid rather than diploid. In diploid species, allele dosage is easily determined for microsatellite loci, but determining allele dosage and genotype frequencies for polyploids is much more complex (De Silva *et al.* 2005; Obbard *et al.* 2006). If only one size of PCR product is observed, the individual is assumed to be homozygous for that allele, whereas if two product sizes are observed, the individual is determined to be heterozygous with 1:1 allele dosage. The situation in the tetraploid, where observation of up to four alleles is possible, becomes more complicated when two or three allele sizes are detected because the genotype underlying such a pattern could be any one of three possible allele dosage ratios. For example, if alleles 1 and 2 were observed, the possible dosages of alleles would be 1,1,1,2 or 1,1,2,2 or 1,2,2,2. The amount of PCR product produced for a particular allele measured by band intensity or peak height can be unreliable estimators of allele dosage (Obbard *et al.* 2006; Vinson *et al.* 2009). In the absence of pedigree or segregation information to help determine dosage, a conservative

approach of using a combination of genotype and allele phenotype data was employed. Genotypes were used for individuals having either one or four alleles at a locus and phenotypes were used for individuals with either two or three allele size classes at a locus. The unknown allele or alleles were coded as missing data. This allele phenotype approach has been employed in other studies of polyploid species using microsatellite data (e.g., Hamilton & Eckert 2007; Marrs *et al.* 2008).

A further complication for data analysis arose when the tetraploid nature of this species was discovered because it is unknown whether the species is autotetraploid with polysomic inheritance or allotetraploid with disomic inheritance, or a segmental allopolyploid in which segregation patterns vary among loci. In order to explore genetic diversity and population structure in this species, approaches for both autopolyploids and allopolyploids were employed, as has been done in previous studies where the mode of inheritance in a polyploid was unknown (e.g., González-Pérez *et al.* 2009). To this end SPAGeDi 1.2 (Hardy & Vekemans 2002) was used for analyses assuming autopolyploidy/polysomic inheritance and F-DASH (Obbard *et al.* 2006) was used for analyses assuming allopolyploidy/disomic inheritance.

Genetic diversity statistics were calculated for wild, *ex situ*, and restored populations. SPAGeDi 1.2 was used to calculate the average number of alleles (A_a), the number of private alleles per population, and gene diversity corrected for sample size under a polysomic model of inheritance (H_E; Nei 1987). Observed heterozygosity was calculated by hand. The number of allele phenotypes per population (A_P) and allelic phenotype diversity (H'; Obbard *et al.* 2006), which is related to the concept of Nei's

gene diversity except that similarities rather than differences are considered, were calculated using F-DASH. Averages across populations were calculated excluding the two samples comprised of single individuals (Wild D and Snapper Creek), which do not represent true population samples. The significance of differences among populations for these values (A_a, A_P, H_E, H') was tested using 1-way ANOVA and the Tukey-Kramer honestly significant difference test ($\alpha = 0.05$) to correct for multiple tests and implemented in JMP 8.0 (SAS Institute, Inc.). If only two populations were being compared, a student's t-test ($\alpha = 0.05$) was employed in JMP 8.0 to determine if differences between them were significant.

In order to assess population differentiation and structure among the remaining wild populations, global and pairwise F_{ST} values (Weir & Cockerham 1984) were calculated using SPAGeDi 1.2 and the significance of population differentiation was assessed using permutation tests of 20,000 replicates. F-DASH was used to calculate F'_{ST}, an analog of F_{ST} calculated using the H' allelic-phenotype diversity statistic (Obbard *et al.* 2006); significance was assessed using permutation tests of 20,000 replicates. The program STRUCTURE 2.3.2 (Pritchard *et al.* 2000; Falush *et al.* 2003, 2007; Hubisz *et al.* 2009) was also used to detect population structure among two samples: 1) the wild populations, and 2) the wild populations plus extirpated populations from the *ex situ* collection. Preliminary runs to ascertain the range of K values (number of populations) for more extensive analyses were run using the admixture model and correlated allele frequencies (Falush *et al.* 2003) and extension of the method to handle tetraploid data (Falush *et al.* 2007). The use of allele phenotypes instead of genotypes was accounted for by using the recessive alleles option set to the missing data value (Pritchard *et*

al. 2009). These runs consisted of five chains for each value of K (K = 1 – 6 for wild only; K = 1 – 8 for wild plus extirpated) run for 500,000 generations following a burn-in of 200,000 generations. Preliminary runs were conducted both with consideration of source population information to assist in detecting weak population structure (Hubisz *et al.* 2009) and without consideration of source population information. Final STRUCTURE analyses were conducted using the same options as the preliminary runs, except analyses consisted of ten chains for each value of K (K = 1 – 5 for wild only and wild plus extirpated) and run for 1,000,000 generations following a burn-in of 500,000 generations. Values of K were chosen after considering both the *ad hoc* ln $\Pr(X|K)$ method presented by Pritchard *et al.* (2000) and the ΔK method of Evanno *et al.* (2005).

RESULTS

Genotyping Success and Ploidy of *Amorpha crenulata*

Individuals were successfully genotyped for eight of twelve microsatellite loci developed for *A. georgiana* (Table 5.2). Genotyping revealed the presence of more than the two alleles expected for diploids for at least some loci for all but one individual, a member of the Restored Z population that had two alleles for the seven loci successfully genotyped. In general individuals had between one and four alleles per locus, indicating that *Amorpha crenulata* is likely a tetraploid species. Several other *Amorpha* species are known or suspected tetraploids (Straub *et al.* 2009; Straub *et al.* in press). The discovery of the tetraploid nature of this species led to the use of allele phenotypes rather than genotypes in some cases, therefore some of the

results obtained must be interpreted cautiously because this analytical approach could bias some calculations, such as for gene diversity (H_E).

Overall Genetic Diversity in *Amorpha crenulata*

Across all populations genetic diversity in *Amorpha crenulata* was high with an average of 18.75 alleles per locus and total of 151 different alleles observed for the eight sampled microsatellite loci. On average, 89.304 different allele phenotypes were observed per locus. Average observed heterozygosity (H_O) was 0.868 and average expected heterozygosity (H_E) was 0.772. Allelic phenotype diversity (H'), "the average number of alleles by which pairs of individuals differ at a single locus" (Obbard *et al.* 2006, p. 298), across all loci and populations was 3.206. Numerous rare alleles were observed when all population samples were compared, except in Wild D and Snapper Creek, the two populations represented by a single individual (Table 5.3). When single genetic sources were considered, the extirpated Railroad population had more private alleles than all other populations combined (Table 5.3).

Genetic Diversity and Population Structure in Wild *Amorpha crenulata* Populations

When samples of the four wild populations were considered, genetic diversity in general, including the number the private of alleles, was higher in the two larger populations, Wild A and Wild B, than the smaller Wild C population (Table 5.4). Wild A had significantly higher allelic diversity than Wild C and both Wild A and Wild B had significantly higher numbers of allele phenotypes than Wild C. However, the populations were

Table 5.3. Genetic differentiation among populations in terms of private alleles for all populations, single genetic source population, and same genetic source population comparisons for *Amorpha crenulata*. See Table 5.1 for explanation of genetic sources.

<table>
<tr><th>Population</th><th>All</th><th>Single Genetic Source</th><th>Same Genetic Source</th></tr>
<tr><td>Wild A</td><td>5</td><td>7</td><td>N/A</td></tr>
<tr><td>Wild C</td><td>1</td><td>4</td><td>N/A</td></tr>
<tr><td>Wild D</td><td>0</td><td>0</td><td>N/A</td></tr>
<tr><td>Wild B</td><td>6</td><td rowspan="2">11</td><td>26</td></tr>
<tr><td>ex situ Wild B</td><td>1</td><td>10</td></tr>
<tr><td>Snapper Creek Canal</td><td>0</td><td>0</td><td>N/A</td></tr>
<tr><td>84th St.</td><td>1</td><td>2</td><td>N/A</td></tr>
<tr><td>ex situ Railroad</td><td>2</td><td rowspan="2">27</td><td>9</td></tr>
<tr><td>Restored Railroad</td><td>5</td><td>18</td></tr>
<tr><td>Restored Y</td><td>2</td><td>N/A</td><td>N/A</td></tr>
<tr><td>Restored Z</td><td>2</td><td>N/A</td><td>N/A</td></tr>
</table>

indistinguishable in terms of observed and expected heterozygosity, as well as allelic phenotype diversity.

At the Wild C population, samples were collected from ten plants, representing approximately fifty percent of the remaining individuals at that site. The number of genetic individuals at this site is questionable because three of the ten plants shared a single genotype and two others differed from this genotype at only a single locus. Two other plants shared a second genotype and two additional plants shared very similar genotypes differing at a single locus. The similarity of genotypes found at the Wild C population could have been due to that population consisting mainly of a few family groups of closely related individuals, or to allelic dropout, meaning that the

Table 5.4. Genetic diversity observed in the four remaining wild populations of *Amorpha crenulata*. Values not connected by the same letter are significantly different (α = 0.05). Wild D was not included in calculations of significance because this population was represented by a single individual.

Population	$\mathbf{A_a}$	$\mathbf{A_p}$	$\mathbf{H_O}$	$\mathbf{H_E}$	**H'**
Wild A	11.000^{A}	15.883^{A}	0.864^{A}	0.804^{A}	3.323^{A}
Wild B	10.000^{AB}	14.714^{A}	0.823^{A}	0.758^{A}	2.972^{A}
Wild C	5.125^{B}	3.625^{B}	0.857^{A}	0.752^{A}	2.583^{A}
Wild D	2.375	1.000	0.875	N/A	N/A

A_a: average number of alleles per locus; A_p: average number of allele phenotypes per locus; H_O: observed heterozygosity; H_E: gene diversity; H': allelic phenotype diversity

similar plants were actually clones and there were even fewer genets present at the site.

Two to three distinct populations were detected using STRUCTURE 2.3.2. For analyses that used collection location as a prior, both the *ad hoc* ln $Pr(X|K)$ method and ΔK method suggested that K = 2. For analyses without the location prior, the *ad hoc* ln $Pr(X|K)$ method indicated that either K = 2 or 3 and the ΔK method indicated that K = 2. Results for both K = 2 and K = 3 are shown in Figure 5.2. Calculation of F_{ST} and F'_{ST} supported the STRUCTURE results with the greatest amount of differentiation being detected between the Wild B and Wild C populations and a smaller amount of differentiation between other population combinations (Table 5.5).

Table 5.5. Differentiation among wild populations of *Amorpha crenulata* based on F_{ST} under a model of polysomic inheritance and F'_{ST} under a model of disomic inheritance. Values below are F_{ST} and values above are F'_{ST} for pairwise combinations. F_{ST} values are significantly different from zero ($P < 0.001$) and F'_{ST} values are not significantly different from each other based on 20,000 permutations.

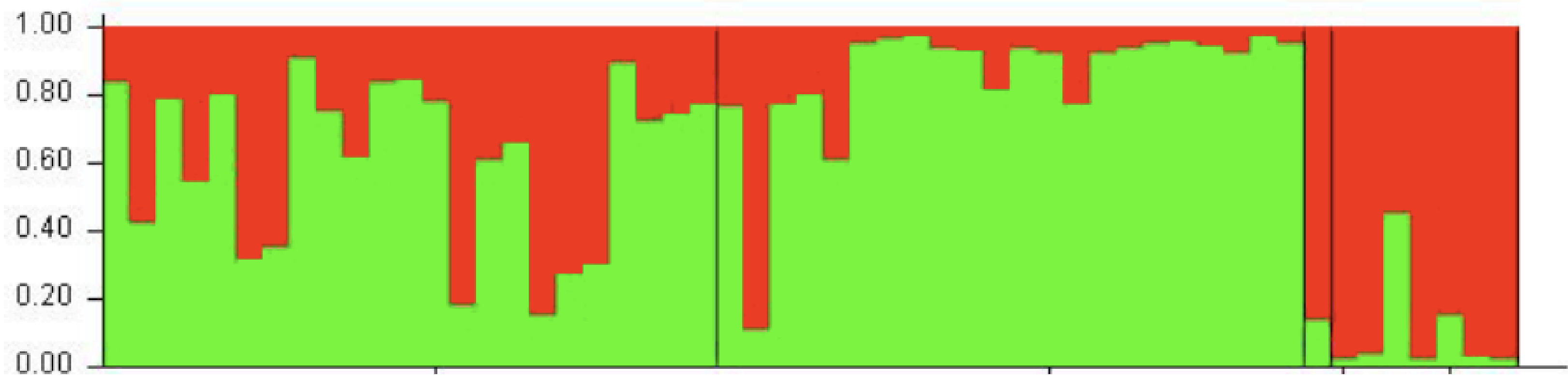
	Wild A	**Wild B**	**Wild C**
Wild A	-	0.04	0.01
Wild B	0.03	-	0.09
Wild C	0.05	0.11	-

Genetic Diversity in and Distinctiveness of Extirpated *Amorpha crenulata* Populations

The *ex situ* populations of *Amorpha crenulata* were similar for several measures of genetic diversity, but the *ex situ* Railroad population did have significantly more alleles per locus than the *ex situ* 84th St. population and significantly more allele phenotypes per locus than both the *ex situ* 84th St.

Figure 5.2. Population structure among the remaining wild populations of *Amorpha crenulata* inferred through Bayesian analysis of microsatellite data using STRUCTURE 2.3.2. Results without consideration of individuals' collection location are shown above for each value of K and results with consideration of collection location are shown below. Each color represents a genetic cluster and each column represents one individual. The amount of each color within a column represents the membership coefficient (Q), or proportion of the individual's genotype assigned to each of the inferred genetic clusters.

K = 2



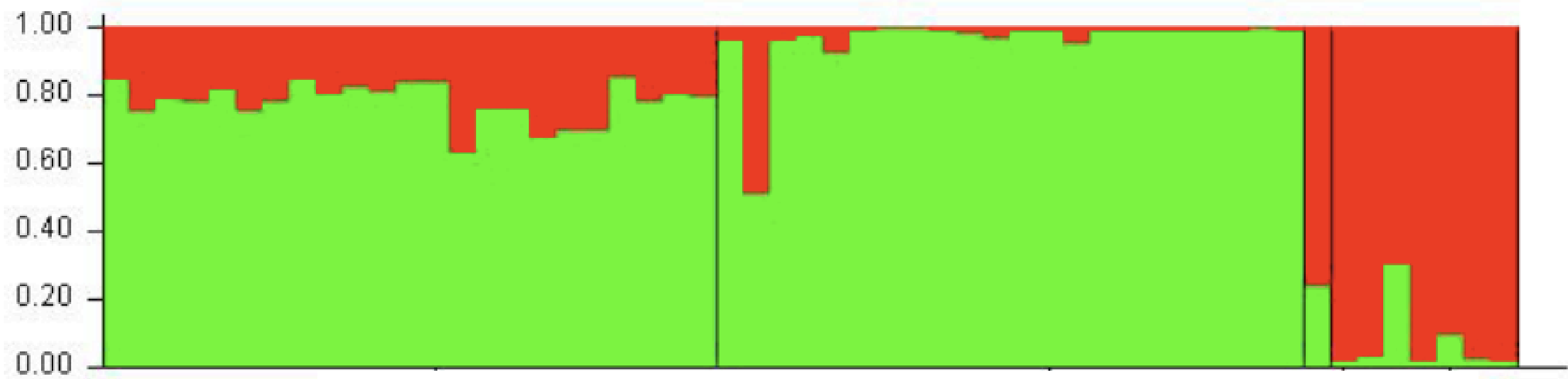
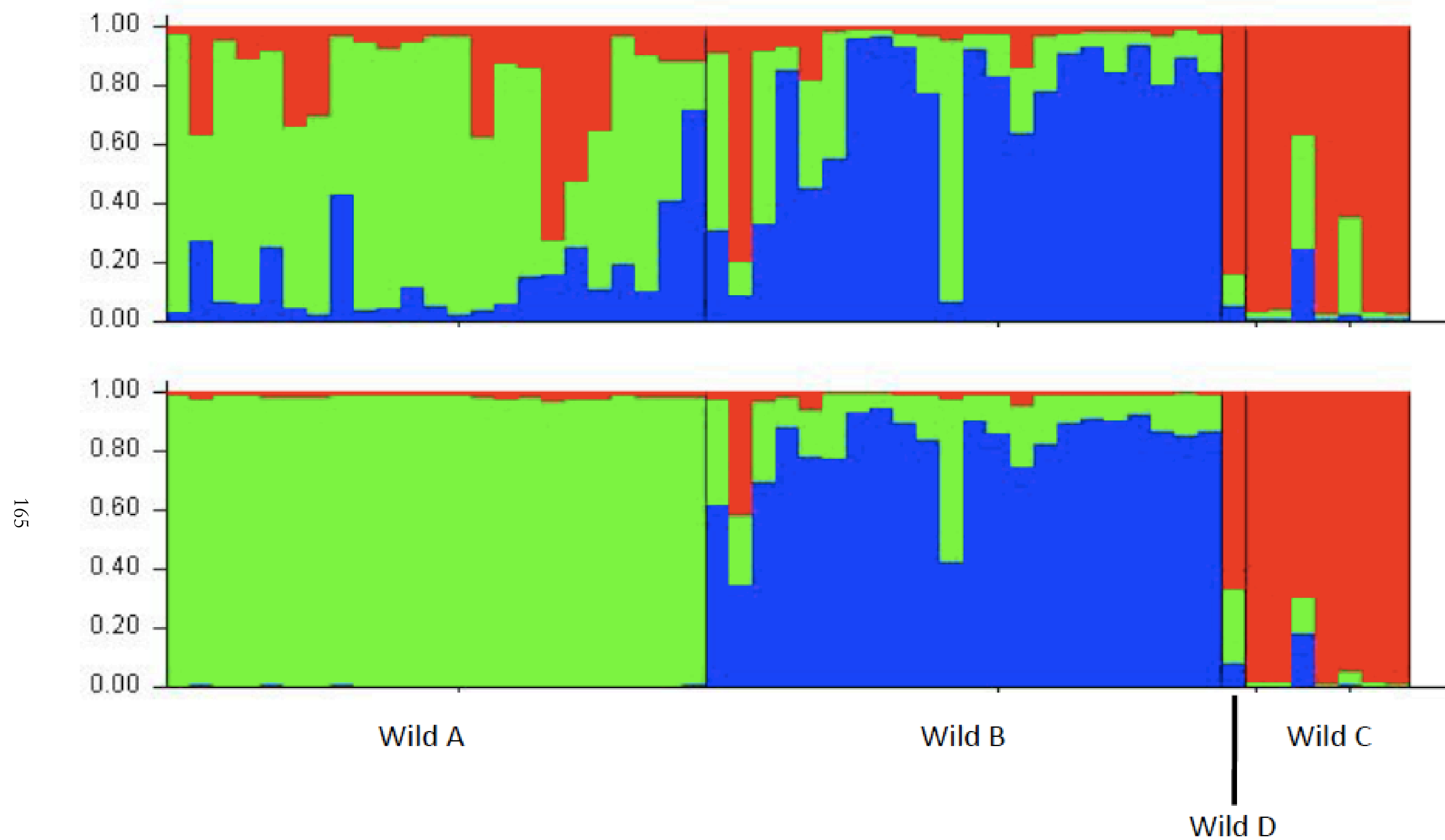


Figure 5.2 (Continued)



and *ex situ* Wild B populations (Table 5.6). Compared to wild populations, only the *ex situ* Railroad population had significantly more alleles per locus on average (P = 0.0009) and number of allele phenotypes (P < 0.0001) than the Wild C population. In a comparison with its genetic source, the *ex situ* Wild B population was similar in most measures of genetic diversity to the Wild B population, except that the wild population had a significantly higher number of allele phenotypes (P = 0.004). The wild population sample also had 2.6 times as many private alleles as the *ex situ* collection sample (Table 5.3). All other population combinations were similar for measures of genetic diversity.

Three genetic clusters were detected when analyses to detect population structure among all natural genetic sources, including wild and extirpated populations preserved in the *ex situ* collection and outplantings, were conducted (Figure 5.3). The choice of K = 3 was supported by both the *ad hoc* ln $\Pr(X|K)$ method and ΔK method in analyses with and without consideration of the collection location of individuals.

Genetic Diversity of Restored *Amorpha crenulata* Populations

High genetic diversity was observed in two of the three restored populations of *Amorpha crenulata* (Table 5.7). Overall, the most genetically diverse restored population was Restored Railroad and the least genetically diverse was Restored Z, which had significantly lower average number of alleles per locus, expected heterozygosity, and allelic phenotype diversity than the Restored Railroad population. In comparison to the remaining wild populations, both the Restored Railroad and Restored Y populations had a significantly higher average number of alleles per locus (P = 0.002, P = 0.003

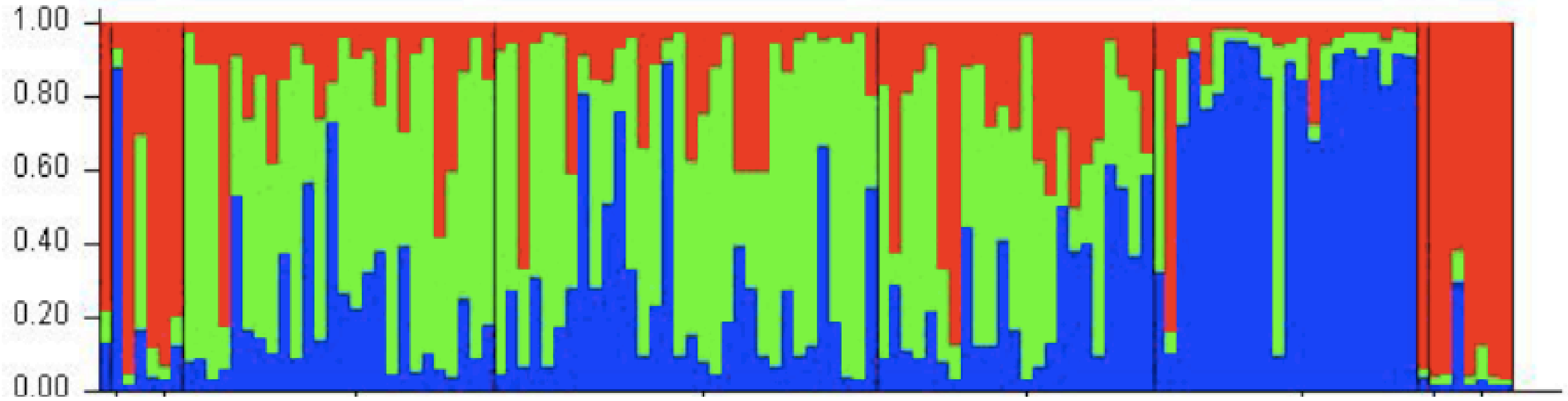
Table 5.6. Genetic diversity observed in the *ex situ* conservation collection of *Amorpha crenulata* at Fairchild Tropical Garden, FL. Values not connected by the same letter are significantly different ($\alpha = 0.05$). Snapper Creek was not included in calculations of significance because this population was represented by a single individual.

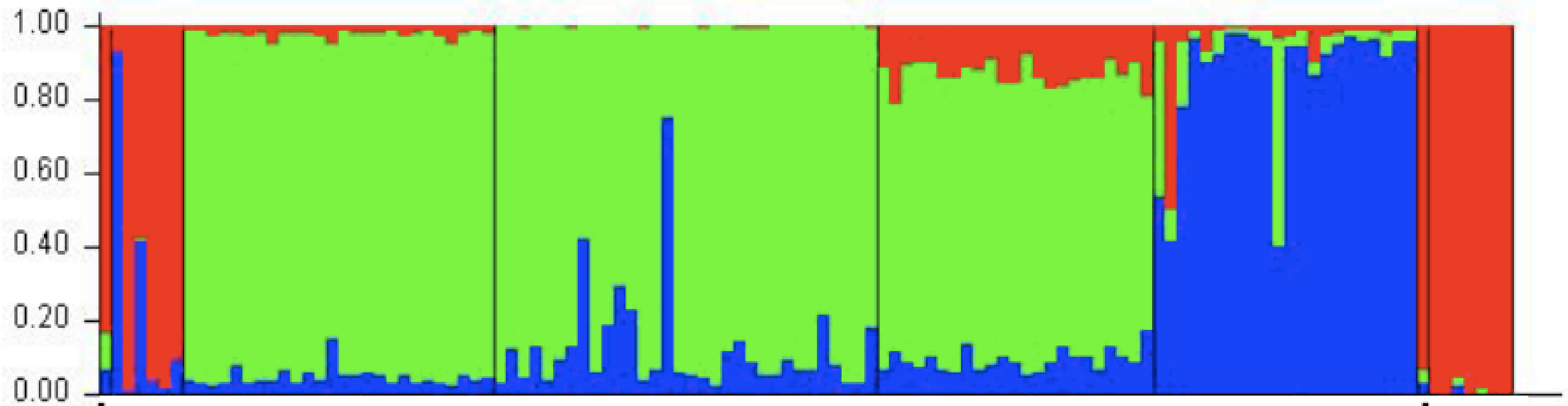
Population	A_a	A_p	H_O	H_E	H'
ex situ Wild B	8.000^{AB}	8.250^{B}	0.898^{A}	0.779^{A}	2.936^{A}
84th St.	6.250^{B}	5.125^{B}	0.875^{A}	0.795^{A}	3.142^{A}
ex situ Railroad	12.125^{A}	17.795^{A}	0.923^{A}	0.817^{A}	3.262^{A}
Snapper Creek Canal	2.125	1.000	1.000	N/A	N/A

A_a: average number of alleles per locus; A_p: average number of allele phenotypes per locus; H_O: observed heterozygosity; H_E: gene diversity; H': allelic phenotype diversity

Figure 5.3. Population structure among the remaining wild populations of *Amorpha crenulata* and the recently extirpated populations of this species represented in the *ex situ* conservation collection at Fairchild Tropical Garden. The Restored Railroad population was also included because it consists of individuals rescued from the extirpated Railroad population. Results without consideration of individuals' collection location are shown above and results with consideration of collection location are shown below. Each color represents a genetic cluster and each column represents one individual. The amount of each color within a column represents the membership coefficient (Q), or proportion of the individual's genotype assigned to each of the inferred genetic clusters.

K = 3





84th St. *ex situ* Railroad Railroad Restored Wild A Wild B Wild C

Snapper Creek Canal Wild D

Table 5.7. Genetic diversity observed in restored populations of *Amorpha crenulata*. Values not connected by the same letter are significantly different (α = 0.05).

Population	A_a	A_p	H_o	H_e	H'
Restored Railroad	13.250^{A}	21.846^{AB}	0.902^{A}	0.810^{A}	3.374^{A}
Restored Y	12.625^{A}	28.144^{A}	0.854^{A}	0.788^{AB}	3.162^{AB}
Restored Z	5.875^{B}	14.163^{B}	0.811^{A}	0.650^{B}	1.759^{B}

A_a: average number of alleles per locus; A_p: average number of allele phenotypes per locus; H_O: observed heterozygosity; H_E: gene diversity; H': allelic phenotype diversity

respectively) and average number of allele phenotypes than the Wild C population ($P < 0.0001$ for both comparisons). Additionally, the Restored Y population had a significantly higher average number of allele phenotypes per locus than the Wild B ($P = 0.003$) and Wild C ($P < 0.0001$) populations.

When the Restored Railroad population was compared with the *ex situ* Railroad population, no differences in genetic diversity were found. However, the two did differ in their complements of rare alleles, even though they were derived from a single genetic source population. Over the eight loci, twice as many private alleles were observed for the Restored Railroad population as were found in the *ex situ* Railroad population (Table 5.3).

DISCUSSION

Genetic Diversity and Population Structure in Wild Populations of *Amorpha crenulata*

As a species, *Amorpha crenulata* does not appear to be suffering the effects of small population size despite very low census size and highly fragmented and degraded habitat. This species displayed high overall genetic diversity and heterozygosity. Larger wild populations retained more genetic diversity and had more private alleles than the small wild populations, which is not unexpected (Honnay & Jacquemyn 2007). The two large wild populations also exhibited little evidence for clonality, a plant survival strategy in stressful environmental conditions, such as degraded habitat (Honnay & Bossuyt 2005). However, clonality was evident in the extremely small Wild C population, where plants thought to be genets turned out to be ramets, further decreasing its population size in terms of genetic individuals.

Another factor contributing to the maintenance of genetic diversity in *A. crenulata* may be the longevity of individual plants. Genets of this species may survive for many years and adults in the remaining populations may capture a snapshot of genetic diversity that predates habitat degradation because long generation times slow the erosion of genetic diversity by genetic drift (Young *et al.* 1996; Honnay & Bossuyt 2005, but see Honnay & Jacquemyn 2007). A similar pattern of high genetic diversity despite severe habitat degradation and population reduction was observed in *Amorpha georgiana*, another endangered species of the genus with similar life history traits (Straub & Doyle 2009).

The putative tetraploid nature of *A. crenulata* must contribute significantly to both the observed high genetic diversity and high heterozygosity. Under an autopolyploid model, there can be up to four alleles at each locus, rather than the two alleles possible for diploids, improving the chances for increased maintenance of genetic diversity and heterozygosity and slowing the process of genetic drift (Moody *et al.* 1993; Soltis & Soltis 2000). If this species is of an allopolyploid nature, heterozygosity may even be fixed due to fixed differences between homoeologous loci which are disomically inherited, and thus contain variation that does not segregate (Soltis & Soltis 2000, 2009).

That observed heterozygosity calculated under a model of polysomic inheritance for tetraploids was greater than the expected heterozygosity for *A. crenulata*, as a whole and in each population, could perhaps be evidence that this species is in fact of allopolyploid origin. Even in very small populations, like Wild C, where inbreeding and consequent increases in homozygosity would be most likely, heterozygosity is high. In an

allopolyploid, loci would be disomically inherited, leading to greater than expected heterozygosity than if all loci were polysomically inherited. The fact that many individuals are homozygous at multiple loci could be due to an allopolyploid origin involving closely related diploid progenitors that shared alleles; alternatively, size homoplasy is a possible explanation if more divergent progenitors were involved. Moreover, it is possible that *A. crenlulata* is a segmental allopolyploid, with fixed heterozygosity at loci such as Age 24 and Age 30, at which 90% of individuals possessed more than two alleles, and multisomic inheritance at other loci, such as Age 13, where only around 35% of individuals had more than two alleles (Table 5.2). The allopolyploid hypothesis is further supported by phylogenetic evidence, which suggest that *A. herbacea* and *A. fruticosa* may be the progenitors of *A. crenulata* (Chapter 2).

Genetic diversity statistics (Obbard *et al.* 2006) for *Amorpha crenulata* were high even when compared to other polyploids. The observed range of H' among wild populations in this study of 2.583 - 3.383, relative to a theoretical maximum of 8 for tetraploids, was much greater than observed in other allotetraploids (*Acacia koa* 0.566 – 0.988, Fredua-Agyeman *et al.* 2008; *Dipteryx odorata* 0.91 - 0.98, Vinson *et al.* 2009; *Plantago hawaiensis* 0.09 – 1.5 and *Plantago pachyphylla* 0.45 – 1.82, Wolff *et al.* 2009) and even allohexaploid species (*Geum triflorum* 0.72 – 2.75, Hamilton & Eckert 2007; *Festuca arundinacea* 0.561 – 1.848, Tehrani *et al.* 2009) that are not endangered.

Population structure was observed among wild populations of *A. crenulata*. The two larger populations, Wild A and Wild B, were most similar to each other and the two smaller populations, Wild C and Wild D, were most similar to each other (Figure 5.2). There was clear differentiation

between these two groups at K = 2 with an additional layer of population structure differentiating Wild A and Wild B at K = 3. This pattern was evident even when population origin was used as a prior in the analysis, but became clearer when origin was considered. The population structure among wild populations appeared to be correlated with geography because the two larger populations are in close proximity, as are the two smaller populations; although, less than 10 km separate the two groups (Figure 5.1). This indicated that there might have been at least two larger areas of gene flow in the past before human development fragmented the habitat.

The Genetic Success of Conservation Efforts for *Amorpha crenulata*

The major goals of most conservation projects involving reintroduction are to achieve abundance, extent, resilience, and persistence for the endangered species (Pavlik 1996). Therefore it seemed reasonable to evaluate the success of conservation efforts for *Amorpha crenulata* from a genetic perspective using the criteria that might have been used to design the conservation collections had genetic factors been used in planning. First, *ex situ* and restored populations were evaluated to determine whether they had captured an amount of genetic diversity similar to that of wild populations if those populations were not genetically depauperate (Lacy 1994; Falk *et al.* 2001; Hufford & Mazer 2003; Menges 2008). Second, these populations were evaluated for their genetic accuracy or authenticity in terms of capturing allele complements seen in natural populations (Clewell 2000; Falk *et al.* 2001). Third, population differentiation and structure among the genetic sources were assessed in order to evaluate the chances that mixing of these genetic sources could lead to outbreeding depression due to differences

among them, such as local adaptations or coadapated gene complexes (Fenster & Dudash 1994, Pavlik 1996; Montalvo & Ellstrand 2001; Hufford & Mazer 2003; Allendorf & Luikart 2007). Mixing of sources has in some cases been observed to have a negative effect (e.g., Montalvo & Ellstrand 2001; Sanders & McGraw 2005), while in others this strategy has been essential or beneficial, especially when available source populations were genetically depauperate (e.g., Demauro 1993; Gustafson *et al.* 2002).

The *ex situ* collection of plants from extirpated populations and the extant Wild B population captured a high amount of genetic diversity, similar to that found in wild populations, except in cases where the sample size was necessarily small because of the number of plants available in the collection or remaining in a wild population (i.e. Wild C). Of the restored populations, Restored Railroad and Restored Y captured a high amount of genetic diversity, but the Restored Z outplanting was relatively genetically depauperate compared to these other two. The design of the Restored Y outplanting, which involved propagules of multiple genetic sources, was a success in that it caused more genetic diversity, by some measures, to be captured in this population than in some of the wild populations (Wild B and Wild C). The greatest success, in terms of preserving genetic diversity, was the rescue of the Railroad population. This genetic population, split among the *ex situ* collection and the Restored Railroad outplanting, was especially genetically diverse and without its salvage a valuable genetic resource for conservation of *A. crenulata* would have been lost.

Another way the amount of genetic diversity captured by the various populations was assessed was to determine which populations had private (unique) alleles because preservation of these alleles adds to the total genetic

diversity retained for the species. Again, plants rescued from the extirpated Railroad population provided the best resource for preserving unique variation, as it had many more private alleles than any of the other populations and more than all other populations combined. Even the small Wild C and *ex situ* 84th St. populations retained unique alleles, increasing their value in terms of preserving overall genetic diversity for the species. The two populations represented by single individuals had no unique alleles and thus added very little to increasing the genetic diversity observed for *A. crenulata* at the eight loci sampled. However, it should be noted that rare alleles may not be crucial for species conservation because they will contribute little to the evolutionary potential and adaptation to changing environmental conditions of the species due to their low frequencies (Brown & Briggs 1991; Holsinger *et al.* 1999). However, it is better to err on the side of conserving more rather than less genetic diversity in *ex situ* conservation and restoration efforts (Center for Plant Conservation 1991).

Overall the most success in terms of capturing high genetic diversity was through larger sample size (Honnay & Jacquemyn 2007). The larger wild populations and larger conservation collection and restored populations had the most genetic diversity and highest numbers of private alleles. However, this underscores the failure of the Restored Z outplanting based on the genetic diversity measure because it is one of the largest populations of *A. crenulata* and was the population with the greatest sampling in this study. So, although it might be judged successful using the abundance criterion, it is likely not a resilient population because the choice of founding individuals has created a genetic bottleneck (Hufford & Mazer 2003; Menges 2008) and the likelihood of its future persistence is questionable. The utility of the small

populations of the *ex situ* collection as future restoration sources may be limited due to low census and consideration of the limited success of using seedlings as propagules in reintroductions of *A. crenulata* as opposed to the success of transplantation of whole plants (Wendelberger *et al.* 2008). However, the large and genetically diverse *ex situ* Railroad population will be a great resource for future restorations.

The accuracy of replication of natural genetic diversity in reconstructed populations was able to be evaluated in the two cases where two samples from a single genetic source were obtained: Wild B vs. *ex situ* Wild B and *ex situ* Railroad vs. Restored Railroad. Taking into account genetic diversity measures as well as private alleles, genetic accuracy was good, but not perfect. In both cases, genetic diversity on the whole was similar except for the number of allele phenotypes in the case of the Wild B samples. The fact that only half as many individuals were sampled from the *ex situ* collection as the wild population might account for this discrepancy. However, this pattern might also be a permanent feature of the *ex situ* collection because its total size is approximately 10% of the natural population. Accuracy judged by private alleles was much worse than when measured using genetic diversity. In both cases many of the alleles observed in one sample of the gene pools were not present in the other sample (Table 5.3). More than twice as many alleles were captured in the Wild B population as in the *ex situ* collection and roughly twice as many were captured in Restored Railroad as opposed to the *ex situ* Railroad collection.

Detection of population structure was useful in identifying differentiated populations, perhaps having different evolutionary trajectories, which should be considered in conservation planning. Bayesian clustering

analysis including both the wild and extirpated populations extended the geographical clustering of sampled populations and did not reveal any additional genetic clusters over the three clusters observed when wild individuals were analyzed alone. The Railroad population, which was originally located in very close geographical proximity to the Wild A population at the northernmost edge of the species range formed a cluster with that population with majority assignment of almost all individuals to the green cluster (Figure 5.3). The Wild B population was differentiated from this group and the southernmost geographic group, which clustered genetically with majority assignment of individuals from Wild C, Wild D, Snapper Creek Canal, and 84th St. to the red cluster. The individual in the Wild B population that showed a high proportion of membership in the green cluster and the individual in the Restored Railroad population that showed a high proportion of membership in the blue cluster might be examples of gene flow among the Wild B population and the Wild A/Railroad populations, most likely through movement of pollen. The individual in the *ex situ* 84th St. collection that showed a high proportion of membership in the Wild B cluster could be an example of gene flow between the northern and southern groups of populations, or alternatively a possible example of switched labels on plants in the *ex situ* collection.

Analysis of population structure in *A. crenulata* revealed three differentiated genetic groups even though the farthest distance between populations is less than 10 km. Genetic differentiation among these populations suggested that these populations could be adapted to local conditions in the northern or southern part of the species range. If so, outbreeding depression from crosses among groups could be a problem for

reintroductions involving individuals from multiple genetic sources. Recruitment is essential for restoration success (Menges 2008), so outbreeding depression could be especially detrimental for a species, such as *A. crenulata*, that may have difficulty achieving persistence and abundance due to poor recruitment (Wendelberger & Maschinski 2009).

In their evaluation of the success of the different propagule types used in the Restored Y outplanting, Wendelberger *et al.* (2008) found that seedlings grown from seeds originating from open pollination in the *ex situ* collection, and thus presumably having mixed genetic sources in many cases, had the highest mortality, with 0% survival compared to survival of seedlings collected at wild sites, which had a survival rate of over 50%. This could indicate that outbreeding depression, rather than heterosis, is a more likely outcome of crossing of genetic clusters in *A. crenulata*; although it is possible that some new genetic combinations could thrive. The long-term outlook for Restored Y may be poor because disparate gene pools will likely be mixed in progeny at this site and contribute to low recruitment (see Wendelberger *et al.* 2008); whereas the outlook for restored populations established from a single genetic source, such as Restored Railroad, is more positive and could be considered more successful conservation efforts. Further study of the effects of combining gene pools on establishment and fitness of *A. crenulata* is warranted (Hufford & Mazer 2003).

In summary, conservation efforts for *Amorpha crenulata* have by and large been successful from a genetic standpoint. High genetic diversity was captured, although genetic accuracy was questionable. The most successful action was rescue of the Railroad population to the *ex situ* collection and Restored Railroad outplanting. Future monitoring of the Restored Railroad

and Restored Y populations will help determine how the use of one or multiple genetic sources affects the long-term success of reintroduced populations and which strategy will be considered most successful. The least successful conservation effort was the Restored Z outplanting, which is genetically depauperate.

Implications for Future Conservation Management of *Amorpha crenulata*

The primary goals of preventing the extinction and achieving stabilization of *Amorpha crenulata* have been furthered by development of the *ex situ* conservation collection as a resource for restoration efforts, and by creation of restored populations as a first step to increasing this rare species' presence in natural habitats. For future conservation efforts, consideration of genetic and ecological factors will be important for recruitment and progression of restored populations to become self-sustaining. The genetic source used in population restoration can determine whether or not that population will become self-sustaining (Falk *et al.* 2001). The observation that single genetic sources have high genetic diversity eliminates the concern that mixing of source populations is needed to increase genetic diversity. The presence of population structure among *A. crenulata* populations also argues that single genetic sources should be used in future reintroductions of this species in order to avoid outbreeding depression. Use of single genetic sources will also increase the accuracy of genetic replication of wild populations (Falk *et al*. 2001), but it should be noted that translocated propagules might have local adaptation to sites other than the restored site. If populations are established far from source populations, as is likely to be the case for *A. crenulata* due to near total habitat destruction in its native

range (e.g., Wendelberger *et al.* 2008), creation of some populations of mixed ancestry might be advised to allow for new combinations of traits that could spur adaptation to the new sites. If additional reintroductions occur within the native range, source material might best be chosen to represent the genetic cluster nearest that locality for the best chance of individuals to be well suited for the local environment (McKay *et al.* 2005).

Detection of three genetic clusters also argues that future conservation efforts should involve all three of these unique groups. The *ex situ* collection will be an invaluable resource in restoring the genetic group comprised of the southern populations, which are represented by fewer than twenty individuals in natural populations and whose total population size was doubled through rescue of individuals from the 84th St. population. Efforts to increase numbers for the Wild B genetic cluster have benefitted by population replication in the *ex situ* collection as a measure to buffer it against environmental and demographic stochasticity. Further augmentation of the *ex situ* Wild B collection would provide benefits in terms of genetic diversity and authenticity. The largest remaining genetic cluster of individuals comprised of the Wild A and Railroad populations is already well represented in wild, *ex situ*, and restored populations and will be an important resource for future conservation work. An additional recommendation stemming from these findings is to ensure that all genetic clusters are well represented in the *ex situ* seed bank for *A. crenulata.*

As with any conservation program, preservation of *Amorpha crenulata* will depend on an integrated approach (e.g., Williams 2001; Nishihiro *et al.* 2009) involving full consideration of ecological, demographic, and genetic factors. Much progress has been made in evaluating various aspects of the

biology of this species (Roncal *et al*. 2006; Wendelberger *et al*. 2008; this study). As more information about the habitat requirements and conditions necessary for recruitment of this species are learned and additional reintroduction efforts are made with consideration of genetic parameters, progress will be made towards the goals of maintaining or increasing abundance, extent, resilience and persistence for *A. crenulata*.

REFERENCES

Allendorf FW, Luikart G (2007) *Conservation and the Genetics of Populations.* Blackwell Publishing Ltd, Malden, Massachusetts.

Barrett SCH, Kohn JR (1991) Genetic and evolutionary consequences of small population size in plants: Implications for conservation. In: *Genetics and Conservation of Rare Plants* (eds. Falk DA, Holsinger KE), pp. 3-30, Oxford University Press, New York, New York.

Brown AHD, Briggs JD (1991) Sampling strategies for genetic variation in *ex situ* collections of endangered plant species. In: *Genetics and Conservation of Rare Plants* (eds. Falk DA, Holsinger KE), pp. 99-119, Oxford University Press, New York, New York.

Center for Plant Conservation (1991) Genetic sampling guidelines for conservation collections of endangered plants. In: *Genetics and Conservation of Rare Plants* (eds. Falk DA, Holsinger KE), pp. 225-238, Oxford University Press, New York, New York.

Clewell AF (2000) Restoring for natural authenticity. *Ecological Restoration*, **18**, 216-217.

De Silva HN, Hall AJ, Rikkerink E, McNeilage MA, Fraser LG (2005) Estimation of allele frequencies in polyploids under certain patterns of inheritance. *Heredity*, **95**, 327-334.

Demauro MM (1993) Relationship of breeding system to rarity in the lakeside daisy (*Hymenoxys acaulis* var. *glabra*). *Conservation Biology*, **7**, 542-550.

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*, **19**, 11-15.

Ellstrand NC, Elam DR (1993) Population genetic consequences of small population-size - Implications for plant conservation. *Annual Review of Ecology and Systematics*, **24**, 217-242.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611-2620.

Falk DA, Knapp EE, Guerrant EO (2001) *An introduction to restoration genetics*. [online data], Society for Ecological Restoration International, Tucson, Arizona. Available at: http://www.ser.org/reading_resources.asp. Accessed 8 December 2009.

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics*, **164**, 1567-1587.

Falush D, Stephens M, Pritchard JK (2007) Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Molecular Ecology Notes*, **7**, 574-578.

Fant JB, Holmstrom RM, Sirkin E, Etterson JR, Masi S (2008) Genetic structure of threatened native populations and propagules used for restoration in a clonal species, American Beachgrass (*Ammophila breviligulata* Fern.). *Restoration Ecology*, **16**, 594-603.

Federal Register (1984) Endangered and threatened wildlife and plants: Proposed endangered status for five Florida pine rockland plants. *Federal Register*, **49**, 217, 44507-44512.

Federal Register (1985) Endangered and threatened wildlife and plants: Endangered and threatened status for five Florida pine rockland plants. *Federal Register*, **50**, 138, 29345-29349.

Fenster CB, Dudash MR (1994) Genetic considerations for plant population restoration and conservation. In: *Restoration of Endangered Species: Conceptual Issues, Planning, and Implementation* (eds. Bowles ML, Whelan CJ), pp. 34-62, Cambridge University Press, Cambridge.

Frankham R (2003) Genetics and conservation biology. *Comptes Rendus Biologies*, **326** (Suppl 1), S22-S29.

Frankham R, Ralls K (1998) Conservation biology - Inbreeding leads to extinction. *Nature*, **392**, 441-442.

Fredua-Agyeman R, Adamski D, Liao RJ, Morden C, Borthakur D (2008) Development and characterization of microsatellite markers for analysis of population differentiation in the tree legume *Acacia koa* (Fabaceae: Mimosoideae) in the Hawaiian Islands. *Genome*, **51**, 1001-1015.

González-Pérez MA, Dolores Lledo M, Lexer C, *et al.* (2009) Genetic diversity and differentiation in natural and reintroduced populations of *Bencomia exstipulata* and comparisons with *B. caudata* (Rosaceae) in the Canary Islands: An analysis using microsatellites. *Botanical Journal of the Linnean Society*, **160**, 429-441.

Gustafson DJ, Gibson DJ, Nickrent DL (2002) Genetic diversity and competitive abilities of *Dalea purpurea* (Fabaceae) from remnant and restored grasslands. *International Journal of Plant Sciences*, **163**, 979-990.

Hamilton JA, Eckert CG (2007) Population genetic consequences of geographic disjunction: A prairie plant isolated on Great Lakes alvars. *Molecular Ecology*, **16**, 1649-1660.

Hardy OJ, Vekemans X (2002) SPAGeDi: A versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes*, **2**, 618-620.

Hedrick PW (2001) Conservation genetics: Where are we now? *Trends in Ecology & Evolution*, **16**, 629-636.

Helenurm K, Parsons LS (1997) Genetic variation and the reintroduction of *Cordylanthus maritimus* ssp. *maritimus* to Sweetwater Marsh, California. *Restoration Ecology*, **5**, 236-244.

Holsinger KE, Mason-Gamer RJ, Whitton J (1999) Genes, demes, and plant conservation. In: *Genetics and the extinction of species: DNA and the conservation of biodiversity* (eds. Landweber LF, Dobson AP), pp. 23-46, Princeton University Press, Princeton, New Jersey.

Honnay O, Bossuyt B (2005) Prolonged clonal growth: Escape route or route to extinction? *Oikos*, **108**, 427-432.

Honnay O, Jacquemyn H (2007) Susceptibility of common and rare plant species to the genetic consequences of habitat fragmentation. *Conservation Biology*, **21**, 823-831.

Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources*, **9**, 1322-1332.

Huenneke LF (1991) Ecological implications of genetic variation in plant populations. In: *Genetics and Conservation of Rare Plants* (eds. Falk DA, Holsinger KE), pp. 31-44, Oxford University Press, New York, New York.

Hufford KM, Mazer SJ (2003) Plant ecotypes: Genetic differentiation in the age of ecological restoration. *Trends in Ecology & Evolution*, **18**, 147-155.

Isely D (1986) Notes about *Psoralea* sensu Auct., *Amorpha, Baptisia, Sesbania,* and *Chamaecrista* Leguminosae in the Southeastern USA. *Sida Contributions to Botany*, **11**, 429-440.

Isely D (1998) *Native and Naturalized Leguminosae (Fabaceae) of the United States.* Brigham Young University, Provo, Utah.

Lacy RC (1994) Managing genetic diversity in captive populations of animals. In: *Restoration of endangered species: Conceptual issues, planning, and implementation* (eds. Bowles ML, Whelan CJ), pp. 63-89, Cambridge University Press, New York, New York.

Lande R (1999) Extinction risks from anthropogenic, ecological, and genetic factors. In: *Genetics and the Extinction of Species: DNA and the Conservation of Biodiversity* (eds. Landweber LF, Dobson AP), pp. 1-22, Princeton University Press, Princeton, NJ.

Le Gouar P, Rigal F, Boisselier-Dubayle MC, *et al.* (2008) Genetic variation in a network of natural and reintroduced populations of Griffon vulture (*Gyps fulvus*) in Europe. *Conservation Genetics*, **9**, 349-359.

Lesica P, Allendorf FW (1999) Ecological genetics and the restoration of plant communities: Mix or match? *Restoration Ecology*, **7**, 42-50.

Marrs RA, Sforza R, Hufbauer RA (2008) Evidence for multiple introductions of *Centaurea stoebe micranthos* (spotted knapweed, Asteraceae) to North America. *Molecular Ecology*, **17**, 4197-4208.

McKay JK, Christian CE, Harrison S, Rice KJ (2005) "How local is local?" - A review of practical and conceptual issues in the genetics of restoration. *Restoration Ecology*, **13**, 432-440.

Meirmans PG, van Tienderen PH (2004) GENOTYPE and GENODIVE: Two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes*, **4**, 792-794.

Menges ES (2008) Restoration demography and genetics of plants: When is a translocation successful? *Australian Journal of Botany*, **56**, 187-196.

Montalvo AM, Williams SL, Rice KJ, *et al.* (1997) Restoration biology: A population biology perspective. *Restoration Ecology*, **5**, 277-290.

Montalvo AM, Ellstrand NC (2001) Nonlocal transplantation and outbreeding depression in the subshrub *Lotus scoparius* (Fabaceae). *American Journal of Botany*, **88**, 258-269.

Moody ME, Mueller LD, Soltis DE (1993) Genetic variation and random drift in autotetraploid populations. *Genetics*, **134**, 649-657.

Nei M (1987) *Molecular Evolutionary Genetics.* Columbia University Press, New York, New York.

Nishihiro J, Uesugi R, Takagawa S, Washitani I (2009) Toward the restoration of a sustainable population of a threatened aquatic plant, *Nymphoides peltata*: Integrated genetic/demographic studies and practices. *Biological Conservation*, **142**, 1906-1912.

Obbard DJ, Harris SA, Pannell JR (2006) Simple allelic-phenotype diversity and differentiation statistics for allopolyploids. *Heredity*, **97**, 296-303.

Pavlik BM (1996) Defining and measuring success. In: *Restoring Diversity: Strategies for Reintroduction of Endangered Plants* (eds. Falk DA, Millar CI, Olwell M), pp. 127-155, Island Press, Washington, D.C.

Possley J, Fellows M, Wendelberger K (2004) *Conservation Action Plan – Public Version Amorpha herbacea var. crenulata* [online data]. Fairchild Tropical Botanic Garden, Coral Gables, Florida. Available at: http://www.fairchildgarden.org/indes.cfm?section=centerfortropicalplantconservation&subsection=speciesandhabitatconservation&page=speciesweworkwith. Accessed 6 January 2010.

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

Prichard JK, Wen X, Falush D (2009) *Documentation for structure software: Version 2.3.* [online data], Available at: http://pritch.bsd.uchicago.edu/structure.html. Accessed 16 December 2009.

Ramp JM, Collinge SK, Ranker TA (2006) Restoration genetics of the vernal pool endemic *Lasthenia conjugens* (Asteraceae). *Conservation Genetics*, **7**, 631-649.

Roncal J, Fisher JB, Fellows MQN, Wendelberger KS, Maschinski J, Fidelibus MW (2006) Propagation protocol for the endangered Crenulate Lead Plant *Amorpha herbacea* var. *crenulata*. *Native Plants Journal*, **7**, 89-93.

Saccheri I, Kuussaari M, Kankare M, Vikman P, Fortelius W, Hanski I (1998) Inbreeding and extinction in a butterfly metapopulation. *Nature*, **392**, 491-494.

Sanders S, McGraw JB (2005) Population differentiation of a threatened plant: Variation in response to local environment and implications for restoration. *Journal of the Torrey Botanical Society*, **132**, 561-572.

Seddon PJ, Armstrong DP, Maloney RF (2007) Developing the science of reintroduction biology. *Conservation Biology*, **21**, 303-312.

Soltis PS, Soltis DE (2009) The role of hybridization in plant speciation. *Annual Review of Plant Biology*, **60**, 561-588.

Soltis PS, Soltis DE (2000) The role of genetic and genomic attributes in the success of polyploids. *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 7051-7057.

Spielman D, Brook BW, Frankham R (2004a) Most species are not driven to extinction before genetic factors impact them. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 15261-15264.

Spielman D, Brook BW, Briscoe DA, Frankham R (2004b) Does inbreeding and loss of genetic diversity decrease disease resistance? *Conservation Genetics*, **5**, 439-448.

Straub SCK, Reveal JL, Weakley AS (in press) *Amorpha*. In: *Flora of North America North of Mexico* (eds. Flora of North America Editorial Committee) 12+ vols. New York and Oxford.

Straub SCK, Bogdanowicz SM, Doyle JJ (2009) Characterization of twelve polymorphic microsatellite markers for Georgia false indigo (*Amorpha georgiana* Wilbur var. *georgiana*), an endangered species, and their utility in other dwarf *Amorpha* L. species. *Molecular Ecology Resources*, **9**, 225-228.

Straub SCK, Doyle JJ (2009) Conservation genetics of *Amorpha georgiana* (Fabaceae), an endangered legume of the Southeastern United States. *Molecular Ecology*, **18**, 4349-4365.

Tani N, Yoshimaru H, Kawahara T, Hoshi Y, Nobushima F, Yasui T (2006) Determination of the genetic structure of remnant *Morus boninensis* Koidz. trees to establish a conservation program on the Bonin Islands, Japan. *BMC Ecology*, **6**, 14.

Tehrani MS, Mardi M, Sahebi J, Catalan P, Diaz-Perez A (2009) Genetic diversity and structure among Iranian tall fescue populations based on genomic-SSR and EST-SSR marker analysis. *Plant Systematics and Evolution*, **282**, 57-70.

Travis SE, Proffitt CE, Lowenfeld RC, Mitchell TW (2002) A comparative assessment of genetic diversity among differently-aged populations of *Spartina alterniflora* on restored versus natural wetlands. *Restoration Ecology*, **10**, 37-42.

Vinson CC, Ribeiro DO, Harris SA, Sampaio I, Ciampi AY (2009) Isolation of polymorphic microsatellite markers for the tetraploid *Dipteryx odorata*, an intensely exploited Amazonian tree species. *Molecular Ecology Resources*, **9**, 1542-1544.

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, **38**, 1358-1370.

Wendelberger KS, Fellows MQN, Maschinski J (2008) Rescue and restoration: Experimental translocation of *Amorpha herbacea* Walter var. *crenulata* (Rydb.) Isely into a novel urban habitat. *Restoration Ecology*, **16**, 542-552.

Wendelberger KS, Maschinski J (2009) Linking geographical information systems and observational and experimental studies to determine optimal seedling microsites of an endangered plant in a subtropical urban fire-adapted ecosystem. *Restoration Ecology*, **17**, 845-853.

Williams SL (2001) Reduced genetic diversity in eelgrass transplantations affects both population growth and individual fitness. *Ecological Applications*, **11**, 1472-1488.

Wolff K, Houston K, Dunbar-Co S (2009) Cross-species amplification of primers developed from *Plantago major* and *P. intermedia* in two Hawaiian *Plantago* species from the section *Plantago*. *Molecular Ecology Resources*, **9**, 981-984.

Young A, Boyle T, Brown T (1996) The population genetic consequences of habitat fragmentation for plants. *Trends in Ecology & Evolution*, **11**, 413-418.

Young TP (2000) Restoration ecology and conservation biology. *Biological Conservation*, **92**, 73-83.

APPENDIX A

A SURVEY OF GENOME SIZE IN *AMORPHA FRUTICOSA* USING FLOW CYTOMETRY

A flow cytometry study was undertaken in order to estimate genome size and explore genome size variation in *Amorpha fruticosa*, a highly morphologically variable polyploid species. Genome size was surveyed in 14 accessions of *Amorpha fruticosa* obtained from the USDA germplasm collection. Collection sites of these accessions spanned much of the wide range of the species. From one to three plants per seed accession were grown at Cornell University in 2007.

For flow cytometry, a procedure modified from Arumuganathan and Earle (1991) and Meagher and Costich (1996) was used. All samples were prepared and incubated on ice. Approximately 50 mg of fresh leaf tissue were finely sliced in petri dishes containing 1 mL of cold, freshly-prepared buffer (10 mM MgSO4, 50 mM KCl, 5mM Hepes, 0.099% dithiothreitol, 0.247% Triton X-100). The homogenate was then filtered through a 35 micron nylon mesh cell strainer cap into a flow-cytometry collection tube (BD Biosciences). Released nuclei in each sample were then treated with 10 µl of propidium iodide (5 mg/ml) and 5 µL RNase (10 mg/ml) and incubated for at least 15 min. Trout erythrocyte nuclei TEN cytometry control (BioSure) was used as a size standard. Samples were run on a FACS Caliber flow cytometer (BD Biosciences) at the Cornell University College of Veterinary Medicine following manufacturer's recommendations. This process was repeated for each sample on three different days. DNA content

was calculated using the formula given in Dolezel and Bartos (2005) and average DNA content calculated for each accession.

No variation in ploidy and little variation in nuclear DNA content was detected for *A. fruticosa*, whose average 1C value was 1.465 pg (Table A1). Using the conversion factor of Dolezel et al. (2003), the genome size of *A. fruticosa* was estimated to be approximately 1.43 Gbp.

Table A1. Estimates of genome size for *Amorpha fruticosa*. The 2C value, or nuclear DNA content of unreplicated DNA of somatic cells, is reported. Measurements represent averages and standard deviations calculated for the three replicates per individual and 2 – 3 individuals per accession. If no standard deviation is given, only one individual was available for that accession.

USDA Accession	Provenance	2C Value (pg of DNA)
PI 303182	Idaho, U.S.A.	2.99 (±0.09)
PI 314213	Former U.S.S.R. (cultivated)	2.84 (±0.09)
PI 372505	Ontario, Canada (cultivated)	2.83 (±0.10)
PI 436707	Texas, U.S.A.	3.00 (±0.10)
PI 436708	Oklahoma, U.S.A.	2.96 (±0.10)
PI 436709	Oklahoma, U.S.A.	2.95 (±0.14)
PI 436710	Oklahoma, U.S.A.	2.99 (±0.13)
PI 436711	Oklahoma, U.S.A.	3.00 (±0.06)
PI 436712	Oklahoma, U.S.A.	2.96 (±0.05)
XDL 90-0274	Colorado, U.S.A. (cultivated)	2.89 (±0.16)
XDL 91-0432	Virginia, U.S.A.	2.92 (±0.07)
XDL 92-0268	Virginia, U.S.A.	2.94
XDL 94-0175	New Mexico, U.S.A.	2.89 (±0.12)
XDL 94-0176	New Mexico, U.S.A.	2.87
	Species Average :	**2.93 (±0.12)**

REFERENCES

Arumuganathan K, Earle ED (1991) Estimation of nuclear DNA contents of plants by flow cytometry. *Plant Molecular Biology Reporter*, **9**, 3, 229-241.

Dolezel J, Bartos J (2005) Plant DNA flow cytometry and estimation of genome size. *Annals of Botany*, **95**, 99-110.

Dolezel J, Bartos J, Voglmayr H, Greilhuber J (2003) Nuclear DNA content and genome size of trout and human. *Cytometry*, **51**, 127–128.

Meagher TR, Costich DE (1996) Nuclear DNA content and floral evolution in *Silene latifolia*. *Proceedings of the Royal Society of London B*, **263**, 1455 – 1460.