ABSTRACT

Title of Document:	ELUCIDATING THE MACRO- AND MICRO-EVOLUTIONARY RELATIONSHIPS OF THE FEDERALLY LISTED ENDANGERED SPECIES *AGALINIS ACUTA* (OROBANCHACEAE).
	James Beaton Pettengill, Ph.D. 2010
Directed By:	Associate Professor, Dr. Maile C. Neel, Department of Plant Sciences and Landscape Architecture and Department of Entomology

Agalinis acuta (Orobanchaceae) is a federally listed endangered plant species native to the mid-Atlantic and northeastern coastal plains of the United States. Due to morphological ambiguity and molecular similarity between *A. acuta* and *Agalinis tenella* and *Agalinis decemloba* a conservation priority is to determine whether *A. acuta* represents an evolutionarily distinct entity worthy of protection under the Endangered Species Act. To resolve this question, a phylogenetic study was first conducted based on seven chloroplast DNA loci and the nuclear DNA locus ITS from 79 individuals representing 29 *Agalinis* species. A study evaluating the utility of those cpDNA loci and three analytical techniques for the purpose of DNA barcoding was also conducted. The phylogenetic study indicated that *A. acuta* was perhaps evolutionarily indistinct from *A. decemloba* and *A. tenella*. Based on the results of subsequent analyses of 21 microsatellite loci and morphological data evaluated under myriad species concepts, *A.*

acuta, *A. decemloba*, and *A. tenella* best represent a single species with two subspecies; the former two putative species would constitute a subspecies called *A. decemloba* ssp. *decemloba* and *A. tenella* would be *A. decemloba* ssp. *tenella*.

With evolutionary distinct entities described, a phylogeographic study was conducted to determine the extent to which historical processes rather than contemporaneous events can explain extant patterns of genetic and phenotypic diversity within *A. decemloba*. The dispersal of a few individuals out of southern refugial populations likely represents the process through which northern populations were established; however, recent anthropogenic effects that disproportionately affected northern populations may have also contributed to extant patterns of diversity. Neutral or adaptive explanations for phenotypic variation among populations are also investigated.

The conservation implications of population genetic analyses were assessed for members of *A. decemloba* ssp. *decemloba*. Despite the evidence that this taxon is self-compatible, the high levels of inbreeding and low levels of heterozygosity are of such a magnitude in certain populations that genetic factors may be negatively impacting fitness. Because of the small effective population sizes and degree of isolation, all populations should be managed to reduce the risk of extinction associated with demographic and environmental stochasticity.

ELUCIDATING THE MACRO- AND MICRO-EVOLUTIONARY
RELATIONSHIPS OF THE FEDERALLY LISTED ENDANGERED SPECIES
AGALINIS ACUTA (OROBANCHACEAE)

By

James Beaton Pettengill

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2010

Advisory Committee:
Professor Dr. Maile C. Neel, Chair
Dr. Sheri A. Church
Dr. Michael P. Cummings
Dr. David J. Hawthorne
Dr. Charles Mitter

PREFACE

This dissertation contains an overall abstract, introduction, general objectives, five chapters, and three appendices. Each chapter contains an abstract and is represented in manuscript form; background information and methods may be repeated and pronoun usage reflects manuscript authorship. The tables and figures appear at the end of each chapter and the numbering of those also reflects the syntax associated with a manuscript (i.e., the "S" in the title of some tables or figures identifies them as representing supplemental material). A bibliography that includes all references cited throughout the dissertation is found at the very end.

DEDICATION

To my parents, John and Marian Pettengill, for their unwavering support and acceptance.

ACKNOWLEDGMENTS

I would first like to thank my advisor Dr. Maile C. Neel for providing me the opportunity to conduct the research described in this dissertation; I am indebted to her for the support and guidance she provided over the years and thank her for instilling in me the rigor required to be a successful research scientist. I would also like to thank the members of committee: Dr. Michael P. Cummings for his guidance, support, and excellent understanding of molecular evolution; Dr. Charlie Mitter for introducing me to the study of phylogenetics and always encouraging me to think harder about the fundamentals; and, Drs. Sheri Church and Dave Hawthorne for their flexibility in joining my committee and insightful comments about how I could make my research stronger.

I also acknowledge the support provided by the environment fostered within the Neel lab. In particular, I thank Michael Lloyd for sharing his expertise in programming, discussion of evolutionary issues in general, and his overall camaraderie. I appreciate the help provided by Robert Burnett in accomplishing laboratory work under a strict deadline, and I am grateful for the company he provided during my time at the University of Maryland.

I am thankful for the academic community that is the Behavior, Ecology, Evolution and Systematics (BEES) Graduate Program at the University of Maryland, which exposed me to the multifaceted field of evolutionary biology. I also thank members of Dr. Charles Delwiche's lab for allowing me to spend countless hours using the proprietary software located in that lab; I thank Dr. Ruth Timme and Dr. John Hall for the help and insight provided during those hours. I thank the weekly Systematics Seminar at the University of Maryland that has over the past years exposed me to many important

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

GENERAL INTRODUCTION

Given that species are the fundamental unit of study within many disciplines (e.g., conservation biology), a great deal of literature has been published regarding what constitutes a species, with much of the contemporary discussion rooted in the merging of genetics and evolution during the Modern Evolutionary Synthesis (Dobzhansky 1937; Mayr 1942; Stebbins 1950). There are also a plethora of empirical methods that have been advocated as appropriate for delimiting the taxonomic boundaries associated with species (Sites & Marshall 2004). However, the numerous species concepts and delimitation methods that have been developed make determining when a collection of individuals or populations warrants recognition as a species challenging as well as highly contentious (Coyne & Orr 2004). Despite this controversy, the potentially negative consequences associated with inaccurate taxonomy (e.g., the importance of accurate estimates of species richness in ecology and inefficient use of funds for conservation management due to erroneous taxonomy) illustrate the importance of being able to resolve such issues (Isaac et al. 2004). As a means of overcoming the controversy, de Quieroz (2007) has suggested that there is an underlying concept that unifies the myriad species concepts: that species are "separately evolving segments of metapopulation lineages". In contrast to early typological views of species, de Queiroz (2007) embraces the fact that in absence of gene flow, there is a continuum of evolutionary differentiation that proceeds after a speciation event (Cummings et al. 2008).

Given that alternative species concepts emphasize different characteristics that will be acquired at different times following a speciation event, adherence to different concepts will potentially result in different species boundaries (de Queiroz 2007).

Further, different types of data will be informative at different stages of differentiation (Baum 1998; Marshall et al. 2006). Given this continuum of variation and the fact that investigators have little idea about where putative species may lie on that continuum, it is not possible to recommend a single type of data or exact characteristic that will define a species in all cases. However, a strategic and efficient approach to resolving questions of taxonomic uncertainty is to begin with applying the most restrictive definitions and delimitation methods (e.g., genealogical exclusivity; Baum and Shaw 1995). Additional analyses that can detect more subtle differences would be conducted only if the more restrictive definitions are not met. Results of subsequent analyses would be evaluated in light of the properties associated with a range of commonly used species concepts and delimitation criteria. Analyses would also include taxa beyond the specific entities of concern and sampling intensively within the entities of concern. This combination of broad and deep sampling provides context that is needed to understand the magnitude of differences that distinguish closely related entities and allows assessment of whether those differences are of sufficient magnitude to warrant the taxonomic rank of species (Baum 2009).

The ability to determine the accuracy of historical taxonomic alignments is particularly important within the field of conservation biology where it is assumed that a species is evolutionarily distinct upon being listed under the Endangered Species Act (ESA) (U.S.C. 1973). Issues of taxonomic uncertainty are among the concerns for many rare species and have also played a large role in the delisting of species. As of April 3, 2007 1,326 species were listed as endangered and 41 of these have been delisted. That 8 of those 41 delistings were due to taxonomic revisions illustrates the importance of

adequate methods for delimiting species. Given the direct impact taxonomic status has on the listing of a species (Fallon 2007) and that a taxonomic revision has on the allocation of limited funds for the purposes of conservation management, it is important to have a strategic and efficient approach to resolving questions of taxonomic uncertainty.

As well as determining the accuracy of historical taxonomic alignments ascribing individuals to a species, there is the need to be able to quickly identify the species to which an individual of unknown taxonomic identity belongs. Given the increasing ease with which sequence data can be obtained from a range of taxa, DNA barcoding has emerged as a potential method to determine the taxonomic identity of an individual by sequencing a small portion of its genome and comparing this nucleotide sequence with sequences in a reference database (e.g., Hebert et al. 2003). DNA barcoding should be a valuable tool to a number of disciplines including conducting rapid biodiversity assessments, forensics, detecting illegal wildlife trade, identifying species during cryptic life stages, and monitoring invasive species (Armstrong & Bar 2005; Darling & Blum 2007; Dawnay et al. 2007; Little & Stevenson 2007; Vogler 2006). Barcoding has also been shown to be a useful tool in the discovery of new species (Burns et al. 2008; Murray et al. 2008; Yassin et al. 2008). However, the grand promises made by proponents of DNA barcoding have generated concern and there is extensive debate over exactly what it can contribute to various disciplines (e.g., Rubinoff 2006; Trewick 2008; Will et al. 2005). In particular, DNA barcoding may be problematic when used to differentiate among sequences representing closely related species where the issues of inaccurate taxonomy and incomplete lineage sorting are likely to be most prevalent.

With taxonomic questions resolved, the extent to which observed patterns of intraspecific genetic diversity are the result of historical processes related to the establishment of populations (e.g., founder or vicariance events) rather than recent anthropogenically induced changes (Eckert et al. 2008) can be investigated. The likelihood of historical events, and the time over which they may have occurred, is strongly dependent on past environmental conditions. For example, with regard to eastern North America, the environment associated with the Wisconsin glaciation (maximum at about 18,000 yr BP) during the Pleistocene was vastly different than current conditions where ice sheets reached as far south as 40º N (e.g., New York, NY) and tundra and boreal forest habitats extended even further south (Hewitt 2000; Lomolino et al. 2006). Therefore, extant species or populations may have migrated relatively recently into northern regions that historically were inhabitable (i.e., "the leading edge hypothesis;" Cwynar & Macdonald 1987; Hewitt 1996; Soltis et al. 1997) or northern populations persisted throughout the ice ages of the Pleistocene as refugia (i.e., "north-south recolonization hypothesis"; Soltis et al. 1997). In addition to understanding the processes responsible for extant patterns, differentiating between historical and anthropogenic hypotheses as causes for extant patterns is particularly important to conservation geneticists (e.g., Crandall et al., 2000; Eckert et al. 2008).

Once issues of taxonomic uncertainty have been resolved regarding putative taxa of conservation concern, population genetic analyses can be conducted to determine whether a taxon faces an increased extinction risk as a result of genetic factors. Such information is important because it can be used to better ensure the efficient use of management resources (e.g., Haig 1998). Through the analysis of molecular variation

(e.g., microsatellite loci), population genetics can help quantify the amount of inbreeding, the degree of isolation, and whether populations have experienced a bottleneck, all of which are related to the probability of persistence of a population or species (Ellstrand & Elam 1993; Frankham 1995; Frankham et al. 2002; Luikart & Cornuet 1998). Molecular data can also provide a measure of the genetic diversity within a population, which is indicative of a population's ability to adapt to environmental changes (Frankham et al. 2002). However, patterns of genetic diversity that are often associated with an increased extinction risk may also be the result of a species' life-history characteristics (e.g., self-compatible short-lived gravity dispersed species exhibit low levels of allelic diversity and isolation among populations; Nybom 2004). Based on theoretical and empirical research, self-compatible species may also have purged the deleterious recessive alleles known to cause inbreeding depression (Hedrick 1994; Holsinger 1988). Consequently, rather than genetic factors, protecting against the negative consequences of demographic and environmental stochasticity may represent the dominant conservation priority for self-compatible species.

OBJECTIVES

The primary objective of my research was to evaluate the evolutionary distinctiveness of the federally listed plant species *Agalinis acuta* Pennell (Orobanchaceae), which is native to eastern North America where it is found on the coastal plain in eastern Massachusetts, Rhode Island, Connecticut, New York, and the piedmont in Maryland. The species was listed as endangered in 1987 due to conversion of its grassland habitat to agricultural, residential, and commercial uses, which were also cited as the most serious threats to the persistence of the species (U. S. Fish and Wildlife Service 1988). Vegetation succession to closed-canopy forest has also caused habitat loss and is a continuing threat at all remaining sites (U. S. Fish and Wildlife Service 1988). However, due to morphological and molecular similarity with putative heterospecific individuals (Neel & Cummings 2004; U.S. Fish and Wildlife Service 1989b), resolving taxonomic uncertainty became a major objective since a species is assumed to be evolutionarily distinct upon being listed under the Endangered Species Act.

My strategy for accomplishing this primary objective was to begin with conducting a phylogenetic analysis of multiple individuals from the majority of North American *Agalinis* species. Through the phylogenetic analysis I was able to assess whether putative *A. acuta* individuals exhibited the expected criteria of genealogical exclusivity. The dense sampling design allowed me to determine the level of differentiation that supports the taxonomic status of other species within the genus and identify the species from which *A. acuta* may not be evolutionarily distinct. The next step was to conduct analyses using microsatellite loci and morphological data assayed from multiple individuals from a number of populations representing each of the species with which *A. acuta* was

polyphyletic. A conclusion was then made regarding the taxonomic status and rank of *A. acuta* by considering the results from the analyses of multiple sources of data in light of numerous species concepts and delimitation methods.

In addition to identifying the taxonomic status of *A. acuta*, I conducted studies within the realms of DNA barcoding, phylogeography, and conservation genetics. A brief description follows of the studies that were conducted to elucidate the evolutionary relationships and conservation concerns of *A. acuta*.

1. A phylogenetic hypothesis was constructed based on molecular DNA sequence variation assayed from putative *A. acuta* individuals and an additional 28 congeneric species, all but two of which were represented by more than one accession. The purpose of the phylogenetic study was to test section and subsection levels of classification within the genus and provide a measure of differentiation that is characteristic within and among *Agalinis* species. The phylogenetic study also was used to identify the species from which *A. acuta* may not be distinct based on the criteria of genealogical exclusivity.

2. Based on the results from the phylogenetic study and other sources suggesting taxonomic uncertainty, the purpose my second chapter was to analyze DNA sequence, microsatellite, and morphological variation sampled from putative individuals of *A. acuta*, *A. decemloba*, *A. obtusifolia*, *A. skinneriana*, and *A. tenella*. To determine the taxonomic status and rank of those putative species, the results were considered in light of multiple species

concepts. The conservation status of the taxon that would include *A. acuta* is also discussed.

3. The purpose of my third chapter was to evaluate the efficacy of DNA barcoding in correctly identifying closely related species. Specifically, I conducted a DNA barcoding study to assess the utility of seven chloroplast loci and three analytical techniques (i.e., genetic distance, tree-based, and diagnostic characters) in discriminating among sequences that represent 29 congeneric species, 27 of which were represented by multiple accessions. The results highlight the issues of inaccurate taxonomy and incomplete lineage sorting of ancestral polymorphisms, which will decrease the efficacy of DNA barcoding and are most likely to be of concern when closely related species are considered. The application of DNA barcoding to conservation biology is also discussed.

4. The primary objective of my fourth chapter was to differentiate among alternative phylogeographic hypotheses in explaining the distribution and relationships among populations of the species that includes individuals that were what were historically ascribed to *Agalinis acuta*. In addition to historical processes related to the Pleistocene, I also considered whether more contemporaneous events associated with anthropogenic activities that have increased the degree habitat fragmentation and isolation among populations could explain extant patterns.

5. The goal of my fifth chapter was to evaluate the conservation implications of population genetic analyses based on microsatellite variation assayed from

individuals of the taxon that includes *A. acuta.* A range exists among the populations in the estimates of population genetic parameters, some of which are quite extreme and suggest that individuals within those populations are suffering a reduction in fitness due to genetic factors. I also discuss the extent to which patterns of genetic diversity may be explained by the species life history characteristics and, therefore, might not be indicative of genetic factors having a detrimental effect on fitness.

CHAPTER 1: PHYLOGENETIC PATTERNS AND CONSERVATION AMONG NORTH AMERICAN MEMBERS OF THE GENUS *AGALINIS* (OROBANCHACEAE)

ABSTRACT

North American *Agalinis* Raf. species represent a taxonomically challenging group and there have been extensive historical revisions at the species, section, and subsection levels of classification. The genus contains many rare species, including the federally listed endangered species *Agalinis acuta*. In addition to evaluating the degree to which historical classifications at the section and subsection levels are supported by molecular data sampled from 79 individuals representing 29 *Agalinis* species, we assessed the monophyly of 27 species by sampling multiple individuals representing different populations of those species. Twenty-one of these species are of conservation concern in at least some part of their range. Phylogenetic relationships estimated using maximum likelihood analyses of seven chloroplast DNA loci (aligned length = 11 076 base pairs (bp) and the nuclear ribosomal DNA ITS (internal transcribed spacer) locus (733 bp); indicated no support for the historically recognized sections except for Section Erectae. Our results suggest that North American members of the genus comprise six major lineages, however we were not able to resolve branching order among many of these lineages. . The North American *Agalinis* species sampled form a well supported, monophyletic group within the family Orobanchaceae relative to the outgroups sampled. Monophyly of 24 of the 29 sampled species was supported based on significant branch lengths of and high bootstrap support for subtending branches. Lack of support for monophyly of *Agalinis acuta* leaves the important question regarding its taxonomic status

unanswered. Lack of resolution is potentially due to incomplete lineage sorting of

ancestral polymorphisms among recently diverged species; however the gene regions

examined did distinguish among almost all other species in the genus. Due to the

important policy implications of this finding we are further evaluating the evolutionary

distinctiveness of *A. acuta* using morphological data and loci with higher mutation rates.

INTRODUCTION

The increase in use of molecular systematics in studies of angiosperm evolution has resulted in numerous phylogenies describing relationships across a range of evolutionary history (Soltis 2000). Studies of closely related species (Beardsley et al. 2004) are particularly important for filling in the tips on the angiosperm tree of life (Palmer et al. 2004). Phylogenetic hypotheses of the evolutionary relationships among members of the same genus provide frameworks for comparative research on mechanisms of diversification and speciation (Barraclough & Nee 2001). These phylogenies are also valuable resources for people concerned with conservation in that they provide a relatively objective means of quantifying evolutionary distinctiveness and resolving taxonomic ambiguities involving rare taxa (Andreasen 2005; Crandall et al. 2000; Fallon 2007; Soltis & Gitzendanner 1999). It is this application to identifying lineages that are sufficiently distinct to warrant taxonomic status and thus are eligible for legal protection (i.e., species, subspecies, and varieties) that greatly interests us.

A basic assumption of many species concepts (e.g. Baum & Shaw 1995; Donoghue 1985; Mallett 1995; Sokal & Crovello 1970; e.g. Wu 2001) and operational species delimitation methods (Sites & Marshall 2004) is that individuals of one species share common ancestry to the exclusion of members of other species. This shared common ancestry, which is a logical consequence of reproductive isolation between two groups, is expected to ultimately be reflected by genealogical exclusivity or monophyletic relationships inferred from phylogenetic analysis of DNA sequence or fragment data (Baum & Shaw 1995). However, the rapidly accumulating phylogenies of congeneric taxa with mismatches between gene trees and an expected species tree (Funk & Omland

2003; Syring et al. 2007) is yielding a startling picture of the extent to which the expectation of monophyly is not met. Such mismatches can indicate imperfect taxonomy, lack of sufficient variation to detect differentiation, incomplete lineage sorting of shared ancestral polymorphisms, or contemporary hybridization or introgression. The amount of evolutionary time required for mutations to accumulate and for shared ancestral polymorphisms to sort out after speciation events (Hudson & Coyne 2002; Tajima 1983; Takahata 1989) can make distinguishing among recently diverged taxa quite challenging. Coalescent theory predicts that it will take on the order of ~8.7 N_e generations for reciprocal monophyly of neutral, biparentally inherited loci to evolve in diverging lineages (Hudson & Coyne 2002; Rosenberg 2003). Thus, although the degree and duration of isolation necessary to achieve monophyly (especially across multiple loci) guarantees the evolutionary independence of monophyletic operational taxonomic units, absence of evidence for such independence, however, cannot automatically be assumed to mean that two entities are not reproductively isolated (Knowles & Carstens 2007). In these cases, additional evidence will be required to resolve ambiguities.

In this study, we examined phylogenetic relationships among 29 North American *Agalinis* (Raf.) species. This genus of flowering plants is restricted to the Western Hemisphere where approximately 40 species occur in the eastern and central United States and Canada and approximately 30 species are found in South America, Mexico, and Central America (Canne-Hilliker 1988; Missouri Botanical Garden 2007; Pennell 1928, 1929, 1935; USDA NRCS 2008). Due to taxonomic uncertainties, the exact number of species in the genus is unclear; acceptance of particular taxa varies across authors and taxonomic revision is in progress. Historically, *Agalinis* was considered to be

part of the family Scrophulariaceae but multiple phylogenetic analyses support placement in the family Orobanchaceae (dePamphilis et al. 1997; Olmstead et al. 2001; Olmstead & Reeves 1995). This plant family was traditionally associated with holoparasitism; however the broadened concept includes a number of autotrophic genera, such as *Agalinis*, that are hemiparasitic.

The majority of North American *Agalinis* species grow on the coastal plains of southern and southeastern North America. A secondary concentration of species occurs in the midwestern part of the continent and an even smaller number of species extend to the piedmont and to the coastal plains of the Mid-Atlantic, New England and the southern reaches of the Maritime Provinces in Canada (Pennell 1913a, 1913b, 1928, 1929, 1935). Throughout this geographic range, habitats occupied by *Agalinis* species are typically grasslands and savannas, grassy openings in woodlands and forests, or other herb dominated habitats. Soil moisture requirements vary greatly across taxa, ranging from inundated wetlands (including bogs, streams, ponds, and salt marshes), to wetland edges, to dry uplands. Because *Agalinis* species thrive in relatively open sites with no or low cover of shrubs and trees, many of them are found in early successional habitats and are most abundant following fire or other disturbance events. Due to overall declining trends in grassland extent and condition resulting from both development and lack of natural disturbance, a number of *Agalinis* species are increasingly restricted to forest edges and anthropogenically maintained openings such as utility corridors, and road verges. Although the more ruderal species can be extremely abundant and widespread in these highly modified habitats, our observation indicated that some less abundant and more geographically restricted species are susceptible to mowing during the reproductive

season, insufficient disturbance to remove woody vegetation, herbicide applications, and invasions of aggressively competitive non-native species.

General characteristics of the genus *Agalinis* include membranaceous, ephemeral corollas mostly with red-purple spots and yellow guide lines and wingless seeds that have variously reticulate seed coats (Canne 1979, 1980; Pennell 1929, 1935). Beyond the above characteristics, life form, morphology, anatomy, and floral form and color are variable, particularly in South American taxa. Unfortunately, relationships among the South American taxa are poorly understood, they are not included in any published classification schemes for the genus, and we were unable to obtain material to include them in this work. With exception of the perennial species *A. linifolia*, all North American species are annual herbs and all but three species (*A. auriculata*, *A. densiflora*, and *A. heterophylla*) have linear to filiform or scale-like leaves. Although mating systems have not been described for all members of the genus, the species that have been investigated include obligate outcrossing (*A. strictifolia* (Dieringer 1991)), mixed mating (*A. acuta* (Neel 2002), *A. skinneriana* (Dieringer 1999), *A. obtusifolia* (Snider 1969), *A. decemloba* (Snider 1969), and *A. auriculata* (Mulvaney et al. 2004)), and predominantly selfing due to cleistogamy (*A. neoscotica* (Stewart et al. 1996)).

The genus is taxonomically difficult and there have been numerous revisions of species and subspecies. In addition to taxonomic uncertainties, relationships among *Agalinis* species have been enigmatic and section-level classifications have been anything but stable. Pennell (1929) originally suggested five sections within the genus but later suggested only three sections with five subsections (Pennell 1935). Work based on seed, stem and leaf, and seedling characteristics as well as karyotypes (Canne-Hilliker 1987;

Canne-Hilliker & Kampny 1991; Canne 1979, 1981, 1983, 1984; Stewart & Canne-Hilliker 1998) yielded revisions to Pennell's classification that recognized five sections (Erectae, Heterophyllae, Linifolieae, Purpureae, and Tenuifolieae) and three subsections within the Purpureae (Pedunculares, Purpureae, and Setaceae). Previous phylogenetic analysis of 15 *Agalinis* species based on 7323 aligned bp of nucleotide sequence variation at three cpDNA loci (*rbc*L, *ndh*F, and *mat*K) (Neel & Cummings 2004) did not fully support either Pennell's or Canne-Hilliker's section-level classifications, although one section and some subsections suggested by Canne-Hilliker appeared to represent natural groups. Specifically, monophyly of Section Erectae was supported but Sections Purpureae and Heterophyllae were polyphyletic. Subsection Pedunculares was monophyletic but did not appear to be related to other Section Purpureae taxa as had been presumed. Limited taxon sampling and relatively low cpDNA sequence variation in that study prevented more thorough evaluation of relationships among Sections Linifoliae and Tenuifolieae and other subsections within the Purpureae.

In the present study, we provide a more comprehensive phylogenetic treatment of the genus by examining 29 North American *Agalinis* species using 7 cpDNA loci and 1 nuclear locus. Our specific objectives included simultaneously evaluating the monophyly of sections, subsections, and species that have been named solely based on anatomy and morphology. Every polytypic section and subsection is represented by multiple species and 27 species are represented by multiple individuals. In contrast to traditional sampling approaches in systematics studies that include only one representative per species (Syring et al. 2007; Wiens & Servedio 2000), we were able to treat species labels as testable hypotheses (Baum 1998). The extensive sampling also provides a genus-wide context in

which to evaluate the amounts and patterns of divergence among putative species that can

be detected using the loci we sampled. This context is particularly critical for interpreting

cases in which we fail to detect differentiation.

In addition to describing the evolutionary relationships among the sampled

individuals, this study has important implications for conservation. We sampled 21

species that are considered imperiled (S2) or critically imperiled (S1) in at least 1 state in

which they occur; 6 of these species are also globally vulnerable (G3 or G3-G4) and 3 are

critically imperiled (G1) (Table 1.1; NatureServe 2007). Data on the divergence of and

relationships among such a large number of species of conservation concern can help

prioritize rare species for conservation (Moritz 1995; Redding & Mooers 2006) by

estimating their degree of uniqueness within the genus. We were specifically interested in

addressing questions regarding the taxonomic status of three sets of species whose

distinctiveness from one another and thus conservation status had previously been

questioned: *A. acuta* and *A. tenella*, *A. tenella* and *A. obtusifolia*, *A. decemloba* and *A.*

obtusifolia. The status of *A. acuta* has been questioned previously and Neel and

Cummings (2004) found only a single nucleotide difference between *A. acuta* and *A.*

tenella across 4048 bp of cpDNA that included *rbc*L and *mat*K. *Agalinis acuta* occurs on

the coastal plain in eastern Massachusetts; Rhode Island; on Long Island, New York; and

in Maryland. *Agalinis tenella* occurs on the coastal plain from South Carolina south to

Florida and west to Alabama (Pennell 1935). Morphologically, *A. acuta* is distinguished

from *A. tenella* by having a shorter corolla, smaller seeds, and shorter pedicels (Pennell

1935). We were interested in the other two sets of species because the current taxonomic

treatment in the USDA PLANTS database (USDA NRCS 2008) suggests that *A. tenella*

and *A. decemloba* are synonymous with *A. obtusifolia* (Kartesz 1999). If this taxonomic

treatment is accurate and *A. acuta* is also not distinguishable, combining all four taxa

would be appropriate and there would be important conservation policy consequences. As

originally described, *A. decemloba* grows on the piedmont in Virginia, North Carolina,

and South Carolina (Pennell 1929, 1935). *Agalinis obtusifolia* is known from collections

from Maryland south to Florida and then west through Georgia to Louisiana on both the

piedmont and the coastal plain. Clarifying the taxonomic status of *A. acuta* (U. S. Fish

and Wildlife Service 1988) is essential because if it is not a species, subspecies, or variety

it is not eligible for listing under the U.S. Endangered Species Act (U.S.C. 1973). If it is

synonymous with other species, the status of *A. acuta* would need to be revised based on

the distribution, abundances, and threats of the populations representing those other

species.

METHODS

Taxon sampling

A total of 79 individuals representing 29 out of the ~40 putative North American

Agalinis species were included in this study (Table 1.1). The sampled species represented

all North American sections and subsections and all polytypic groups were represented

by more than one species. The number of individuals per species ranged from 1-9 and

when multiple individuals were used, they were from different populations. Sample

locations for most species were selected somewhat opportunistically and often coincided

with locations sampled for anatomical and morphological work by Dr. J. Canne-Hilliker.

We attempted to include samples from geographically distinct locations for each species

in order to capture the potential range of within-species variation (Appendix A). Samples

of *A. acuta* represent all geographic regions from which this species is known, and include most extant populations (Appendix A). Samples of *A. obtusifolia* and *A. tenella* were also distributed to represent the range of each species (Appendix A). The two *A. decemloba* populations were from the north central portion of the range. One representative of each of four outgroup species was also sampled: *Aureolaria pedicularia* (L) Raf., *Aureolaria pectinata* (L) Raf., *Brachystigma wrightii* (A. Gray) Pennell, and *Dasistoma macrophylla* (Nutt.) Raf. Fifteen of the *Agalinis* individuals and three of the four outgroup individuals were the same as those used in the previous phylogenetic study of the genus and related genera (*Aureolaria pectinata* is new and *Seymeria pectinata* Pursh was not included) (Neel & Cummings 2004). Vouchers are located at University of Guelph, University of Maryland, Iowa State University, and University of Texas Austin. Specific information on the location of particular specimens is available on request. We did not collect voucher specimens from the endangered *A. acuta* because these populations are well documented by state Natural Heritage Programs and the U.S. Fish and Wildlife Service.

DNA isolation, amplification, and sequencing

Total genomic DNA was isolated from fresh or frozen (-80°C) leaves and flower buds by grinding 50-100 mg of tissue to powder in liquid nitrogen with a mortar and pestle, and then using GenElute Plant Genomic DNA Kits (Sigma Chemical Company, St. Louis, Missouri, USA) or Qiagen DNEasy Kits (Qiagen Corporation, Valencia, California USA) following manufacturer's instructions.

We analyzed sequences from seven chloroplast gene regions (*mat*K, *rbc*L, *ndh*F, *trn*T (UGU)-*trn*F (GAA), *rps*2, *rpo*B, and *psb*A-*trn*H) and the nuclear DNA (nDNA)

locus ITS (18S-5.8S-26S). The first three cpDNA loci were used in the previously

mentioned study (Neel & Cummings 2004) and they represent relatively slowly evolving

portions of the chloroplast genome. Although there is rate variation among sites within

these loci (Neel & Cummings 2004) that inform different levels of the phylogeny, they

are most useful for resolving more ancestral relationships. The other chloroplast loci and

the nuclear ITS locus were chosen because they have been shown to be informative at

distinguishing among recently diverged taxa and even among populations within species

(Baldwin 2007; Shaw et al. 2005; Small et al. 1998). Our strategy was to assay *rbc*L and

*mat*K from at least one individual of each species to resolve the deeper relationships

within the genus. We then attempted to sequence the other five loci from all sampled

individuals. All but two of these loci were amplified using a single forward and reverse

primer pair. The exceptions were *trn*T-*trn*F which required two PCR reactions per

individual using *trn*T-a/*trn*L-d and *trn*L-c/*trn*F-f (Taberlet et al. 1991). The *rps*2 locus

was problematic for certain species but amplifications using the alternative forward

primer *rps*2-47F, instead of *rps*2-18F, were successful. Details of amplification and

sequencing for *rbc*L and *mat*K are given in Neel and Cummings (2004). In previous work

*ndh*F was extremely difficult to amplify from a number of *Agalinis* species and although

we did not pursue additional *ndh*F sequences, we used the ones available from Neel and

Cummings (2004) in our analysis.

Despite the well documented problems with using ITS for phylogenetic analyses,

due to high copy number and difficulty optimizing PCR (e.g., Alvarez & Wendel 2003),

we reliably obtained sequences using two primer pairs (ITS4 and ITS5 or ITS1 and

ITS4). These primers did yield multiple PCR products and attempts to design species-

specific primers for these taxa did not sufficiently reduce the number of copies. We therefore extracted the desired PCR product, (identified as the brightest band nearest to the target size), from an agarose gel using Qiagen's QIAquick Gel Extraction Kit according to the manufacturer's protocol. Inspection of the sequence trace curves confirmed that only a single copy had been sequenced.

All polymerase chain reactions (PCR) were done with Eppendorf MasterTaq PCR kits (Brinkman, Westbury, New York, USA) on MJ Research PTC-200 Thermal Cyclers. In general, the PCR temperature profile was 30 cycles of 94°C for 60 s, annealing temperature set approximately 5°C below the lower of the two primer melting temperatures for 90 s, 72°C for 150 s, and a final 15 min elongation period at 72°C. Amplified DNA fragments were purified using the Qiagen QIAQuick PCR Purification Kit according to manufacturer's instructions, unless noted otherwise.

Because many of the individuals and species we investigated were closely related and thus sequence variation was likely to be low, four replicate sequencing reactions were carried out for both forward and reverse primers for a given locus, resulting in eight-fold coverage across most regions of all loci. This conservative sequencing strategy ensured accuracy and prevented erroneous base calls associated with sequencing error that can cause serious issues when only single sequences are analyzed. Sequencing reactions were conducted with BigDye Terminator v3.1 Cycle Sequencing chemistry (Applied Biosystems, Foster City, California, USA) with reactions set up in 96-well microtiter plates. Total reaction volume was 7 µl (1-3 µl DNA template, 1.5 µl 5X Sequencing Buffer, 1 µl primer [25 µM], 0.5 µl BigDye Terminator, and 1-3 µl ddH2O). Cycle sequencing of purified PCR product was performed on an MJ Research PTC-200

Thermal Cycler and subsequent cleanup and preparation for sequencing was performed according to the manufacturer's protocol.

Data analysis

The program Sequencher v4.6 (Gene Codes Corporation, Ann Arbor, Michigan, USA) was used for base calling, quality assignments, and assembling consensus sequences for each sample from the replicate bi-directional sequence reads. Contigs for each locus exported from Sequencher were aligned using the default settings of MUSCLE (Edgar 2004b). BioEdit (Hall 1999) was used to manually edit alignments of the cpDNA loci *rps*2, *trn*T-*trn*F, and *psb*A-*trn*H which had numerous insertions/deletions. Alignments were exported as FASTA files and then converted to non-interleaved NEXUS files using MacClade v4.06 (Maddison & Maddison 2001). Three different data matrices were created: 1) cpDNA only, 2) nuclear ITS sequences only, and 3) a concatenation of all sequences.

To evaluate the variability of each locus, we calculated the number of characters that were constant, parsimony informative, and autapomorphic using the default parsimony settings in PAUP* (Swofford 2003). We also estimated the maximum likelihood pairwise distances between sampled individuals within and among *Agalinis* species for each locus separately. Nucleotide substitution model parameters for the maximum likelihood distance measures were chosen using MODELTEST (Posada & Crandall 1998). MODELTEST evaluates the likelihood scores of the same neighbor-joining tree for each of the 56 nucleotide substitution models calculated using PAUP* and the best fitting model was chosen using Akaike's Information Criterion (AIC).

Phylogenetic analyses were performed using the program GARLI v0.951 (Genetic Algorithm for Rapid Likelihood Inference) (Zwickl 2006). GARLI performs heuristic phylogenetic searches under the GTR + Γ + I (General Time Reversible with Gamma distributed rate heterogeneity and a proportion of invariant sites (Gu et al. 1995; Yang 1994)) nucleotide substitution model where topologies are evaluated based on their likelihood. The program calculates the maximum likelihood of a topology in the same manner as PAUP* but uses a genetic algorithm (Lewis 1998) to more efficiently evaluate alternative topologies. For each dataset, the best tree was found by running GARLI on the original data matrix with the default settings. We used likelihood ratio tests as implemented in PAUP* to assess whether branch lengths associated with the best topology inferred with GARLI were significantly greater than zero. To estimate the support for each node, phylogenies were created for 1000 bootstrap replicates of each dataset. A 50% majority rule consensus tree of the 1000 bootstrap replicates from GARLI was then created using PAUP*. The support values at each node on the consensus tree were added to the best tree found by GARLI, which allowed us to display both node support values and branch lengths.

To decrease the computational time required to complete the bootstrap replicates we reduced the number of generations that were performed without finding a better scoring topology before a replicate was terminated from 10 000 to 5000. To complete the bootstrap analyses for the cpDNA and all loci combined datasets in a relatively short time we used Grid computing through The Lattice Project (Bazinet & Cummings in press). The GARLI executable was converted to a Grid service such that batches of bootstrap replicates were distributed among hundreds of computers where they were conducted

asynchronously in parallel (Bazinet et al. 2007). The 1000 bootstrap replicates for the smaller ITS dataset were accomplished on a single desktop computer.

We used the approximately unbiased (AU) test (Shimodaira 2002) as implemented in the program CONSEL (Shimodaira & Hasegawa 2001) to evaluate whether a tree that constrained both *A. acuta* and *A. tenella* to be monophyletic was significantly worse than the best tree from an unconstrained analysis using the same data set. We repeated this test for each of the three data matrices. The AU test calculates a probability value of different topologies from bootstrap replicates of the site-likelihoods (Shimodaira 2002). We also used the AU test to determine the influence of missing data on phylogenies inferred from the cpDNA and ITS datasets and to assess the degree of congruence between the phylogeny based on cpDNA loci and the phylogeny based on the complete data set. It is not possible to directly assess the incongruence between the topologies from the concatenated cpDNA dataset and the ITS locus because the data matrices differed in the number of individuals. However, given that the cpDNA dataset and the all-loci-combined dataset differed only in the inclusion of ITS, we used the AU test to compare these two topologies as a means to estimate the incongruence with the cpDNA phylogeny introduced by the ITS locus.

RESULTS AND DISCUSSION

Characteristics of the sampled loci

Despite extensive efforts, it was not possible to obtain sequences of all loci for all species (Table 1.2 & Appendix A). Total aligned length of the cpDNA dataset was 11 076 bp and the total aligned length for ITS was 733 bp including only a few small (tri- or

tetranucleotide) insertions. The aligned concatenated dataset of ITS and the 7 cpDNA loci was 11 809 bp (Table 1.2).

The percent of constant characters among *Agalinis* species varied from 68.76% - 93.95% for ITS and *ndh*F respectively (Table 1.2). After *ndh*F, *rbc*L had the largest percentage of constant characters (90.53%). The number of parsimony informative sites for individual loci ranged from 52 (*rpo*B) to 320 (*trn*T-*trn*F) (Table 1.2). ITS exhibited the widest range of pair-wise maximum likelihood distances among species within the genus, ranging from 0.14% - 21.26%. The most conserved locus was *rbc*L with pairwise distances among *Agalinis* species ranging from 0 - 3.07% and averaging 1.07%; *psb*A-*trn*H had the largest range of among-species pairwise maximum likelihood distance of all the cpDNA loci, ranging from 0 - 20.50% and averaging 7.75% (Table 1.2).

Levels of variation we observed were similar to those found in other phylogenetic studies of congeneric species. The extensive length variation we observed in *trn*T-*trn*F (shortest sequence length of 1228 bp compared to the length of the alignment of 1868 bp) has also been observed within the confamilial genus *Pedicularis* (Yang & Wang 2007). A study of *Mimulus* (Phrymaceae) (Beardsley et al. 2004) in which only the *trn*L-*trn*F portion of *trn*T-*trn*F was sampled found a similar degree of variability expressed as the percent of parsimony informative characters (20.7% compared to 17.13% observed in this study). In *Lymania* (Bromeliaceae) 577 of 602 (96%) bases of *psb*A-*trn*H were constant (de Oliveira et al. 2007) compared to 669 of 884 (75.68%) constant characters within this study. The maximum level of variation we observed at the nuclear ITS locus (ML distance = 21.26%) is similar to that found in other genera within the

Orobanchaceae (*Pedicularis* (Yang et al. 2003) and *Orobanche* (Schneeweiss et al. 2004)).

General phylogenetic hypotheses

The phylogenies inferred from the three data matrices differed in tree shape and support for specific relationships (Figs 1.1 - 1.3). Results of the AU test (Shimodaira 2002) suggested that topologies derived from the cpDNA and the complete data set (Figs. 1.1 & 1.3) were significantly different from one another ($P < 0.05$). To rule out the possibility that samples missing from the ITS dataset (Appendix A) were causing some of the incongruence with the cpDNA phylogeny a reduced data matrix of the cpDNA loci was created that included only those samples also present in the ITS dataset. Results of the AU test ($P < 0.05$), indicated that the resulting topology (data not shown) was similar to the one from the complete cpDNA dataset, suggesting that missing individuals are not responsible for the incongruence between the nuclear and chloroplast DNA datasets.

Incongruences between phylogenies based on nDNA and cpDNA are not uncommon (e.g., Baumel et al. 2002; e.g., Levin et al. 2004; Soltis & Kuzoff 1995) and can indicate specific biological processes in species evolution. For example, hybridization has often been posited as an explanation for incongruence (Holder et al. 2001; Kimball et al. 2003). Alternatively, differences between the topologies might simply reflect the stochastic nature of the coalescent process (Kaplan et al. 1988). The lack of bootstrap support for many of the internal nodes on the phylogenies (particularly those on the phylogeny inferred with the ITS dataset) prevent us from making strong statements regarding the meaning of the incongruences. Our discussion of relationships among putative taxa relies primarily on the full and chloroplast data sets because they tended to provide better

support for inferred relationships. We point out specific instances where the estimates of relationships are different and well supported in the ITS data set.

All three topologies we examined provided strong statistical support for the monophyly of the sampled *Agalinis* species relative to the sampled outgroup species. Species now recognized as *Agalinis* have variously been included in the genera *Gerardia, Tomanthera,* and *Virgularia. Gerardia* had previously been applied to another taxon and the name was abandoned in favor of *Agalinis* (Pennell 1913a); *Tomanthera* and *Virgularia* are now synonymous with *Agalinis. Aureolaria*, *Brachystigma*, *Dasistoma*, *Seymeria*, and *Esterhazya* are considered close allies and at times have been considered congeneric with *Agalinis* (Pennell 1929; 1935). Morphological evidence suggested *Agalinis* was a distinct genus from *Brachystigma* and *Aureolaria* (Canne 1980), which our results clearly support. Phylogenetic analysis of the Orobanchaceae based on a single locus (phytochrome A) (Bennett & Mathews 2006) indicates that the South American genus *Esterhazya* may be more closely related to *Agalinis* than are *Aureolaria* or *Seymeria*. South American species of *Agalinis* have not been included in any systematic studies and the only publicly available sequence from *Esterhazya* represents a locus we did not sample (Bennett & Mathews 2006). Sampling additional *Esterhazya* species and South American *Agalinis* species will be essential to fully understanding evolutionary relationships in this group as a whole and to confirm the monophyly of the genus.

Section-level hypotheses

Agalinis linifolia is the only perennial *Agalinis* species in North America and has additional distinguishing characters that have resulted in placement in its own monotypic section (Table 1.1) that has been suggested to be basal to the rest of the species. The

unique ensemble of traits includes the type of thickenings on the inner walls of the seed

coat cells (Canne 1979), lack of yellow lines on the corolla, dense pubescence at the

bases of the posterior corolla lobes, presence of aerenchyma in stems, conspicuous

endodermis in roots, and palisade tissue in leaves that is developed more strongly towards

the lower surface (Pennell 1929). Monophyly of *A. linifolia* was supported; however, its

placement within the genus remains ambiguous due to lack of support for surrounding

nodes in all trees (Figs. 1.1 - 1.3). Despite this ambiguity, there is no evidence that this

taxon is basal because it is placed within a relatively derived clade that is well supported.

Further, a tree placing *A. linifolia* as basal had a significantly worse likelihood score than

the best tree from the unconstrained analyses ($P < 0.05$). Thus, the perennial growth habit

appears to be derived within this genus of otherwise annual species.

Members of Section Heterophyllae have also been suggested to be basal within the

genus based on having relatively large, broad, lanceolate leaves; leaf-like calyx lobes that

are longer than the calyx tube; oblong or ovoid-oblong capsules, and glabrous stems

(Pennell 1935). In particular, Pennell (1935) suggested that *A. auriculata* most closely

resembled the ancestral state of the genus based on also having relatively large corollas

with pubescence limited to the area below the posterior sinus, and having posterior anther

cells that are smaller than the anterior cells in addition to the characters described above

(Pennell 1929). Diagnostic aspects of leaf anatomy include thickened epidermal cell

walls, bands of sclerids and fibers between the cortex and phloem, patterns of the

subepidermal collenchyma on the leaf midribs, and lack of specialized trichomes (Canne-

Hilliker & Kampny 1991; Canne 1984). We found only two of the three species

hypothesized to comprise this section (*A. heterophylla* and *A. calycina*) to be

monophyletic, and this well supported clade is indeed basal within the genus in the cpDNA and full data sets. Of these two species, we could obtain ITS sequence only for *A. heterophylla*, which was placed sister to the Subsection Pedunculares clade, but with low bootstrap support. The third species, *A. auriculata,* is not closely related to this group in any of the trees (Figs. 1.1 – 1.3). *Agalinis auriculata* was known to differ from *A. heterophylla* in leaf and stem pubescence (Canne-Hilliker & Kampny 1991), and the perceived importance of differences suggested by those features are supported by our molecular data (Figs. 1.1 - 1.3).

Section Tenuifolieae has long been taxonomically problematic (Canne-Hilliker & Kampny 1991). Pennell (1929) united *Agalinis tenuifolia*, *A. divaricata* and *A. filicaulis* in this section based on lack of pubescence on the posterior corolla and upper corolla lobes being arched forward rather than erect or reflexed back as is seen in the rest of the genus. Canne-Hilliker and Kampny (1991) placed *A. tenuifolia* in Section Purpureae based on morphological and anatomical features, while retaining *A. divaricata* and *A. filicaulis* in Section Tenuifolieae. Most obviously, the upper corolla lobes in *A. divaricata* and *A. filicaulis* are less than 1/3 the length of the lower lobes and the corolla is greatly flattened, occluding the opening to the throat. In contrast, the upper corolla lobes of *A. tenuifolia* are more equal in length to the lower lobes and the corolla throat is closer to round in cross section. *Agalinis divaricata* and *A. filicaulis* also share peculiar seedling and trichome types (Canne 1983) and stem anatomy (Canne-Hilliker & Kampny 1991) that are not similar to any other *Agalinis* species and thus their placement has been challenging. High bootstrap support and the relatively long branch length supporting this clade in both the cpDNA and nDNA trees (Figs. 1.1 & 1.2) strongly support a sister

relationship between *A. divaricata* and *A. filicaulis*. At the same time, branch lengths

separating these two species are the longest of any sister-taxon pairs in the data set (Fig.

1.3). Relationships of this clade to other members of the genus depicted in the cpDNA

tree conflict with those in the ITS tree. The cpDNA sequence data indicate that the most

likely placement of the *A. divaricata*/*A. filicaulis* clade is sister to a clade including

Section Purpureae (Fig. 1), and in the phylogenies from the ITS and the full data set these

species have a more basal placement within the genus (Figs. 1.2 & 1.3).

Relationships of *A. tenuifolia* to other taxa are ambiguous; the cpDNA phylogeny

supports a sister relationship of *A. tenuifolia* with all *Agalinis* species except the *A.*

heterophylla/A. calycina clade (bootstrap support = 100%) (Fig. 1.1). Phylogenies based

on the ITS and full data sets indicate an alliance with *A. maritima* and *A. gattingeri*

(bootstrap support = 84%) (Figs. 1.2 & 1.3). In no case, however, does this species

appear to be closely related to *A. divaricata* and *A. filicaulis*.

With the exception of *A. gattingeri*, which is found within the clade discussed above,

the monophyly of Section Erectae is strongly supported in the full data tree (bootstrap

support = 95%) (Fig. 1.3). This section is united by the following genetic, anatomical,

and morphological characters: chromosome number of $n = 13$ (Canne 1984), yellow-

green colored foliage that does not blacken upon drying, small flowers that have

relatively short corolla tubes and reflexed corolla lobes, pedicels longer than the calyx

tube and light brown seeds (Canne-Hilliker & Kampny 1991; Canne 1979). Lack of

blackening upon drying is thought to be due to low concentrations of aucubosides (Snider

1969) that are at higher concentrations in other members of the genus. Placement of *A.*

gattingeri apart from other members of the Erectae is problematic because it contradicts

evidence that suggests close evolutionary relationships based on chromosome number (Canne 1981) and the unique seed type (Canne 1979) shared by other members of the section. However, *A. gattingeri* was always considered peripheral within Section Erectae due to its lack of anatomical features of the stem that are characteristic of the rest of the group (Canne 1984). Additional sampling is necessary to determine if this placement outside the Erectae is accurate or due to misidentification of the collections we sequenced or misinterpretation of the anatomical and morphological features.

Section Purpureae as defined by both Canne-Hilliker and Pennell was the largest section in the genus and it has been considered to have 3-5 subsections. Members were united by having globose capsules, dark brown seeds, narrow leaves that turn black upon drying and calyx lobes that are shorter than the calyx tube. We found little support for any of the historical concepts of this section or the majority of the recognized subsections (Figs. 1.1 -1.3). Only Subsection Pedunculares appears to be a natural group (Figs. 1.1 & 1.3); however, this subsection is sister to taxa comprising Section Erectae rather than to other taxa considered to be in the Purpureae. Subsection Pedunculares was considered to be distinct from the Erectae based on corolla form and pubescence patterns, seed color and surface patterns, and stem and leaf anatomy (Canne 1984). Neel and Cummings (2004) had previously suggested a sister relationship between the Pedunculares and the Erectae but their results were based on fewer species and the relationship did not have strong bootstrap support. Aligning the Pedunculares with the Erectae unites all the taxa with 13 chromosomes except *A. gattingeri,* which is placed with species considered to be in Section Purpureae. If not for the problematic placement of *A. gattingeri*, it would

appear that $n = 14$ was ancestral and the haploid chromosome number of 13 arose only once in the genus.

We found no support for the monophyly of Subsections Purpureae or Setaceae and many nodes supporting members that have been recognized to comprise these groups have weak bootstrap support (Fig. 1.3). Thus, despite the fact that most species are separated by well supported branches with non-zero lengths (exceptions will be discussed below), higher level relationships among species remain unclear. We did, however, find support for some hypothesized relationships. For example, our data support close relationships of *A. fasciculata*, *A. purpurea*, and *A. paupercula* in all trees (Fig 1.3). We also found support for a sister relationship between *A. setacea* and *A. plukenettii* in the full data set (Fig. 1.3) that had previously been hypothesized based on both species having acute trichomes (Canne 1983). We found no support for a close relationship of *A. laxa* to these two taxa in the full data set but *A. laxa* and *A. plukenettii* were part of a poorly supported clade that also comprised the *A. tenuifolia*/*A. gattingeri*/*A.maritima* clade in the ITS tree. Although it had been classified with *A. setacea* and *A. plukenettii* in Subsection Setaceae, *Agalinis laxa* was known to differ in having capitate trichomes on the hypocotyls and lacking acute trichomes (Canne 1983).

Overall, our results suggest that North American members of the genus comprise six major lineages, however we were not able to resolve branching order among many of these lineages. We propose that Section Heterophyllae consisting of *A. calycina*, *A. heterophylla* (and potentially *A. densiflora* but we did not sample this species) represents the basal group. Following the divergence of Section Heterophyllae, a rapid diversification resulted in five additional primary lineages. These lineages include one

comprising what have been considered Section Erectae and Subsection Pedunculares, two unrelated monospecific lineages (one comprising *A. auriculata* and the other *A. linifolia*), a fourth lineage corresponding roughly to Section Tenuifolieae, and a fifth consisting of the remaining taxa that have been included in Section Purpureae. We further recognize Section Erectae (sans *A. gattingeri*) and what was Subsection Pedunculares as distinct sister lineages that are relatively derived within the genus.

One potential explanation for lack of bootstrap support for the more basal relationships in the genus is a rapid diversification of lineages (i.e., a hard polytomy) (e.g., Page & Holmes 1998). This explanation is also supported by presence of comparatively short branches towards the base of the phylogeny (Fig. 1.3). Alternatively our data may simply not be sufficient to determine the order of branching (i.e., a soft polytomy) (e.g., Page & Holmes 1998). Although there is the potential that sequencing additional nuclear loci may be able to resolve the branching order at interior nodes on the phylogeny, we believe that a soft polytomy seems unlikely given the amount of DNA sequence we sampled and the levels of variation we observed in those sequences. Because our objectives included estimating both deep and shallow relationships, we specifically chose an array of loci that were expected to be useful for estimating relationships across the ranges of divergence anticipated.

Testing Species-level hypotheses

Our results corroborate most of the species designations in the genus and clarify some previous taxonomic ambiguities. Based on likelihood ratio tests, 83% (24 of 29) sampled species in the genus have significant (non-zero) branch lengths and 78% of species with multiple samples have bootstrap support > 98%. There are also multiple

cases in which branch lengths between conspecific individuals are greater than zero (e.g. the two *A. aphylla* samples (Fig. 1.3)) indicating that there is substantial differentiation among conspecific populations. Although we do not believe that there is a single particular amount of differentiation that determines a cutoff for recognizing a species-level distinction, we do expect populations within species to lack strong hierarchical structure due to tokogenetic processes (Posada & Crandall 2001). The hierarchical structure indicated by significant branch lengths and high bootstrap support within species (e.g., *A. skinneriana*, *A. decemloba*, *A. oligophylla*, *A. fasciculata*, and *A. tenuifolia*) indicates the need for closer examination of the biological basis of the observed patterns. Sampling additional loci and populations and using phylogeographic analytical methods would contribute to understanding whether these populations actually represent different species.

Exceptions to overall pattern of monophyly described above are the apparent polyphyly of *A. harperi*; the lack of differentiation between *A. purpurea* and *A. paupercula*; and the lack of differentiation among *A. decemloba*, *A. tenella*, and *A. acuta* (Fig. 1.3). All rare species that are of conservation concern except *A. acuta* and *A. paupercula* were supported as distinct.

The two sampled *A. harperi* individuals had identical ITS sequences (Fig. 1.2) but were polyphyletic based on cpDNA data (Fig. 1.1). Both individuals were part of a moderately supported clade (bootstrap support = 83%) consisting of representatives of putative *A. purpurea* and *A. paupercula*, *A. fasciculata* and *A. gattingeri* individuals. However, *A. harperi* 13FL is most closely related to *A. gattingeri* and the other appears sister to the *A. fasciculata*/*A. purpurea*/*A. paupercula* clade (Fig. 1.1). Reamplification

and sequencing of *rpo*B, *rps*2, and *trn*T-*trn*F loci from the two *A. harperi* samples yielded sequences that were identical to those used in constructing the phylogenies, thus ruling out the possibility that samples were mishandled. Therefore, the difference between the cpDNA and nDNA may be best explained by hybridization or introgression from another species that is represented by chloroplast capture (Tsitrone et al. 2003). Although it is not possible to say with much certainty given the lack of statistical support for the relationship between the *A gattingeri* samples and *A. harperi* 13FL, *A. gattingeri* may be the species from which cpDNA has introgressed into the *A. harperi* collection from Florida. Sampling from more *A. harperi* and *A. gattingeri* individuals and populations is required to resolve this issue.

 Agalinis purpurea and *A. paupercula* have been the subject of debate, with taxonomic hypotheses ranging from treating them as two species, as two varieties of *A. purpurea*, or synonymizing them under a single species. Pennell suggested relatively recent divergence related to the last ice age (Pennell 1935). These putative taxa differ from one another in that *A. paupercula* is reported to have smaller corollas (10-20 mm) and broader calyx lobes that are greater than half the length of the calyx tube (Pennell 1935). *Agalinis purpurea* has corollas ranging from 18-38 mm long and narrow calyx lobes that are less than half the length of the tube. Although they share many features during floral ontogeny, they do differ in *A. paupercula* var. *borealis* having different anther orientation, filament insertion points closer to the ovary height, later stigma initiation, and less exsertion of the stigma at anthesis than *A. purpurea* (Stewart & Canne-Hilliker 1998). The effect of these characteristics on mating system or reproductive isolation is unknown. It is also not known if these characteristics extend to other *A.*

paupercula varieties. Our data do not support recognizing *A. paupercula* as a distinct taxon. However, we did not thoroughly sample from a large number of putative populations of *A. paupercula* and it remains possible that some populations that have been attributed to that species represent a distinct entity. Further it is possible that higher resolution markers would allow us to differentiate *A. paupercula* and *A. purpurea* as is discussed below for *A. acuta*.

One of the primary objectives of this study was to evaluate the evolutionary distinctiveness of the federally listed endangered species *Agalinis acuta*. Potential synonymy of *A. acuta* with *A. tenella* was raised by Neel and Cummings (2004) due to lack of sequence divergence in two cpDNA loci between two individuals. Sampling nine representatives of *A. acuta* and five of *A. tenella* in this study allowed us to more thoroughly examine this issue. Previous taxonomic revisions (Kartesz 1999) that synonymized *A. tenella* and *A. decemloba* with *A. obtusifolia* necessitated inclusion of accessions attributed to the latter two species. Rather than clarifying relationships among these taxa, our results show a more convoluted situation than was previously thought to exist. The ITS phylogeny shows *A. tenella*, *A. acuta*, *A. decemloba* and *A. obtusifolia* to be polyphyletic (Fig. 1.3); however, there is little support for this topology. The phylogenies based on cpDNA loci alone and all loci combined (Figs. 1.1 & 1.3, respectively) show *A. tenella* to be monophyletic and subtended by a branch with a length that is significantly different from zero based on the likelihood ratio test. There is, however, no bootstrap support for this clade and it is nested within a clade that includes *A. acuta* and *A. decemloba*. On both topologies, *A. acuta* and *A. decemloba* are polyphyletic (Figs. 1.1 & 1.3) and an AU test indicated that forcing the monophyly of *A.*

acuta and *A. tenella* yielded a topology that was significantly worse than the best topology for all three datasets ($P < 0.05$). Although *A. acuta, A. decemloba*, and *A. tenella* form a highly supported monophyletic clade, one accession of *A. decemloba* (6VA) is distinguished from all other accessions of these three taxa based on branch lengths that are significantly different than zero and 98% bootstrap support (Fig. 1.3). This differentiation is the result of differences within the *trn*T-*trn*F locus. These include numerous single nucleotide differences and a 16 bp deletion, the majority of which are also present in the *A. obtusifolia* samples.

Regardless, our results do not provide statistical support for separate species status for *A. acuta*, *A. decemloba*, and *A. tenella* under the criteria of either a phylogenetic species concept (de Queiroz & Donoghue 1988) or a genealogical species concept (Baum & Shaw 1995). *Agalinis obtusifolia* comprises a monophyletic clade that is sister to the clade containing *A. skinneriana*, and *A. tenella, A. acuta,* and *A. decemloba*, thus strongly refuting the recent taxonomic revision synonymizing both *A. decemloba* and *A. tenella* with *A. obtusifolia* (e.g., Kartesz 1999; USDA NRCS 2008).

Lack of monophyly of even morphologically well defined species can result from incomplete lineage sorting of shared ancestral polymorphism or contemporary gene flow (Broughton & Harrison 2003; Funk & Omland 2003; Hudson & Coyne 2002). Given that it takes on the order of ~8.7 N_e generations for an 0.95 probability of reciprocal monophyly to evolve at a single locus after speciation events (Hudson & Coyne 2002; Rosenberg 2003), it can be challenging to distinguish among closely related taxa using phylogenetic methods. It is also possible that the DNA sequences we examined do not have sufficient mutation rates to have accumulated nucleotide differences in the time

since divergence. However, the loci sampled appear to have a sufficient amount of

variation to distinguish ~83% of the 29 sampled species in the genus, some of which are

likely to have recently diverged from a common ancestor. Due to the important policy

implications of combining *A. acuta*, *A. tenella*, and *A. decemloba* into a single taxon,

additional research is being conducted on the morphological and genetic differences,

using more variable loci, from samples collected from throughout the range of each

species.

As mentioned above, we found that *A. skinneriana* formed a well supported clade

that was sister to the clade containing *A. tenella*, *A. decemloba*, and *A. acuta* (Figs. 1.1 &

1.3). Prior to this work, taxonomic boundaries and phylogenetic affinities of *A.*

skinneriana were not understood. Additionally, the Maryland populations that we

sampled were problematic for experts to identify because these populations were beyond

the known range for the species at the time they were discovered. The morphological

characteristics of *A. skinneriana* most closely matched these populations, but there was

some lingering question as to their identity. Our results confirm that these populations are

sister to the *A. skinneriana* sample from Missouri and they represent an extension of this

otherwise Midwestern prairie taxon to the grasslands of the Atlantic coastal plain.

However, the branch separating the Maryland populations from the Missouri population

is significantly different from zero indicating that further investigation of the

phylogeography of this putative species may be warranted to determine if the Maryland

populations are actually an unrecognized species. Clarifying these relationships is

important because this species is considered rare in the state of Maryland and correct

identification is essential for both protecting a rare entity and not imposing restrictions for something that does not warrant them.

CONCLUSIONS

In conclusion, the sampled *Agalinis* species form a well supported, monophyletic group relative to the other genera sampled from within the family Orobanchaceae. Despite the well known taxonomic difficulty in this genus, 24 of the 29 the species we sampled that had been recognized based on anatomy and morphology were well supported. We confirmed the monophyly of 19 rare species, thus supporting their eligibility for receiving conservation attention. The species that do not form well supported clades based on DNA sequence data include the federally listed species *A. acuta* and the state-rare species *A. paupercula*. Although we were able to resolve some relationships among these species, most notably that the synonymization of the latter two with *A. obtusifolia* is unwarranted, a number of ambiguities remain. Due to the important policy implications raised by this finding, we are examining relationships among *A. acuta, A. decemloba*, and *A. tenella* further by sampling more individuals and populations using higher resolution molecular markers and morphological data. It is clear that most hypotheses regarding section- and subsection-level relationships based on morphology are not supported and taxonomic revisions are warranted.

Table 1.1. North American *Agalinis* species including the number of individuals (N) and conservation status of all species included in this study.

Taxon[1]	N[2]	Status[3]
Section Erectae (*n = 13*)		
A. acuta	9	G1/S1
A. aphylla	2	G3-G4/S2
A. decemloba	2	NR
A. gattingeri	3	G4/S1
A. obtusifolia	5	G4-G5-Q/S1
A. oligophylla	3	G4/S1
A. skinneriana	3	G3/S1
A. tenella	6	NR
A. viridis	2	G4/S1
A. keyensis	NS	
Section Heterophyllae (*n = 14*)		
A. auriculata	2	G3-G4/S1
A. calycina	1	G1/S1
A. heterophylla	3	G4-G5/S1
A. densiflora	NS	
Section Linifoliae (*n = 14*)		
A. linifolia	2	G4?/S1
Section Purpureae		
Subsection Pedunculares (*n = 13*)		
A. edwardsiana	1	G4/S4
A. homalantha	2	G5/S1
A. pulchella	2	G4-G5/S3?
A. strictifolia	2	G4/SNR
A. navasotensis	2	G1/S1
A. peduncularis	NS	
A. aspera	NS	
Subsection Purpureae (*n= 14*)		
A. fasciculata	3	G5/S1
A. harperi	2	G4?/SNR
A. maritima	2	G5/S2
A. paupercula	2	G5/S1
A. purpurea	4	G5/S1
A. tenuifolia	3	G5/S1
A. pinetorum	NS	
A. neoscotica	NS	
A. virgata	NS	
Subsection Setaceae (*n=14*)		
A. laxa	2	G3-G4/S3?
A. plukenettii	2	G3-G5/S1
A. setacea	2	G5?
A. stenophylla	NS	
A. filifolia	NS	
Section Tenuifolieae (*n=14*)		
A. filicaulis	2	G3-G4/S1
A. divaricata	2	G3?/S1
A. nutallii	NS	
Outgroup Species		
Aureolaria pectinata	1	G5?
Aureolaria pedicularia	1	G5
Brachystigma wrightii	1	G4
Dasistoma macrophylla	1	G4

[1]Chromosome counts represent those known for the section or subsection based on extensive species sampling (Canne 1983; Stewart and Canne-Hilliker 1998). [2]NS = Not Sampled. [3]Conservation Status: global ranking (G1=critically imperiled; G2=imperiled; G3=vulnerable to extinction or extirpation; G4=apparently secure; G5=demonstrably secure or widespread)/highest state ranking for each species (S1-S5 are equivalent to the global scale but applied to within a single state) (USA); when a range or question mark (?) is given the precise conservation status is uncertain.; NR and SNR= not ranked.

Table 1.2. Summary of the cpDNA loci and the nDNA locus ITS used in this study. N = the number of *Agalinis* species for each locus. Pairwise distances were calculated using *Agalinis* species only and do not include outgroup taxa.

Locus	N	Aligned Length (bp)	Characters Constant (percent)	Parsimony Informative Characters (percent)	Autapo-morphies	Nucleotide Substitution Model[1]
*mat*K	21	3822	3379 (88.40%)	141 (3.68%)	302	GTR+Γ+I
*ndh*F	6	2131	2002 (93.95%)	66 (3.10%)	63	TVM
*rbc*L	37	1331	1205 (90.53%)	53 (4.00%)	73	GTR+I
*rpo*B	78	375	306 (81.60%)	52 (13.87%)	17	GTR+Γ
*rps*2	77	665	520 (78.20%)	135 (20.30%)	10	TVM+Γ+I
*trn*T-*trn*F	79	1868	1479 (79.68%)	320 (17.13%)	69	TVM+Γ+I
*psb*A-*trn*H	79	884	669 (75.68%)	189 (21.38%)	26	TVM+Γ
All cpDNA Loci	79	11076	9592 (86.51%)	950 (8.57%)	545	TVM+Γ+I
ITS	68	733	504 (68.76%)	175 (23.87%)	54	GTR+Γ+I
All Loci	79	11809	10096 (85.41%)	1125 (9.52%)	599	GTR+Γ+I

[1] Nucleotide substitution model as selected using MODELTEST.
[2] Pairwise differences are based on the maximum likelihood distances calculated using the nucleotide substitution parameters associated with the best fitting model identified using MODELTEST

Table 1.2. Continued.

	Average Pairwise Difference (range)[2]		
Locus	Within Species	Among Species	Primer Source
*mat*K	n/a	3.12%	see (Neel and Cummings 2004)
		(0 – 6.20%)	
*ndh*F	n/a	2.97%	see (Neel and Cummings 2004)
		(0.42 – 5.00%)	
*rbc*L	0.17%	1.07%	see (Neel and Cummings 2004)
	(0– 0.39%)	(0 – 3.07%)	
*rpo*B	0.31%	2.15%	http://www.kew.org/barcoding/update.html
	(0-3.19%)	(0 – 5.40%)	
*rps*2	0.11%	4.37%	de Pamphilis et al. 1997
	(0-2.84%)	(0 – 8.59%)	
*trn*T-*trn*F	0.29%	3.24%	Taberlet et al. 1991
	(0-3.29%)	(0 – 6.04%)	
*psb*A- *trn*H	0.20%	7.75%	Sang et al. 1997; Tate and Simpson 2003
	(0-2.94%)	(0 – 20.50%)	
All cpDNA Loci	0.31%	3.82%	
	(0-2.0%)	(0 – 7.40%)	
ITS	0.75%	6.51%	White et al. 1990
	(0-3.93%)	(0.14 -21.26%)	
All Loci	0.36%	4.05%	
	(0.02-1.94%)	(0.04-7.99%)	

Figure 1.1. Phylogenetic tree depicting evolutionary relationships among sampled taxa based on seven cpDNA loci. Branch lengths depict the inferred number of nucleotide substitutions per site. Numerals at nodes represent the percent of 1000 bootstrap replicates supporting that clade. The ln likelihood of the tree is -30816.271.

0.0060 inferred nucleotide substitutions/site
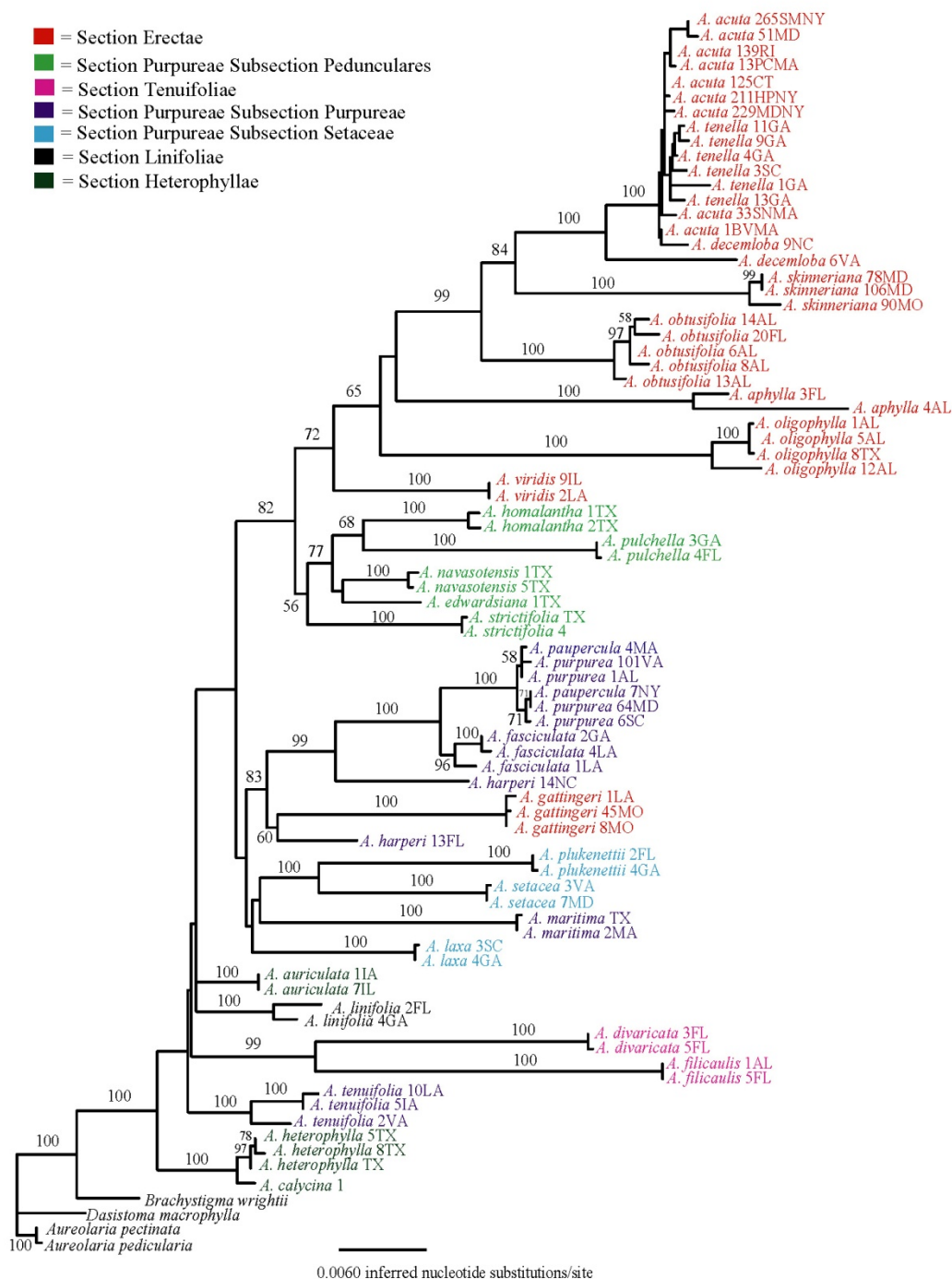
Figure 1.2. Phylogenetic tree depicting evolutionary relationships among sampled taxa based on the nDNA ITS locus. Branch lengths depict the inferred number of nucleotide substitutions per site. Numerals at nodes represent the percent of 1000 bootstrap replicates supporting that clade. The ln likelihood of the tree is -4250.1813.
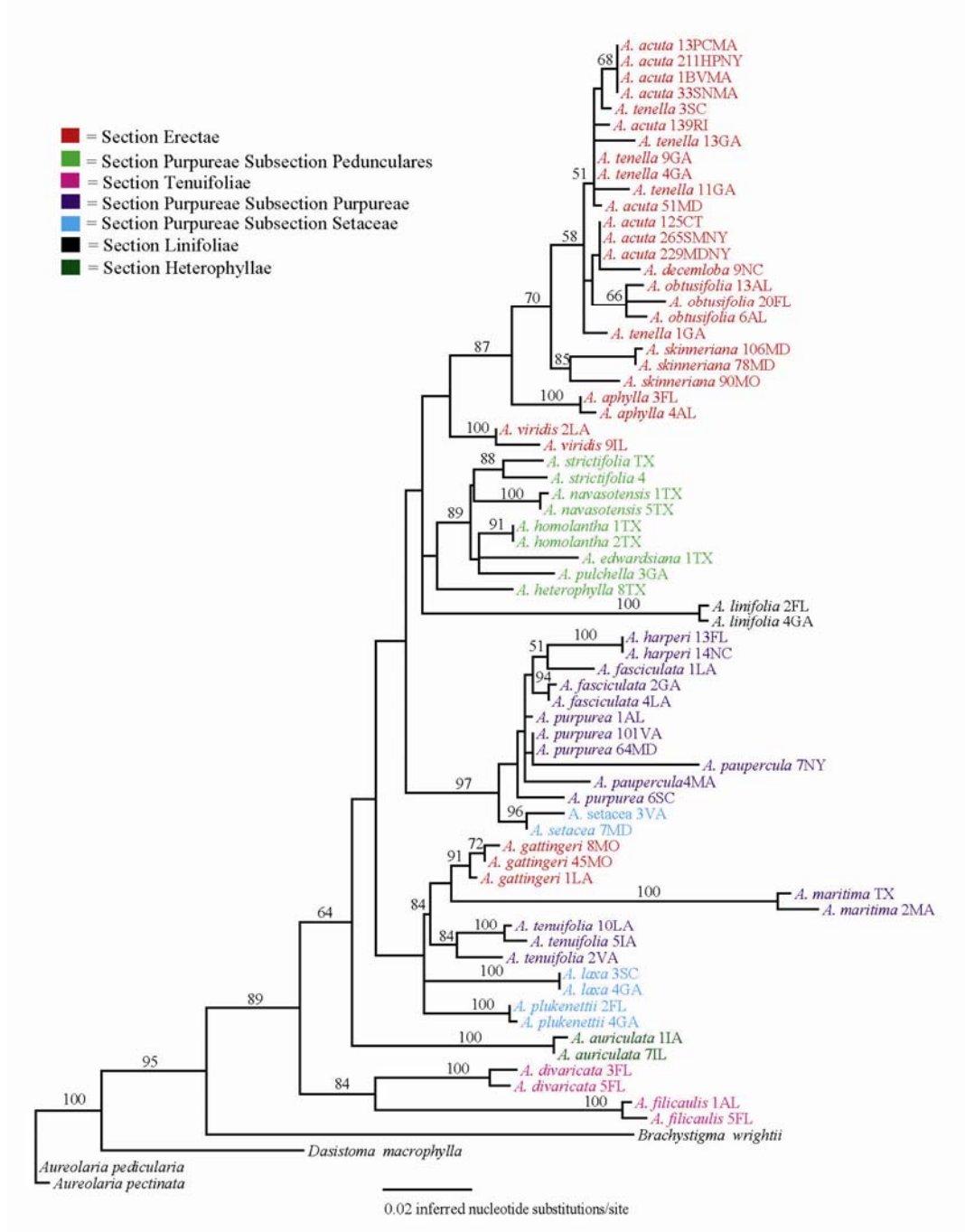
Figure 1.3. Phylogenetic tree depicting evolutionary relationships among sampled taxa based on a concatenated dataset of the seven cpDNA loci and the nDNA ITS locus. Branch lengths depict the inferred number of nucleotide substitutions per site. Numerals at nodes represent the percent of 1000 bootstrap replicates supporting that clade. The ln likelihood of the tree is -35900.524.
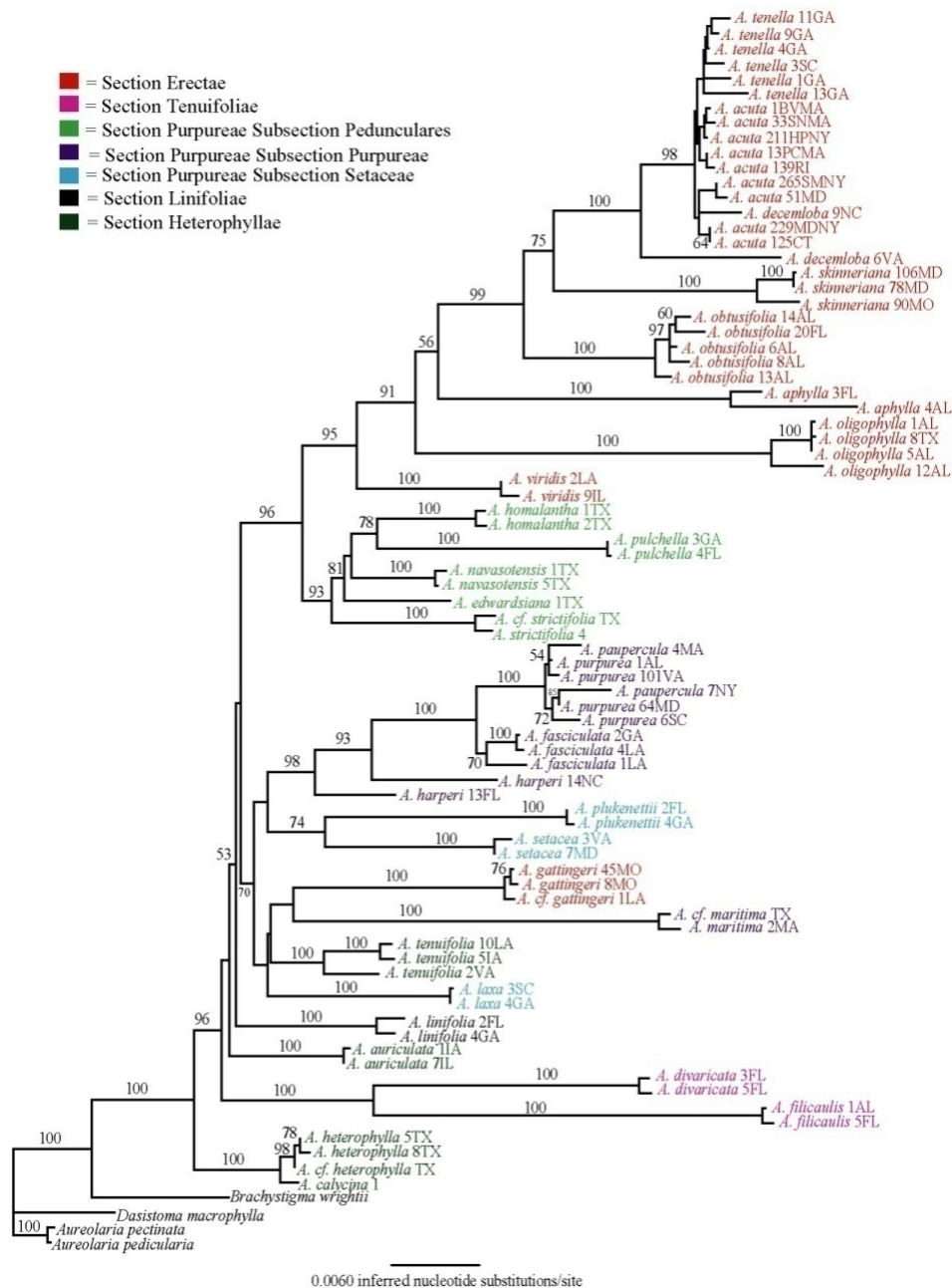
CHAPTER 2: COMPREHENSIVE GENETIC AND MORPHOLOGICAL ANALYSES DO NOT SUPPORT THE TAXONOMIC RANK OF SPECIES FOR THE FEDERALLY LISTED ENDANGERED PLANT *AGALINIS ACUTA* (OROBANCHACEAE)

ABSTRACT

Agalinis acuta (Orobanchaceae) is a federally listed endangered plant species that is at risk of extinction due to habitat loss and degradation. The taxonomic status of *A. acuta* has been questioned based on DNA sequence and morphological similarity with *A. decemloba* and *A. tenella*. However, those latter two species have been synonymized with *A. obtusifolia*. To better understand the evolutionary distinctiveness and phylogenetic affinities of these putative species and an additional closely related congener, *A. skinneriana*, we sampled six chloroplast DNA loci from representatives of 35 different populations representing the five putative species, characterized variation at 14 microsatellite loci across 20 populations representing *A. acuta*, *A. decemloba*, *A. tenella*, and *A. obtusifolia*, and measured 61 morphological characters assayed from multiple individuals from 18 populations. These different datasets provide evidence for five identifiable lineages that correspond to each of the five putative species. However, the magnitude and patterns of differences observed do not support the taxonomic rank of species for all taxa investigated. There is strong support for species status for both *A. obtusifolia* and *A. skinneriana*, but there is insufficient evidence to warrant the designation of species for the putative samples of the listed species *A. acuta*. The most appropriate taxonomic alignment is treating *A. acuta* and *A. decemloba* as one subspecies and *A. tenella* as another subspecies of a single species. Nomenclatural precedence

dictates that the former taxon would be *Agalinis decemloba* ssp. *decemloba* and the latter would be *Agalinis decemloba* ssp. *tenella*. We also discuss whether *Agalinis decemloba* ssp. *decemloba* represents a taxon deserving of federal protection under the Endangered Species Act.

INTRODUCTION

The fundamental importance of species in biology is reflected in the extensive literature regarding the conceptual basis of this taxonomic and evolutionary category (Baum & Shaw 1995; Cracraft 1989; Mallett 1995; Mayr 1942; Nixon & Wheeler 1990; Van Valen 1976). A large body of literature also describes empirical methods that have been advocated for delimiting species (Knowles & Carstens 2007; Sites & Marshall 2004; Wiens 2007). The numerous species concepts and delimitation methods make determining when a collection of individuals or populations warrants recognition as a species challenging as well as highly contentious (Coyne & Orr 2004). Yet understanding this fundamental aspect of biological diversity is essential (Isaac et al. 2004). Wrongly assuming that the historical taxonomy ascribing individuals to species is accurate can have negative consequences for understanding patterns of biodiversity in that incorrect species circumscription will yield inaccurate species richness estimates and erroneous estimation of macroecological patterns (Isaac et al. 2004). More importantly, taxonomic misunderstandings can have negative ecological and evolutionary consequences as happened when invasion of vernal pool habitat in California by the non-native low mannagrass *Glyceria declinata* went undetected because it was erroneously synonymized with the native western mannagrass *Glyceria occidentalis* (Gerlach et al. 2009). This invasion negatively impacted the vernal pool ecosystem and increased the extinction risk of protected species.

Taxonomic uncertainty also directly affects the status of federally listed plant species in that improved taxonomy has caused downlisting of one and delisting of seven plant taxa (U. S. Fish and Wildlife Service 1996; U.S. Fish and Wildlife Service 1989a, 1999).

Only evolutionarily distinct species, subspecies, or varieties (and distinct population segments of vertebrates) that are at elevated risk of extinction due to anthropogenic effects are eligible for listing under the Endangered Species Act (ESA; U.S.C. 1973). Listing entities that are later synonymized wastes time and resources that could be spent conserving truly imperiled taxa. At the same time, failing to recognize a taxon due to cryptic variation can result in the unwitting loss of significant evolutionary diversity if extinction results due to lack of protection. In spite of controversy due to different species concepts and delimitation methods, rigorous and objective methods must be utilized to resolve taxonomic uncertainty and reduce the negative consequences that such uncertainty can have for species of conservation concern.

To resolve the controversy surrounding which species concept one should adhere to, de Quieroz (2007) has suggested that there is an underlying concept that unifies the myriad species concepts: that species are "separately evolving segments of metapopulation lineages". Although his use of the term metapopulation is not in accordance with its original definition (Levins 1969), the concept of species comprising sets of interacting populations in space and time that are related to one another through ancestry and descent is perhaps useful. In contrast to early typological views of species, de Queiroz (2007) embraces the fact that in absence of gene flow, there is a continuum of evolutionary differentiation that proceeds after a speciation event (Cummings et al. 2008). Characteristics emphasized under alternative species concepts will be acquired at different times following such an event, thus adherence to different concepts or use of different types of data to detect differences can result in different species boundaries (Baum 1998; de Queiroz 2007; Marshall et al. 2006). Therefore, it is most productive to

use multiple methods and data sources to determine where putative species lie on the divergence continuum; such a sampling strategy will likely be required to differentiate among recently diverged species (e.g., Leache et al. 2009). The importance of using multiple lines of evidence is well accepted in evolutionary biology (Shaw 2002) and biogeography (Petit et al. 2003; Petit et al. 2002; Schaal et al. 1998) and has recently been emphasized as being essential for conservation decisions (Fallon 2007).

Beyond simply using multiple types of data, we suggest that it is most cost effective and efficient to use a sequential process starting with the most restrictive definitions and delimitation methods (e.g., genealogical exclusivity); additional analyses that can detect more subtle differences would be conducted only if the more restrictive definitions are not met. Therefore, rather than requiring one predetermined characteristic or property for species status, we evaluated which properties were present against a range of commonly used species concepts and delimitation criteria. We also suggest taxonomic sampling beyond the specific entities of concern and sampling intensively within the entities of concern. This combination of broad and deep sampling provides context that is needed to understand the magnitude of differences that distinguish closely related entities and allows assessment of whether those differences are of sufficient magnitude to warrant the taxonomic rank of species based on biological significance, utility, predictive power, robustness, and precedence (Baum 2009). Such an approach will potentially use data and analytical techniques associated with both systematics and population genetics. Although techniques from these two fields have different assumptions derived from different theoretical frameworks, both are necessary to quantify patterns at species boundaries

because prior to completion of an investigation it is not clear whether the individuals of interest represent a single species or different species.

We illustrate the power of this sequential approach to determining where a putative taxon falls along the continuum of divergence with a study of the taxonomic boundaries of the federally listed endangered plant species *Agalinis acuta* Pennell (Family Orobanchaceae) in which we sequentially assayed DNA sequences, microsatellite allele frequencies, and morphology. The observed patterns at those different data types were then compared with expectations from different species concepts. *Agalinis acuta* is a fall blooming annual plant native to eastern North America where it is found on the coastal plain in eastern Massachusetts, Rhode Island, Connecticut, and New York, as well as on the piedmont in Maryland (Fig. 2.1). The type locality was a historic location near Edgartown, Massachusetts on the island of Martha's Vineyard (Pennell 1929). The species was listed under the Endangered Species Act (ESA; U.S.C. 1973) in 1987 primarily due to threats associated with declining grassland habitat (U. S. Fish and Wildlife Service 1988). A recovery plan approved in 1989 required meeting three criteria (20 stable and wild occurrences, protection of at least 15 of these occurrences, and ability to propagate plants from seed) in order to downlist the species to threatened status (U.S. Fish and Wildlife Service 1989b). The recovery plan also identified the need to resolve uncertainty in the distinctness of *A. acuta* from closely related taxa due to morphological similarities (Table 1). Based on the recovery plan and prior work, the other species we considered were *Agalinis decemloba* Greene (Pennell), *Agalinis tenella* Pennell, *Agalinis obtusifolia* Raf., and *Agalinis skinneriana* (Alph. Wood) Britton.

Agalinis decemloba has been recorded from Pennsylvania south to South Carolina with a type locality in the District of Columbia (Pennell 1935). The Pennsylvania populations are thought to be extirpated, as is the type locality. *Agalinis tenella* ranges from North Carolina south to northern Florida and west to Alabama, with a type locality in Thomas County, Georgia (Pennell 1935). In 1994, *A. decemloba* and *A. tenella* were synonymized with *A. obtusifolia* (Kartesz 1994; USDA NRCS 2004). However, recent phylogenetic analyses showed that *A. obtusifolia* was monophyletic (Figs. 1 & 3 in Pettengill & Neel 2008 or Figs. 1.1 & 1.3 in Chapter 1), which strongly refutes the taxonomic revision of Kartesz (1994). *A. obtusifolia* is known from the eastern seaboard from Maryland south to Florida and west to Mississippi on the Gulf coast. *Agalinis skinneriana* (Alph. Wood) Britton is primarily recognized as a Midwestern prairie taxon and, thus, the putative identification of this species in Maryland would represent a range expansion that includes the grasslands of the Atlantic coastal plain. Despite the broad geographic range represented by these putative species, they all occupy similar habitats: open canopied sites with nutrient poor soils that are dominated by bunchgrasses and other herbaceous species.

To be considered a separate species, we expect *A. acuta* to be reproductively isolated and somehow recognizably different from other species, whether that recognition is based on morphology, ecological function, or genetic diversity. Below we briefly describe the characteristics and associated techniques that scientists have advocated to be appropriate for determining if entities meet these expectations and discuss the results of previous studies of *A. acuta* relevant to those characteristics and techniques. We then describe how we further evaluated the evolutionary distinctiveness of *A. acuta* in light of these

criteria using DNA sequence data, microsatellite allele frequencies, and morphological

variation.

Reproductive isolation

Historically, the biological species concept has been a dominant definition of

whether a group of individuals represents a species. Under this concept a species is a

group of potentially interbreeding individuals that are reproductively isolated relative to

other putatively heterospecific individuals (Mayr 1942). The classic way to demonstrate

reproductive isolation in plants is to directly compare reproductive success of crosses

among versus within putative species (Stebbins 1950; Turesson 1922). Alternatively,

reproductive isolation can be inferred indirectly based on patterns of genetic

differentiation, with the null hypothesis that sampled individuals representing different

putative species constitute a relatively homogenous gene pool (e.g., King et al. 2006). If

that null hypothesis is refuted, reproductive isolation can be inferred.

Unfortunately, attempts at germinating *A. acuta* seeds did not yield sufficient

germination to grow plants for crossing experiments (unpublished data); other

investigators have also had low germination success for *A. acuta* seeds (Brumback &

Kelley 1990). Instead of directly quantifying the reproductive compatibility of putative *A.*

acuta individuals and putative heterospecifics, we indirectly assessed the degree of

reproductive isolation using an array of techniques from the fields of systematics and

population genetics as described below.

Genealogical exclusivity

According to the original genealogical species concept (Baum & Shaw 1995), when

using DNA sequence data all putative conspecific individuals should be monophyletic

(i.e., they should share a most recent common ancestor to the exclusion of putative heterospecific individuals). This genealogical exclusivity reflects long-term reproductive isolation, thus fulfilling the criterion of the biological species concept. The biological species concept does not require monophyly; rather monophyly based on DNA sequence data signifies that the groups in question have been reproductively isolated for a substantial amount of time.

Phylogenetic analysis of relationships among *Agalinis* species based on 11,076 base pairs (bp) of cpDNA and 733 bp of nuclear DNA indicated that *A. acuta* was not monophyletic with respect to *A. decemloba* and *A. tenella* (Pettengill & Neel 2008). Although reciprocal monophyly definitively indicates reproductive isolation, its absence does not necessarily indicate that putative heterospecifics are interbreeding; rather it may be due to lack of sufficient time for the sorting of shared ancestral polymorphism to have occurred (Cummings et al. 2008; Hudson & Coyne 2002; Knowles & Carstens 2007). Many empirical studies have found that entities that are considered 'good species' based on behavioral or morphological criteria are paraphyletic or polyphyletic based on sequence data (Carstens & Knowles 2007; Funk & Omland 2003). Further, due to the stochastic nature of the coalescent process, individual gene trees may not be concordant with the species tree expected based on a taxonomic classification and different gene trees from within one species may not be concordant with one another (Hudson & Coyne 2002; Maddison 1997; Maddison & Knowles 2006). The loci we examined did yield monophyletic groups of 24 of 29 of the *Agalinis* species included in a previous phylogenetic study (Pettengill & Neel 2008). Sampling broadly in the genus thus allows us to know that failure to detect differences among *A. acuta*, *A. tenella* and *A. decemloba*

is not due to inability of these loci to detect species level differences in *Agalinis* more generally.

It is still possible that the lack of monophyly observed among *A. acuta*, *A. tenella*, and *A. decemloba* reflects recent divergence. Because it can take on the order of 8.7 N_e generations after speciation events for mutations to accumulate and for shared ancestral polymorphisms to sort out at a single neutral locus many recently diverged species will remain polyphyletic (Hudson & Coyne 2002; Rosenberg 2003). Short of monophyly, determining what threshold level of concordance among gene trees supports a taxonomic designation is somewhat arbitrary. Baum (2009) has proposed that a taxon should show genealogical exclusivity for a plurality of sampled gene regions and should show more exclusivity than any other conflicting grouping of individuals. Cummings et al. (2008) have developed a method for quantifying the degree of exclusive ancestry of groups along the continuum of divergence, thus departing from the typological view of relationships being either monophyletic or not.

The previous phylogenetic results combined with morphological and ecological similarities led us to further evaluate the nature and degree of polyphyly among putatively heterospecific populations of *A. acuta*, *A. decemloba*, *A. tenella*, and *A. obtusifolia*. We also included *Agalinis skinneriana* (Alph. Wood) Britton because it is morphologically similar to *A. acuta* and *A. tenella* (Table 2.1) and its taxonomic boundaries and phylogenetic affinities had not been investigated. For this examination, we incorporated previously unsampled populations (i.e., those not sampled in Neel & Cummings 2004 or Pettengill & Neel 2008) into additional phylogenetic analyses. We

also used the method developed by Cummings et al. (2008) to quantify the degree of non-random genealogical exclusivity among groups.

Diagnostic molecular characters

Under the phylogenetic species concept as conceived by Cracraft (1989), a species is an irreducible cluster of organisms that is diagnosably distinct from other such clusters and within which there is a pattern of ancestry and descent. Diagnostic characters are taken as evidence of reproductive isolation because even small amounts of gene flow among putative heterospecific populations would disrupt a diagnostic pattern. We examined the two most variable cpDNA loci we sampled for diagnostic insertions/deletions or single nucleotide polymorphisms corresponding with the putative species boundaries (Goldstein & DeSalle 2000; Goldstein et al. 2000). We do not consider single base changes alone to be sufficient to delineate a species, but rather a collection of such polymorphisms or diagnostic insertions/deletions would provide evidence of reproductive isolation.

Evidence of differentiated gene pools

Differentiation of gene pools is often documented using data derived from allele or haplotype frequencies among individuals and populations as measured by, for example, measures of pairwise genetic distances (Sites and Marshall 2004). Generally, differences based on allele frequencies alone are not considered sufficient to support species level designations. Rather, they are thought to reflect geographic structuring of populations within species, and thus to support designation as management units or possibly subspecies (Goldstein et al. 2000; King et al. 2006; Moritz 1994). However, in conjunction with other information such as ecological or behavioral differences, allele

frequency differences could provide support for considering species status if recent divergence is suspected.

In this study, we used microsatellite data collected from multiple individuals and populations per species to assess whether there was evidence that putatively heterospecific populations differed in terms of allele frequencies and gene pools. High mutation rates in microsatellites allow differences to arise relatively quickly after a speciation event, making them a useful complement to sequence data for investigations of recently diverged taxa. Most of the population genetic methods we have employed allow inferences about the degree of reproductive isolation from analysis of genetic distances or deviations from Hardy-Weinberg equilibrium and linkage equilibrium among putative conspecific versus heterospecific individuals or populations. Because these different measures are all derived from the same allele frequency data, they do not provide independent lines of evidence. Rather, the different analytical approaches provide multiple perspectives on the patterns in the data.

Morphological differentiation

Morphological differentiation can be detected based on fixed, categorical character differences (Goldstein et al. 2000) or based on significant differences in continuous traits (Wiens & Servedio 2000). Ideally species will be well differentiated with no or extremely few individuals having intermediate morphology (Mallett 1995). Studies quantifying patterns of morphological variation in plants are best carried out in common garden or reciprocal transplant experiments to isolate genetic from environmentally dependent phenotypic variation (Claussen et al. 1940). Unfortunately, we were unable to conduct a common garden study due to the low germination rates discussed previously.

Rather we present the analysis of a suite of morphological traits that were measured on individual plants in the field.

Together, these comprehensive analyses provide a rigorous framework against which we can evaluate where along the continuum of diversification the putative *Agalinis* taxa exist. Although the final assignment of a collection of individuals to any particular taxonomic rank can be subjective (Baum 2009), we have attempted to limit that subjectivity and to make the rationale for such assignments transparent by evaluating data against established species concepts and associated delimitation criteria.

MATERIALS AND METHODS

Phylogenetic and population sampling

Multiple samples representing different and often geographically distant populations of each species of interest were included in a phylogenetic analysis (10 *A. acuta*, 5 *A. decemloba*, 9 *A. obtusifolia*, 3 *A. skinneriana*, and 8 *A. tenella*) (Table 2.S1). Fourteen of these 35 samples were not part of previously published phylogenetic treatments (Neel & Cummings 2004; Pettengill & Neel 2008). Based on Pettengill and Neel (2008) we chose single representative of *A. calycina* to serve as the outgroup.

Samples from populations used in the analysis of microsatellite and morphological variation were collected during the blooming season from 2003 through 2008. We collected tissue for genomic DNA extraction from approximately 30 individuals from 11 populations of *A. acuta*, 4 populations of *A. tenella*, 2 populations of *A. decemloba*, 3 populations of *A. skinneriana*, and 3 populations of *A. obtusifolia* (Table 2.2). When population sizes were <30, we sampled all individuals we could locate. From these sampling efforts, we acquired DNA samples from 662 individuals from 24 populations

representing five putative *Agalinis* species (Table 2.2). As allowed by phenology, we sampled morphological traits from 395 individuals in 18 of the populations of the five species from which we collected DNA (Table 2.2). Because we collected material throughout the growing season in three years, we encountered a wide range of phenological conditions during our visits and we could not measure all morphological variables on all individuals. In particular, flowers were not always present and thus floral characters were not measured for many individuals.

DNA extraction

Total genomic DNA for all samples used in the phylogenetic and population genetic analyses was isolated by grinding 50-100 mg fresh or frozen (-80°C) leaves and flower buds to powder in liquid nitrogen with a mortar and pestle, and then using GenElute Plant Genomic DNA Kits (Sigma Chemical Company, St. Louis, Missouri, USA) or Qiagen DNEasy Kits (Qiagen Corporation, Valencia, California USA) following manufacturer's instructions. Some extractions were carried out on a Qiagen BioSprint 96 robotic workstation using Qiagen's BioSprint 96 DNA Plant Kit, which is a high throughput extraction procedure that processes samples in 96-well microtiter plates using proprietary magnetic particle technology (MagAttract).

Phylogenetic loci and analyses

The phylogenetic analyses were based on six chloroplast loci (*trn*T-*trn*F, *rpo*B, *rps*2, *psb*A-*trn*H, *rbc*L, and *mat*K) following protocols used Pettengill and Neel (2008). However, in this study we used more recently developed universal primers to amplify *rbc*L (Kress & Erickson 2007) and *mat*K (Cuénoud et al. 2002). All sequences were aligned using the program MUSCLE (Edgar 2004a, 2004b) with the default settings.

The number of constant characters and variable characters at each locus and in the concatenation of all loci were calculated using PAUP v4.0 (Swofford 2003). We estimated intra- and interspecific Kimura-2-paramater (K2P) distances (Kimura 1980) using the program MEGA (Kumar et al. 2004). These analyses were conducted excluding the outgroup *A. calycina*.

Phylogenetic inference was conducted in the same manner described in detail in Pettengill and Neel (2008), which was based on a maximum-likelihood method implemented in the computer program GARLI (Genetic Algorithm for Rapid Likelihood Inference; Zwickl 2006). Statistical support for topological relationships was assessed through 1000 bootstrap replicates. To evaluate the range of topologies and associated likelihood scores across independent runs, we conducted 1000 analyses of the original non-bootstrapped dataset. Because the topologies from these independent runs differed in terms of monophyly of *A. tenella*, we used the approximately unbiased test (AU-test) (Shimodaira 2002) as implemented in the program CONSEL (Shimodaira & Hasegawa 2001) to evaluate whether the likelihoods of these alternative topologies were significantly different from one another. For the AU-test, the site-likelihoods were estimated using PAUP v4.0 where the appropriate nucleotide substitution model was determined with MODELTEST (Posada & Crandall 1998) under Akaike's information criterion (AIC; Akaike 1974).

We calculated the genealogical sorting index (*gsi*) (Cummings et al. 2008) for each of the putative species to provide a measure of the genealogical exclusivity among putative conspecific individuals when species were not reciprocally monophyletic. The *gsi* represents the ratio of the minimum number of nodes necessary to unite all

conspecific individuals into a monophyletic group to the actual number of nodes that unite all conspecifics on a rooted phylogeny. This normalized index ranges between 0 and 1, with 0 representing a total lack of monophyly or a random arrangement of putative conspecifics on the tree and 1 representing monophyly. We calculated the ensemble gsi_T statistic for the 1000 trees from the bootstrap analyses in GARLI; each topology was weighted by the proportion of times it was observed among the 1000 replicates. Significance of the gsi_T statistic was assessed through 1000 permutations of each topology where the labels on the tips of each tree were randomized (Cummings et al. 2008).

Diagnostic cpDNA characters

Two of the cpDNA loci that were used for the phylogenetic study (*psb*A-*trn*H and *trn*T-*trn*F) were further evaluated to determine whether they possessed diagnostic characters that could be used to discriminate among the species of interest based on the 35 individuals included in the phylogenetic analysis. These regions were chosen because they were the most informative in terms of percentage of variable sites in Pettengill and Neel (2008) and they have been advocated in the plant DNA barcoding literature as having the most power to distinguish closely related species (Kress et al. 2005; Shaw et al. 2005).

The program DNAsp v4.10.8 (Rozas et al. 2003) was used to calculate the number of segregating sites that were fixed between each pair of putative species with gaps in sequences treated as missing data. We also manually inspected the alignments for substantial insertion/deletion events that were diagnostic of putative species. We admittedly did not sample sufficiently to guarantee that characters were truly fixed within

species (Wiens & Servedio 2000); however, the presence of such characters in relatively slowly evolving DNA regions across geographically distant populations provides compelling evidence of their utility in delimiting species.

Microsatellite genotyping

Twenty-one microsatellite loci that were developed from four microsatellite libraries corresponding to the repeat motifs CA-, AAG-, CAG-, and TACA- (Appendix B; Pettengill et al. 2009) were assayed across all populations of putative *A. acuta*, *A. decemloba*, *A. tenella*, *A. obtusifolia*, and *A. skinneriana*. Amplification success varied across species and populations; virtually all microsatellites failed to amplify in the *A. skinneriana* populations and as a result, that species was not included in the population genetic analyses. Additionally, due to amplification failures, the data set with all 21 loci and 20 populations representing *A. acuta*, *A. decemloba*, *A. tenella*, and *A. obtusifolia* was missing 9.4% of the total possible genotypes (Tables 2.S2 & 2.S3). Seventy-five percent of the missing genotypes were from putative *A. obtusifolia* individuals. Because this degree of missing data is problematic for the analyses we wished to use, we constructed a reduced dataset that consisted of 14 loci and included all 20 populations. The 14-locus data set used in all subsequent analyses was missing 4.6% of the total possible genotypes, 57% of which were from *A. obtusifolia* individuals (Table 2.3).

Population genetic analyses

The program ARLEQUIN (Excoffier et al. 2005) was used to calculate observed (H_o) and expected (H_e) heterozygosity for each species and for each population within species. The program FSTAT (Goudet 1995) was used to estimate allelic richness (A_R) using rarefaction to account for differences in sample size. The program GENEPOP

(Raymond & Rousset 1995) was used to estimate inbreeding within populations as measured by R_{IS} (Rousset 1996). We also calculated Jost's D_{EST} (Jost 2008) to estimate population differentiation using SMOGD v1.2.3 (Software for the Measurement of Genetic Diversity) (Crawford submitted). Jost's D_{EST} is proposed to be a more accurate measure of population differentiation than those used previously (e.g., G_{st} and F_{st}) in that it does not produce erroneous results that suggest low levels of differentiation among populations with few or no alleles in common or underestimate the degree of differentiation when hypervariable loci (e.g., microsatellites) are used (Hedrick 1999; Heller & Siegismund 2009; Jost 2008). For comparison, we also calculated the traditional measure of population differentiation, F_{st}, using GENEPOP.

We used Analysis of Molecular Variance (AMOVA), as implemented by ARLEQUIN to assess the statistical support for different taxonomic alignments. One model included four groups corresponding to each of the putative species *A. acuta, A. decemloba, A. tenella*, and *A. obtusifolia*. We also constructed three models that excluded *A. obtusifolia* to help determine the most likely grouping among the three putative species that were not monophyletic in Pettengill and Neel (2008): 1) three different groups corresponding to *A. acuta, A. decemloba*, and *A. tenella*; 2) two groups total, with *A. acuta* and *A. decemloba* representing one group and *A. tenella* representing the second group; and 3) two groups total, with *A. decemloba* and *A. tenella* representing one group, *A. acuta* representing the second group. Support was assessed by quantifying the amount of variance in genetic distance that was explained at each level in each model: among individuals within populations, among populations within putative species, and among putative species. We expect that a model that best reflects the correct species

classification will have the largest amount of variation explained at the species level when compared with competing models (e.g., King et al. 2006). Because the same populations are grouped differently among the models, we were not able to use the AIC as a statistical method of determining which AMOVA model is best supported by the data (Burnham & Anderson 1998)

We calculated Nei's estimate of genetic distance (D_a) (Nei et al. 1983) among all pairs of populations with the program DISPAN (Ota 1993); DISPAN was also used to construct a tree based on the neighbor-joining algorithm (Saitou & Nei 1987) where node support was calculated from 1000 bootstrap replicates. We chose D_a because simulation and empirical studies have shown that this distance exhibits a better linear relationship with evolutionary time when the groups of interest have only recently diverged and that it is more accurate at recovering the correct topology than alternative distance measures (Paetkau et al. 1997; Takezaki & Nei 1996).

Because pooling individuals into populations can mask similarities among heterospecific individuals and artificially inflate apparent differences among entities, we constructed a neighbor-joining tree based on genetic distances (D_a) among all individuals using the program Populations 1.2.30 (Langella 2002). Due to the large number of missing genotypes, putative *A. obtusifolia* individuals were excluded from this analysis.

We used the Bayesian program STRUCTURE (Falush et al. 2003; Pritchard et al. 2000) to investigate group assignments of individuals. This model-based clustering method does not incorporate a priori knowledge about the group membership of individuals; rather, individuals are assigned to groups based on maximizing the degree of Hardy-Weinberg equilibrium and linkage equilibrium. Operationally, independent runs

of STRUCTURE are carried out varying the number of clusters (k) to which individuals are assigned and likelihood scores for these different values of k are evaluated to make inferences about population structure and genetic similarity. We also employed the statistic Δk to help identify the optimum value of k based on changes in the likelihood score among successive values of k (Evanno et al. 2005). We conducted 100 replicate runs at each value of k from 1 to 21 and present the results from the replicate with the highest likelihood score at each k. The analyses were performed using a model that allows for admixture and correlated allele frequencies; Markov Chain Monte Carlo sampling consisted of 50000 generations for burnin and an additional 250000 generations. Our primary focus was the composition of the groups at $k = 4$ because we were interested in testing the hypothesis that the individuals we sampled represent four species.

We used the program FreeNA (Chapuis & Estoup 2007) to estimate the frequency of null alleles, which can occur within microsatellite loci due to preferential amplification of one allele or mutations in the priming site of certain alleles (Chapuis & Estoup 2007). Using the dataset created by FreeNA that corrects for the presumed presence of null alleles, we conducted a subset of the analyses described above to determine if the patterns and amounts of differentiation differed substantially from those associated with our observed genotypes.

Morphological analyses

Data for 61 morphological characters (Appendix C) were collected from 395 individuals in 18 populations of five species in Section Erectae (Table 2). The morphological characters measured were selected either because they represented traits

that potentially influence the deposition of pollen and pollinator behavior and thus could contribute to reproductive isolation or because the characters were previously identified as diagnostic for the species of concern (Table 1). Descriptions of all variables including details of how they were measured and statistical properties (e.g., means and standard deviations) are provided in Tables 2.S4 & 2.S5. To avoid biasing our conclusions due to strong correlations among variables, if two variables were correlated at $r^2 > 0.6$ we included only one of the two in the analyses. A variable was chosen preferentially for removal if it was highly correlated with many other variables or was found, based on exploratory analyses, to contribute little information in differentiating groups.

We used linear discriminant function analysis (DFA) to assess whether *A. acuta*, *A. tenella*, *A. decemloba*, *A. obtusifolia*, and *A. skinneriana* could be distinguished based on morphology and, if so, which morphological characters best discriminated them. Specifically, we constructed two DFA models, one in which species was the grouping variable and another using population as the grouping variable. In all cases DFA was performed using all sampled plants that had complete data (i.e., plants with missing data were casewise deleted). This casewise deletion yielded a data set with 294 individuals in 18 populations. We chose to use DFA because the technique is designed specifically to distinguish between two or more groups using more than one variable at a time (Neff & Marcus 1980; Statsoft 2004). Incorporating multiple variables simultaneously is desirable because often taxa or populations are distinguished by suites of characters rather than a single character (James & McCulloch 1990).

We used forward stepwise procedures to enter variables in each model where the tolerance for a variable to enter was set at 0.01 to prevent redundancy of variables; F-to-

Enter the model was set between 2 and 2.8. Statistical significance of the discriminatory power of the model was then assessed based on Wilks' lambda and its associated F and p values. We examined partial Wilks' lambda for each variable to understand its independent contribution to discrimination among the groups. We used the eigenvalue of and cumulative proportion of variance explained by each discriminant function to understand its unique contribution to discriminating among groups in each model; factor structure coefficients and standardized canonical coefficients were used to assign meaning to each of the DFA axes.

Group separation was assessed by classification success which is a function of pairwise Mahalanobis distances (d^2) (Titus et al. 1984). Cross-validation for the comparisons in which species was the grouping variable was used to provide a more conservative evaluation of classification success. We created 10 independent data sets in which half of the individuals in each population were randomly assigned to a training set that was used to build the DFA model. The remaining individuals were assigned to a test set that was then classified using the model created with the training set. Next, we calculated the median number of individuals classified into each species and the median percent correct classification for each species across the 10 runs. Small sample sizes in a number of populations precluded use of cross validation when population was used as the grouping variable.

RESULTS

Phylogenetic analysis and monophyly

We obtained sequences of *trn*T-*trn*F, *rpo*B, and *psb*A-*trn*H, from all sampled individuals (Table 2.4 and 2.S1). We obtained reliable sequences of *mat*K from all

individuals except *A. decemloba* 1VA; of *rbc*L from all but five samples (*A. obtusifolia* 18FL, *A. obtusifolia* 10AL, *A. decemloba* 19NC, *A. tenella* 79GA, and *A. tenella* 91GA) (Tables 2.4 & 2.S1); and sequences of *rps*2 from all but two samples (*A. obtusifolia* 8AL and *A. obtusifolia* 14AL).

Concatenating the six loci produced an alignment of 4799 bp that had 216 variable sites (4.71%). The number of polymorphic sites at a single locus ranged from 16 sites (4.56%) among the *rpo*B sequences to 79 sites (8.33%) among the *psb*A-*trn*H sequences (Table 2.4). The average intraspecific K2P distances was 0.001, which was an order of magnitude smaller than the average interspecific distances (0.018) (Table 2.4). *mat*K exhibited the smallest average interspecific distance (K2P = 0.009 across all five species) and was monomorphic within three putative species (*A. acuta*, *A. decemloba*, and *A. tenella*); *rps*2 had the highest average interspecific distance (K2P = 0.025). Across all loci, *Agalinis skinneriana* had the highest intraspecific distance (K2P = 0.002); *A. obtusifolia* and *A. skinneriana* had the largest average interspecific distances (K2P = 0.028 and 0.025, respectively) (Table 2.4).

Pairwise interspecific K2P distances based on cpDNA loci among all *Agalinis* species previously examined (i.e., samples in Pettengill and Neel 2008 and Pettengill and Neel in review) except *A. acuta*, *A. decemloba*, and *A. tenella* (26 species; 57 accessions) averaged 0.0292 (st. dev. = 0.0092). Intraspecific distances within those same species averaged 0.0021 (st. dev. = 0.0041). In contrast, interspecific genetic distances among only *A. acuta*, *A. decemloba*, and *A. tenella* averaged 0.0011 (st. dev. = 0.0005) while intraspecific distances averaged 0.0008 (st. dev. = 0.0022) (Table 2.5). *Agalinis*

obtusifolia and *A. skinneriana* differed from one another and from *A. acuta*, *A. decemloba*, and *A. tenella* by distances of 0.021 – 0.031.

In general, the results of the phylogenetic analysis including additional samples representing populations not previously assayed were congruent with the observations of Pettengill and Neel (2008). Specifically, monophyly of both *A. skinneriana* and *A. obtusifolia* was strongly supported (bootstrap support = 100%) (Fig. 2.2) and both *A. acuta* and *A. decemloba* were polyphyletic. The tree with the best likelihood score (-ln 8729.079657) from the 1000 independent runs of the non-bootstrapped dataset showed *A. tenella* to be polyphyletic. However, an alternative topology showing *A. tenella* to be monophyletic had a statistically indistinguishable likelihood score (-ln 8734.027845; AU-test $P >> 0.05$). Because additional information based on microsatellite loci supports the monophyly of *A. tenella* (discussed below) and the difference between alternative topologies lacked significance, we present a topology in which *A. tenella* is monophyletic (Fig. 2.2).

Given that both *A. obtusifolia* and *A. skinneriana* were monophyletic, gsi_T for both species was 1.0. The gsi_T statistic was significant for the other three species ($P << 0.05$) but was substantially less than 1.0 (*A. acuta*, $gsi_T = 0.394$; *A. decemloba*, $gsi_T = 0.284$; and *A. tenella*, $gsi_T = 0.415$). If *A. acuta and A. decemloba* are combined, $gsi_T = 0.689$.

Diagnostic cpDNA characters

We found fixed nucleotide differences in *psbA-trnH* and *trnT-trnF* between either *A. obtusifolia* and all other species, *A. skinneriana* and all other putative species, or these two putative species and the other three species (Table 2.S6). There were no fixed differences that distinguished *A. acuta*, *A. decemloba*, and *A. tenella* from each other

(Table 2.S6). Of the two loci, *psb*A-*trn*H had more fixed differences among groups and had an 18bp insertion that was diagnostic of *A. obtusifolia*. There was a ~200bp insertion in *psb*A-*trn*H present in both *A. obtusifolia* and *A. skinneriana*. Within this large insertion, there was a 5bp insertion unique to *A. obtusifolia* and a 21bp insertion within *A. skinneriana* that distinguished between the two species. The *trn*T-*trn*F locus had two insertion/deletion events, one of which was diagnostic of *A. obtusifolia* and one that was present in both *A. skinneriana* and *A. obtusifolia*.

Microsatellite amplification success

We successfully genotyped nearly 100% of the 14 microsatellite loci in all *A. acuta* and *A. decemloba* populations; amplification across loci in *A. tenella* samples was ~95% successful and in *A. obtusifolia* was ~86% successful (Table 2.3). The differential success is not too surprising given that the loci were developed from genomic DNA for *A. acuta* and amplification across species is known to be challenging (Barbara et al. 2007). At the same time, if *A. acuta*, *A. tenella*, and *A. decemloba* all represent one species, cross-amplification should not be an issue. Therefore, amplification success provides useful information about the evolutionary relationships in that individuals that can be genotyped at a given locus are more likely more closely related to one another than they are to individuals in which the same locus could not be genotyped. Patterns of amplification failure indicate the highest degree of similarity between *A. acuta* and *A. decemloba*, somewhat less similarity between these two putative taxa and *A. tenella*, and *A. obtusifolia* being the least similar (Tables 2.S2 & 2.S3).

Levels of genetic diversity

   Agalinis acuta had by far the largest number of individuals sampled and yet we detected 113 alleles at the 14 loci compared to 79 in *A. decemloba*, 271 in *A. tenella*, and 227 in *A. obtusifolia* (Table 2.3). When rarefaction, as implemented in FSTAT, was used to account for sample size differences, A_R per locus at the species level ranged from 5.12 in *A. acuta*, to 12.88 in putative *A. tenella* (Table 2.3). The high A_R in *A. tenella* was due in large part to one population (ATE-LCGA) that had large numbers of alleles at multiple loci, many of which were exclusive to that population (Table 2.3).

   Species-level H_e ranged from 0.51 in *A. decemloba* to 0.79 in *A. tenella* (Table 2.3). The average H_o within species was substantially lower in *A. acuta* and *A. decemloba* (H_o = 0.10 and H_o = 0.18, respectively) than it was in *A. tenella* (H_o = 0.60), and *A. obtusifolia* (H_o = 0.34; Table 2.3). Consequently, both *A. acuta* and *A. decemloba* had high values of R_{IS} (0.92 and 0.70, respectively).

   Population-level allele richness mirrored species level patterns in that *A. acuta* had the lowest average A_R per locus (A_R =1.78), with 8 populations averaging < 2.0 alleles per locus (Table 2.3). *Agalinis tenella* had the highest average A_R within populations (A_R =5.04). The proportion of polymorphic loci within populations was substantially lower in *A. acuta* than it was for populations of the other species, ranging from 0.07-0.93. The only other population that had a value lower than the highest value for *A. acuta* was one *A. obtusifolia* population (AOB-DCSC; *P*=0.86). The low levels of polymorphism for *A. acuta* are particularly striking because sampling was biased towards polymorphic loci for this species.

Levels of within-population H_e were <0.2 in all but 1 population of *A. acuta* and in 1 of the 2 populations of *A. decemloba* (Table 2.3). H_o within *A. acuta* was exceptionally low in the populations at Shadmoor (AAC-SHNY), Montauk Downs (AAC-MDNY), and Soldiers Delight (AAC-SDMD) (Table 2.3). The low heterozygosity in the latter population is surprising given its size and extent. Soldiers Delight also had the highest R_{IS} value of any population.

Across all four putative species, mean pairwise intraspecific Jost's D_{EST} values for individual populations were lower than mean pairwise interspecific values (Table 2.3). Although intraspecific levels of differentiation were lowest among the *A. tenella* populations (intraspecific $\overline{D_{EST}} = 0.384$) and greatest among the *A. obtusifolia* populations (intraspecific $\overline{D_{EST}} = 0.647$) (Table 2.3), these levels of intraspecific differentiation suggest that conspecific populations are substantially different from one another. *Agalinis obtusifolia* populations were also the most differentiated from putative heterospecific populations followed by *A. acuta* populations with *A. decemloba* populations being the least differentiated from putative heterospecific populations.

Heterozygosity estimates based on the synthetic data set that we created using FreeNA (Chapuis & Estoup 2007) to correct for potential presence of null alleles were larger than those observed in the original data, with averages across populations ranging from $H_o = 0.31$ in *A. acuta* to $H_o = 0.77$ in *A. tenella* (Table 2.3). Despite these large changes in estimates of observed heterozygosity, there was little difference in measures of differentiation based on our observed data (overall $F_{ST} = 0.45$) versus the data set that was calculated with FreeNA (overall $F_{ST} = 0.43$).

Analysis of molecular variance

In the AMOVA model treating each of the four putative species as a separate group, the largest amount of variance in genetic distances was explained by the species level (55.37%; *P* <0.001). Substantially less variation was explained by differences among populations within species and among individuals within populations (22.52% and 22.11%, respectively; *P* <0.001) (Table 2.6a).

Although the three models excluding putative *A. obtusifolia* populations from the analyses also had a significant portion of variance in genetic distances explained by the 'species' level, the magnitude was approximately 50% less than when that taxon was included (Table 2.6). Of the three models excluding *A. obtusifolia*, the largest amount of variance explained at the 'species' level was found in the model treating *A. acuta* and *A. decemloba* as one group and *A. tenella* as a second group (31.12%) (Table 2.6c). The model treating *A. acuta*, *A. decemloba*, and *A. tenella* each as a separate 'species' only explained 22.93% of the variance in genetic distance at the 'species' level (Table 2.6b). The final model combining *A. decemloba* and *A. tenella* into one group and treating *A. acuta* as a separate group explained the smallest amount of genetic variance (20.9%) at the 'species'. These results indicate that *A. obtusifolia* is responsible for the majority of the differences among the four putative species and when it is not included, the model treating *A. acuta* and *A. decemloba* as one group explained the most amount of variance at the 'species' level (Table 2.6).

Genetic distance among populations and individuals based microsatellites

Pairwise interspecific D_a distances between *A. acuta* and *A. decemloba* populations averaged 0.47, which was identical to intraspecific distances among *A. acuta* populations

and slightly higher than intraspecific distances between the two *A. decemloba* populations (Table 2.5). Distances between these two species and *A. tenella* were larger ($D_a = 0.57$-0.59; Table 2.5). Distances between both *A. acuta* and *A. decemloba* combined and *A. obtusifolia* was ~0.74 and between *A. tenella* and *A. obtusifolia* D_a was ~0.65. All interspecific distances except those between *A. acuta* and *A. decemloba* were substantially larger than intraspecific distances (Table 2.5).

The neighbor-joining tree inferred from D_a distances placed all *A. acuta* populations in the same clade with moderate support (bootstrap value of 73%) (Fig. 2.3). The two putative *A. decemloba* populations form a clade with 91% bootstrap support. Together these two clades form a strongly supported group (bootstrap value of 94%; Fig. 2.3).

All of the putative *A. tenella* populations also formed a strongly supported clade (bootstrap value of 99%). However, there was substructuring within the clade, with ATE-CCSC placed outside the group containing the other three populations (ATE-BCGA, ATE-LCGA, and ATE-GCGA), which were grouped together with bootstrap support of 95% (Fig. 2.3). The *A. obtusifolia* populations were united into a strongly supported clade to the exclusion of the other three putative taxa.

The relationships depicted on the neighbor-joining trees based on D_a distances calculated among individuals were congruent with the results based on distances among populations. Although the *A. obtusifolia* populations could not be included in the analysis because of the substantial amount of missing data, the other three putative taxa each formed clusters consisting of only putative conspecifics (Fig. 2.4). The sister relationship between the clade consisting of the two *A. decemloba* populations and the

clade representing all *A. acuta* populations detected at the population level was also recovered.

The neighbor-joining tree based on D_a distances calculated from the data set created by FreeNA had nearly identical topology to the tree based on uncorrected data (Figs. 2.3 & S1) with differences being in the relationships among conspecific rather than heterospecific populations and in having lower bootstrap support values. Thus, if the null alleles were present in our dataset, the resulting bias does not dramatically affect the relationships we inferred with our observed data.

Population structure and assignment tests

Assignment of individuals using STRUCTURE when $k = 4$, which corresponds to the number of putative species we sampled, yielded monospecific clusters of *A. obtusifolia* and *A. tenella* individuals with minimal amounts of admixture and two groups composed of combinations of both *A. acuta* and *A .decemloba*. Examination of Δk did not yield an obvious number of most likely populations. Rather multiple values of k showed an appreciable improvement in the likelihood score relative to k -1. At $k = 2$ and 3, Δk was 4.09 and 4.75, respectively, and $k = 5$ and 6 both had values of Δk greater than 1 that represented peaks when Δk was plotted against k. The individuals that formed putatively heterospecific clusters varied among these values of k (Fig. 2.5). At $k = 2$, all *A. obtusifolia* and *A. tenella* individuals occupy one cluster and representatives of the other two species comprise the second cluster. At $k = 3$, all *A. obtusifolia* individuals occupy a single cluster with minimal admixture, *A. acuta* individuals represent a second cluster, and the third includes all *A. decemloba* and *A. tenella* samples. For $k = 5$ and 6, all *A. obtusifolia* and *A. tenella* individuals form monospecific clusters but the apparent group

assignments of *A. decemloba* and *A. acuta* individuals varies. At $k = 5$, AAC-HPNY and AAC-WANY populations are grouped with the two *A. decemloba* populations and at $k = 6$, the *A. decemloba* populations form a single cluster and AAC-HPNY and AAC-WANY are clustered with AAC-BVMA and AAC-PEMA. Similar likelihoods of multiple values of *k* can be explained by hierarchical relationships among the individuals (e.g., differentiation among taxa and differentiation among the populations within those taxa) (Evanno et al. 2005). Even with such a hierarchy, we still expected that at $k = 4$ putative conspecifics would form a single cluster and as *k* increased beyond 4 that conspecifics would either form a single cluster or multiple clusters with no heterospecifics present. This was the case for *A. tenella* and *A. obtusifolia* but not *A. decemloba* or *A. acuta* which at $k = 4$ and 5 resulted in heterospecific groups.

Morphology

Using a DFA model with species as the grouping variable and 19 of the morphological variables (Table 2.S7) we distinguished among the five putative species with greater than 89% accuracy (Table 2.S8) when the same cases were used to build and evaluate the model. Median correct classification rates based on cross-validation of 50% of the cases in each species in 10 independent runs were at least 75% for all species, and were above 95% for *A. acuta* and *A. obtusifolia*.

Although this model yielded four significant canonical axes, the first two axes explained 83% of the variance (Table 2.S9). The first axis accounted for 65% of the explained variance and was most important in distinguishing *A. obtusifolia* and *A. tenella* from the other three species (Fig. 2.6 and Table 2.S9). The most important distinguishing character on this axis was the ratio of pedicel length to bract length, which was larger in

A. obtusifolia and *A. tenella*. These species have comparatively long pedicels (median

length 10.7 mm and 12.9 mm, respectively), short bracts (median length 3.4 mm and 4.8

mm, respectively), and shorter calyx lobes. *Agalinis obtusifolia* tended to have much

wider leaves that were also wide relative to their length. Based on both the standardized

canonical coefficients and the factor structure coefficients the second DFA axis was

explained by increasing values of dorsal anther length, the ratio of pedicel to calyx

length, and decreasing values of stigma length. This axis primarily separates *A.*

decemloba and *A. skinneriana* from the three other taxa (Fig. 2.6). *Agalinis obtusifolia*

and *A. tenella* tended to have much longer dorsal anthers (median length 2.0 mm and 2.3

mm respectively) compared to median lengths of ~ 1.5 mm in *A. decemloba* and *A.*

skinneriana. Dorsal anther length was correlated with ventral anther length and both

dorsal and ventral filament lengths, thus this one variable represents overall stamen size.

A DFA model with population as the grouping variable and 18 predictor variables

yielded 11 significant axes, 4 of which explained a cumulative ~79% of variance (Table

2.S10). This model indicated substantial variation within and among populations within

species. Specifically, 5 populations had correct classification rates of < 75%, rates for 6

populations were between 75% and 90%, and 8 populations had ≥ 90% correct

classification (Table 2.S11). Of the 35 misclassified individuals, 80% were classified

into a conspecific population. All 12 *A. obtusifolia* misclassifications were into

conspecific populations (Table 2.S11). Both misclassified *A. decemloba*, 2 out of 5

misclassified *A. tenella* individuals, and 4 of 14 misclassified *A. acuta* individuals were

assigned to heterospecific populations. As with the species level model (Fig. 2.6), there

was substantial overlap in the morphological variation captured by DFA Axis 1 that

followed primarily a south to north gradient (Fig. 2.7). Axis 2 distinguished the *A. tenella* population from Colleton County, South Carolina (ATE-CCSC) from all other populations.

In addition to the continuous variables discussed above, the absence of pink spots and yellow guidelines from *A. obtusifolia* has been considered diagnostic for this species. Based on our sampling and other field observations we found this to be a consistent diagnostic character for this species.

DISCUSSION

The combined evidence from the analysis of DNA sequence, microsatellite, and morphological variation in putative *A. acuta*, *A. tenella*, *A. decemloba*, *A. obtusifolia*, and *A. skinneriana* suggests there are four recognizable lineages that are congruent with what were historically classified as five species based on morphology and anatomy (Bicknell 1915; Greene 1899; Pennell 1913b, 1929; Rafinesque 1836). Further, not all of those lineages met the criteria we and other investigators require for recognition as species. Only *A. obtusifolia* and *A. skinneriana* were monophyletic based on cpDNA sequences at multiple loci. *Agalinis obtusifolia* and *A. skinneriana* also differed from the other taxa and from each other based on diagnostic insertions/deletions at two of those loci. Amplification failure of microsatellite loci due to loss of priming sites or absence of microsatellites altogether represents additional diagnostic differences. Morphological similarities between *A. obtusifolia* and what has been called *A. tenella* (e.g., Fig. 2.6) have led to difficulty in distinguishing the two taxa in the field (J. Hays and J. Canne-Hilliker, personal communication) and resulted in the erroneous synonymization of *A. obtusifolia* with *A. tenella*, as well as *A. decemloba* (Kartesz 1999). However, we found

the absence of pink spots and yellow guidelines on the corolla of *A. obtusifolia* to distinguish it from the other taxa. This species also has relatively short and wide leaves, long pedicels, and short bracts compared to the other species (Table 2.1).

In contrast, relationships among *A. acuta, A. decemloba*, and *A. tenella* based on cpDNA sequences were polyphyletic (Fig. 2.2) and there were no diagnostic cpDNA characters distinguishing these putative taxa from one another. Absence of reciprocal monophyly and diagnostic characters among those three taxa is in stark contrast to the strongly supported monophyletic relationships and diagnosability of almost all other species in the genus (Pettengill & Neel 2008). Branch lengths separating individuals of these three species were similar to those within other species in the genus, and were substantially shorter than branches subtending species (Fig. 2 and Neel and Pettengill 2008 or Fig. 1.2 Chapter 1). We also found no strong breaks in morphological characters that were concordant with species labels (Fig 2.6). Rather, we saw relatively continuous and overlapping morphological variation among populations along a latitudinal gradient (Table 2.1 and Fig. 2.7). The neighbor-joining tree based on D_a distances among populations calculated using microsatellite allele frequencies also strongly supported *A. acuta*, *A. decemloba*, and *A. tenella* a single clade (Fig. 2.3). Given this combined evidence, the most parsimonious taxonomic alignment is to treat them as a single species. This synonymization is in accordance with Baum (2009) in that the resulting group exhibits more genealogical exclusivity (e.g., $gsi_T = 1$) than does the historical taxonomic alignment or grouping *A. acuta* and *A. decemloba* as one species and *A. tenella* as another. Based on nomenclatural precedence, they would be merged under *Agalinis decemloba*.

Within *A. decemloba*, which would now include what was historically known as *A. acuta*, *A. decemloba*, and *A. tenella*, there is support for infraspecific differences that warrant the designation of two subspecies. *Agalinis acuta* and *A. decemloba* would comprise *A. decemloba* ssp. *decemloba*; *A. tenella* would represent *A. decemloba* ssp. *tenella*. Although recognition of a group of individuals as an infraspecific taxon can be as controversial as species level decisions (e.g., Haig et al. 2006; Paetkau 1999; Ryder 1986), Baum (2009) proposes that infraspecific taxa should be distinguished by, at most, "minor traits." We have employed this reasoning in making our decision that *A. acuta* and *A. decemloba* represent a distinct subspecies from *A. tenella* because it is based primarily on microsatellite allele frequency differences.

Support for subsuming putative *A. acuta* and *A. decemloba* populations within a single subspecies to the exclusion of *Agalinis tenella* comes from near identity at cpDNA loci among individuals from those two species (Table 2.5). Although *A. decemloba* ssp. *tenella* and *A. decemloba* ssp. *decemloba* are differentiated based on those same cpDNA loci, the magnitude of the differences are not similar to what is found among other *Agalinis* species and, therefore, are not indicative of species-level differences. Microsatellite-based distances between these two taxa (alone or combined) and *A. tenella* are greater than infraspecific distances in *A. tenella*. However, distances of *A. tenella* from *A. acuta* and *A. decemloba* were lower than distances observed among other well-supported species (Table 2.5). Differential amplification success among the three putative species across the 21 microsatellite loci provides additional support for separate subspecific rank for *A. tenella*. *Agalinis acuta* and *A. decemloba* had nearly 100% amplification success while *A. tenella* had only 95% success (Tables 2.S2 & 2.S3),

suggesting that there is a greater evolutionary and genetic similarity between the former two taxa than between either and *A. tenella*. The substantially higher failure rates for *A. obtusifolia* (36.64%) and *A. skinneriana* (~100%) illustrate that although *A. tenella* differs from *A. acuta* and *A. decemloba*, the magnitude of that difference is substantially less than we observed among other species. The neighbor-joining tree based on D_a distances among populations calculated using microsatellite allele frequencies also strongly supported two distinct clades, one which includes *A. acuta* and *A. decemloba* and the other *A. tenella* (Fig. 2.3). One could argue that the well-supported clades on the neighbor-joining tree support that *A. acuta*, *A. decemloba*, and *A. tenella* should all be subspecies (Fig. 2.3), we suggest that the other analyses of microsatellite allele frequencies indicate those groupings best represent phylogeographic substructure.

The lack of genealogical exclusivity of *A. acuta* and *A. decemloba* based on cpDNA is mirrored in the heterospecific clusters at $k = 4$ and $k = 5$ in the STRUCTURE analyses based on microsatellites. At the same time, all *A. tenella* individuals were placed in a separate cluster at those values of k (Fig. 2.5). At $k = 3$, however, *A. decemloba* and *A. tenella* were grouped together. The AMOVA analyses evaluating the alternative taxonomic groupings of putative *A. acuta*, *A. decemloba*, and *A. tenella* populations, indicate a higher proportion of variance in genetic distances when *A. acuta* and *A. decemloba* represent a single group than when all three putative species are treated separately (Table 2.6). Among *A. acuta* populations, the average D_a distance was 0.489, which is similar to interspecific differentiation between *A. acuta* and *A. decemloba* ($D_a = 0.476$) (Table 2.5). Thus, many *A. acuta* populations are more different from one another than these populations are from putative *A. decemloba* populations. Consequently, some

interspecific distances among putative *A. decemloba* and *A. acuta* populations are less than intraspecific distances within either putative species; such a pattern is not expected for different species or even subspecies.

This comprehensive investigation evaluating the evolutionary distinctiveness of *A. acuta* shows the power of sequentially analyzing multiple data types in context of different species concepts to determine the degree of differentiation among putative species. The negative consequences of inaccurate taxonomy render it essential to apply such a rigorous approach to test taxonomic hypotheses. Phylogenetic methods provide an objective test of the accuracy of species designations if monophyly based on sequence data is observed. Because that criterion is an extreme requirement that takes a significant amount of evolutionary time to acquire and represents the end of the continuum of diversification in the speciation process (Cummings et al. 2008; de Queiroz 2007), there will be many cases in which historically accepted species are neither monophyletic nor diagnosable (e.g., Comes & Abbott 2001; Dueck & Cameron 2008; e.g., Edwards et al. 2008a; Funk & Omland 2003; Syring et al. 2007; Yi et al. 2008). In these cases, the point at which an entity will be considered discreet or differentiated enough to warrant species status will be subjective. However, the rationale for making a determination needs to be transparent.

Because genealogical exclusivity in DNA sequences may not be present for recently diverged species (Baum 2009; Funk & Omland 2003; Hudson & Coyne 2002), lack of monophyly should trigger analysis with more rapidly evolving genomic regions, a more comprehensive sampling of the genome, extensive population-level sampling (e.g., Edwards et al. 2008a) and sampling morphological traits (e.g., Olfelt et al. 2001).

Although questions of evolutionary distinctiveness have resulted from a lack of confidence in the morphological characters upon which taxonomic classifications of a number of rare species were based (e.g., Bacon & Bailey 2006; e.g., Brunsfeld et al. 1991; Case et al. 1998; Mymudes & Les 1993; Nicole et al. 2007; Wood & Nakazato 2009), there have been few comprehensive studies investigating the evolutionary distinctiveness of species listed under the Endangered Species Act. In one such study, six morphologically distinct species in the genus *Conradina* (Lamiaceae) species native to the southeastern United States were found to be paraphyletic based on 3642 bp of combined chloroplast and nuclear DNA sequence (Edwards et al. 2008a; Edwards et al. 2008b). Subsequent analyses employing 10 microsatellite loci found patterns of differentiation that the authors considered sufficient to support recognition of 6 putative species (e.g., STRUCTURE analyses showed all conspecifics forming single clusters at $k = 6$ with minimal or no admixture) (Edwards et al. 2008a; Edwards et al. 2008b). Interestingly, Edwards et al. (2008) also suggest that three populations of *Conradina canescens* warrant species status because they corresponded to three clusters at $k = 8$, although those cluster each had an appreciable amount of admixture; a similar claim was made regarding the genetic distinctness of a group of recently discovered populations from Santa Rosa, FL (Edwards et al. 2008b). We have employed a stricter requirement for species level distinctiveness (e.g., Goldstein et al. 2000) and do not consider microsatellite variation alone to be sufficient to support species level designations.

In contrast, evolutionary distinctiveness of the federally listed narrowly endemic species *Spiranthes parksii* (Orchidaceae) from the more widespread *Spiranthes cernua* was not supported when examined using multiple lines evidence (Dueck & Cameron

2008; Walters 2005). The two species are often sympatric, possess many morphological

similarities, and the differences that have been used historically to differentiate *S. parksii*

were considered reflect local phenotypes of the more widespread and highly variable *S.*

cernua (Dueck & Cameron 2008). Phylogenetic analyses of 3191 bp of DNA sequence

at 4 loci showed *S. parksii* to be nested within *S. cernua*. The patterns of relatedness

among conspecific individuals based on AFLP and microsatellite variation showed little

concordance with the pattern expected of species and as provided no evidence of

substructure that would indicate infraspecific taxonomic groups. After explicitly

considering different species concepts (e.g., biological, morphological/phenetic, and

phylogenetic) in light of the patterns of relatedness they observed, they recommended

that the two species be synonymized.

Conservation implications

The ESA allows listing of taxa below the rank of species (U.S.C. 1973), thus *A.*

decemloba ssp. *decemloba* would represent a listable entity. However, to receive federal

protection, it must either be in danger of extinction throughout all or a significant portion

of its range (endangered) or at risk of becoming endangered in the foreseeable future

(threatened) (U.S.C. 1973). The southward range extension from northern Maryland to

central North Carolina, and possibly to South Carolina that results from merging *A.*

decemloba and *A. acuta* has three potential outcomes: 1) the extinction risk to *A.*

decemloba ssp. *decemloba* remains high enough to warrant continued listing as

endangered; 2) the increased range decreases the threat to the taxon such that it is at risk

of endangerment rather than immediate extinction and thus is more appropriately listed as

threatened; or 3) the taxon is not at risk of immediate extinction or of endangerment in

the foreseeable future and no longer warrants listing. Determining which of these alternatives is true is the purview of the USFWS and would require a full status review, but observations during our sampling efforts provide some insight into the status of the taxon.

We obtained locations of 24 documented occurrences of *A. decemloba* from local and state agencies, major herbaria, and taxonomic experts in the genus. Across multiple years, we visited 22 of these occurrences but found plants at only 3 of them. One of these locations supported only a single individual and thus was not included in our samples. We also searched appropriate habitat at many other locations and found no additional occurrences. Many of the potential explanations for the absence of populations at sites we visited (e.g., succession of grassland habitat to forest, mowing, herbicide spraying on road verges and utility corridors, road construction, and conversion of habitat to agricultural and commercial development) would have been less likely to occur if these populations had received federal protection. Our inability to locate populations may also have been the result of drought conditions, thus some populations we were unable to locate might still be extant. In some cases, we found other *Agalinis* species at the sites indicating potential misidentifications in the original collections. Taxonomic confusion was evident on old herbarium specimens. For example, *A. setacea* specimens were often originally erroneously attributed to *A. decemloba* (Pettengill and Neel, in review). Due to uncertainty in identification and the fact that *A. decemloba* ceased to be tracked by natural heritage programs when it was synonymized with *A. obtusifolia*, we are not able to determine how many populations of what would be attributable to what we suggest be considered *A. decemloba* ssp. *decemloba* actually exist.

Due to extirpation or original misidentification, very few of the documented *A. decemloba* sites in the portion of the range beyond what was recognized as the listed entity *A. acuta* remain extant. We anticipate that focused surveys would locate additional populations, but there is no way to know how many. At the same time, all remaining populations of what was known as *A. acuta* in the northeast are small and isolated and are vulnerable to extirpation if habitat protection and management were not ongoing through implementation of the recovery plan (U.S. Fish and Wildlife Service 1989b). The two populations of *A. decemloba* we sampled exist along utility rights of way (one electrical transmission line and one petroleum pipeline). The transmission line is located within William B. Umstead State Park (ADE-WCNC), which affords a measure of protection and management. The other population from Virginia (ADE-L1VA), which exists along a petroleum pipeline, is unprotected and is likely at risk of extirpation due to an irregular disturbance regime. The third occurrence consisting of a single individual was in a small opening along a trail in an undeveloped portion of the North Carolina Zoological Park. Furthermore, populations ADE-L1VA and ADE-WCNC have high inbreeding coefficients and low levels of allelic richness (Table 2.3) that appear to be of such a magnitude that genetic issues could negatively affect fitness and adaptive potential (Frankham et al. 2002). All populations that would represent *A. decemloba* ssp. *decemloba* are also highly differentiated from one another and all but one possess unique allelic diversity such that extirpation of any single population would reduce the amount of genetic diversity within the subspecies (Table 2.3). Based on all this evidence combined, it appears that the range extension that results from merging *A. acuta* and *A.*

decemloba into *A. decemloba* ssp. *decemloba* does not substantially reduce the risk of extinction.

Conclusions

Using a sequential approach with decreasing levels of restrictiveness provides a rigorous framework in which to assess three different issues that are crucial to determining the evolutionary distinctiveness of *Agalinis acuta* from other species. We were able to 1) identify where along the divergence continuum this entity lies, 2) rigorously compare the observed patterns against the criteria of multiple species concepts, and 3) assign the taxonomic rank for the groupings best supported by the data (Baum 2009; de Queiroz 2007). Including additional putative species beyond the few that were our primary focus provided perspective that the degree of differentiation among other closely related *Agalinis* species was substantially higher than among *Agalinis tenella*, *Agalinis acuta*, and *Agalinis decemloba* (e.g., Baum 2009). Using this comprehensive sequential approach, we have determined that *A. acuta* does not represent a distinct taxon and recommend that it be synonymized with what have been called *A. decemloba* and *A. tenella* into the species *A. decemloba.* We further consider *A. acuta* and *A. decemloba* populations to form the subspecies *A. decemloba* ssp. *decemloba*, which is a distinct taxon that is a listable entity under the Endangered Species Act.

Whether the newly defined *A. decemloba* ssp. *decemloba* still meets the definition of endangerment according to the five listing factors under Section 4 of the Endangered Species Act is ultimately a decision for the United States Fish and Wildlife Service. Although we visited a large number of historic locations that we compiled from herbaria, natural heritage programs, and experts in the genus, a full status review would require a

more thorough search for historic locations and a level of field survey that was beyond

the scope of this study. Based on the following three observations, we suggest that

despite the range expansion associated with creation of *A. decemloba* ssp. *decemloba* the

taxon is still at risk of extinction throughout its range: 1) we could only locate a small

number of putative *A. decemloba* populations (i.e., 3) that would be attributed to *A.*

decemloba ssp. *decemloba*, 2) all 12 populations of *A. acuta* in the northeast require

ongoing management to avoid extirpation, and 3) the estimates of population genetic

parameters associated with all populations we sampled suggests that genetic factors may

have a negative impact on levels of fitness.

Table 2.1. Morphological characteristics that have been used historically to differentiate among *A. acuta*, *A. decemloba*, *A. tenella*, *A. obtusifolia*, and *A. skinneriana*. All measurements are in millimeters.

Species	Corolla Length	Floral Characteristics					
		Calyx Tube Length	Calyx Teeth/ Lobe Length	Yellow Lines on Corolla	Pedicel Length to Calyx Ratio	Pedicel to Bract Ratio	Pedicel Length (at anthesis)
A. acuta	10-13[‡]	3[‡]	0.5-1.0 [‡]	Present[*‡]	1-3:1[‡]	1-2:1[‡]	5-15[‡]
A. decemloba	10-15[*‡]	2.5-3[‡]	0.05-0.2[‡]	Present[*]	2.5:1[*]	2-3:1[‡]	5-20[*‡]
A. tenella	15-20[*‡]	1.5-2.0[‡]	1.5-2.0[‡]	Present[*‡]	2.5:1[*]	3-8:1[‡]	5-20[*], 8-20[‡]
A. obtusifolia	10-15[†], 12-16[‡]	1.8-3[†], 2.0-2.5[‡]	0.1-0.4[†]	Absent[*]	2.5:1[*]	3-5:1[‡]	5-20[*], 3-18[‡]
A. skinneriana	10-15[†], 12-15[‡]	2.5-3.5[†‡]	0.5[†], 0.3-0.8 [‡]	Present[†]	>1[†], 1.5:1[‡]	1-2:1[‡]	5-20[*], 5-17[‡]

[*] Weakley (2008)

[†] Brown and Brown (1984)

[‡] Pennell (1929)

Table 2.2. Location of sampled populations and the number of samples for microsatellites and morphological characters obtained from five *Agalinis* species. *Agalinis skinneriana* populations were not used in analyses because microsatellites failed to amplify.

Species – population location	Population Code	# of Individuals (Morphology)	Latitude	Longitude
Agalinis acuta				
Bay View, Barnstable Co., MA	AAC-BVMA	29 (27)	+41°34'53.68"	-70°31'35.48"
Percival Cemetery, Barnstable Co., MA	AAC-PEMA	29 (14)	+41°41'13.00"	-70°27'22.00"
Scrubby Neck, Dukes Co., MA	AAC-SNMA	16 (0)	+41°21'0.65	-70°36'55.80
Richmond Cemetery, Washington Co., RI	ACC-RCRI	20 (0)	+41°29'41.70"	- 71°37'16.44""
Eppley Wildlife Sanctuary, Washington Co., RI	AAC-EPRI	30 (29)	+ 41°31'41.62"	- 71°34'35.99""
Plainfield Cemetery, Windham Co., CT	AAC-PCCT	30 (30)	+41°41'12.98"	-71°55'22.01"
Montauk Downs, Suffolk Co., NY	AAC-MDNY	30 (14)	+ 41° 3'6.53"N	-71°56'20.32"
Shadmoor State Park, Suffolk Co., NY	AAC-SHNY	24 (10)	+ 41° 2'20.34"	- 71°55'37.23"
Hempstead Plains, Nassau Co., NY	AAC-HPNY	31 (28)	+40°44'2.18"	-73°35'9.77"
Warhol Preserve, Suffolk Co., NY	AAC-WANY	9 (0)	+41° 2'49.40""	- 71°53'31.22"
Soldiers Delight, Baltimore Co., MD	AAC-SDMD	30 (30)	+39°25'2.26"	-76°49'47.37"
Total: 11 populations		Total: 278 (182)		
Agalinis tenella				
Colleton Co. SC	ATE-CCSC	12 (12)	+32°56'4.00"	-80°37'22.22"
Lowndes Co. GA	ATE-LCGA	30 (30)	+30°47'45.42"	-83°22'49.86"
Brooks Co, GA	ATE-BCGA	31 (6)	+30°42'47.76"	-83°29'24.06"
Grady Co., GA	ATE-GCGA	30 (30)	+30°45'4.86"	-84°5'11.22"
Total: 4 populations		Total: 103 (78)		
Agalinis decemloba				
Lunenberg Co. VA	ADE-L1VA	24 (8)	+36° 58'17.46"	-78° 21'30.42"
Wake Co. NC	ADE-WCNC	33 (29)	+35° 52'36.12"	-78° 45'58.80"
Total: 2 populations		Total: 57 (37)		
Agalinis obtusifolia				
Lee Co. SC	AOB-LCSC	30 (30)	+34°2'37.96"	-80°8'14.74"
Dorchester Co. SC	AOB-DCSC	24 (17)	+33°2'56.04"	-80°23'12.84"
Liberty Co. FL	AOB-LCFL	35 (30)	+30°5'1.56"	-85°3'22.92"
Total: 3 populations		Total: 89 (77)		
Agalinis skinneriana				
Chesapeake Forest, Dorchester Co., MD	ASK-CFMD	82 (0)	+38°32'51.43"	-75°47'58.44"
Barton Co., MO	ASK-BCMO	4 (0)		
Vernon Co., MO	ASK-VCMO	4 (0)		

Species – population location	Population Code	# of Individuals (Morphology)	Latitude	Longitude
Prince Georges Co., MD	ASK-PGMD	45 (21)	+38°45'15.19"	-76°51'55.80"
Total: 4 populations		135 (21)		
Grand Total: 24 populations		Total: 662 (395)		

Table 2.3. Population genetic characteristics of species and populations based on microsatellite loci. Values within species are calculated based on 14 loci. P = proportion of polymorphic loci; A = number of alleles; A_P = number of private alleles; A_R = allelic richness as calculated in FSTAT using a rarefaction method; H_o^* = observed heterozygosity based on the FreeNA dataset; intraspecific and interspecific estimates of Jost's D_{EST}.

Species Population	N	Amplification Success (%)	P	A	A_P	A_R	H_o	H_o^*	H_E	R_{IS}	D_{EST} (intra)	D_{EST} (inter)
Agalinis acuta	271	98.08	1.0	113	22	5.12	0.10	0.29	0.60	0.92	n/a	0.680
AAC-BVMA	29	97.29	0.71	29	1	1.58	0.07	0.26	0.19	0.59	0.533	0.870
AAC-SNMA	16	100.00	0.50	23	2	1.43	0.08	0.20	0.16	0.45	0.576	0.844
AAC-PEMA	29	99.01	0.57	26	3	1.59	0.07	0.26	0.20	0.74	0.493	0.856
AAC-PCCT	29	98.52	0.77	38	5	1.94	0.16	0.38	0.32	0.82	0.503	0.799
AAC-EPRI	24	92.26	0.93	50	3	2.52	0.18	0.54	0.45	0.65	0.526	0.793
AAC-RCRI	20	98.93	0.79	37	2	2.28	0.28	0.45	0.40	0.11	0.511	0.828
AAC-HPNY	31	99.08	0.64	27	3	1.76	0.13	0.35	0.29	0.53	0.543	0.761
AAC-MDNY	30	99.76	0.07	15	0	1.04	0.01	0.02	0.01	0.48	0.561	0.839
AAC-SHNY	24	97.92	0.50	21	3	1.19	0.04	0.06	0.05	0.03	0.520	0.821
AAC-WANY	9	93.65	0.79	37	2	2.45	0.13	0.52	0.44	0.77	0.462	0.730
AAC-SDMD	30	99.29	0.64	31	7	1.82	0.05	0.37	0.30	0.94	0.579	0.843
Mean of populations	**24.64**	**97.79**	**0.63**	**30.4**	**2.8**	**1.78**	**0.11**	**0.31**	**0.25**	**0.56**	**0.528**	**0.817**
Agalinis decemloba	57	99.37	1.0	79	7	4.61	0.18	0.49	0.51	0.70	n/a	0.678
ADE-WCNC	33	99.35	0.93	54	7	2.32	0.21	0.43	0.36	0.63	0.420	0.722
ADE-L1VA	24	99.40	1.00	49	8	2.52	0.15	0.57	0.47	0.69	0.420	0.709
Mean of populations	**28.5**	**99.38**	**0.96**	**51.5**	**7.5**	**2.42**	**0.18**	**0.50**	**0.41**	**0.66**	**0.420**	**0.716**
Agalinis tenella	103	95.08	1.00	271	126	12.88	0.60	0.77	0.79	0.31	n/a	0.724
ATE-GCGA	30	95.48	1.00	148	21	5.20	0.57	0.77	0.75	0.26	0.367	0.805
ATE-CCSC	12	93.45	1.00	87	11	4.52	0.61	0.76	0.73	0.37	0.478	0.815
ATE-LCGA	30	96.19	1.00	179	46	5.59	0.60	0.77	0.75	0.24	0.323	0.792
ATE-BCGA	31	94.24	1.00	126	17	4.86	0.60	0.77	0.74	0.14	0.367	0.831
Mean of populations	**25.75**	**94.48**	**1.00**	**135**	**23.8**	**5.04**	**0.60**	**0.77**	**0.74**	**0.25**	**0.384**	**0.811**
Agalinis obtusifolia	89	85.71	1.00	227	120	11.13	0.34	0.69	0.77	0.60	n/a	0.797
AOB-LCFL	35	79.59	1.00	146	52	4.98	0.42	0.77	0.74	0.56	0.709	0.880
AOB-LCSC	30	90.00	1.00	104	26	4.05	0.30	0.68	0.62	0.51	0.730	0.912
AOB-DCSC	24	89.29	0.86	76	25	3.52	0.27	0.60	0.54	0.41	0.501	0.917
Mean of populations	**29.66**	**86.29**	**0.95**	**108. 7**	**34.3**	**4.18**	**0.33**	**0.68**	**0.63**	**0.49**	**0.647**	**0.903**

Table 2.4. Characteristics of the sampled cpDNA loci used in the phylogenetic analyses based on 35 samples representing *A. acuta* (n = 10), *A. decemloba* (n = 5), *A. obtusifolia* (n = 9), *A. tenella* (n = 8), and *A. skinneriana* (n = 3). Intraspecific and average interspecific K2P distances are also shown.

Locus	N	Aligned Length*	Constant Characters (%)	Variable Characters (%)
*mat*K	34	872	853 (97.82)	19 (2.23)
*psb*A-*trn*H	35	585	540 (92.31)	45 (8.33)
*rpo*B	35	367	351 (95.64)	16 (4.56)
*rps*2	35	651	614 (94.32)	37 (6.03)
*rbc*L	31	617	597 (96.76)	20 (3.35)
*trn*T-*trn*F	35	1707	1628 (95.37)	79 (4.85)
Combined	35	4799	4583 (95.50)	216 (4.71)

Table 2-4 (continued)

| Locus | N | Species Level K2P Distances | | | | | | | | | | | |
| | | A. acuta | | A. decemloba | | A. tenella | | A. obtusifolia | | A. skinneriana | | Total | |
		intra	inter	intra	inter	intra	inter	intra	inter	intra	inter	intra	inter
*mat*K	34	0	0.006	0	0.006	0	0.006	0.001	0.013	0.002	0.014	0.001	0.009
*psb*A-*trn*H	35	0.002	0.0143	0	0.011	0.001	0.014	0.002	0.026	0.001	0.023	0.001	0.018
*rpo*B	35	0.001	0.011	0.002	0.011	0.002	0.012	0.002	0.024	0.004	0.019	0.002	0.015
*rps*2	35	0.0	0.017	0.002	0.018	0.001	0.018	0.001	0.045	0.003	0.030	0.002	0.025
*rbc*L	31	0	0.001	0.001	0.009	0.003	0.008	0.003	0.007	0	0.015	0.002	0.010
*trn*T-*trn*F	35	0.001	0.014	0	0.011	0.001	0.014	0.002	0.026	0.001	0.023	0.001	0.018
Combined	35	0.001	0.012	0.001	0.012	0.001	0.014	0.001	0.028	0.002	0.025	0.001	0.018

Table 2.5. Interspecific and intraspecific genetic distances among five putative *Agalinis* species. D_a distances based on 14 microsatellite loci are above the diagonal and K2P distances based on the cpDNA loci are below the diagonal. On the diagonal are the intraspecific D_a distances followed by K2P distances within putative species. Intraspecific D_a distances are the average among putative conspecific populations and interspecific D_a distances are based on combining putative conspecific individuals together into one group and comparing with the other three groups.

	A. acuta	*A. decemloba*	*A. tenella*	*A. obtusifolia*	*A. skinneriana*
A. acuta	0.489/0.00062	0.476	0.568	0.743	n/a
A. decemloba	0.000477	0.386/0.000729	0.589	0.749	n/a
A. tenella	0.001443	0.001253	0.386/0.001055	0.646	n/a
A. obtusifolia	0.025936	0.025784	0.028107	0.474/0.001395	n/a
A. skinneriana	0.022109	0.021481	0.023264	0.031711	(n/a)/0.001956

Table 2.6. Results from analysis of molecular variance based on different taxonomic hypotheses (* denotes statistical significance at *P* < 0.001).
    6a. The full model including all four putative species treated separately.

Source of Variation	d.f.	Sum of Squares	Variance Components	Percentage of Variation
Among putative species	3	1853677.390	2487.15409	55.37*
Among populations within putative species	16	836913.486	1011.69325	22.52*
Within populations	1020	1012776.967	992.91859	22.11*
Total	1039	3703367.842	4491.76593	

6b. *Agalinis tenella*, *A. acuta* and *A. decemloba* comprise separate species; *A. obtusifolia* is not included in the model.

Source of Variation	d.f.	Sum of Squares	Variance Components	Percentage of Variation
Among putative species	2	339760.1	492.552	22.93*
Among populations within putative species	14	704192.2	1001.143	46.6*
Within populations	845	553140.8	654.6045	30.47*
Total	861	1597093	2148.3	

6c. *A. acuta* and *A. decemloba* comprise one taxon and *A. tenella* represents another; *A. obtusifolia* is not included in the model.

Source of Variation	d.f.	Sum of Squares	Variance Components	Percentage of Variation
Among putative species	1	289085.4	743.5039	31.12*
Among populations within putative species	15	754866.9	991.1479	41.48*
Within populations	845	553140.8	654.6045	27.4*
Total	861	1597093	2389.256	

6d. *Agalinis tenella* and *A. decemloba* comprise one 'taxon' and *A. acuta* represents a separate taxon; *A. obtusifolia* is not included in the model.

Source of Variation	d.f.	Sum of Squares	Variance Components	Percentage of Variation
Among putative species	1	240974	451.5327	20.9*
Among populations within putative species	15	802978.3	1054.82	48.81*
Within populations	845	553140.8	654.6045	30.29*
Total	861	1597093	2160.957	

Table 2.S1. Samples of *A. acuta*, *A. decemloba*, *A. obtusifolia*, *A. skinneriana,* and *A. tenella* that were included in the phylogenetic analyses and the loci for which sequences were obtained. The EU numbers below certain loci refer to the Genebank accession numbers and those loci that have a YES will be deposited into Genbank; those samples for which a sequence could not be obtained for a given locus are denoted with NO. Information regarding the other samples included in the phylogenetic analyses (Fig. 2.1) can be found in Pettengill and Neel (2008)

Taxon	Sampled Location	Population Code for Microsatellite Analyses	*mat*K	*rbc*L	*rps*2	*trn*T -*trn*F	*trn*H-*psb*A	*rpo*B
Focal Species								
A. acuta 125CT	Windham Co., CT	AAC-PCCT	YES	YES	EU827965	YES	EU827882	EU828046
A. acuta 139RI	Washington Co., RI	AAC-EPRI	YES	YES	EU827966	EU828129	EU827883	EU828047
A. acuta 13PCMA	Sandwich, Barnstable Co. MA	AAC-PEMA	YES	YES	EU827967	EU828130	EU827884	EU828048
A. acuta 1BVMA	Waquoit Bay, Barnstable Co. MA	AAC-BVMA	YES	YES	EU827968	EU828131	EU827885	EU828049
A. acuta 211HPNY	Nassau Co., NY	AAC-HPNY	YES	YES	EU827969	EU828132	EU827886	EU828050
A. acuta 229MDNY	Montauk Downs, Suffolk Co. NY	AAC-MDNY	YES	YES	EU827970	EU828133	EU827887	EU828051
A. acuta 265SMNY	Shadmoor,Suffolk Co., NY	AAC-SHNY	YES	YES	EU827971	EU828134	EU827888	EU828052
A. acuta 33SNMA	Dukes Co., MA	AAC-SNMA	YES	YES	EU827972	EU828135	EU827889	EU828053
A. acuta 51MD	Baltimore Co., MD	AAC-SDMD	YES	YES	EU827973	EU828136	EU827890	EU828054
A. acuta 161RI	Washington Co., RI	AAC-RCRI	YES	YES	YES	YES	YES	YES
A. decemloba 1VA	Lunenberg Co., VA		NO	YES	YES	YES	YES	YES
A. decemloba 6VA	Lunenberg Co., VA		YES	YES	EU827982	EU828145	EU827899	EU828063
A. decemloba 45VA	Lunenberg Co., VA	ADE-L1VA	YES	YES	YES	YES	YES	YES
A. decemloba 19NC	Wake Co., NC	ADE-WCNC	YES	NO	YES	YES	YES	YES
A. decemloba 9NC	Randolph Co., NC		YES	YES	EU827983	EU828146	EU827900	EU828064
A. obtusifolia 177SC	Dorchester Co., SC	AOB-DCSC	YES	YES	YES	YES	YES	YES
A. obtusifolia 169SC	Lee Co., SC	AOB-LCSC	YES	YES	YES	YES	YES	YES
A. obtusifolia 18FL	Fl.		YES	NO	YES	YES	YES	YES
A. obtusifolia 10AL	Baldwin Co., AL		YES	NO	YES	YES	YES	YES
A. obtusifolia 13AL	Geneva Co., AL		YES	YES	EU828008	EU828171	EU827925	EU828089
A. obtusifolia 14AL	Mobile Co., AL		YES	YES	YES	EU828172	EU827926	EU828089
A. obtusifolia 20FL	Liberty Co., FL	AOB-LCFL	YES	YES	EU828010	EU828174	EU827928	EU828092
A. obtusifolia 6AL	Mobile Co., AL		YES	YES	EU828011	EU828175	EU827929	EU828093
A. obtusifolia 8AL	Geneva Co., AL		YES	YES	YES	EU828176	EU827930	EU828094

Taxon	Sampled Location	Population Code for Microsatellite Analyses	*mat*K	*rbc*L	*rps*2	*trn*T -*trn*F	*trn*H-*psb*A	*rpo*B
A. skinneriana 106MD	Prince Georges Co., MD		YES	YES	EU828028	EU828193	EU827947	EU828110
A. skinneriana 78MD	Dorchester Co., MD		YES	YES	EU828029	EU828194	EU827948	EU828111
A. skinneriana 90MO	Vernon Co., MO		YES	YES	EU828030	EU828195	EU827949	EU828112
A. tenella 1GA	Ware Co., GA		YES	YES	EU828009	EU828173	EU827927	EU828091
A. tenella 79GA	Brooks Co., GA	ATE-BCGA	YES	NO	YES	YES	YES	YES
A. tenella 11GA	Lowndes Co., GA	ATE-LCGA	YES	YES	EU828032	EU828197	EU827951	EU828114
A. tenella 13GA	Grady Co., GA		YES	YES	EU828033	EU828198	EU827952	EU828115
A. tenella 91GA	Grady Co., GA	ATE-GCGA	YES	NO	YES	YES	YES	YES
A. tenella 3SC	Colleton Co., SC	ATE-CCSC	YES	YES	EU828034	EU828199	EU827953	EU828116
A. tenella 4GA	Ware Co., GA		YES	YES	EU828035	EU828200	EU827954	EU828117
A. tenella 9GA	Lanier Co., GA		YES	YES	EU828036	EU828201	EU827955	EU828118
Outgroup								
A. calycina	Pecos Co., TX		YES	YES	EU828039	EU828144	EU827895	EU828247

Table 2.S2. The number of genotypes per locus resolved from populations of *Agalinis acuta*.

						Agalinis acuta						
Locus	AAC-SDMD (N = 30)	AAC-WANY (N = 9)	AAC-PCMA (N =29)	AAC-BVMA (N = 29)	AAC-MDNY (N =30)	AAC-SHNY (N =24)	AAC-HPNY (N = 31)	AAC-EPRI (N =24)	AAC-PCCT (N = 29)	AAC-RCRI (N =20)	AAC-SNMA (N = 16)	% Success
1) Agac.M1*	29	9	28	28	27	24	29	24	29	20	16	97.05
2) Agac.M6*	30	8	27	29	30	22	31	24	29	19	16	97.79
3) Agac.M33	30	9	29	26	30	24	31	18	26	20	16	95.57
4) Agac.M11	29	9	29	29	29	24	30	24	29	20	16	98.89
5) Agac.M14	29	8	29	29	30	23	31	22	28	20	16	97.79
6) Agac.M28	30	9	28	27	30	24	31	23	28	20	16	98.15
7) Agac.M42	30	7	28	28	30	23	31	23	29	20	16	97.79
8) Agac.M46	29	7	28	27	30	24	31	23	28	20	16	97.05
9) Agac.ca 11	30	9	29	29	30	24	31	24	29	20	16	100.00
10) Agac.ca20	30	9	29	29	30	21	30	24	29	20	16	98.52
11) Agac.ca26	30	9	29	29	30	23	30	21	29	19	16	97.79
12) Agac.aag46	30	6	28	29	30	23	31	19	29	18	16	95.57
13) Agac.ca10	30	9	29	29	30	24	30	18	29	20	16	97.42
14) Agac.aag29*	30	9	29	29	30	23	31	24	29	20	16	99.63
15) Agac.ca48*	30	9	29	29	30	24	31	24	29	20	16	100.00
16) Agac.ca 21	30	9	29	29	30	24	31	24	29	20	16	100.00
17) Agac.taca12*	30	9	29	29	30	24	31	24	29	20	16	100.00
18) Agac.taca45*	30	9	29	29	30	24	31	23	29	20	15	99.26
19) Agac.taca04	30	9	29	29	30	24	31	23	29	20	16	99.63
20) Agac.ca45*	30	8	29	29	30	24	31	24	29	20	16	99.63
21) Agac.ca33	30	9	29	26	30	24	31	24	29	20	16	98.89
% Success	99.37	94.71	98.85	98.03	99.37	98.02	99.08	94.64	99.01	99.05	99.70	

* Due to the large number of missing genotypes these loci were not included in the analyses.

Table 2.S3. The number of genotypes per locus resolved from populations of *Agalinis decemloba, Agalinis tenella*, and *Agalinis obtusifolia*. Populations with genotypes missing from >50% of individuals sampled for a locus are indicated in bold.

Locus	*Agalinis decemloba*		*Agalinis tenella*			
	ADE-L1VA (N = 24)	ADE-WCNC (N = 33)	ATE-CCSC (N = 12)	ATE-GCGA (N = 30)	ATE-LCGA (N =30)	ATE-BCGA (N = 31)
1) Agac.M1*	24	33	9	14	27	13
2) Agac.M6*	24	33	12	30	28	31
3) Agac.M33	24	33	12	29	29	31
4) Agac.M11	24	33	12	30	27	31
5) Agac.M14	24	32	12	20	27	25
6) Agac.M28	24	33	11	24	25	19
7) Agac.M42	24	33	7	29	30	31
8) Agac.M46	24	33	11	29	29	31
9) Agac.ca 11	24	33	12	30	30	31
10) Agac.ca20	24	33	12	30	29	29
11) Agac.ca26	22	33	12	30	30	31
12) Agac.aag46	24	33	10	30	30	29
13) Agac.ca10	24	33	10	30	29	31
14) Agac.aag29*	24	27	12	30	30	31
15) Agac.ca48*	24	33	4	30	25	30
16) Agac.ca 21	24	33	12	30	29	31
17) Agac.taca12*	23	15	9	9	13	0
18) Agac.taca45*	24	33	12	30	29	30
19) Agac.taca04	24	31	12	30	30	29
20) Agac.ca45*	24	33	12	29	26	31
21) Agac.ca33	24	33	12	30	30	30
% Success	99.40	96.10	90.07	90.95	92.38	88.33

Table 2.S3 continued.

Locus	*Agalinis obtusifolia*			% Success
	AOB-LCFL (N = 35)	AOB-LCSC (N = 30)	AOB-DCSC (N = 24)	
1) Agac.M1*	13	1	0	53.82
2) Agac.M6*	0	0	22	72.29
3) Agac.M33	5	9	9	72.69
4) Agac.M11	29	29	24	95.98
5) Agac.M14	33	30	23	90.76
6) Agac.M28	26	22	23	83.13
7) Agac.M42	18	26	23	88.76
8) Agac.M46	29	30	24	96.39
9) Agac.ca 11	34	30	24	99.60
10) Agac.ca20	25	26	20	91.57
11) Agac.ca26	25	29	21	93.57
12) Agac.aag46	34	30	24	97.99
13) Agac.ca10	34	29	23	97.59
14) Agac.aag29*	0	0	0	61.85
15) Agac.ca48*	1	16	24	75.10
16) Agac.ca 21	33	30	22	97.99
17) Agac.taca12*	0	15	7	36.55
18) Agac.taca45*	0	0	0	63.45
19) Agac.taca04	33	30	24	97.59
20) Agac.ca45*	0	0	0	62.25
21) Agac.ca33	32	28	16	94.38
% Success	54.97	65.07	70.04	

* Due to the large number of missing genotypes these loci were not included in the analyses.

Table 2.S4. Description of the morphological characters measured for each plant. N = the total number of individuals sampled from each species. Population assignments to species are according to the original identifications. The actual number for which there are data varies by character

Character	Short Name	Measurement Description	Units	Mean (standard deviation)				
				AAC N=175	ADE N=67	ATE N=78	AOB N=76	ASK N=21
Plant height	Height	Distance from ground to highest point on plant	cm	21.26 (7.24)	32.17 (13.17)	52.79 (10.53)	53.30 (14.74)	13.55 (2.12)
Plant width	Width	Average of two measurements: widest distance across the plant canopy and distance perpendicular to the widest point.	cm	5.72 (3.24)	13.04 (11.41)	33.19 (14.16)	19.86 (12.59)	2.42 (1.01)
Height to width ratio	H/W	Height / Width	none	4.41 (1.83)	4.14 (2.88)	1.86 (0.78)	4.06 (3.25)	6.50 (2.73)
Number of primary branches	nmbrnch	Number of primary branches originating from near the base of a plant	count	3.52 (2.15)	4.91 (4.33)	9.17 (4.72)	4.67 (3.39)	2.48 (1.47)
Leaf length	**LeafL**	Distance from the point at which the leaf inserts on the stem to the tip of the leaf. Measurements from five leaves per plant averaged to give one value per plant.	mm	14.15 (3.03)	16.02 (6.74)	12.69 (2.72)	10.20 (1.86)	8.47 (1.33)
Leaf width	LeafW	Distance across the widest point of a leaf. Measurements from five leaves per plant averaged to give one value per plant.	mm	1.05 (0.17)	0.88 (0.17)	0.93 (0.24)	1.20 (0.19)	0.99 (0.14)
Leaf length to width ratio	**LeafL/W**	Length / Width	none	13.63 (2.92)	18.51 (7.50)	14.03 (2.77)	8.55 (1.32)	8.69 (1.80)
Pedicel length at anthesis	Pedicel	Distance from the point at which the pedicel inserts on the stem to the base of the calyx. Measurements from five pedicels per plant averaged to give one value per plant.	mm	9.86 (4.80)	8.76 (4.60)	13.12 (2.48)	10.79 (2.53)	4.77 (1.06)
Bract length	**Bract**	Distance from the point at which the bract subtending a pedicel inserts on the stem to the tip of the bract.	mm	6.94 (1.93)	7.25 (5.25)	4.22 (1.29)	3.42 (0.95)	4.63 (2.23)

Character	Short Name	Measurement Description	Units	Mean (standard deviation)				
				AAC N=175	ADE N=67	ATE N=78	AOB N=76	ASK N=21
Pedicel to bract ratio	**Ped/Bract**	Pedicel / Bract	none	1.43 (0.57)	1.55 (0.82)	3.30 (0.86)	3.25 (0.66)	1.29 (0.69)
Adaxial corolla tube length	**CorollaAd**	Distance from the base of the corolla to the sinus between corolla lobes on the adaxial side of the corolla (the side towards the centerline of the plant, (i.e., the top of the flower)).	mm	6.03 (1.24)	6.93 (1.49)	6.60 (1.18)	6.77 (1.06)	5.57 (0.87)
Abaxial corolla tube length	CorollaAb	Distance from the base of the corolla to the sinus between corolla lobes on the abaxial side of the corolla (away from the centerline of the plant, (i.e., the bottom of the flower)).	mm	8.50 (1.53)	10.58 (1.74)	10.68 (1.65)	9.90 (1.34)	6.91 (0.75)
Corolla tube height	**TubeDiH**	Diameter of the corolla tube, measured from top to bottom at the mid-point of the distance between the base of the corolla tube and the base of the corolla lobes.	mm	3.25 (0.66)	3.53 (0.68)	3.61 (0.57)	3.72 (0.72)	2.67 (0.41)
Corolla tube width	TubeDiW	Diameter of the corolla tube, measured from side to side at the mid-point of the distance between the base of the corolla tube and the base of the corolla lobes.	mm	4.25 (0.87)	5.26 (0.92)	5.24 (0.70)	5.30 (1.02)	3.85 (0.61)
Corolla tube height to width ratio	**CorTubH/W**	Corolla tube height / Corolla tube width	none	0.78 (0.14)	0.67 (0.09)	0.70 (0.13)	0.71 (0.11)	0.71 (0.14)
Corolla throat height	**ThroatDiH**	Diameter of the corolla throat opening measured top to bottom.	mm	4.18 (1.05)	5.06 (0.93)	5.21 (0.71)	5.27 (0.79)	3.52 (0.64)
Corolla throat width	ThroatDiW	Diameter of the corolla throat opening measured side to side.	mm	4.63 (1.28)	6.62 (1.21)	7.06 (1.03)	7.02 (1.11)	4.75 (1.01)
Corolla throat height to width ratio	**ThroatH/W**	Throat diameter height / Throat diameter width.	none	0.92 (0.18)	0.77 (0.10)	0.74 (0.09)	0.76 (0.12)	0.75 (0.12)
Upper corolla lobe length	**UprLobeL**	Distance from the base of the sinus between corolla lobes to the tip of an upper corolla lobe.	mm	4.17 (1.10)	4.71 (1.02)	5.55 (1.07)	4.73 (0.86)	3.28 (0.48)
Upper corolla lobe width	UpLobeW	Widest distance across the upper corolla lobe.	mm	4.22 (0.90)	5.44 (0.97)	6.71 (1.24)	5.81 (1.09)	3.65 (0.52)

				Mean (standard deviation)				
Character	**Short Name**	**Measurement Description**	**Units**	**AAC** **N=175**	**ADE** **N=67**	**ATE** **N=78**	**AOB** **N=76**	**ASK** **N=21**
Upper corolla lobe length to width ratio	**CoUprLobL/W**	Upper corolla lobe length / width	none	1.00 (0.25)	0.88 (0.20)	0.84 (0.16)	0.83 (0.18)	0.91 (0.16)
Lower corolla lobe length	lowLobeL	Distance from the base of the sinus between corolla lobes to the tip of a lower corolla lobe.	mm	3.89 (0.87)	4.58 (0.87)	5.28 (0.96)	4.55 (0.96)	2.85 (0.53)
Lower corolla lobe width	lowLobeW	Widest distance across the upper corolla lobe.	mm	4.70 (0.99)	5.59 (0.81)	6.64 (1.27)	6.15 (1.12)	3.76 (0.57)
Lower corolla lobe length to width ratio	**CoLLobeL/W**	Lower corolla lobe length / width	none	0.85 (0.18)	0.82 (0.13)	0.82 (0.22)	0.75 (0.11)	0.76 (0.13)
Distance from corolla tube to upper lobe tip	**DistTubeLobe**	Straight-line distance from top surface of the corolla tube to the tip of the upper corolla lobe. Measures the degree to which the upper corolla is reflexed backwards.	mm	3.06 (1.16)	2.65 (0.56)	3.40 (1.04)	3.31 (0.86)	2.01 (0.70)
Style length	Style	Distance from point at which style inserts onto the ovary to the point at which the stigmatic surface begins; does not include the length of the stigmatic surface.	mm	6.74 (1.17)	8.85 (2.48)	8.40 (1.11)	7.68 (0.81)	5.44 (0.59)
Stigma length	**Stigma**	Length of the stigmatic surface.	mm	1.28 (0.42)	1.67 (0.44)	1.90 (0.43)	1.54 (0.33)	1.70 (0.42)
Length of Style+Stigma	**stylstig**	Style+Stigma	mm	8.02 (1.34)	10.52 (2.69)	10.30 (1.27)	9.22 (0.94)	7.14 (0.79)
Ratio of length of the stigma/style complex to corolla tube length	**stigexsert**	Stylestig/Corollab	none	1.37 (0.33)	1.53 (0.26)	1.60 (0.31)	1.39 (0.21)	1.30 (0.18)
Dorsal anther length	**DAntherL**	Length of one of the two dorsal anthers	mm	1.65 (0.28)	1.53 (0.29)	2.34 (0.35)	1.99 (0.26)	1.43 (0.11)
Dorsal anther width	DAntherW	Width of one of the two dorsal anthers	mm	0.87 (0.24)	0.92 (0.20)	1.25 (0.24)	1.08 (0.20)	0.78 (0.19)
Dorsal anther length to width ratio	**DAntherL/W**	Dorsal anther length / width	none	2.02 (0.62)	1.70 (0.34)	1.94 (0.42)	1.89 (0.40)	1.92 (0.42)
Dorsal stamen filament length	DFilament	Length of dorsal stamen filament from point of insertion on the corolla tube to insertion on the anther	mm	3.97 (0.96)	4.50 (0.84)	4.85 (0.86)	4.35 (0.56)	3.79 (0.64)

Character	Short Name	Measurement Description	Units	Mean (standard deviation)				
				AAC N=175	ADE N=67	ATE N=78	AOB N=76	ASK N=21
Dorsal anther to filament ratio	**Danth/Fil_ratio**	Dorsal anther length / filament length	none	0.44 (0.13)	0.35 (0.07)	0.50 (0.11)	0.46 (0.08)	0.39 (0.07)
Ventral anther length	VAntherL	Length of one of the two ventral anthers	mm	1.63 (0.24)	1.61 (0.34)	2.09 (0.33)	1.82 (0.24)	1.42 (0.11)
Ventral anther width	**VAntherW**	Width of one of the two ventral anthers	mm	0.99 (0.29)	1.05 (0.17)	1.42 (0.38)	1.30 (0.28)	0.95 (0.18)
Ventral anther length to width ratio	VAntherL/W	Ventral anther length / width	none	1.76 (0.56)	1.56 (0.36)	1.58 (0.47)	1.47 (0.38)	1.54 (0.29)
Ventral stamen filament length	VFilament	Length of ventral stamen filament from point of insertion on the corolla tube to insertion of the anther	mm	5.47 (1.18)	6.69 (1.05)	6.71 (1.34)	6.43 (0.68)	4.40 (0.65)
Ventral anther to filament ratio	**Vanth/Fil**	Ventral anther length / filament length	none	0.31 (0.08)	0.24 (0.04)	0.33 (0.09)	0.29 (0.05)	0.33 (0.05)
Adaxial calyx tube length	**CalyxAdaxial**	Distance from the base of the calyx to the sinus between calyx lobes on the adaxial side of the calyx (the side towards the centerline of the plant, (i.e., the top of the flower)).	mm	3.52 (0.67)	3.80 (0.75)	3.59 (0.82)	3.14 (0.37)	3.60 (0.60)
Abaxial calyx tube length	CalyxAbaxial	Distance from the base of the calyx to the sinus between calyx lobes on the abaxial side of the calyx (away from the centerline of the plant, [i.e., the bottom of the flower]).	mm	3.28 (0.65)	3.89 (0.78)	3.52 (0.82)	3.17 (0.42)	3.21 (0.47)
Ratio of pedicel length to calyx length	**Ped/Cal**	Pedicel/(CalyxAbaxial+CalyxlowL)	none	2.36 (1.04)	1.93 (0.94)	3.14 (0.75)	2.88 (0.82)	1.24 (0.28)
Calyx diameter height	**CalyxdiH**	Distance across the opening of the calyx, measured from top to bottom	mm	1.77 (0.28)	2.05 (0.25)	2.18 (0.28)	1.87 (0.27)	1.76 (0.26)
Calyx diameter width	CalyxDiamterW	Distance across the opening of the calyx, measured from side to side	mm	1.76 (0.29)	2.05 (0.25)	2.16 (0.29)	1.87 (0.27)	1.77 (0.22)
Calyx diameter height to width ratio	**CaDiL/W**	Calyx diameter height / width	none	1.02 (0.17)	1.00 (0.00)	1.01 (0.05)	1.00 (0.00)	0.99 (0.07)
Calyx upper lobe length	**CalyxupL**	Distance from base of sinus between calyx lobes to the tip of an upper lobe	mm	1.03 (0.26)	0.93 (0.18)	0.84 (0.29)	0.77 (0.20)	0.87 (0.19)

Character	Short Name	Measurement Description	Units	Mean (standard deviation)				
				AAC N=175	ADE N=67	ATE N=78	AOB N=76	ASK N=21
Calyx upper lobe width	CalyxupW	Distance from sinus to sinus at the base of an upper calyx lobe	mm	1.18 (0.27)	1.37 (0.20)	1.34 (0.34)	1.23 (0.23)	1.25 (0.17)
Calyx upper lobe length to width ratio	**CaUpLobeL/W**	Upper calyx lobe length / width	none	0.90 (0.27)	0.69 (0.14)	0.76 (0.63)	0.63 (0.16)	0.70 (0.15)
Calyx lower lobe length	CalyxlowL	Distance from base of sinus between calyx lobes to the tip of a lower lobe	mm	0.97 (0.24)	0.82 (0.20)	0.74 (0.31)	0.72 (0.17)	0.81 (0.22)
Calyx lower lobe width	CalyxlowW	Distance from sinus to sinus at the base of a lower calyx lobe	mm	1.21 (0.31)	1.55 (0.23)	1.51 (0.43)	1.24 (0.17)	1.31 (0.19)
Calyx lower lobe length to width ratio	CaLLobeL/W	Lower calyx lobe length / width	none	0.82 (0.24)	0.53 (0.12)	0.65 (0.67)	0.59 (0.15)	0.62 (0.17)

Table 2.S5. Medians and quartile values for characters measured for five *Agalinis* species. AAC = *A. acuta*, ADE = *A. decemloba*, ATE = *A. tenella*, AOB = *A. obtusifolia*, and ASK = *A. skinneriana*. Population assignments to species are according to the original identifications.

Character	AAC N=175	ADE N=67	ATE N=78	AOB N=76	ASK N=21
Height Median	20.5	32.0	53.0	54.5	13.0
Height quartiles	16 to 26	23.0 to 42	47.0 to 59	43.0 to 65	12.5 to 14.5
Width Median	5.0	9.0	30.0	18.8	2.3
Width quartiles	3.3 to 7.5	3.5 to 20.5	22.0 to 44.5	11.0 to 26.3	1.5 to 3
H/W Median	4.0	3.0	1.7	3.0	6.0
H/W quartiles	3.1 to 5	2.1 to 5.3	1.3 to 2.2	2.0 to 4.4	4.3 to 8.6
nmbrnch Median	3.0	3.0	8.0	4.0	3.0
nmbrnch quartiles	1.0 to 5	2.0 to 8	6.0 to 11	2.0 to 6	1.0 to 3
LeafL Median	14.2	15.1	12.4	10.1	8.3
LeafL quartiles	12.3 to 16.1	10.9 to 19.7	10.8 to 14.6	8.7 to 11.4	7.4 to 9.5
LeafW Median	1.1	0.9	0.9	1.2	1.0
LeafW quartiles	0.9 to 1.1	0.8 to 1	0.7 to 1.1	1.1 to 1.3	0.9 to 1.2
LeafL/W Median	13.7	16.7	13.5	8.6	9.0
LeafL/W quartiles	11.8 to 15.2	13.0 to 22.8	11.9 to 15.7	7.7 to 9.3	7.1 to 9.7
Pedicel Median	9.1	8.1	13.2	10.8	4.7
Pedicel quartiles	6.7 to 11.4	5.5 to 11.1	11.3 to 14.7	8.8 to 12.3	4.0 to 5.8
Bract Median	7.1	5.0	4.1	3.3	4.6
Bract quartiles	5.8 to 8	3.1 to 10.8	3.4 to 4.9	2.8 to 3.9	3.1 to 5.9
Ped/Bract Median	1.3	1.4	3.2	3.2	1.1
Ped/Bract quartiles	1.1 to 1.7	0.9 to 1.9	2.7 to 3.8	2.8 to 3.7	0.8 to 1.6
CorollaAd Median	6.0	7.0	6.7	6.8	5.7
CorollaAd quartiles	5.4 to 6.6	5.7 to 8.1	6.0 to 7.5	6.1 to 7.2	5.1 to 6.2
CorollaAb Median	8.4	10.8	10.8	9.7	7.0
CorollaAb quartiles	7.6 to 9.4	9.7 to 11.7	10.1 to 11.9	8.9 to 10.9	6.2 to 7.5
TubeDiH Median	3.3	3.5	3.6	3.7	2.5
TubeDiH quartiles	2.7 to 3.6	3.2 to 4	3.3 to 4	3.1 to 4.2	2.4 to 3
TubeDirW Median	4.1	5.2	5.3	5.2	4.0
TubeDirW quartiles	3.6 to 4.7	4.7 to 5.8	4.8 to 5.8	4.6 to 6	3.3 to 4.4
CorTubH/W Median	0.8	0.7	0.7	0.7	0.7
CorTubH/W quartiles	0.7 to 0.9	0.6 to 0.7	0.6 to 0.8	0.6 to 0.8	0.6 to 0.8
ThroatDiH Median	4.0	4.9	5.3	5.3	3.5
ThroatDiH quartiles	3.5 to 5	4.3 to 5.7	4.8 to 5.7	4.8 to 5.8	2.9 to 4

Character	AAC N=175	ADE N=67	ATE N=78	AOB N=76	ASK N=21
ThroatDiW Median	4.4	6.7	7.1	7.1	4.7
ThroatDiW quartiles	3.7 to 5.4	5.8 to 7.5	6.4 to 7.7	6.3 to 7.6	4.1 to 5.6
ThroatH/W Median	0.9	0.8	0.7	0.7	0.7
ThroatH/W quartiles	0.8 to 1	0.7 to 0.8	0.7 to 0.8	0.7 to 0.8	0.7 to 0.8
UprLobeL Median	4.2	4.7	5.6	4.6	3.5
UprLobeL quartiles	3.3 to 4.9	4.1 to 5.4	4.9 to 6.5	4.1 to 5.2	2.9 to 3.6
UpLobeW Median	4.2	5.4	6.7	5.7	3.5
UpLobeW quartiles	3.6 to 4.9	4.7 to 6.2	5.7 to 7.7	5.2 to 6.5	3.4 to 3.9
CoUprLobL/W Median	1.0	0.8	0.8	0.8	0.9
CoUprLobL/W quartiles	0.8 to 1.1	0.7 to 1.1	0.7 to 1	0.7 to 0.9	0.8 to 1
lowLobeL Median	3.8	4.7	5.2	4.3	3.0
lowLobeL quartiles	3.3 to 4.5	3.9 to 5.2	4.8 to 5.7	3.8 to 5.3	2.4 to 3.3
lowLobeW Median	4.6	5.7	6.5	6.1	3.7
lowLobeW quartiles	4.0 to 5.4	5.0 to 6.2	5.7 to 7.8	5.3 to 6.8	3.4 to 4.2
CoLLobeL/W Median	0.8	0.8	0.8	0.8	0.7
CoLLobeL/W quartiles	0.7 to 0.9	0.7 to 0.9	0.7 to 0.9	0.7 to 0.8	0.7 to 0.9
DistTubeLobe Median	3.0	2.8	3.2	3.2	2.0
DistTubeLobe quartiles	2.2 to 3.8	2.3 to 3.1	2.7 to 3.8	2.7 to 4	1.2 to 2.5
Style Median	6.8	8.0	8.5	7.7	5.5
Style quartiles	5.9 to 7.6	6.8 to 11.4	7.8 to 9.1	7.1 to 8.2	5.1 to 5.9
Stigma Median	1.2	1.7	1.8	1.5	1.7
Stigma quartiles	1.0 to 1.5	1.4 to 1.9	1.6 to 2.2	1.3 to 1.7	1.5 to 2
stigstyle Median	7.9	9.8	10.3	9.2	7.3
stigstyle quartiles	7.1 to 9	8.5 to 13.2	9.8 to 11.1	8.5 to 9.9	6.4 to 7.8
stigexsert Median	1.3	1.5	1.6	1.4	1.3
stigexsert quartiles	1.2 to 1.5	1.4 to 1.7	1.4 to 1.8	1.2 to 1.5	1.2 to 1.4
DAntherL Median	1.6	1.5	2.3	2.0	1.5
DAntherL quartiles	1.5 to 1.9	1.3 to 1.7	2.1 to 2.6	1.8 to 2.2	1.4 to 1.5
DAntherW Median	0.9	0.9	1.2	1.1	0.8
DAntherW quartiles	0.7 to 1	0.8 to 1	1.1 to 1.4	1.0 to 1.2	0.6 to 0.9
DAntherL/W Median	1.9	1.7	2.0	1.9	1.9
DAntherL/W quartiles	1.6 to 2.3	1.5 to 1.9	1.6 to 2.2	1.6 to 2.2	1.6 to 2.2
DFilament Median	3.8	4.5	4.5	4.4	3.6
DFilament quartiles	3.4 to 4.5	4.0 to 5.2	4.3 to 5.3	4.0 to 4.6	3.5 to 4.3
Dant/Fil_ratio Median	0.4	0.3	0.5	0.5	0.4
Dant/Fil_ratio quartiles	0.4 to 0.5	0.3 to 0.4	0.4 to 0.6	0.4 to 0.5	0.3 to 0.4

Character	AAC N=175	ADE N=67	ATE N=78	AOB N=76	ASK N=21
VAntherL Median	1.6	1.7	2.1	1.9	1.4
VAntherL quartiles	1.5 to 1.8	1.3 to 1.8	1.9 to 2.3	1.7 to 2	1.4 to 1.5
VAntherW Median	1.0	1.1	1.4	1.3	1.0
VAntherW quartiles	0.8 to 1.1	0.9 to 1.2	1.2 to 1.7	1.1 to 1.5	0.8 to 1
VAntherL/W Median	1.7	1.5	1.5	1.4	1.5
VAntherL/W quartiles	1.4 to 2	1.3 to 1.8	1.2 to 1.9	1.2 to 1.7	1.3 to 1.7
VFilament Median	5.3	6.7	6.9	6.4	4.5
VFilament quartiles	4.8 to 6.1	5.9 to 7.4	6.2 to 7.6	5.9 to 6.9	4.2 to 4.8
Vanth/Fil Median	0.3	0.2	0.3	0.3	0.3
Vanth/Fil quartiles	0.3 to 0.4	0.2 to 0.3	0.3 to 0.4	0.3 to 0.3	0.3 to 0.3
CalyxAdaxial Median	3.5	3.9	3.5	3.2	3.6
CalyxAdaxial quartiles	3.0 to 3.9	3.3 to 4.3	3.0 to 4.4	2.9 to 3.3	3.3 to 3.9
CalyxAbaxial Median	3.3	3.9	3.5	3.2	3.2
CalyxAbaxial quartiles	2.8 to 3.7	3.3 to 4.4	3.0 to 4.1	2.9 to 3.5	3.0 to 3.5
Ped/Cal Median	2.1	1.8	3.1	2.8	1.3
Ped/Cal quartiles	1.6 to 2.8	1.2 to 2.6	2.6 to 3.5	2.3 to 3.3	1.0 to 1.4
CalyxdiH Median	1.8	2.0	2.2	1.9	1.7
CalyxdiH quartiles	1.5 to 2	1.9 to 2.3	2.0 to 2.3	1.7 to 2	1.5 to 2
CalyxDiamterW Median	1.7	2.0	2.2	1.9	1.7
CalyxDiamterW quartiles	1.5 to 2	1.9 to 2.3	2.0 to 2.3	1.7 to 2	1.6 to 2
CaDiL/W Median	1.0	1.0	1.0	1.0	1.0
CaDiL/W quartiles	1.0 to 1.1	1.0 to 1	1.0 to 1	1.0 to 1	0.9 to 1
CalyxupL Median	1.0	1.0	0.8	0.7	0.9
CalyxupL quartiles	0.9 to 1.2	0.8 to 1.1	0.7 to 1	0.6 to 0.9	0.7 to 1
CalyxupW Median	1.2	1.3	1.4	1.2	1.3
CalyxupW quartiles	1.0 to 1.3	1.2 to 1.5	1.2 to 1.6	1.1 to 1.4	1.2 to 1.3
CaUpLobeL/W Median	0.9	0.7	0.5	0.6	0.7
CaUpLobeL/W quartiles	0.7 to 1	0.6 to 0.8	0.4 to 0.7	0.5 to 0.7	0.6 to 0.8
CalyxlowL Median	1.0	0.8	0.6	0.7	0.8
CalyxlowL quartiles	0.8 to 1.1	0.7 to 1	0.5 to 0.8	0.6 to 0.8	0.7 to 1
CalyxlowW Median	1.2	1.5	1.6	1.2	1.3
CalyxlowW quartiles	1.0 to 1.4	1.4 to 1.7	1.4 to 1.7	1.1 to 1.4	1.2 to 1.5
CaLLobeL/W Median	0.8	0.5	0.4	0.6	0.6
CaLLobeL/W quartiles	0.7 to 0.9	0.5 to 0.6	0.3 to 0.6	0.5 to 0.7	0.5 to 0.7

Table 2.S6. The number of segregating sites that are fixed within each putative species for a) *psb*A-*trn*H and b) *trn*T-*trn*F. Numbers in parentheses represent the number of variable nucleotides that are shared between putative groups and the numbers on the diagonal are the number of nucleotides that are variable within each group.

a)

	A. acuta	A. decemloba	A. tenella	A. obtusifolia	A. skinneriana
A. acuta	1	0 (1)	0 (1)	14 (16)	19 (21)
A. decemloba		0	0 (0)	14 (15)	19 (20)
A. tenella			0	14 (15)	19 (20)
A. obtusifolia				1	34 (37)
A. skinneriana					2

b)

	A. acuta	A. decemloba	A. tenella	A. obtusifolia	A. skinneriana
A. acuta	0	0 (0)	0 (1)	12 (13)	16 (19)
A. decemloba		0	0 (1)	10 (14)	14 (16)
A. tenella			1	6 (6)	12 (15)
A. obtusifolia				1	27 (30)
A. skinneriana					3

Table 2.S7. Summary of 19 variables entered into the discriminant function model with species as the grouping variable and including all five species: *A. acuta*, *A. decemloba*, *A. tenella*, *A. obtusifolia*, and *A. skinneriana*.

Variable	Wilks' Lambda	Partial Lambda	F-remove	p-level	Toler.	1-Toler. (R-Sqr.)
Ped/Bract	0.022011	0.877632	9.37665	0.000000	0.321348	0.678652
DAntherL	0.025314	0.763113	20.87586	0.000000	0.803159	0.196841
LeafL/W	0.020037	0.964085	2.50527	0.042581	0.083257	0.916743
Bract	0.020126	0.959826	2.81477	0.025782	0.308507	0.691493
CalyxdiH	0.023155	0.834282	13.35823	0.000000	0.751041	0.248959
LeafW	0.022509	0.858215	11.11032	0.000000	0.102208	0.897792
CalyxupL	0.023893	0.808503	15.92840	0.000000	0.714342	0.285658
Stigma	0.024940	0.774575	19.57183	0.000000	0.563059	0.436941
UprLobeL	0.020560	0.939558	4.32621	0.002083	0.445640	0.554360
LeafL	0.021536	0.896987	7.72321	0.000007	0.072539	0.927461
Ped/Cal	0.021080	0.916407	6.13445	0.000097	0.269076	0.730924
CalyxAdaxial	0.020282	0.952443	3.35791	0.010540	0.597827	0.402173
Vanth/Fil	0.020338	0.949840	3.55140	0.007640	0.775474	0.224526
stigstyle	0.020621	0.936777	4.53872	0.001455	0.441775	0.558225
CorTubH/W	0.020931	0.922932	5.61558	0.000235	0.830105	0.169895
H/W	0.020416	0.946192	3.82437	0.004842	0.906629	0.093371
DistTubeLobe	0.020263	0.953336	3.29174	0.011763	0.710476	0.289524
CorollaAd	0.020565	0.939331	4.34347	0.002023	0.534778	0.465222
ThroatDiH	0.020522	0.941336	4.19103	0.002616	0.467072	0.532928

Table 2.S8. Classification matrix for discriminant function analysis for five *Agalinis* species AAC=*A. acuta*, ADE=*A. decemloba*, ATE=*A. tenella*, AOB=*A. obtusifolia*, and ASK=*A. skinneriana*. The value p is the prior probability of group membership based on the proportion of the total number of individuals attributable to a species.

Classified From	Percent Correct	Classified Into				
		AAC p=0.503	ADE p=0.096	ATE p=0.140	AOB p=0.195	ASK p=0.065
AAC (n=148)	98.65	146	0	0	0	2
ADE (n=27)	89.29	2	25	0	0	1
ATE (n=42)	90.48	0	0	38	4	0
AOB (n=58)	96.55	0	0	2	56	0
ASK (n=19)	100.0	0	0	0	0	19
Total (n=294)	96.27	148	25	40	60	22

Table 2.S9. Standardized coefficients, eigenvalues, and cumulative variance explained for discriminant functions for five *Agalinis* species: *A. acuta*, *A. decemloba*, *A. tenella*, *A. obtusifolia*, and *A. skinneriana*. These coefficients reflect the change in the canonical scores per unit change in the standardized independent variables. Therefore, these coefficients may be compared in order to determine the magnitudes and directions of the unique contributions of the variables to each canonical function.

Variable	Axis 1	Axis 2	Axis 3	Axis 4
Ped/Bract	0.642403	0.013663	-0.22705	0.033991
DAntherL	0.437788	0.385772	0.03390	-0.410813
LeafL/W	0.417725	0.169043	-0.65884	0.336543
Bract	-0.287743	0.118611	-0.26566	-0.229688
CalyxdiH	0.322809	-0.282887	0.37135	0.163636
LeafW	0.518859	0.253336	-1.41990	0.180586
CalyxupL	-0.528154	0.169779	0.11513	0.147049
Stigma	-0.216532	-0.632211	0.21334	-0.618533
UprLobeL	0.255026	0.063978	0.33717	0.257003
LeafL	-0.802219	0.318072	1.21144	-0.015347
Ped/Cal	-0.009853	0.668685	0.23500	0.100136
CalyxAdaxial	-0.267214	-0.053810	0.09677	-0.226819
Vanth/Fil	0.108525	0.050508	-0.00556	-0.483228
stigstyle	0.314276	0.306380	0.01021	0.045113
CorTubH/W	-0.159067	0.302105	-0.15825	0.057251
H/W	-0.076725	-0.097514	-0.29037	0.093483
DistTubeLobe	0.075778	0.271060	-0.14981	0.110198
CorollaAd	-0.321105	0.056607	-0.20262	-0.062627
ThroatDiH	0.243649	-0.268398	-0.04217	0.358623
Eigenval	5.803960	1.641988	1.22570	0.293857
Cum.Prop	0.647366	0.830511	0.96722	1.000000

Table 2.S10. Summary of variables entered into the discriminant function model with 18 populations of five *Agalinis* species: *A. acuta*, *A. decemloba*, *A. tenella*, *A. obtusifolia*, and *A. skinneriana*

Variable	Wilks' Lambda	Partial Lambda	F-to-remove	p-level	Toler.	1-Toler. (R-Sqr.)
Ped/Bract	0.000172	0.746729	5.12751	0.000000	0.414763	0.585237
UprLobeL	0.000182	0.707471	6.25092	0.000000	0.694464	0.305536
Bract	0.000148	0.866721	2.32470	0.002606	0.345977	0.654023
CaUpLobeL/W	0.000243	0.529907	13.41123	0.000000	0.553112	0.446888
LeafL/W	0.000230	0.559707	11.89228	0.000000	0.568269	0.431731
DAntherL	0.000207	0.622336	9.17413	0.000000	0.514110	0.485890
Vanth/Fil	0.000197	0.654900	7.96627	0.000000	0.751000	0.249000
CalyxupL	0.000186	0.691370	6.74857	0.000000	0.476574	0.523426
LeafL	0.000195	0.659821	7.79411	0.000000	0.524061	0.475939
DistTubeLobe	0.000186	0.692861	6.70151	0.000000	0.825337	0.174663
CalyxdiH	0.000174	0.737567	5.37900	0.000000	0.805398	0.194602
Stigma	0.000179	0.720569	5.86250	0.000000	0.691770	0.308230
Dant/Fil_ratio	0.000171	0.754375	4.92232	0.000000	0.525997	0.474003
Ped/Cal	0.000175	0.734754	5.45747	0.000000	0.434070	0.565930
DAntherL/W	0.000170	0.759072	4.79832	0.000000	0.766103	0.233897
CalyxAdaxial	0.000163	0.788098	4.06480	0.000000	0.716577	0.283423
stigstyle	0.000166	0.775930	4.36563	0.000000	0.633813	0.366187
CorollaAd	0.000163	0.788720	4.04968	0.000000	0.820391	0.179609

Table 2.S11. Classification matrix for discriminant function analysis of 18 populations in five *Agalinis* species: *A. acuta*, *A. decemloba*, *A. tenella*, *A. obtusifolia*, and *A. skinneriana*. The value p is the prior probability of group membership based on the proportion of the total number of individuals attributable to a species. Gray shading highlights putatively conspecific populations. Population codes are given in Table 2.3.

| Classified From | % Correct | Classified Into | | | | | | | |
		AAC-BVMA p=.0764	AAC-PEMA p=.0255	AAC-EPRI p=.0764	AAC-PCCT p=.0924	AAC-HPNY p=.0064	AAC-MDNY p=.0350	AAC-SMNY p=.0796	AAC-SDMD p=.0764
AAC-BVMA	96.00	24	0	0	0	0	0	0	0
AAC-PEMA	70.00	3	7	0	0	0	0	0	0
AAC-EPRI	92.00	0	0	23	0	1	0	0	0
AAC-PCCT	89.65	0	0	3	26	0	0	0	0
AAC-HPNY	88.00	1	0	2	0	22	0	0	0
AAC-MDNY	83.33	0	0	0	0	2	10	0	0
AACSMPNY	33.33	1	0	0	0	0	0	1	0
AAC-SDMD	96.00	0	0	0	0	0	0	0	24
ADE-L1VA	100.00	0	1	0	0	0	0	0	0
ADE-WCNC	100.00	0	0	0	0	0	0	0	0
ATE-CCSC	100.00	0	0	0	0	0	0	0	0
ATE-BCGA	85.71	0	0	0	0	0	0	0	0
ATE-LCGA	89.47	0	0	0	0	0	0	0	0
ATE-GCGA	73.68	0	0	0	0	0	0	0	0
AOB-LCFL	91.66	0	0	0	0	0	0	0	0
AOB-LCSC	68.75	0	0	0	0	0	0	0	0
AOB-DCSC	100.00	0	0	0	0	0	0	0	0
ASK-PGMD	96.00	0	0	0	0	0	0	0	0
Total	88.49	29	8	28	26	25	10	1	24

Table 2.S11. Continued.

		Classified Into									
Classified From	% Correct	ADE-L1VA p=.0223	ADE-WCNC p=.0669	ATE-CCSC p=.0191	ATE-BCGA p=.0573	ATE-LCGA p=.0414	ATE-GCGA p=.0127	AOB-LCSC p=.0764	AOB-DCSC p=.0509	AOB-LCFL p=.0541	ASK-PGMD p=.0605
AAC-BVMA	96.00	1	0	0	0	0	0	0	0	0	0
AAC-PEMA	70.00	0	0	0	0	0	0	0	0	0	0
AAC-EPRI	92.00	0	0	0	0	0	0	0	0	0	1
AAC-PCCT	89.65	0	0	0	0	0	0	0	0	0	0
AAC-HPNY	88.00	0	0	0	0	0	0	0	0	0	0
AAC-MDNY	83.33	0	0	0	0	0	0	0	0	0	0
AAC-SMNY	33.33	1	0	0	0	0	0	0	0	0	0
AAC-SDMD	96.00	1	0	0	0	0	0	0	0	0	0
ADEL1VA	71.43	5	0	0	0	0	0	0	1	0	0
ADEWCNC	100.00	0	21	0	0	0	0	0	0	0	0
ATECCSC	100.00	0	0	7	0	0	0	0	0	0	0
ATEBCGA	85.71	0	0	0	4	0	0	0	0	0	0
ATELCGA	89.47	0	0	0	0	12	2	0	0	0	0
ATEGCGA	73.68	0	0	0	0	1	17	0	1	0	0
AOBLCFL	91.66	0	0	0	0	0	0	14	1	4	0
AOBLCSC	68.75	0	0	0	0	0	0	0	22	2	0
AOBDCFL	100.00	0	0	0	0	0	0	2	3	11	0
ASKPGMD	96.00	0	0	0	0	0	0	0	0	0	19
Total	88.49	8	21	7	4	13	19	16	28	17	20

Figure 2.1. Locations of the 20 populations representing four putative species, *A. acuta, A. decemloba, A. tenella,* and *A. obtusifolia.* Top map shows the northeastern populations and the bottom map shows the locations of the southeastern populations.
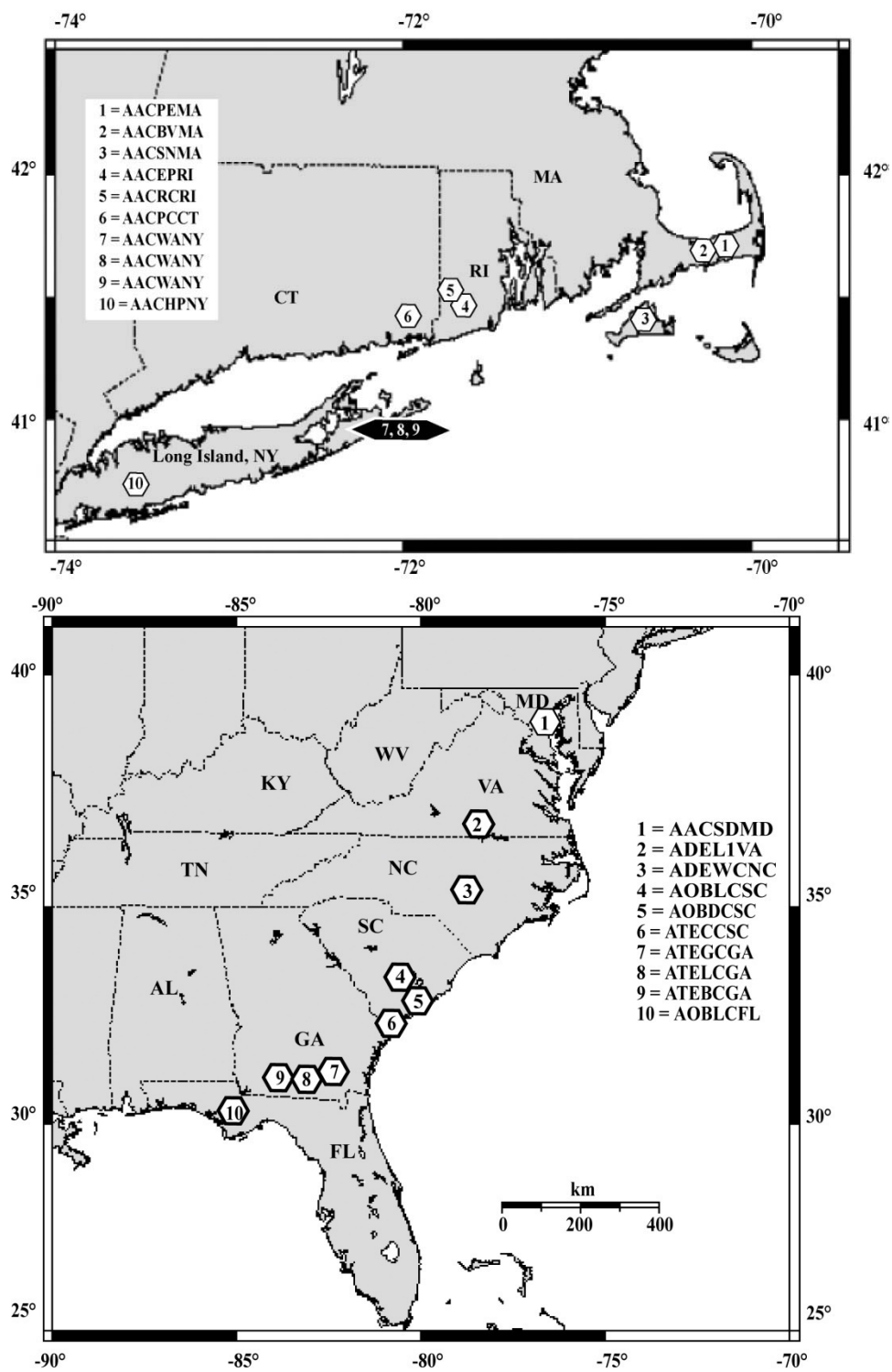
Figure 2.2. Phylogenetic tree depicting evolutionary relationships among sampled taxa based on six cpDNA loci. Samples in red are the representatives of the populations used in the microsatellite analyses. Branch lengths depict the inferred number of nucleotide substitutions per site. Numerals above or below branches represent the percent of 1000 bootstrap replicates supporting that clade. The ln likelihood of the tree is -ln 8734.027845.
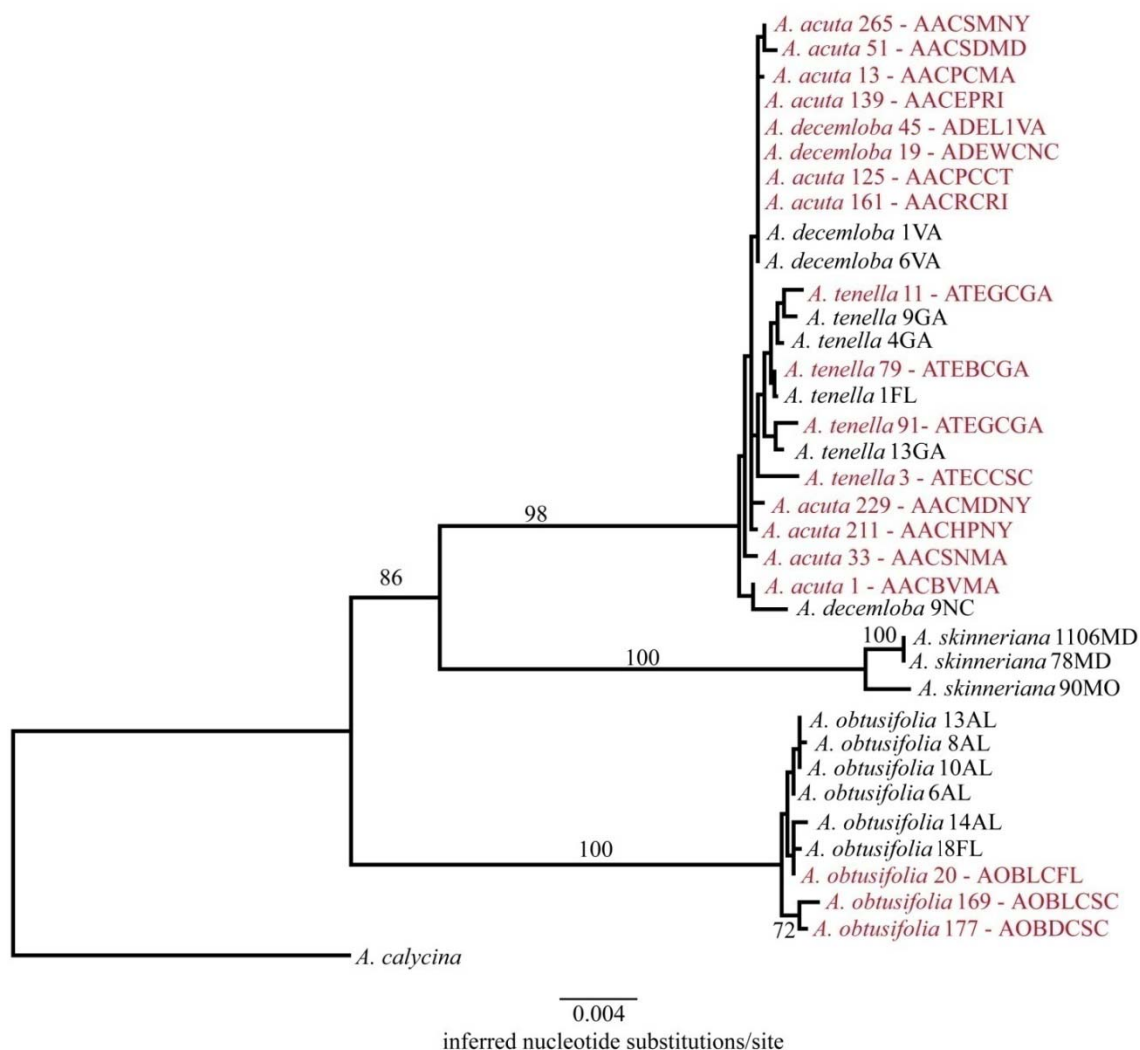
Figure 2.3. Neighbor-joining tree based on the genetic distance D_a (Nei et al. 1983) among populations. Branch lengths are proportional to D_a units and numerals above the branches represent the percent of 1000 bootstrap replicates supporting that cluster. Bootstrap replicates are based on the resampling, with replacement, of loci.
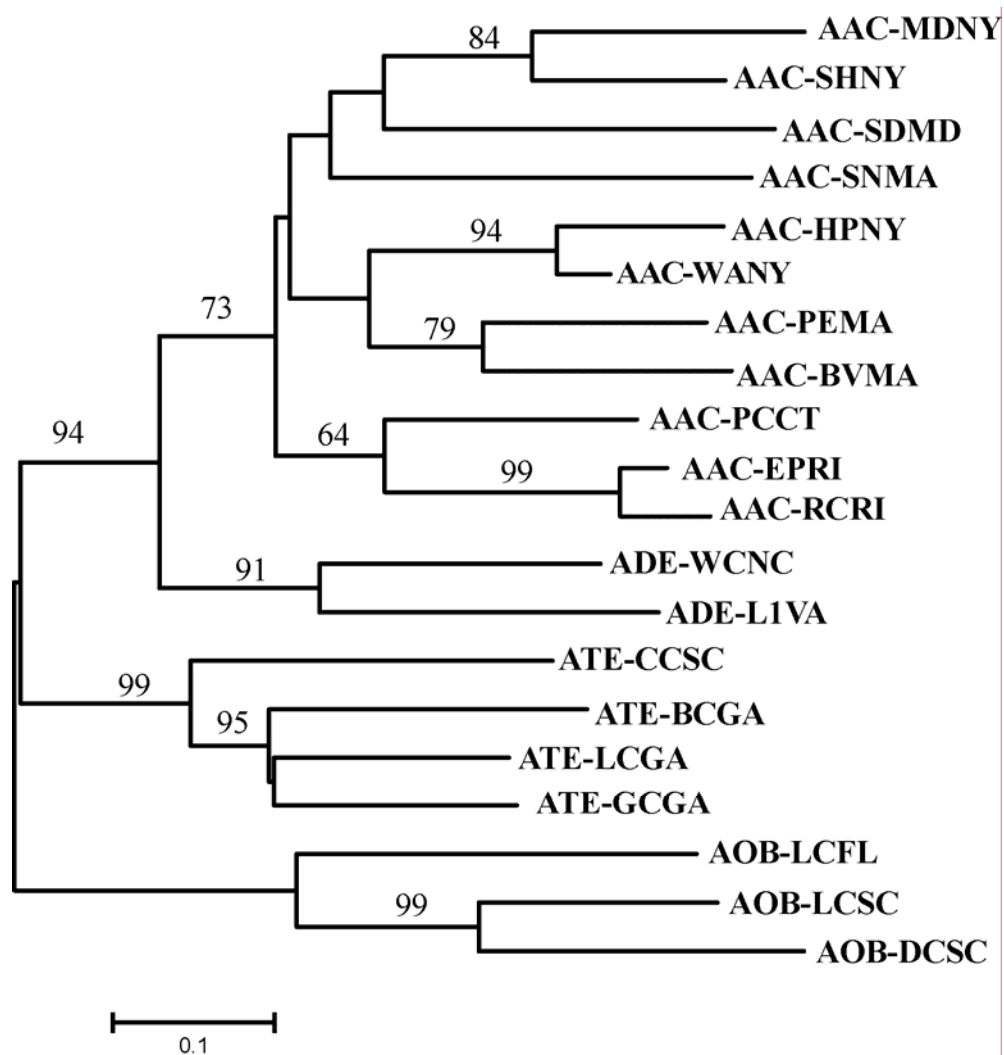
Figure 2.4. Unrooted Neighbor-joining tree based on the genetic distance D_a (Nei et al. 1983) among individuals. Branch lengths are proportional to D_a units.
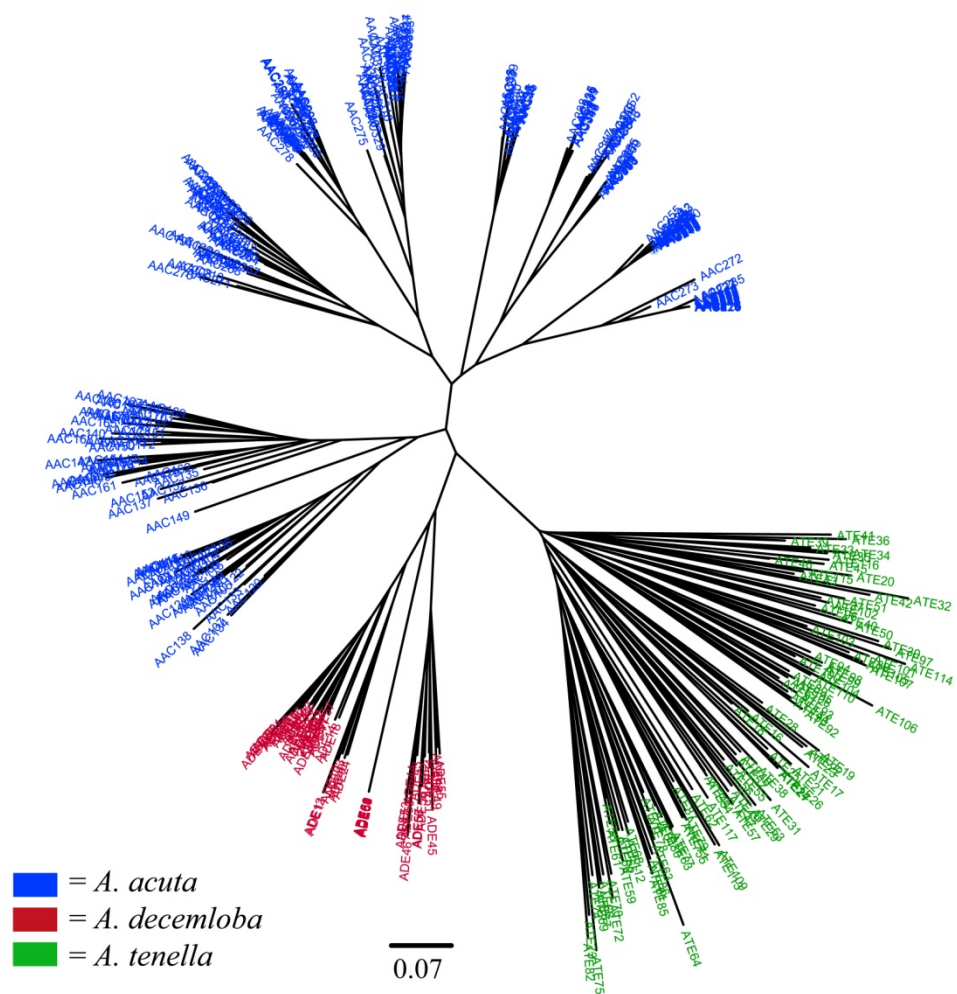
Figure 2.5. Distruct (Rosenberg 2004) diagram representing the assignment of individuals to clusters inferred using STRUCTURE. The results associated with the replicate at each value of *k* that had the highest likelihood score are presented. Each line represents an individual and each block represents a population. The degree to which a line has multiple colors is indicative of the degree of admixture.
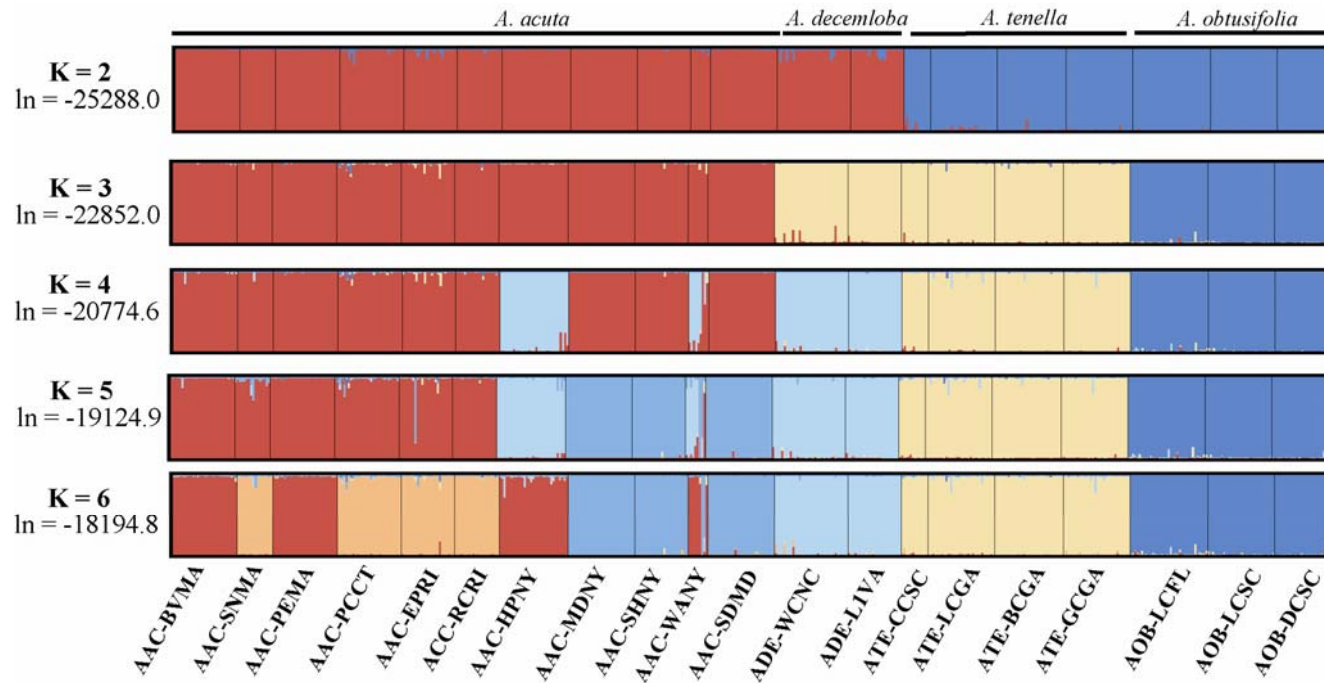
Figure 2.6. Scatterplot of discriminant function scores on DFA Axis 1 and DFA Axis 2 for five *Agalinis* species: AAC = *A. acuta*, ADE = *A. decemloba*, ATE = *A. tenella*, AOB = *A. obtusifolia*, and ASK = *A. skinneriana*.
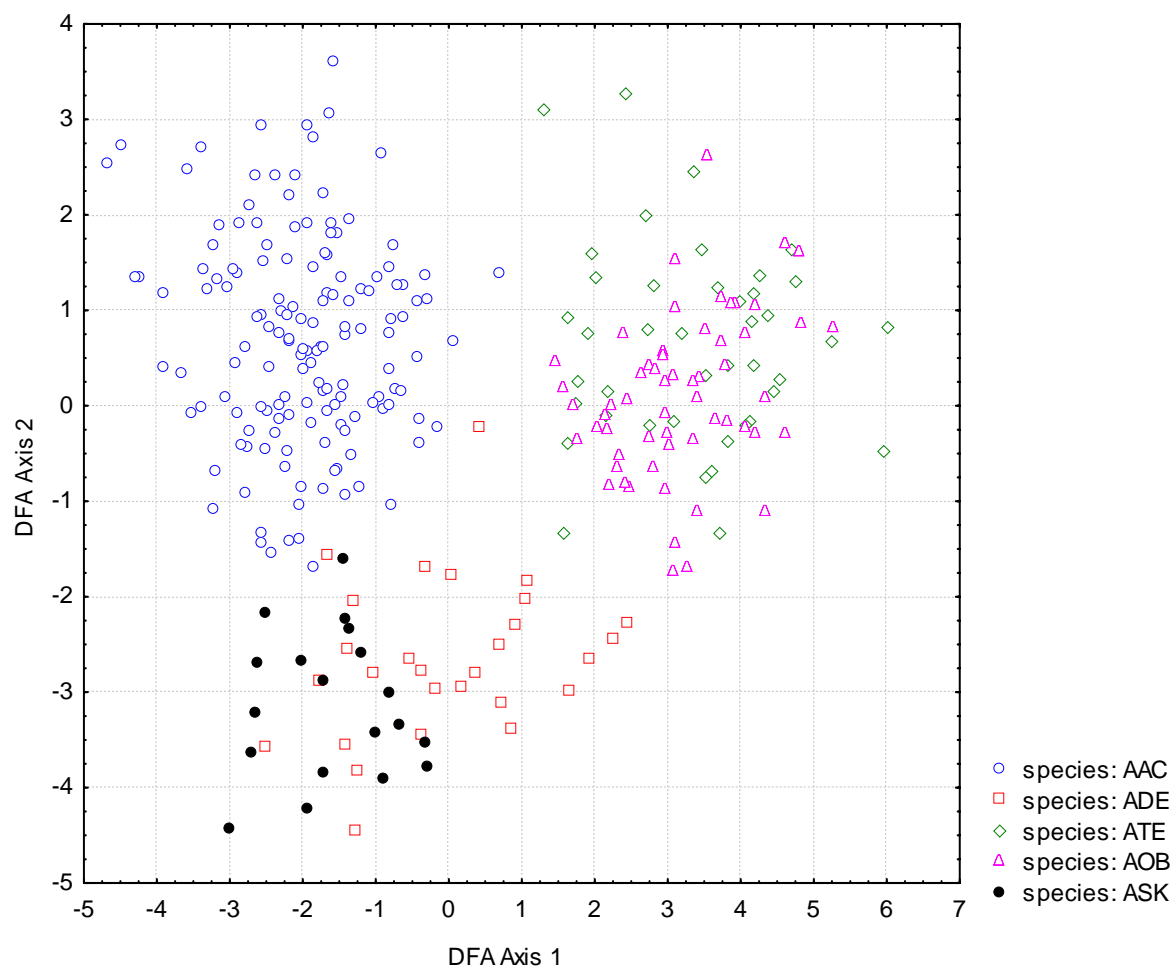
Figure 2.7. Scatterplot of discriminant function scores for 18 populations of five species: AAC = *A. acuta*, ADE = *A. decemloba*, ATE = *A. tenella*, AOB = *A. obtusifolia* and ASK = *A. skinneriana*. Population codes are provided in Table 2.3.
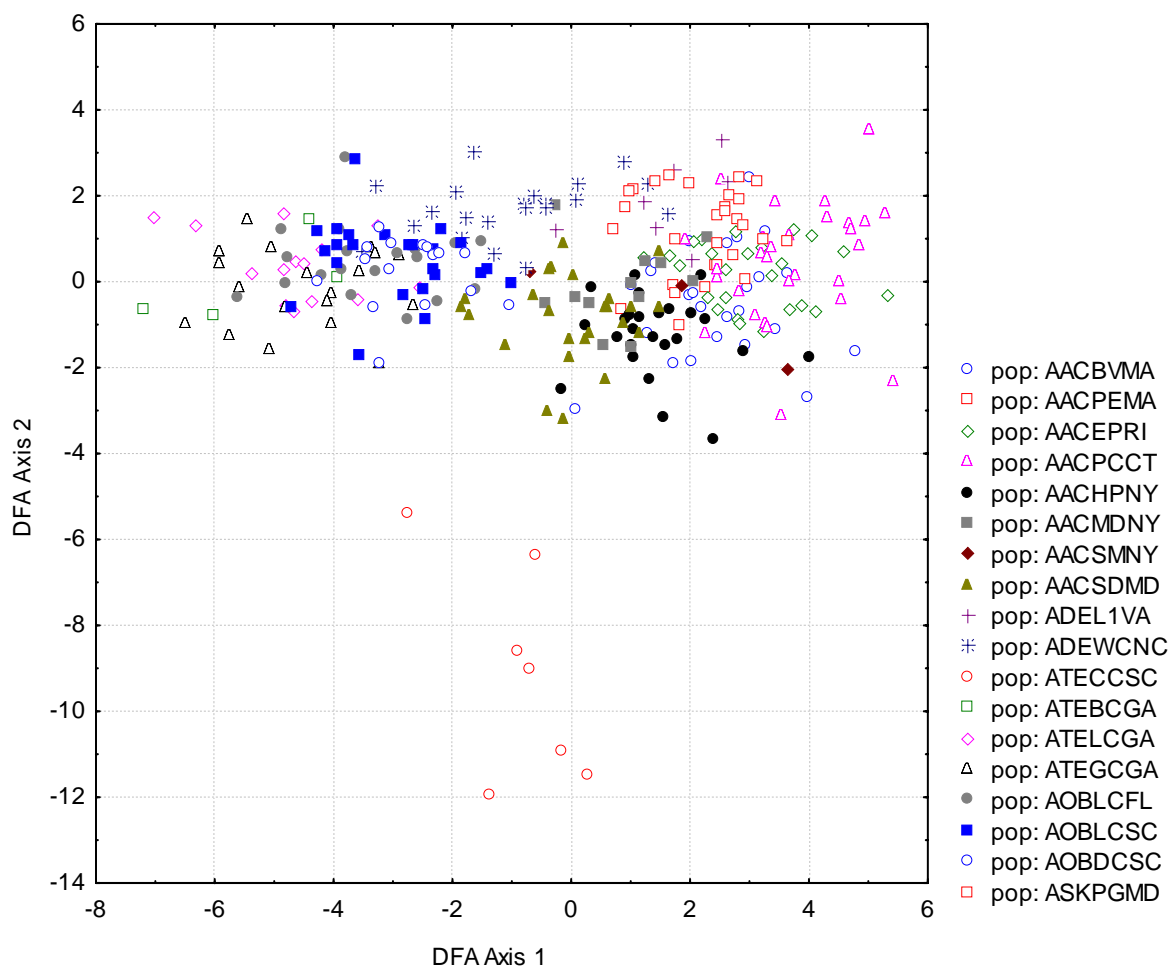
Figure 2.S1. Neighbor-joining tree based on the genetic distance D_a (Nei et al. 1983) derived from a data set created using FreeNA to account for potential null alleles. Branch lengths are proportional to D_a units and numerals above or below the branches represent the percent of 1000 bootstrap replicates supporting that cluster.
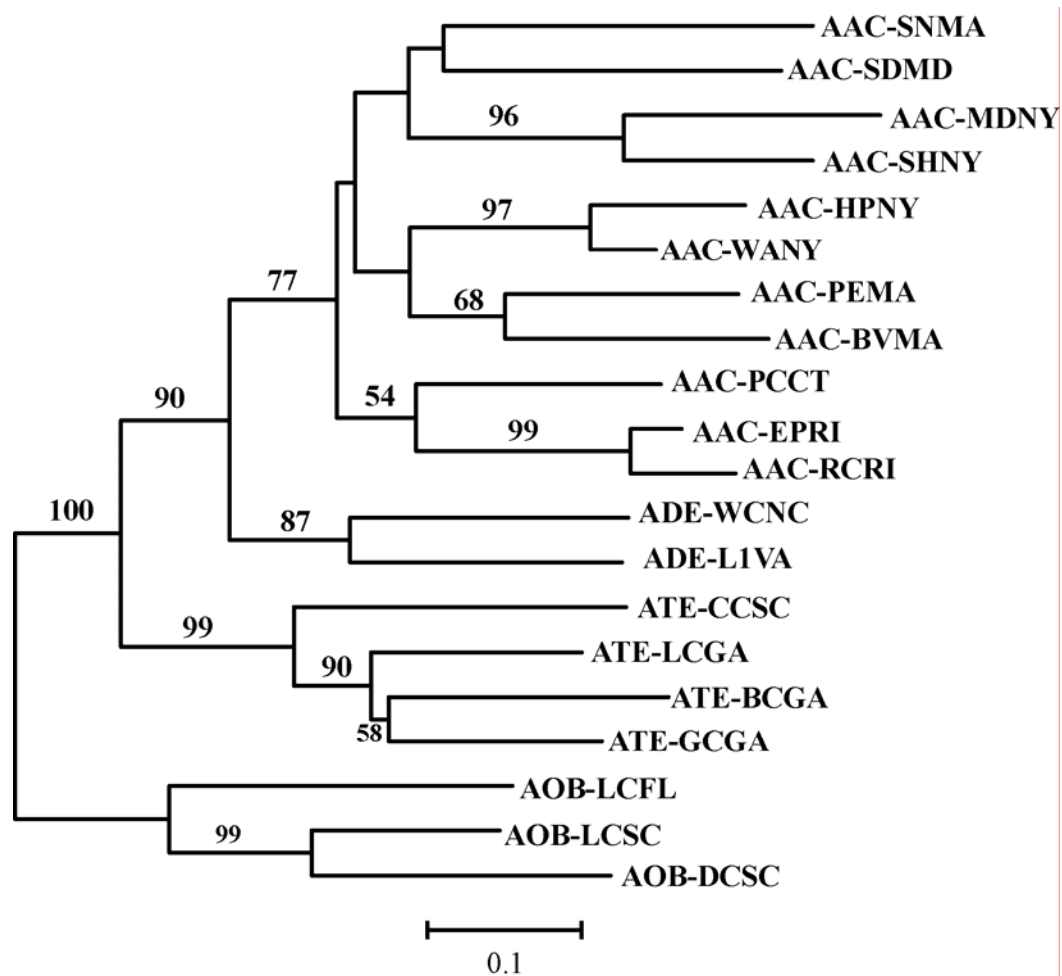
CHAPTER 3: AN EVALUATION OF CANDIDATE PLANT DNA BARCODES AND ASSIGNMENT METHODS IN DIAGNOSING 29 SPECIES IN THE GENUS *AGALINIS* (OROBANCHACEAE)

ABSTRACT

Application of DNA barcoding to plants has focused on evaluating the success of candidate barcodes across a broad spectrum of evolutionary divergence while less attention has been paid to the performance when distinguishing closely related species or to differential success of analytical techniques. Lack of monophyly that indicates incomplete lineage sorting in closely related species or inaccurate taxonomy will pose challenges for barcoding and dictate the degree to which query sequences can be accurately assigned. We examine these issues using 11 candidate barcodes and three analytical methods (i.e., genetic distances, hierarchical tree-based, and diagnostic character differences) in 92 samples representing 29 putative species in the genus *Agalinis* (Orobanchaceae). Based on questions of evolutionary distinctiveness raised by previous studies, we evaluated success under two taxonomic hypotheses. We found that *psb*A-*trn*H and *trn*T-*trn*L in conjunction with the 'best close match' distance-based method best met the objectives of DNA barcoding. Successful assignment was also a function of the taxonomy used. In addition to accurately identifying the taxonomy of a query sequence, our results show that DNA barcoding is useful for detecting taxonomic uncertainty, but determining whether erroneous taxonomy or incomplete lineage sorting is the cause requires additional information provided by traditional taxonomic approaches.

INTRODUCTION

DNA barcoding promises to provide a fast and reliable way to determine the taxonomic identity of an individual by sequencing a small portion of its genome and comparing this nucleotide sequence with sequences in a reference database (e.g., Hebert et al. 2003). This promise along with the increasing ease with which sequence data can be obtained from a range of taxa has spawned the ambitious goal of creating a genetic catalog of the world's biodiversity (Chase et al. 2005; Hebert et al. 2003; Kress et al. 2005; Savolainen et al. 2005). Such a reference collection can be an important tool for a broad range of applications including conducting rapid biodiversity assessments, forensics, detecting illegal wildlife trade, identifying species during cryptic life stages, and monitoring invasive species (Armstrong & Bar 2005; Darling & Blum 2007; Dawnay et al. 2007; Little & Stevenson 2007; Vogler 2006). Barcoding has also been shown to be a useful tool in the discovery of new species (Burns et al. 2008; Murray et al. 2008; Yassin et al. 2008). However, the grand promises made by proponents of DNA barcoding have generated concern and there is extensive debate over exactly what it can contribute to various disciplines (e.g., Rubinoff 2006; Trewick 2008; Will et al. 2005).

The application of DNA barcoding requires that a reference database of DNA sequences for many species be developed. It is most desirable to have a single locus or a few loci that have highly conserved universal primer sequences and at the same time exhibit sufficient nucleotide variation to diagnose species. Due to exceedingly low rates of nucleotide substitution and high rates of chromosomal rearrangements within the plant mitochondrial genome (Palmer 1985) and extensive gene duplication in the nuclear

genome (Alvarez & Wendel 2003), plastid loci are the most promising candidate

barcodes for plant species (e.g., Chase et al. 2005; Kress et al. 2005; Lahaye et al. 2008).

Candidate barcodes from the plastid genome include both slowly evolving coding regions

(e.g., *rbc*L, *rpo*B, or *mat*K) and more rapidly evolving loci (e.g., *rps*2, *psb*A-*trn*H,and

*trn*T-*trn*F). However, when used alone the more conserved loci may not possess enough

nucleotide variation to discriminate among closely related species and the more variable

loci may be problematic because of homoplasy, lack of conserved priming regions, and

difficulties in alignment among distantly related species. To overcome these problems a

multi-locus DNA barcoding system has been suggested (CBOL Plant Working Group

2009; Chase et al. 2005; Fazekas et al. 2008; Kress & Erickson 2007) in which slowly

evolving loci delineate individuals into families, genera, or groups within genera and the

more rapidly evolving loci differentiate species within those higher groups.

In addition to a barcoding locus, application of DNA barcoding requires analytical

methods that accurately assign query sequences of unknown taxonomic identity to

species based on the sequences contained in the reference database. The Barcode of Life

Data Systems (BOLD), which is an online initiative to facilitate the implementation of

DNA barcoding, assigns taxonomy to a query sequence based on a pairwise genetic

distance threshold (i.e., of 1% for animals) below which signifies conspecific

relationships (Ratnasingham & Hebert 2007). Pairwise genetic distances were also used

to in the analysis that resulted in determining *the* plant DNA barcode of *rbc*L coupled

with *mat*K (CBOL Plant Working Group 2009). The primary criticism of a distance

based method is that there will undoubtedly be instances where intraspecific and

interspecific distances will exceed and be below, respectively, the distance threshold

value (DeSalle et al. 2005; Ferguson 2002; Little & Stevenson 2007; Meyer & Paulay

2005). Methods based on the clustering of individuals on a phylogenetic tree have been

suggested as alternatives to distance methods because through the bootstrap a measure of

statistical support can be achieved and the resulting tree provides a visual representation

of the affinities of the query sequence (Armstrong & Bar 2005). However, they have also

been criticized because of the difficulties in assigning taxonomy to a query sequence

based on a hierarchical tree-like structure (DeSalle et al. 2005) and the fact that low

levels of divergence at a candidate barcode may be able to differentiate among sequences

despite its lack of ability to sufficiently estimate phylogenetic relationships (Erickson et

al. 2008). An additional method identifies combinations of nucleotide character states

that are unique to members of a species (i.e., diagnostic characters) and, therefore, can be

used to assign taxonomy if a query sequences possess that combination of nucleotide

characters. However, the recently developed character-based method (Sarkar et al. 2008)

relies on a phylogenetic tree and, therefore, suffers to some extent from the concerns

associated with tree based methods. In addition to these drawbacks, different

implementation of assignment techniques can also result in different correct classification

rates when evaluating the same DNA barcode. As such, the efficacy of DNA barcoding

may be highly dependent on the assignment method, which has likely contributed to the

lack of consensus in identifying a suitable locus to serve as the plant DNA barcode

(Erickson et al. 2008).

Although the search for loci and the development of analytical techniques has been crucial, most of the barcoding studies in plants have focused on reliable amplification across the evolutionary breadth of land plants (e.g., Chase et al. 2005; Kress et al. 2005) or within geographically or ecologically defined communities (Lahaye et al. 2008). Much less effort has been spent assessing potential to distinguish closely related species (for an exception see Fazekas et al. 2009). Regardless of the analytical methods used, DNA barcoding will be highly successful in divergent taxa that have much smaller intraspecific than interspecific differences (i.e., the 'barcoding gap'; Meyer & Paulay 2005), are reciprocally monophyletic, and have diagnostic character differences (Fazekas et al. 2009). However, when the time since divergence between species is too short for sorting of shared ancestral polymorphisms (e.g., 8.7 N_e (effective population size) generations; Hudson & Coyne 2002; Rosenberg 2003), using short sequences of DNA to diagnose species will potentially be challenging regardless of the assignment method employed (Ross et al. 2008). Simulations show failure rates become excessive when overlap between intra- and interspecific distances is >10% and when <80% of species are monophyletic (Ross et al. 2008). Lack of differentiation can be due to inaccurate taxonomy or low levels of divergence, both of which result in non-monophyletic relationships. Polyphyly and paraphyly have been observed in 23% of 2319 animal species (Funk & Omland 2003), suggesting that such overlap in distances may be widespread. Paraphyly may be even more extensive among plant species (e.g., Crisp & Chandler 1996; Fazekas et al. 2009; Rieseberg & Brouillet 1994). The frequency of occurrence of such relationships is also likely to be greatest among closely related

species; therefore, it is critical to understand the degree to which DNA barcoding can distinguish sequences representing recently diverged species (Hollingsworth et al. 2009). Performance of DNA barcoding in this realm has only begun to be tested (Fazekas et al. 2009; Meyer & Paulay 2005; Newmaster et al. 2008).

The purpose of the present research was to evaluate the ability to distinguish 29 putative species in the plant genus *Agalinis* (family Orobanchaceae) using DNA barcoding. We evaluated the utility of the 11 barcodes (8 single-locus and 3 2-locus barcodes) first by considering how easy it was to obtain a full sequence across all individuals. We then evaluated the combined utility of loci under each of three assignment techniques as a barcode's ability to classify individuals to species correctly and to determine accurately when a sequence had no conspecific in the reference database. The first method was based on genetic distances (Meier et al. 2006), the second was based on the relationships depicted when a hierarchical clustering algorithm was applied to a matrix of pairwise distances (Ross et al. 2008), and the third was based unique combinations of nucleotide character states that differentiate taxa (i.e., diagnostic nucleotide character differences; Sarkar et al. 2008).

The genus *Agalinis* provides an excellent test case for barcoding. It is well studied taxonomically (Canne-Hilliker 1987; Canne-Hilliker & Dubrule 1993; Canne-Hilliker & Kampny 1991; Canne 1979, 1980, 1981, 1982, 1983, 1984; Kampny & Canne-Hilliker 1987; Pennell 1913a, 1913b, 1928, 1929, 1935), and the array of evolutionary divergence is relatively well understood and ranges from taxa that are highly differentiated to those that may not be evolutionarily distinct (Neel & Cummings 2004; Pettengill & Neel 2008).

As such, it reflects the array of evolutionary distinctiveness one expects in plant genera. Lack of distinction among several taxa indicates that some taxonomic alignments may not be warranted (Neel & Cummings 2004; Pettengill & Neel 2008), which allows us to investigate the consequences of taxonomic uncertainty for barcoding. By focusing on the practical application of our results to the conservation of rare *Agalinis* species, we are also able to evaluate the utility of DNA barcoding to conservation biology. Twenty-one of the sampled species are considered imperiled (S2) or critically imperiled (S1) in at least 1 state in which they occur; 7 of these species are also globally vulnerable (G3 or G3-G4) and 4 are critically imperiled (G1) (Table 3.1; NatureServe 2007). Because many of the species are particularly difficult to distinguish in the field, we are interested in knowing whether DNA barcoding can serve as a tool for field botanists from state and federal agencies to identify quickly the taxonomy of individuals that represent species that are challenging to classify based on morphology. Greater certainty in identification can increase understanding of the true distribution and abundance of putatively rare species and facilitate appropriate management and priority setting.

MATERIALS AND METHODS

Sampling

The database used to assess performance of how the different loci and methods consisted of sequences from 92 individuals representing 29 out of the ~40 putative North American *Agalinis* species (Table 3.1); no specimens were available of the approximately 30 additional *Agalinis* species that exist in the South America, Mexico, and Central America. The sampled species represented all North American sections and subsections

of the genus and 27 species were represented by more than one accession. The number

of individuals per species ranged from 1-10. When multiple individuals were sampled

for a species, they were from different populations that often encompassed a majority of

the species range and therefore may provide a means of assessing how well DNA

barcoding performs when there is the potential for substantial intraspecific differentiation

due to geographic isolation. The two species represented by a single individual allow us

to assess the performance of methods when there is no conspecific reference sequence in

the database, which has been identified as problematic when using methods that assign

taxonomy based on the closest match in the reference database (Ratnasingham & Hebert

2007). We did not include representatives of other genera in our database because there

is strong support for the monophyly of the genus (Neel & Cummings 2004; Pettengill &

Neel 2008).

Previous study of the phylogenetic relationships among the sampled individuals

showed that *Agalinis decemloba*, *Agalinis tenella*, and *Agalinis acuta* were polyphyletic

as were *Agalinis paupercula* and *Agalinis purpurea* (Pettengill & Neel 2008) and

synonymization may be warranted. An investigation quantifying the degree of

differentiation based on phylogenetic, population genetic, and morphological analyses

among the first three of those named species is in progress to determine whether in fact

synonymization is warranted. As a result of this taxonomic uncertainty, we evaluated the

performance of candidate barcodes using two taxonomic hypotheses. The first taxonomic

alignment, which we refer to as the 'historical taxonomy,' treated each of these species

individually and in the second *A. decemloba*, *A. tenella*, and *A. acuta* were synonymized

under *A. decemloba*; *A. paupercula* and *A. purpurea* were synonymized under *A.*

purpurea. Synonymizing yielded a data set with 26 species and between 1 and 23

individuals per species.

DNA extraction, PCR, and sequencing

Total genomic DNA was isolated from fresh or frozen (-80°C) leaves and flower

buds by grinding 50-100 mg of tissue to powder in liquid nitrogen with a mortar and

pestle, and then using GenElute Plant Genomic DNA Kits (Sigma Chemical Company,

St. Louis, Missouri, USA) or Qiagen DNEasy Kits (Qiagen Corporation, Valencia,

California USA) following manufacturer's instructions.

We attempted to analyze sequences for all 92 individuals from six chloroplast gene

regions (4 coding regions (*mat*K, *rbc*L, *rps*2, and *rpo*B) and the non-coding *psb*A-*trn*H

and *trn*T-*trn*F spacer regions) that have been recommended for barcoding in plants. Two

PCR reactions using the primer pairs *trn*T-a/*trn*L-d and *trn*L-c/*trn*F-f (Taberlet et al.

1991) were required to obtain the full length of the *trn*T-*trn*F locus. Given that it is best

to use a single primer pair to acquire a barcode, we divided the *trn*T-*trn*F locus into three

separate regions that represent candidate barcodes evaluated by other studies. These

three regions are defined by the primers used to amplify them (Taberlet et al. 1991): 1)

*trn*T-a/*trn*L-d that we refer to as the *trn*T-*trn*L barcode (evaluated by Edwards et al.

(2008c)), 2) *trn*L-c/*trn*L-d that we refer to as *trn*L-intron (suggested by Taberlet et al.

(2007)), and 3) *trn*L-c/*trn*F-f, which we refer to as the *trn*L-*trn*F barcode (evaluated by

Chase et al. (2005)). This treatment of *trn*T-*trn*F as three separate regions brings our

total assessment to eight single-locus barcodes.

All polymerase chain reactions (PCR) were done with Eppendorf MasterTaq PCR

kits (Brinkman, Westbury, New York, USA) on MJ Research PTC-200 Thermal Cyclers.

In general, the PCR temperature profile was 30 cycles of 94°C for 60 s, annealing

temperature set approximately 5°C below the lower of the two primer melting

temperatures for 90 s, 72°C for 150 s, and a final 15 min elongation period at 72°C.

Amplified DNA fragments were visualized by agarose gel electrophoresis and purified

using the QIAQuick PCR Purification Kit according to manufacturer's instructions

(Qiagen Inc., Valencia, California, USA).

We employed a conservative sequencing strategy in which four replicates of both

forward and reverse sequence reads were obtained per sample, resulting in eight-fold

coverage across the majority of all loci. Although this strategy is not concordant with the

rapid and inexpensive acquisition of sequences promoted as an advantage of DNA

barcoding, we feel it is a necessary step to ensure accuracy of nucleotide sequences when

considering closely related taxa. Sequencing reactions were conducted with BigDye

Terminator v3.1 Cycle Sequencing chemistry (Applied Biosystems, Foster City,

California, USA) and products were sequenced using an Applied Biosystems 3730xl

DNA Analyzer. Total reaction volume was 7 µl (1-3 µl PCR template, 1.5 µl 5X

Sequencing Buffer, 1 µl primer [2.5µM], 0.5 µl BigDye Terminator, and 1-3 µl ddH2O).

The program Sequencher v4.6 (Gene Codes Corporation, Ann Arbor, Michigan,

USA) was used for base calling, quality assignments, and assembling consensus

sequences for each sample from the replicate bi-directional sequence reads. Sequences

were aligned using the default settings of MUSCLE (Edgar 2004). We manually

inspected alignments to ensure that there were no obvious errors but, in general, we avoided manually editing sequences due to the subjectivity and non-reproducibility of such actions (Morrison 2009). A lack of manual editing of sequence alignments is also a desirable property of a DNA barcoding (CBOL Plant Working Group 2009).

Barcoding analyses

In addition to evaluating the performance of the eight single-locus barcodes described above, we also tested three, two-barcode combinations that have either been suggested (*mat*K and *psb*A-*trn*H, Chase et al. 2007; *rbc*L and *psb*A-*trn*H, Kress & Erickson 2007) or have potential given the objectives of a multi-locus barcode (i.e, *rpo*B and *psb*A-*trn*H). The multi-locus barcodes were constructed by concatenating sequences for only those samples that had sequences for both loci. To avoid confounding issues due to DNA sequence alignment differences, multi-locus barcodes were constructed using the same alignment associated with the corresponding single locus barcodes.

We calculated variability of each barcode as the number and percentage of variable characters using PAUP*v4b10 (Swofford 2003) and by the mean and range of intraspecific and interspecific Kimura 2-parameter distances (K2P; Kimura 1980) using MEGA (Kumar et al. 2004).

The software package TAXONDNA (Meier et al. 2006) was used to assess performance of each barcode based on the 'best close match' option using K2P distances. This method reduces the potential for errors of commission by employing a user-specified pairwise distance threshold above which a sequence was classified as 'no match'. Query sequences were classified as 'ambiguous' if they could be assigned to

both the correct and an incorrect species; 'correct,' when pairwise genetic distances between query and reference sequences were below the threshold only for conspecifics; and 'incorrect,' when only heterospecific sequences had pairwise distances with the query sequence that were below the threshold. Performance for each barcode is presented as the percentage of the total number of samples for that barcode that fell into each of these categories.

We evaluated the success of each barcode at three different thresholds (2%, 5%, and a threshold calculated from the observed levels of intra- and interspecific divergence for each barcode). We chose the first two values to test the performance of standard thresholds that have been proposed as suitable for assigning sequences to species for animals (e.g., Blaxter 2004; Hebert et al. 2003). The third threshold was calculated separately for each barcode to represent the pairwise genetic distance below which 95% of all conspecific individuals were correctly classified. This approach to establishing a threshold is appealing because it uses information contained in the reference sequence data, thus accounting for the differences in mutation rate among the various loci and divergence among taxa. However, application of a calculated threshold requires multiple accessions of most species be present in the reference database.

We evaluated the success of both the 'liberal' and 'strict' tree-based methods of assigning sequences to species as described by Ross, Murugan, and Li (2008). With the liberal method, a sequence is assigned to a species if it is either sister to or embedded within a group. Thus, when A is the query sequence and there are two representatives of species Z in the reference database, A is assigned to species Z if it is either ((Z, A), Z) or

((Z, Z), A). The 'strict' tree-based method, which is capable of accounting for sequences that have no conspecific sequence in the database, only assigns a sequence to a species if it is embedded within a group (e.g., only when ((Z, A), Z)). Trees were constructed for each barcode using PAUP*v4b10 based on the neighbor-joining (NJ) algorithm (Saitou & Nei 1987) applied to K2P distances. Success was determined by how each sequence would be classified if it was the query sequence and all other sequences were present in the reference database. Results are presented as the percentage of total sequences that were assigned to the categories correct and incorrect for the liberal method and as correct, incorrect, or ambiguous for the strict method.

Diagnostic nucleotide polymorphisms can potentially provide a character based approach that captures important differences that are not identified by distance- or tree-based barcoding methods (DeSalle et al. 2005). To assess the degree to which a priori defined species groups harbored diagnostic nucleotide differences, we used the program SITES (Hey & Wakeley 1997) to calculate the number of sites that were fixed between species pairs. To determine how well samples of unknown membership could be correctly assigned using diagnostic characters we used the Character Attribute Organization System (CAOS) (Sarkar et al. 2002a; Sarkar et al. 2008; Sarkar et al. 2002b). The CAOS method involves two steps that are accomplished by the programs P-Gnome and P-Elf. A reduced data set that included one sample for each species (i.e., 29 samples representing 29 species) was used to infer a phylogenetic tree for each barcode using two different inference methods: NJ based on K2P distances accomplished using PAUP*v4b10 and maximum-likelihood as implemented in GARLI v0.951 (Genetic

Algorithm for Rapid Likelihood Inference; Zwickl 2006) using the default parameter

settings. The data matrix and associated tree file were then imported into MacClade

where, based on previous phylogenetic work (Pettengill and Neel 2008), *Agalinis*

calycina was placed as sister to the other members of the genus. The resulting NEXUS

file was used in P-Gnome to generate rules for diagnosing the species where the tree

associated with the data matrix served as a guide tree against which the presence and

absence of nucleotides at a given node were evaluated to determine whether they were

diagnostic (i.e., characteristic attributes) of the taxa subtending that node. Samples for

the guide tree were arbitrarily chosen based on the first sequence for each species when

sorted alphabetically by sampling location. The same samples were used across all

barcodes to prevent confounding our interpretations of barcode performance due to

inclusion of different samples among guide trees. The program P-Elf was then used to

assign sequences not included in the original file to species based on the rules generated

by P-Gnome.

RESULTS

Barcode variation

Sequences were obtained from at least one representative of all 29 species for all loci

and sequences for each locus were obtained from at least 92% of the samples (Table 3.2).

Across all loci, 720 out of a total possible 736 sequences were obtained (98%) and most

samples required only a single attempt at PCR and sequencing to obtain a high quality

sequence. One exception was *rps*2, which required the alternative forward primer *rps*2-

47 instead of the typical *rps*2-18F for 10 individuals (dePamphilis et al. 1997). Despite

these efforts, two individuals were still missing sequences at this locus (Table 3.2). Only

91 sequences could be obtained for *rpo*B and the majority of the missing sequences were

due to *mat*K and *rbc*L that failed for 6 and 7 species, respectively. Both of these latter

loci proved problematic and required multiple attempts at PCR and sequencing to obtain

a high quality sequence from samples.

The percent variable characters ranged from 7.71% to 31.52% for *rbc*L and *psb*A-

*trn*H, respectively (Table 3.2). The number of variable sites for individual loci ranged

from 55 (*rpo*B) to 290 (*trn*T-*trn*L) (Table 3.2). The most conserved locus was *rbc*L with

K2P pairwise distances among *Agalinis* species ranging from 0.0 - 2.20% and averaging

1.02% (Fig. 3.1). The *psb*A-*trn*H locus had the largest range of among-species K2P

distances (0.0 – 14.89%) and averaged 6.68% (Fig. 3.1). The combined *rpo*B/*psb*A-*trn*H

barcode had the largest range of and highest average among-species K2P distances of any

two locus barcode (0.08 – 9.47 and 4.47, respectively) (Fig. 3.1). Not surprisingly, the

average and range of pairwise distances within species generally increased after

synonymization while among-species distances decreased (e.g., Fig. 3.1).

Performance of candidate barcodes

Success of each candidate barcode using the 'best close match' distance-based

method differed depending on the threshold value (Fig. 3.2). In all cases the calculated

threshold was substantially less than 2%, varying from 0.25% (*mat*K) to 1.01% (*rbc*L)

under the historical taxonomy and from 0.15% (*mat*K) to 1.01% (*rbc*L) when species

were synonymized (Table 3.3). Threshold distance values for the multi-locus barcodes

were 0.37% (*mat*K/*psb*A-*trn*H), 0.46% (*rpo*B/*psb*A-*trn*H), and 0.77% (*rbc*L/*psb*A-*trn*H)

under the historical taxonomy and were all lower but in the same rank order in the

synonymized data set (Table 3.3). Regardless of the taxonomic hypothesis considered, as

the distance threshold increased from the calculated one to 2% there was a marginal (i.e.,

few percentage points) increase in correct and incorrect classifications with a

corresponding decrease in 'no match' classifications (Fig. 3.2). For *mat*K, *trn*T-*trn*L, and

the multi-locus barcodes this pattern was also observed when comparing performance

between the 2% and 5% thresholds. Ambiguous classifications remained fairly constant

across the three thresholds (Fig. 3.2). Because it generally had the best results, we

discuss the performance of the candidate barcodes considering only the calculated

distance threshold.

Of the single-locus barcodes, *trn*L-*trn*F had the highest correct classification rate and

tied for the lowest incorrect classification rate with *trn*T-*trn*L and the *trn*L-intron under

the historical taxonomy; the *rbc*L region had the lowest correct classification rate and had

the highest percentage of incorrect classifications (Table 3.3). Although the percent

incorrect classification rates were relatively low for all single locus barcodes (1.08-

7.14%), there were between 26.37% (*psb*A-*trn*H) and 40.47% (*rbc*L) ambiguous

assignments using this method (Table 3).

The multi-locus barcodes *rpo*B/*psb*A-*trn*H and *rbc*L/*psb*A-*trn*H provided marginally

higher correct classification than when any of the loci were considered individually

(Table 3.3). The *mat*K/*psb*A-*trn*H barcode had a higher percentage of correct

classification (62.79%) than when *mat*K was used by itself (61.62%) but performance

was worse than when *psb*A-*trn*H was used alone (64.83%) (Table 3.3).

Synonymizing species increased the average correct classification across all candidate barcodes from 61.83% to 89.05% and correct classification was >90% for four of the single-locus barcodes (Table 3.3). *psb*A-*trn*H had the highest correct classification (94.5%) and no misclassification; *rbc*L had the lowest correct classification (79.76%). Performance of *trn*L-*trn*F declined relative to when the historical taxonomy was used in that there was an increase in the number of query sequences determined to have no match (Table 3.3). The most likely reason for this relative poor performance is that upon synonymizing samples with nearly identical sequences (e.g., *A. acuta*, *A. decemloba*, and *A. tenella*) the threshold value associated with 95% correct classification was too low such that more pairwise conspecific comparisons exceeded that threshold. The three multi-locus barcodes performed similarly to one another after synonymizing species; only the *rbc*L/*psb*A-*trn*H multi-locus barcodes had any individuals misclassified (1.17%) (Table 3.3).

The two tree-based methods differed substantially in the degree to which they accurately handled query sequences. The strict method, by default, classified as ambiguous every sequence that represented a species for which we had only sampled two individuals. This was a result of treating one sample as a reference sequence and the other as a query sequence, under which there is no way for a query sequence to be embedded among conspecifics when only one conspecific is in the database (Table 3.4). This conservative approach of the strict method resulted in an average 17.07% correct classification rate that increased to 41.95% when species were synonymized (Table 3.4). However, there were no incorrect classifications using the strict method because query

sequences were never incorrectly embedded in a monospecific clade consisting of heterospecifics. Under the historical taxonomy, when samples of *A. acuta*, *A. decemloba*, *A. tenella*, *A. paupercula*, or *A. purpurea* were treated as query sequences, they were embedded within a clade that contained both heterospecifics and conspecifics and as a result assignments were considered ambiguous rather than incorrect. As a consequence of the strict method, even when species were synonymized, two of the samples were sister to rather than embedded within the synonymized group and were classified as ambiguous.

The liberal tree-based method had an average correct classification of 60.69% under the historical taxonomic framework, and 91.69% after syononymization (Table 3.4). Despite the inability of this method to identify sequences that do not have a conspecific in the database it still had a relatively low incorrect classification rate, which, based on the synonymized taxonomic hypothesis ranged from 4.35% for *psb*A-*trn*H to 28.24% for *rbc*L (Table 3.4).

Under both tree-based methods, incorporating a second locus increased the correct classification rate relative to the performance when the more slowly evolving locus was used alone. For example, using the synonymized data set and the strict tree-based method, the correct classification rates for *rbc*L and *rpo*B were 38.82% and 34.07%, respectively; when they were combined with *psb*A-*trn*H correct classification was 41.18% and 45.05%, respectively (Table 3.4). The combination of *mat*K and *psb*A-*trn*H did not significantly change the correct classification rate; the decrease in correct

classification rate of *mat*K/*psb*A-*trn*H relative to *psb*A-*trn*H is due to differences in samples sizes between the two barcodes (Tables 3.2 & 3.4).

We found diagnostic characters in all barcodes for the majority of putative *Agalinis* species pairs. The average number of fixed differences among species varied from 4.85 in *rbc*L to 30.74 in *trn*T-*trn*L for the single locus barcodes. The *mat*K, *rps*2, *psb*A-*trn*H, and all three multi-locus barcodes had fixed nucleotide differences for 82.75% of the species. The *rbc*L and *rpo*B barcodes discriminated the fewest species based on the presence of fixed nucleotides (58.62% and 72.41%, respectively). When species were synonymized, 3 single-locus and all 3 multi-locus barcodes had fixed nucleotide differences that distinguished all 26 species, although the average number of fixed differences was generally lower than when the historical taxonomy was used.

The number of query sequences used to evaluate the CAOS method ranged from 56 (*rbc*L) to 63 (*psb*A-*trn*H, *trn*T-*trn*L, *trn*L-intron, and *trn*L-*trn*F) individuals, which always increased by three for the analyses in the synonymized data set (Table 3.5). Despite indications of the utility of fixed differences, percent correct classification using this method was most often substantially lower than the other methods evaluated (Tables 3.3 - 3.5), ranging from 17.74% for *rpo*B to 63.49% for *trn*L-*trn*F and *psb*A-*trn*H (Table 5). Misclassifications were due to both incorrect and ambiguous calls. The multi-locus barcodes all performed worse than single locus barcodes. This result is surprising given that chloroplast loci do not assort independently of one another and, thus should not be in conflict. Additionally, when two chloroplast loci are concatenated they possess the combined diagnostic characters and should have more information than each single locus.

Using the synonymized data set greatly increased the number of query sequences that could be correctly classified using the CAOS method with two barcodes (*psb*A-*trn*H and *trn*L-*trn*F) having correct classification rates >87% (Table 3.5). The multi-locus barcodes still had classification success rates below those of the single locus barcode *psb*A-*trn*H, the faster evolving locus of the two-locus system.

In addition to being generally low, barcoding success from CAOS depended on the method of inferring the phylogeny that served as the guide tree (NJ or maximum-likelihood) (Table 3.5). Unfortunately, neither method was consistently better in that the percent correct classification was higher for a tree inferred using the NJ algorithm for *rbc*L, *rpo*B, *trn*L-*trn*F and the maximum likelihood tree yielded better classification for *psb*A-*trn*H; the two methods did not differ for *rps*2, *trn*T-*trn*L, and *trn*L-intron. In general, the maximum likelihood method had fewer ambiguous and more incorrect calls than the neighbor joining method.

Species-specific performances

No DNA barcode locus or analytical technique correctly classified all samples of *A. acuta*, *A. decemloba, A. tenella, A. paupercula, A. purpurea*, or *A. harperi*. When *A. acuta*, *A. decemloba,* and *A. tenella* were synonymized as *A. decemloba* and *A. paupercula* and *A. purpurea* were synonymized as *A. purpurea*, the distance- and tree-based method always classified all relevant sequences into the correct synonymized group. However, the CAOS method did not correctly classify all of the samples for five barcodes (*mat*K/*psb*A-*trn*H, *rbc*L, *rbc*L/*psb*A-*trn*H, *trn*T-*trn*L, and *rpo*B). The two *A. harperi* samples were never correctly classified as conspecifics.

For the distance-based methods, although the results varied depending on the DNA barcode, 10 samples were typically classified as 'no match' because the distances among the conspecific individuals exceeded the designated threshold (e.g., *A. skinneriana* 90MO, the two *A. linifolia* samples, *A. oligophylla* 12AL and 5AL, *A. heterophylla* TX and 5TX, *A. pulchella* 3GA, *A. fasciculata* 1LA, and *A. setacea* 76VA). The two species for which we only had a single representative (i.e., *A. calycina* and *A. edwardsiana*) were correctly treated as having no conspecific match in the database when using *trn*L-*trn*F, *psb*A-*trn*H, *rpo*B/*psb*A-*trn*H, and *mat*K/*psb*A-*trn*H. However, for the other barcodes (i.e., *rps*2, *mat*K, *rbc*L, *rpo*B, *trn*L-intron, *trn*T-*trn*L, and *rbc*L/*psb*A-*trn*H) classifications of *A. calycina* and *A. edwardsiana* were incorrect or ambiguous due to insufficient differentiation from heterospecific samples in the database.

Species-specific misclassification was high for the strict tree-based method where the samples from all but one species (i.e., *A. harperi*) that were represented by two individuals were classified as ambiguous. For the liberal tree-based method, the only incorrect classifications using the best performing locus (i.e., *psb*A-*trn*H) involved the query sequences that did not have conspecifics in the database (i.e., *A. calycina* and *A. edwardsiana*) and the two polyphyletic *A. harperi* samples. The higher incorrect classification rates associated with other loci (e.g., *rbc*L and *rpo*B) was due to lack of sufficient nucleotide variation to depict accurately the relationships among the samples.

Under the diagnostic character based method of CAOS, a number of species-specific classifications depended on the inference method used to create the guide tree. For example, using the *psb*A-*trn*H barcode and the NJ algorithm, seven samples were

misclassified (*A. fasciculata* 2GA, *A. fasciculata* 4LA, *A. harperi* 12NC, *A. setacea* 7MD, and *A. tenuifolia* 5IA were classified as *A. skinneriana*; *A. setacea* 3VA was classified as *A. tenuifolia*; and *A. tenuifolia* 2VA was classified as *A. aphylla*), and for the same barcode, but using a guide tree inferred using a maximum likelihood-based method, there were only four incorrect classifications (*A. setacea* 7MD, *A. setacea* 3VA, and *A. tenuifolia* 5IA were classified as *A. skinneriana* and *A. tenuifolia* 2VA was classified as *A. aphylla*).

DISCUSSION

The need to catalog the earth's rapidly declining biodiversity makes the promise of DNA barcoding highly appealing (Savolainen et al. 2005). In addition to facilitating basic inventorying of poorly studied areas, barcoding has been proposed as an effective means of delineating and monitoring species distributions (DeSalle et al. 2005). Sequence databases created as a result of DNA barcoding have also been advocated as being useful for understanding the evolutionary diversity within rare species (Faith & Baker 2006). Achieving these purposes requires the ability to distinguish among closely related species including sister taxa. Because most barcoding evaluations to date have been in search of universally amplifiable loci, thorough tests of this ability are only beginning to accumulate (e.g., Newmaster & Ragupathy 2009; Starr et al. 2009). Although limited in taxonomic breadth, our analysis of 92 samples representing 29 *Agalinis* species using three analytical techniques is one of the most extensive examinations to date and clearly shows that barcoding has potential for distinguishing among congeners using single barcodes. Although we have examined only one genus,

the range of evolutionary distances within and among the 29 species likely reflect patterns in many angiosperm genera and in certain instances even represent the level of differentiation among members from different genera (e.g., Kress et al. 2005; e.g., Lahaye et al. 2008).

The ability to distinguish closely related species means that barcoding can be useful in conservation biology to confirm identities in taxonomically challenging groups that are difficult to identify in the field. Such information can serve to improve understanding of species distributions. A specific example from our study is the case of the state rare *A. skinneriana*, whose geographic range was thought to not include Mid-Atlantic coastal states. When samples from Maryland were first collected they were tentatively attributed to *A. skinneriana*, but local botanists had concerns because no other populations of this species occur within 400-500 km; barcoding unambiguously confirmed the taxonomic identity of those samples. We were also able to confirm identifications from collection locations from which individuals had originally been identified as *Agalinis decemloba* but later annotated as *A. setacea*. However, as discussed below, successful application of DNA barcoding depends on the barcode locus, the analytical method, and the underlying taxonomy used.

Candidate barcodes

Of the loci we tested, *psb*A-*trn*H and *trn*T-*trn*L most closely met the three requirements of a suitable DNA barcode identified by the Consortium for the Barcode of Life in that those loci were routinely retrievable with a single primer pair, easy to obtain bidirectional sequences, required little to no manual editing of sequence traces, and

provided maximal discrimination among species (CBOL Plant Working Group 2009) (Table 3.2, Fig. 3.1). When evaluating the data set in which phylogenetically dubious taxa had been synonymized, these two loci had a greater than 83% correct classification rate across all methods except the strict tree method. Under the historical taxonomic framework, accuracy was closer to 50-70%, mostly due to ambiguous classifications among the species that were synonymized in the other data set. Although they worked well for *Agalinis* and have been advocated by others (Kress & Erickson 2007)*,* these loci are known to have alignment issues (*psb*A-*trn*H; Lahaye et al. 2008) or suffer from a lack of significant interspecific differentiation (*trn*T-*trn*L; Edwards et al. 2008) in some taxa.

Another barcode that has been advocated, either alone (Lahaye et al. 2008) or as part of multi-locus system (Newmaster et al. 2008), *mat*K, performed well under both the distance- and tree-based methods, but not with the diagnostic character-based system (Tables 3.3 – 3.5). However, a serious issue with using *mat*K is that we were only able to acquire 86 sequences from the 92 samples and multiple sequencing attempts were necessary to obtain high quality reads. Difficulty in amplifying *mat*K has been noted in studies including a broad range of angiosperm orders and families (Fazekas et al. 2008; Sass et al. 2007) but we were surprised that even among closely related species, primer regions were not sufficiently conserved to yield reliable PCR amplification.

We found no benefit to using multi-locus barcodes. At best, they showed marginal improvement, and at worst showed an actual decrease in performance compared to the best performing single locus barcodes (Tables 3.3 – 3.5). Therefore, the additional work and cost necessary to acquire a second locus does not seem warranted. Although one

could argue that the closely related species we investigated do not represent the situation under which a multi-locus barcode may be useful, the benefits have also been shown to be negligible with more evolutionarily divergent samples (e.g., Lahaye et al. 2008). However, multi-locus barcodes have yielded improved performance in some studies (e.g., Fazekas et al. 2008; e.g., Hollingsworth et al. 2009).

Given that different methods and loci have been used across plant DNA barcoding studies (CBOL Plant Working Group 2009; Chase et al. 2005; Fazekas et al. 2008; Kress & Erickson 2007; Lahaye et al. 2008; Newmaster et al. 2008; Sass et al. 2007) it is still not possible to determine for certain if there will be a single locus to barcode all plants. Given the accumulated evidence, it appears unlikely. However, as reference databases become more complete in terms of loci and species and more studies are completed using consistent methods of evaluation, a small number of loci may emerge that work across an array of taxa.

Metrics such as the "probability of correct identification" (PCI) (Erickson et al. 2008) allow comparison of rates of correct classification across barcodes, analytical methods, and databases. Erickson and Kress (2008) suggest that the level of unambiguous correct classification to be expected is on the order of 60%-70%, which is similar to the conclusion of 71% success described in Fazekas et al. (2008). Empirical results have yielded both substantially lower (e.g., 0-~30% success across three genera of land plants; Hollingsworth et al. 2009) and higher (e.g., 94.7% success across congeneric species; Newmaster et al. 2008) correct classification, the latter being similar

to what we observed. Correct classification rates are useful when the identities of taxa are known with confidence, which is the case when developing reference databases.

In real barcoding applications, samples of unknown identity that are compared with reference sequences will either be unambiguously assigned to one species, assigned to multiple species, or assigned as having 'no match' depending on the methods used. Some percentage of unambiguous, apparently correct assignments will in fact be incorrect (i.e., will be false positives). We argue that understanding the potential for false positives resulting from unambiguous assignment to the wrong species is more critical than knowing the potential for correct identifications. A means of estimating this potential is to assume that it will be similar to the percent of incorrect matches among samples of known taxonomic identity (i.e., those in the reference database). Using distance methods, our percentage of false positives was 1.25% when species were synonymized and 3.5% under the historical taxonomy (Table 3.3), which is comparable to levels found by Newmaster et al. (2008). Thus, the low levels of correct classification under the historical taxonomy were due to ambiguous classifications because sequences were assigned to multiple species; a few samples were also classified as 'no match' because they were sufficiently distinct from all species. Methods that provide ambiguous and 'no match' classifications are essential because they disclose patterns that would be masked if query sequences were forced into only species that exist in the database and thus highlight the need for further research to understand the cause of the ambiguities.

Differential success of assignment methods

Although some loci were generally better than others, performance of a given barcode was highly dependent on the analytical method being used. The 'best close match' method as implemented in TAXONDNA (Meier et al. 2006) yielded the best overall results across most loci. In addition to high correct classification rates, most of the samples that were not classified correctly were 'ambiguous' rather than incorrect (Table 3.3). This method also has the beneficial property discussed above of having the potential to avoid errors of commission by identifying sequences as having no match rather than forcing incorrect classification as the closest species. Although there were cases in which samples were erroneously identified as having no conspecific match in the database, these misclassifications represent an exception to the behavior of the majority of conspecifics we examined. Such cases yield opportunities for further examination to determine if they resulted from misidentification, contamination, or something biologically interesting such as presence of phylogeographic structure or presence of cryptic species.

The primary criticism of distance-based methods is that no single distance threshold delineates all species (DeSalle et al. 2005; Ferguson 2002; Little & Stevenson 2007). Calculating thresholds from the data has been proposed as a way to overcome this disadvantage (e.g., Meier et al. 2006). Our calculated values were lower than the arbitrary fixed threshold distances that have been suggested and distances calculated among other taxa. For example, calculated thresholds for *psb*A-*trn*H and *mat*K in the synonymized data set were 0.56% and 0.15%, respectively, whereas Newmaster et al.

(2008) documented thresholds of 2.52% and 0.26%, respectively, for these same loci. Although the differences in calculated threshold reinforce the idea that a single threshold may not be appropriate, there was little practical consequence of the three different threshold values we evaluated. For most loci, both correct and incorrect classification rates increased by only a few percentage points, with a concomitant reduction in ambiguous and 'no match' classifications as the threshold is increased. The exception was *trn*L-*trn*F for which the 2% and 5% thresholds yielded ~10% higher correct classification than did the calculated threshold (Fig. 3.2a and 3.b). Because multiple conspecific accessions are required to calculate thresholds and a 2% universal threshold performed reasonably well, the latter could be used in small reference databases with the caveat that there could be a marginal increase in incorrect classifications at the cost of identifying biologically interesting situations through ambiguous or 'no match' classifications. If a reference database contains sufficient accessions, a calculated threshold is preferred.

Although the liberal tree-based clustering method yielded correct classification rates as high as 71.76%-95.65% in the synonymized data set, the inability of this method to accurately handle sequences that do not have a conspecific in the reference database render it inferior to other methods based on distances and diagnostic character differences. This conclusion is in contrast to Ross et al. (2008) who found the liberal tree-based method superior to distance methods in this regard. The strict method can identify sequences that have no conspecific in the database; however, because a query sequence needs to be embedded in a clade in order for a taxonomic identity to be

assigned to it (Ross et al. 2008), success is highly dependent on the number of reference sequences in the database. Given that one sequence was treated as the reference and the other as a query sequence, the 14 species for which we sampled two individuals represent the extreme situation in which any query sequence will at best be sister to a conspecific and thus will be identified as ambiguous. Given the dependence of the success of this method on the number of conspecifics sampled, it has been suggested that at least five conspecifics should be present in the reference database to ensure accurate identifications (Ross et al. 2008). The authors note that this level of sampling may not be achievable for the majority of species.

The number of segregating sites that were fixed among a priori groups representing putative species suggests that such characters could be useful for diagnosing species. The character-based CAOS method (Sarkar et al. 2008) that uses fixed nucleotide differences is conceptually appealing because it does not require multiple conspecifics to be in the reference database and it can handle query sequences with no conspecific in the database. Unfortunately, this method yielded a wide range in the percent correct classification rate and some barcodes performed dismally compared to their performance using other methods (e.g., *mat*K, *rbc*L, *rpo*B and the multi locus barcodes; Table 3.5). A number of the samples that failed to be correctly classified under the CAOS method were accurately identified by the distance- and tree-based methods and formed strongly supported monophyletic groups in Pettengill and Neel (2008) (e.g., the phylogenetically distinct samples *A. setacea* 7MD and *A. tenuifolia* 5IA were classified as *A. skinneriana* using the *psb*A-*trn*H barcode). These misclassifications might have resulted because basal

relationships among these species lack statistical support such that few diagnostic characters exist within individual barcodes to differentiate such historical relationships (Pettengill & Neel 2008). This lack of resolution is exacerbated in the CAOS method because the guide tree is based on a single locus or two loci that differ in their phylogenetic informativeness and the topologies based on them; for the same barcode topologies also differed between the two inference methods. Another potential problem with the CAOS method is whether enough diagnostic character differences will be present for any one barcode when many species (e.g., thousands) are considered in the P-Gnome process. The issue is similar to the problem observed in phylogenetic studies where the degree of homoplasy increases with increased taxon sampling (Sanderson & Donoghue 1989). In barcoding, such a phenomenon would result in a situation where diagnostic characters that distinguish certain species may no longer be diagnostic when additional species are considered. Given the degree of single nucleotide polymorphisms as well as the number of insertion-deletion polymorphisms we observed (data not presented) we suggest that the utility of loci for plant barcoding would increase if an automated means of identifying and incorporating these polymorphisms into assignments were developed.

Importance of taxonomy

Beyond barcode characteristics and analytical methods, classification success was conditional upon the taxonomic hypothesis used. Results under the historical taxonomy of 29 species and the 'best close match' distance method yielded correct identification of ~60% of *Agalinis* individuals using the loci we sequenced (Table 3.3). Such low correct

classification rates can result from polyphyletic or paraphyletic relationships among species that are in fact reproductively isolated and distinguishable by experts based on morphological features or from incorrect taxonomy. There is no way to distinguish these two possible causes of misclassification without additional information. In the case of *Agalinis*, previous phylogenetic studies (Neel & Cummings 2004; Pettengill & Neel 2008) suggested that separate species status may be unwarranted for two groups of species, first *A. acuta*, *A. decemloba*, and *A. tenella* and second *A. paupercula* and *A. purpurea*). When species in these two groups were synonymized with one another, correct classification exceeds 80% for all loci and was above 90% for the best performing loci (Table 3.3). Given that all three of the barcoding methods we evaluated can be misleading when species are not reciprocally monophyletic (Ross et al. 2008) and the potentially high frequency of such relationships (Crisp & Chandler 1996; Funk & Omland 2003; Rieseberg & Brouillet 1994), there is there is potential that DNA barcoding may challenging in many taxonomic groups (Hollingsworth et al. 2009).

Although DNA barcoding alone cannot resolve taxonomic uncertainty, it may be a useful tool for detecting it. The ability to detect cryptic species with barcoding is well known (e.g., Elias-Gutierrez & Valdez-Moreno 2008; e.g., Ragupathy et al. 2009; Yassin et al. 2008). The technique can also be a useful tool for detecting when two putative entities may not actually be different (Fazekas et al. 2009). The 'best close match' method effectively identified reference samples of *A. acuta*, *A. decemloba*, and *A. tenella* as ambiguous and, therefore, sequences of these species were assignable to more than one reference species. Had we had no other information prior to this study, the ambiguous

classifications would have triggered further investigation. Thus, rather than replacing

traditional taxonomic approaches, DNA barcoding can be seen as part of an iterative

process in which query sequences that do not match sequences in the database or that are

assigned to multiple taxa prompt examination of taxonomic hypotheses used to assign

membership. The results from these additional studies may either support the current

taxonomic hypothesis or suggest taxonomic revisions. There is an undeniably

disconcerting circularity to this logic: if samples that cannot be delineated to a single

species using barcoding are synonymized then those samples become distinguishable

with barcodes. However, not all instances of 'no match' or ambiguous classifications

will result in taxonomic revisions when additional data are examined. The ability to

detect incorrect taxonomy is highly dependent on the density of taxonomic coverage in

the reference database (e.g., Ross et al. 2008) and the frequency at which this iterative

process will be repeated will depend on the degree to which current taxonomic

hypotheses are incorrect and putative species are not reciprocally monophyletic.

Conclusions

Our results show that the success of DNA barcoding varies depending on three

factors: 1) the actual barcode being employed; 2) the analytical method being used to

determine the taxonomic identity of a query sequence; and 3) the related factors of

accuracy of the taxonomy associated with the sequences in the reference database and the

degree to which species are not monophyletic. Within limits imposed by the factors

above, we found two loci (*psb*A-*trn*H and *trn*T-*trn*L) to be effective at distinguishing

among congeners in the genus *Agalinis*. The 'best close match' distance method

generally outperformed other methods due to its correct classification rates, potential for

ambiguous classification, and identification of query sequences with no match in the

database.

This success indicates that DNA barcoding can be useful in a conservation context

by determining the identity of morphologically confusing species or populations that

appear to represent an extension of a species' range. Although it has been argued that

application of DNA barcoding to conservation is dependent on having taxonomically

extensive representation in the reference database (e.g., Rubinoff 2006), our results

suggest this is not the case because methods exist that can accurately identify samples

that have no conspecifics present in the database. Assuming that taxonomic hypotheses

are accurate and given an appropriate method, DNA barcoding will provide a means of

identifying understudied and putatively rare species that warrant additional studies to

evaluate their evolutionary distinctiveness and phylogenetic affinity.

Table 3.1. The number of conspecifics, general locations of sampled individuals, and conservation status for the 29 *Agalinis* species investigated.

Taxon	*N*	State	Status[1]
Section Erectae			
A. acuta	10	CT, RI, MA, NY, MD	G1/S1
A. aphylla	2	FL, AL	G3-G4/S2
A. decemloba	5	VA, NC	NR
A. gattingeri	3	MO, LA	G4/S1
A. obtusifolia	9	AL, SC, FL	G4-G5-Q/S1
A. oligophylla	4	AL, TX	G4/S1
A. skinneriana	3	MD, MO	G3/S1
A. tenella	8	GA, SC	NR
A. viridis	2	LA	G4/S1
Section Heterophyllae			
A. auriculata	2	IA, IL	G3-G4/S1
A. heterophylla	3	TX	G4-G5/S1
A. calycina	1	TX	G1/S1
Section Linifoliae			
A. linifolia	2	FL, GA	G4?/S1
Section Purpureae			
A. edwardsiana	1	TX	G4/S4
A. homalantha	2	TX	G5/S1
A. pulchella	2	GA, FL	G4-G5/S3?
A. strictifolia	2	TX	G4/SNR
A. navasotensis	2	TX	G1/S1
A. fasciculata	3	TX, GA, LA	G5/S1
A. harperi	2	FL, NC	G4?/SNR
A. maritima	2	TX, MA	G5/S2
A. paupercula	2	MA, NY	G5/S1
A. purpurea	4	VA, AL, MD, SC	G5/S1
A. laxa	2	SC, GA	G3-G4/S3?
A. plukenettii	2	FL, GA	G3-G5/S1
A. setacea	4	VA, MD	
Section Tenuifoliae			
A. filicaulis	2	FL, AL	G3-G4/S1
A. divaricata	2	FL	G3?/S1
A. tenuifolia	4	LA, VA, IA, MD	G5/S1

[1]Conservation Status: global ranking (G1=critically imperiled; G2=imperiled; G3=vulnerable to extinction or extirpation; G4=apparently secure; G5=demonstrably secure or widespread)/highest state ranking for each species (S1-S5 are equivalent to the global scale but applied to within a single state) (USA); when a range or question mark (?) is given the precise conservation status is uncertain.; NR and SNR= not ranked

Table 3.2. Summary characteristics of the 11 DNA barcodes evaluated.

Locus	N individuals (N species)	Aligned length (bp)	Variable characters (percent)
*mat*K	86 (29)	899	168 (18.69%)
*rbc*L	85 (29)	1012	78 (7.71%)
*rpo*B	91 (29)	368	55 (14.95%)
*rps*2	90 (29)	660	146 (22.12%)
*trn*T-*trn*L	92 (29)	1397	290 (20.76%)
*trn*L-intron	92 (29)	548	84 (15.32%)
*trn*L-*trn*F	92 (29)	965	182 (18.86%)
*psb*A- *trn*H	92 (29)	809	255 (31.52%)
*rpo*B/ *psb*A-*trn*H	91 (29)	1177	310 (26.33%)
*rbc*L/*psb*A-*trn*H	85 (29)	1821	333 (18.28%)
*mat*K/ *psb*A-*trn*H	86 (29)	1708	422 (24.71%)

Table 3.3. Performance as measured by percent of samples for the candidate barcodes when 'best close match' distance method (Meier et al. 2006) is used to identify sample sequences. Threshold distances were calculated from the observed sequences for each locus. Numbers in parentheses indicate the performance of barcodes when species are synonymized based on the results of Pettengill and Neel (2008).

Barcode	Correct	Ambiguous	Incorrect	No match	Threshold
*mat*K	61.62	31.39	1.16	5.81	0.25
	(93.02)	(0.0)	(1.16)	(5.81)	(0.15)
*rbc*L	50.0	40.47	7.14	2.38	1.01
	(79.76)	(14.28)	(3.57)	(2.38)	(1.01)
*rpo*B	58.42	31.46	4.49	5.61	0.54
	(86.51)	(4.49)	(3.37)	(5.61)	(0.54)
*rps*2	60.0	28.88	6.66	4.44	0.58
	(92.22)	(2.22)	(1.11)	(4.44)	(0.58)
*psb*A-*trn*H	64.83	26.37	3.29	5.49	0.64
	(94.5)	(0.0)	(0.0)	(5.49)	(0.56)
*trn*T-*trn*L	60.86	33.69	1.08	4.34	0.54
	(92.39)	(2.17)	(1.08)	(4.34)	(0.46)
*trn*L-intron	59.78	34.78	1.08	4.34	0.38
	(85.86)	(2.17)	(1.08)	(10.86)	(0.18)
*trn*L-*trn*F	67.39	27.17	1.08	4.34	0.56
	(79.34)	(3.26)	(1.08)	(16.3)	(0.21)
*rpo*B *and*	67.39	21.73	4.34	6.52	0.46
*psb*A-*trn*H	(93.47)	(0.0)	(0.0)	(6.52)	(0.34)
*rbc*L *and*	67.05	23.52	4.7	4.7	0.77
*psb*A-*trn*H	(90.58)	(0.0)	(1.17)	(8.23)	(0.54)
*mat*K *and*	62.79	26.74	3.48	6.97	0.37
*psb*A-*trn*H	(91.86)	(0.0)	(0.0)	(8.13)	(0.2)
Average	**61.83**	**29.65**	**3.50**	**4.99**	**0.55**
	(89.05)	**(2.60)**	**(1.25)**	**(7.10)**	**(0.43)**

Table 3.4. Performance of the candidate barcodes based on the strict and liberal tree based methods of Ross, Murugan, and Li (2008). Numbers in parentheses indicate the performance of barcodes when species are synonymized based on the results of Pettengill and Neel (2008).

Barcode	Strict method		Liberal method	
	Correct	Ambiguous	Correct	Incorrect
*mat*K	17.44%	82.56%	63.95%	36.05%
	(44.19%)	(55.81%)	(95.35%)	(4.65%)
*rbc*L	12.94%	87.06%	52.94%	47.06%
	(38.82%)	(61.18%)	(71.76%)	(28.24%)
*rpo*B	18.68%	81.32%	57.14%	42.86%
	(34.07%)	(65.93%)	(89.01%)	(10.99%)
*rps*2	16.67%	83.33%	55.56%	44.44%
	(45.56%)	(54.44%)	(95.56%)	(4.44%)
*psb*A-*trn*H	18.48%	81.52%	64.13%	35.87%
	(45.65%)	(54.35%)	(95.65%)	(4.35%)
*trn*T-*trn*L	17.39%	82.61%	60.87%	39.13%
	(44.57%)	(55.43%)	(92.39%)	(7.61%)
*trn*L-*intron*	17.39%	82.61%	60.87%	39.13%
	(44.57%)	(55.43%)	(92.39%)	(7.61%)
*trn*L-*trn*F	17.39%	82.61%	60.87%	39.13%
	(44.57%)	(55.43%)	(92.39%)	(7.61%)
*rpo*B/*psb*A-*trn*H	18.68%	81.32%	62.64%	37.36%
	(45.05%)	(65.93%)	(93.41%)	(6.59%)
*rbc*L/*psb*A-*trn*H	15.29%	84.71%	64.71%	35.29%
	(41.18%)	(58.82%)	(95.29%)	(4.71%)
*mat*K/*psb*A-*trn*H	17.44%	82.56%	63.95%	36.05%
	(44.19%)	(55.81%)	(95.35%)	(4.65%)
Average	**17.07%**	**82.93%**	**60.69%**	**39.31%**
	(41.95%)	**(58.05%)**	**(91.69%)**	**(8.31%)**

Table 3.5. Performance of the candidate barcodes in terms of percentage of query sequences (*N*) classified into each category based on the diagnostic character based method CAOS (Sarkar et al. 2008). Numbers in parentheses indicate the performance of barcodes when species are synonymized based on the results of Pettengill and Neel (2008).

Barcode	*N*	Neighbor-joining			Maximum-likelihood		
		Correct	Ambiguous	Incorrect	Correct	Ambiguous	Incorrect
*mat*K	57	29.82	43.86	26.32	36.84	10.53	52.63
	(60)	(68.33)	(6.67)	(25.00)	(68.33)	(5.00)	(26.67)
*rbc*L	56	30.36	25.00	44.64	25.00	28.57	46.43
	(59)	(27.12)	(30.51)	(42.37)	(28.81)	(30.51)	(40.68)
*rpo*B	62	22.58	59.68	17.74	17.74	58.06	24.19
	(65)	(53.85)	(23.08)	(23.08)	(53.85)	(24.62)	(21.54)
*rps*2	61	44.26	45.90	9.84	54.10	22.95	22.95
	(64)	(84.38)	(12.50)	(3.13)	(82.81)	(10.94)	(6.25)
*psb*A-*trn*H	63	58.73	30.16	11.11	63.49	30.16	6.35
	(66)	(87.88)	(1.52)	(10.39)	(87.88)	(1.52)	(10.39)
*trn*T-*trn*L	63	55.56	12.70	31.75	55.56	14.29	30.16
	(66)	(83.33)	(0.0)	(16.67)	(84.85)	(0.00)	(15.159)
*trn*L-intron	63	46.03	50.79	3.17	46.03	50.79	3.17
	(66)	(83.33)	(9.09)	(7.58)	(83.33)	(12.12)	(4.55)
*trn*L-*trn*F	63	63.49	7.94	28.57	61.90	7.94	30.16
	(66)	(87.88)	(6.06)	(6.06)	(84.85)	(13.64)	(1.51)
*rpo*B/*psb*A-*trn*H	62	45.16	19.35	35.48	41.94	9.68	48.39
	(65)	(70.77)	(1.54)	(27.69)	(75.38)	(1.54)	(23.08)
*rbc*L/*psb*A-*trn*H	56	53.57	8.93	37.50	58.93	8.93	32.14
	(59)	(76.27)	(0.0)	(23.73)	(54.24)	(0.0)	(45.76)
*mat*K/*psb*A-*trn*H	57	42.11	22.81	35.09	49.12	19.30	31.58
	(60)	(68.33)	(0.0)	(31.67)	(68.33)	(0.0)	(31.67)
Average	**59.67**	**43.34**	**29.29**	**27.37**	**45.45**	**21.79**	**32.76**
	(63.27)	**(71.95)**	**(8.27)**	**(19.78)**	**(70.24)**	**(9.08)**	**(20.68)**

Figure 3.1. Histograms of the number of intraspecific and interspecific comparisons that fell into 0.5% K2P distance categories for three candidate barcodes under two taxonomic hypotheses.
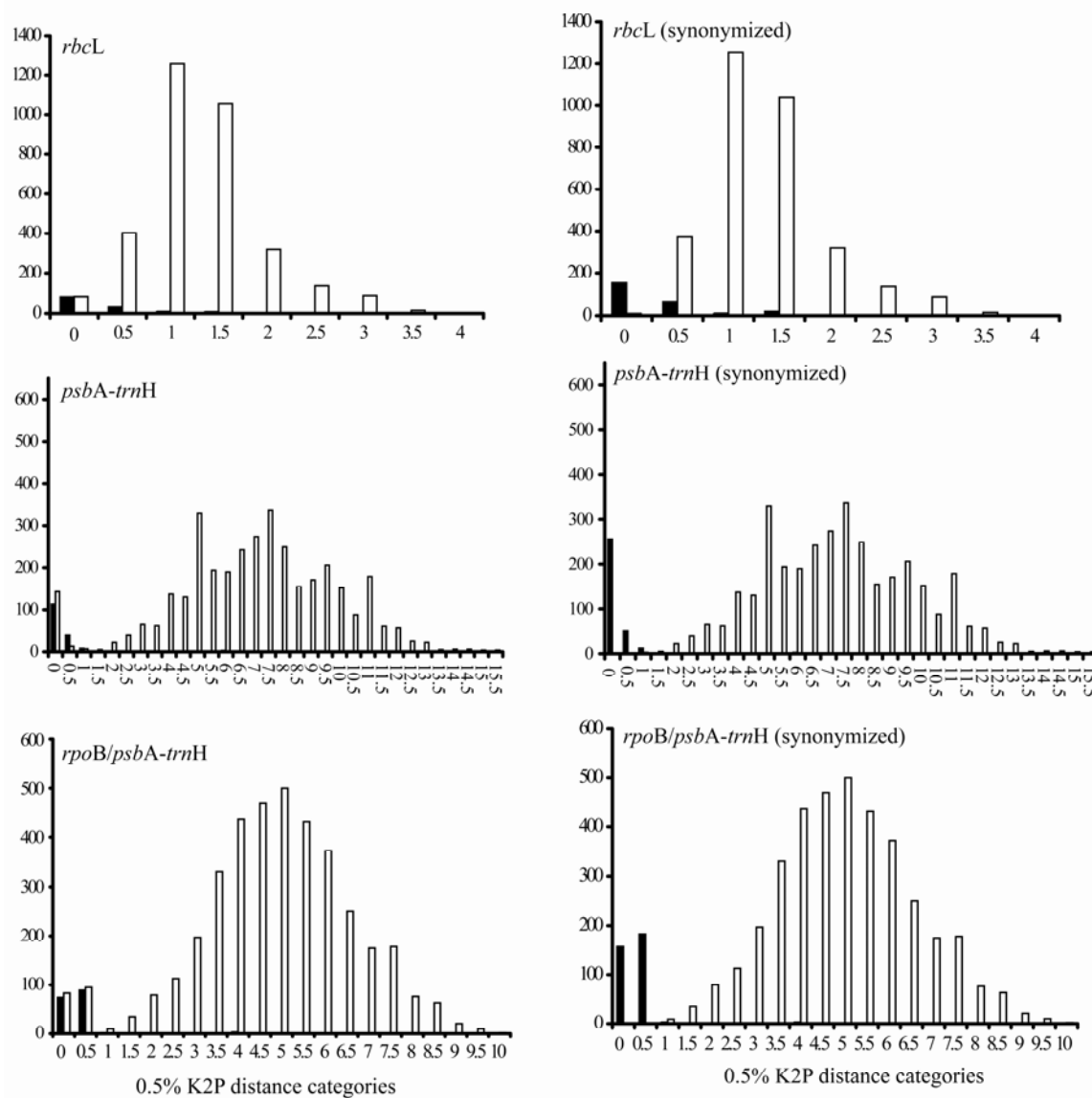
Figure 3.2. Performance of 11 candidate barcodes across three different threshold values for the 'best close match' distance-based method (Meier et al. 2006). Top panel represents classifications under the historical taxonomic alignment and the bottom panel are results after synonymization based on the results of Pettengill and Neel (2008).
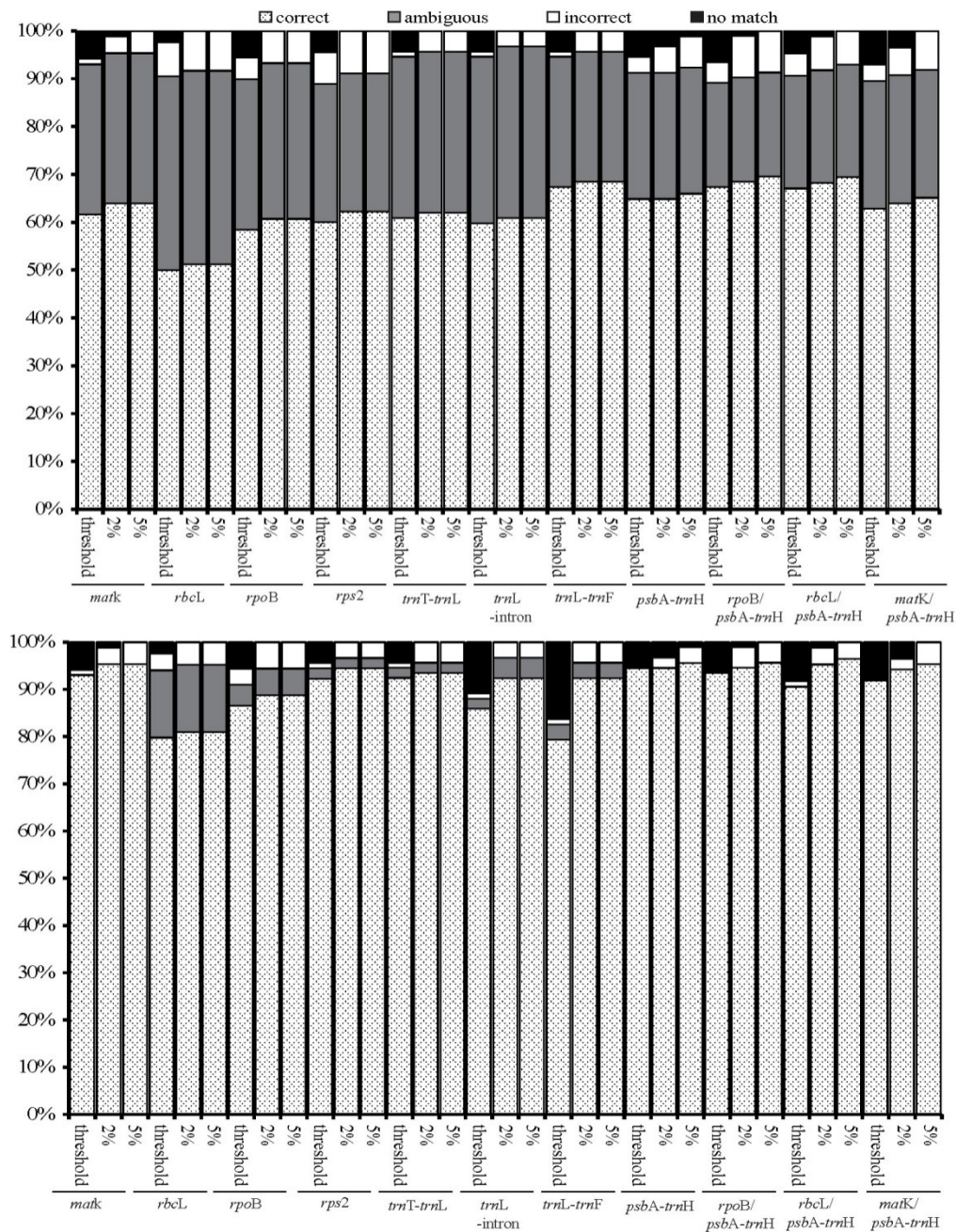
CHAPTER 4: DIFFERENTIATING AMONG ALTERNATIVE PHYLOGEOGRAPHIC HYPOTHESES WITHIN *AGALINIS DECEMLOBA* (OROBANCHACEAE) BASED ON MOLECULAR AND MORPHOLOGICAL VARIATION

ABSTRACT

Differentiating between historical processes related to the establishment of populations and contemporaneous events as explanations for extant patterns of diversity is central to understanding the temporal dynamics of microevolutionary forces. Two dominant phylogeographic hypotheses are that following deglaciation during the Pleistocene northern populations were founded by dispersal events out of southern refugial populations (i.e., the leading edge hypothesis) or that northern populations persisted as refugia throughout the glacial cycles. However, the patterns expected under these hypotheses could potentially be the result of more recent anthropogenic actions that have caused a decrease in suitable habitat or resulted in the extirpation of populations. We evaluated the likelihood of these alternative hypotheses in explaining extant patterns of differentiation at 20 microsatellite loci and a suite of morphological traits sampled from 14 *Agalinis decemloba* Pennell (Orobanchaceae) populations located in eastern North America. We found a significant negative correlation between allelic richness and observed heterozygosity with increasing latitude, which are expected under the leading edge phylogeographic hypothesis. Neighbor-joining trees based on microsatellite and morphological variation were congruent with a cladogram representing the expected northward colonization route under the leading edge hypothesis. Forward in time simulations of each hypothesis under a metapopulation model showed that it is highly

unlikely that northern populations persisted as refugia throughout the glacial cycles of the Pleistocene. Indicative of strong genetic drift, isolation by distance at both morphological and genetic variation was only observed among relatively close populations. We conclude that extant patterns of phenotypic and genetic differentiation within *A. decemloba* are primarily due to the establishment of populations from southern refugia, however, recent anthropogenic events may have also contributed to observed patterns of differentiation.

INTRODUCTION

In addition to contemporary interactions among populations, patterns of differentiation may also be explained by the historical processes associated with the establishment of those populations (e.g., long distance dispersal colonization or vicariance events; Avise et al. 1987; Lomolino et al. 2006; Schaal et al. 1998). The likelihood of those historical events, and the time over which they may have occurred, is strongly dependent on past environmental conditions. Although glacial cycles during the Pleistocene affected species on a global scale, of particular importance to North American taxa are the environmental conditions associated with the Wisconsin glaciation (maximum at about 18,000 years before present (ybp)). During that glacial episode, ice sheets reached as far south as 40º N (e.g., New York, NY) and tundra and boreal forest habitats extended even further south (Hewitt 2000; Lomolino et al. 2006). As a result of these historical conditions that rendered many northern areas inhabitable, species or populations may have migrated relatively recently into northern regions where they currently exist (i.e., the leading edge hypothesis; Cwynar & Macdonald 1987; Hewitt 1996; Soltis et al. 1997). An alternative phylogeographic hypothesis is that isolated areas of suitable habitat remained in otherwise glaciated regions and allowed northern populations to persist throughout the glacial cycles of the Pleistocene (i.e., north-south recolonization hypothesis; Soltis et al. 1997).

Given that different genetic signatures are expected to result from the leading edge and north-south recolonization hypotheses, molecular markers provide a means of determining which is most probable (Comes & Kadereit 1998; Cwynar & Macdonald 1987; Hewitt 2000; Petit et al. 2002). For example, following a south to north range

expansion (i.e., leading edge hypothesis), levels of genetic diversity and heterozygosity are expected to be greatest in the southern portion of a species' range due to northern populations being established by the long distance dispersal of a few individuals. As a consequence of northward dispersal events, phylogenetic reconstructions would show northern populations being successively sister to southern populations (Nason et al. 2002). In contrast, under the north-south recolonization hypothesis recent gene flow among populations that had been isolated during glaciation would result in genetic diversity not being correlated with latitude and populations would show a genetic signature similar to that of hybridization or secondary contact (e.g., Hewitt 2001). The creation and persistence of these historical signatures within molecular variation is determined by the magnitude of ongoing genetic drift, gene flow, and natural selection (Comes & Kadereit 1998).

In addition to historical processes, scenarios under which microevolutionary forces acting over a relatively short time (e.g., hundreds rather than thousands of years) may also explain observed patterns of genetic diversity and differentiation. The rationale for these scenarios is that landscape changes in the eastern United States over the past 2-3 centuries as a result of human activities are well documented (Foster et al. 2002; Hall et al. 2002; Russell et al. 1993) such that anthropogenic actions have likely had a substantial impact on the evolutionary dynamics and population genetic structure of species (e.g., Carroll & Fox 2008; e.g., Crutzen 2002). If anthropogenic induced fragmentation and habitat destruction were more severe in the northern portions of a species' range relative to the southern regions a similar genetic pattern to that of the leading-edge hypothesis could result. We refer to this as the anthropogenic hypothesis.

Historical processes associated with each of the alternative hypotheses may also affect morphological variation in predictable ways (e.g., Cwynar & Macdonald 1987). The degree to which phenotypic and genetic patterns of differentiation are congruent with the same phylogeographic hypothesis and correlated with one another may be indicative of the relative roles of genetic drift, gene flow, and natural selection in determining levels of phenotypic differentiation (Streisfeld & Kohn 2005). If variation at putatively neutral molecular markers and phenotypic traits both show similar degrees of isolation by distance then perhaps the latter is also determined by neutral processes (Gould & Johnston 1972). Although evaluating whether phenotypic isolation by distance is the result of natural selection associated with a latitudinal cline is difficult (e.g., Mitchell-Olds & Schmitt 2006), partial mantel tests among geographic, genetic, and morphological distances provide a potential method for differentiating between genetic drift and natural selection. For example, after accounting for differentiation at neutral loci, a significant correlation between morphological differentiation and geographic distance may be indicative of a latitudinal environmental effect (e.g., Storz 2002). Alternatively, if neutral molecular variation exhibits a pattern of isolation by distance, which is strongly discordant with patterns depicted based on phenotypic variation, there may be evidence for non-neutral processes acting on those quantitative traits. However, there is much debate regarding the interpretation of the magnitude of differentiation between molecular (F_{ST}) and quantitative trait (Q_{ST}) data assayed among wild populations due to the inability to control for an environmental effect on phenotypic variation (Hendry 2002; McKay & Latta 2002; Pujol et al. 2008).

Studies analyzing both phenotypic and neutral genetic variation can also quantify the degree of genetic and ecological exchangeability among different populations, which has been advocated as a means of prioritizing units for conservation (e.g., Crandall et al. 2000). Importantly, ecological exchangeability is rejected if differences among populations in phenotypic traits exist, regardless of whether those differences are the result of natural selection or genetic drift (Crandall et al. 2000). Short-term declines in fitness due to inbreeding depression are often a greater conservation concern for the persistence of a species than long-term erosion of genetic diversity as a result of genetic drift (Amos & Balmford 2001); evaluating the plausibility of historical versus recent phylogeographic hypotheses can inform conservation managers which of those temporally dependent conservation issues is at play. Furthermore, the conservation value of peripheral populations is often debated in that they should be preserved because they are usually genetically and morphological distinct (Channell & Lomolino 2000; Lesica & Allendorf 1995) or should not be a conservation priority because habitat at the margins of a species range may be sub-optimal and therefore conservation efforts should focus on preserving populations at the center of a species' distribution (Griffith et al. 1989). Again, phylogeographic studies can help to determine whether in fact peripheral populations harbor unique genetic diversity and are worthy of conservation efforts (Eckert et al. 2008).

Our primary objective was to determine which of three alternative phylogeographic hypotheses best explains the founding of 14 populations of *Agalinis decemloba*, which represents what has been historically identified as three separate species (i.e., *A. acuta*, *A. decemloba*, and *A. tenella*). The conclusion reached from a comprehensive study

evaluating the evolutionary distinctiveness of these putative plant species was that the levels of differentiation among individuals best represent those found within a single species; the most parsimonious intraspecific taxonomic alignment is that there are two subspecies (Pettengill and Neel, *in prep.*). One subspecies, *Agalinis decemloba* ssp. *tenella*, includes individuals that were historically known as *Agalinis tenella*. The other subspecies is *Agalinis decemloba* ssp. *decemloba* that includes individuals previously ascribed to *A. acuta* and *A. decemloba*. *Agalinis acuta* is currently listed as federally endangered under the Endangered Species Act (ESA; 1973); it was listed in 1987 due to conversion of its grassland habitat to agricultural, residential, industrial, and commercial development; destruction of habitat was also cited as the most serious threat to the persistence of the species (U. S. Fish and Wildlife Service 1988). Based on those reasons for listing the species, it is particularly relevant to differentiate between the anthropogenic and more historical phylogeographic hypotheses in explaining extant patterns of genetic diversity.

Although chloroplast DNA (cpDNA) loci have been the dominant markers in plant phylogeographic studies (Comes & Kadereit 1998), in this study we use microsatellite loci because even the most variable regions within the chloroplast genome lack sufficient variation to discern intraspecific relationships within *Agalinis* species (Pettengill & Neel 2008). To differentiate among the alternative hypotheses we determined which populations had high genetic diversity and, therefore, represent refugial populations. We also tested for congruence between cladograms based on variation at microsatellite loci and phenotypic traits with a cladogram representing the leading edge hypothesis (e.g., Nason et al. 2002). We conducted forward in time simulations of the leading edge,

recolonization, and anthropogenic hypotheses to help determine the most likely phylogeographic scenario responsible for the patterns we observed. Furthermore, to determine whether the observed levels of divergence are the product of historical rather than contemporaneous events, we conducted analyses capable of detecting recent evolutionary events (i.e., migration and/or population bottlenecks). We also investigated the congruence between patterns at microsatellite loci and morphological traits as a means to distinguish between neutral and adaptive processes in explaining phenotypic differences.

MATERIALS AND METHODS

Study system

Agalinis decemloba is a fall blooming annual species where subspecies *decemloba* is native to eastern North America and found on the coastal plain in eastern Massachusetts; Rhode Island; Connecticut; and New York; the piedmont in Maryland; southern Virginia; and central North Carolina (Fig. 4.1). *Agalinis decemloba* ssp. *tenella* is native to the southeastern United States (e.g., South Carolina, Georgia, and Florida) (Fig. 4.1). The species is found in sandplain grassland habitats and in openings in pine–oak forest on nutrient poor soils (e.g., glacial outwash or serpentine). Large fluctuations in population size have been documented; censuses in the two naturally occurring populations on Cape Cod between 1980 and 2000 ranged from 1 to 4253 plants at one site and from 0 to 3674 plants at the other (Neel & Somers 2001). *A. decemloba* lacks an obvious mechanism for long distance dispersal in that seeds are primarily gravity dispersed. The primary pollinators documented among northeastern populations are Bumble Bees (i.e., *Bombus* sp.); other pollinators include members of the Syrphidae (Order Diptera) (Neel 2002).

The degree of self-compatibility within the species is extremely high with 97% of selfed flowers setting fruit, however, fruits resulting from selfed flowers exhibited a 17-20% reduction in seed set (Neel 2002). Geitonogamy (i.e., the transfer of pollen to different flowers on the same individual) and autonomous selfing late in anthesis as corollas dehisce (i.e., 'corolla dragging;' Dole 1990) are likely mechanisms through which selfing naturally occurs.

DNA extraction and microsatellite genotyping

Tissue was collected from 386 individuals representing 14 populations during the fall blooming seasons of 2005-2007 (Table 4.1 & Fig. 4.1). Where possible, we sampled 30 individuals per population and values less than that reflect the total number of individuals we could locate (Table 4.1). Total genomic DNA from all samples was isolated from fresh or frozen (-80°C) leaves and flower buds by grinding 50-100 mg of tissue to powder in liquid nitrogen with a mortar and pestle, and then using GenElute Plant Genomic DNA Kits (Sigma) or QIAGEN DNEasy Kits (QIAGEN) following manufacturer instructions. Extractions were also carried out on a QIAGEN BioSprint 96 robotic workstation using QIAGEN's BioSprint 96 DNA Plant Kit.

All samples were assayed for variation at 20 microsatellite loci. The development, characterization, and amplification protocol of these loci are described in Appendix B (Pettengill et al. 2009).

Morphological measurements

We also measured 61 morphological traits per individual (Appendix C). The traits included those related to the size of the plant, vegetative features such as leaf length, and a suite of floral characters that likely affect pollination. Because we collected throughout

the season, not all individuals were in bloom and some samples for which tissue could be

acquired for molecular analyses were not assayed for all morphological characteristics.

For comparisons with genetic differentiation we would have preferred to conduct

common garden experiments to control for environmental effects, however, due to

difficulties in growing a sufficient number of individuals in a controlled environment of

what was historically known as *A. acuta* (pers. obs.; Brumback & Kelley 1990), we are

limited to inferences made based on phenotypic traits assayed in the wild.

Genetic diversity analysis

For each population we calculated the proportion of polymorphic loci (P), number of

alleles (A), number of private alleles (A_P), and the observed (H_o) and expected (H_e)

heterozygosity using the program POPGENE v1.32 (Yeh et al. 1997); allelic richness

(A_R) was estimated in FSTATv2.9.3.2 (Goudet 1995) using a rarefaction method to

account for differences in sample size. The inbreeding coefficient (R_{IS}) and pairwise R_{ST}

values were estimated with GENEPOP v4.0 (Raymond & Rousset 1995). Pairwise

estimates of Jost's D_{EST} were also calculated using the program SMOGD (Software for

the Measurement of Genetic Diversity; Crawford submitted). Jost's D_{EST} is argued to be

a more accurate measure of population differentiation than those used previously (e.g.,

G_{ST} and F_{ST}) in that it does not produce erroneous results that imply low levels of

differentiation among populations with few or no alleles in common or underestimate the

degree of differentiation when hypervariable loci (e.g., microsatellites) are used (Heller &

Siegismund 2009; Jost 2008).

We used the program DISPAN (Ota 1993) to construct neighbor-joining trees

(Saitou & Nei 1987) based on D_a distances (Nei et al. 1983) among the 14 populations.

Significance of the relationships depicted on the topology was assessed through 1000

bootstrap replicates conducted in DISPAN where a locus was the unit resampled with

replacement.

Morphological analyses

To avoid biasing our conclusions by analyzing highly correlated morphological

traits, we identified pairs of traits that had an $R^2 > 0.6$; of those two traits, the one with

the greater sum of R^2 values across the other traits was removed. From this procedure,

we identified 44 morphological characters that were relatively independent of one

another. To provide a measure of the distance between populations in multivariate space,

we also calculated squared Mahalanobis distances using the General Discriminant

Analysis procedure in STATISTICA v6.1 (Statsoft 2004). A phenogram based on the

pairwise Mahalanobis distances between populations was constructed in MEGAv4.0

(Tamura et al. 2007) using the UPGMA clustering algorithm (Sokal & Michener 1958).

To provide a measure of phenotypic variation that could be compared with measures

of genetic differentiation (e.g., R_{ST}), we first performed two-way ANOVAs among all

pairs of populations (population was a fixed factor) for each of the 44 morphological

traits to obtain within and among population variance components of phenotypic

differentiation. We then calculated the average of the within (σ_w^2) and between (σ_b^2)

population variance components for all 44 traits, which were used to estimate P_{ST} as

$\sigma_b^2/(\sigma_b^2 + 2\sigma_w^2)$ (e.g., Saether et al. 2007; Storz 2002). P_{ST} is a measure of phenotypic

differentiation that can be interpreted as an analog of R_{ST} assuming that populations are in

migration drift-equilibrium and that phenotypic differentiation is due to additive genetic

effects (Lynch & Walsh 1998; Spitze 1993). Because we did not control for the effect of

environment on quantitative traits, we refer to this measure of differentiation as P_{ST} (e.g.,

Pujol et al. 2008) rather than Q_{ST} (Lynch & Walsh 1998; Spitze 1993). In estimating P_{ST},

we used two times the observed variance within populations as the measure of within

population phenotypic variance, which corresponds to a narrow sense heritability of 0.5.

After accounting for the divergence due to drift (i.e., the level of differentiation

inferred with microsatellite loci), the degree to which phenotypic divergence is correlated

with geographic distance may provide an indirect estimate of the potential effects of a

latitudinal gradient on phenotypic variation. We used the program *Le Progiciel R*

(Legendre & Vaudor 1991) to conduct a partial Mantel test (Smouse et al. 1986) among

the three different matrices (i.e., differentiation based on P_{ST} (**P**), genetic differentiation

based on R_{ST} (**G**), and geographic distance (**D**)). Specifically, we calculated the

correlation between a matrix of the residuals, **P'**, from regressing **P** on **G**, and geographic

distances **D**. The correlation was deemed significant if our observed correlation was

greater than 95% of those calculated from 9,999 permutations.

Phylogeographic analyses

We conducted regression analyses of A_R and H_e on the latitude of each population to

determine the statistical support for the expectations under the leading-edge hypothesis

that higher levels of genetic diversity would be present in southern populations than more

northern populations.

We used Mantel tests, performed using GENEPOPv4.0, to determine if pairwise

genetic differentiation (R_{ST}) or pairwise phenotypic differentiation (P_{ST}) was positively

correlated with geographic proximity calculated as the natural log of Euclidean distance

in kilometers among populations (i.e., isolation by distance). We also tested the

significance of the correlation between those two measures of differentiation.
Significance was assessed through 10,000 permutations. To better understand the scale
over which either genetic or morphological differentiation were correlated with
geographic distance, we used the program GenAlExv6.0 (Peakall & Smouse 2006) to
calculate correlation coefficients at 7 distance classes representing 250km increments
according to the method described in Smouse and Peakall (1999). The minimum and
maximum number of pairs that fell into each distance class was 5 and 23, respectively;
the mean was 13 pairs. Significance was based on 9,999 permutations.

To explore the congruence of relationships based on D_a distances and Mahalanobis
distances with those expected under the range expansion hypothesis, we used MacClade
v4.06 (Maddison & Maddison 2001) to construct a cladogram representing one possible
topology expected under that hypothesis (e.g., Nason et al. 2002). The southern
populations in Georgia and South Carolina constituted a single clade and northern
populations were successively nested within one another reflecting sequential
colonization (Fig. 4.1). We then used the program GENETREE v1.3.0 (Page 1998) to
calculate a measure of discordance (i.e., the number of extinct and unsampled lineages or
deletion and duplication events necessary to reconcile two topologies) between the
hypothetical cladogram and each of our observed topologies based on D_a and
Mahalanobis distances. Significance was determined through comparing our observed
measure of discordance to the distribution of discordance values associated with 10,000
randomly generated cladograms. Significance was inferred if no more than 50
comparisons between the randomized topologies and our observed topologies had
discordance values less than our observed value ($\alpha = 0.05$) (e.g., Nason et al. 2002).

To determine the extent to which relatively recent events may be responsible for the observed patterns of diversity and population differentiation we tested for recent bottlenecks and migration events. We used the program BOTTLENECK v1.2.02 (Piry et al. 1999) to determine whether populations had recently experienced a population bottleneck. Tests were carried out under the two-phased mutational model, which is recommended for microsatellite loci (BOTTLENECK documentation) and the one-tailed (for a heterozygote excess) Wilcoxon-sign rank test ($\alpha = 0.05$ based on 1,000 permutations). Second, we used the program GENECLASS v2.0 (Piry et al. 2004) to identify whether individuals represent recent migrants into the populations from which they were sampled. Detection of first generation migrants was based on the Bayesian method of Rannala and Mountain (1997; Soltis et al. 1997) and using the Λ statistic recommended in Paetkau et al. (2004). Significance was assessed through the resampling algorithm of Paetkau et al. (2004), the simulation of 10,000 genotypes, and at $\alpha = 0.01$ (i.e., the null hypothesis that an individual germinated in the population from which it was sampled is rejected if the Λ for that individual is greater than 99% of the distribution of simulated genotypes).

Forward in time simulations

To help determine the most likely scenario responsible for our observed patterns of genetic diversity, we conducted forward in time simulations of the three alternative phylogeographic hypotheses (leading-edge, recolonization, and anthropogenic) using the program *quantiNemo* (Neuenschwander et al. 2008). We employed a metapopulation model in which the total number of patches was 14. Patches were arranged such that migrants could only disperse to the two neighboring patches (i.e., 1-dimensional stepping

stone). The parameter settings were chosen to generally represent those expected for *A. decemloba*. The number of offspring was determined by a logistic growth rate of 0.5^{-2}. The dispersal rate was set to 1.0^{-3} migrants/generation and reflective dispersal boundaries were enforced such that dispersers were not lost within border patches if they dispersed in the opposite direction of the one neighboring patch. The breeding model was set so individuals were hermaphroditic and the proportion of offspring created through selfing was 0.4. We simulated 20 microsatellite loci that could have a maximum of 20 alleles and that mutated according to the *K*-allele model at a rate of 1.0^{-4}. At the onset of the simulations, occupied patches were set to be maximally polymorphic, which is consistent with the idea that ancestral populations should be centers of high diversity (e.g., Hewitt 2000).

To simulate the leading edge hypothesis a single southern border patch was occupied and the 13 other patches were empty; only the 1st and 12th patches were initially occupied under the recolonization hypothesis. Each of these two simulations were run for 12,000 generations and the carrying capacity was 500 within all patches, which is a reasonable estimate based on documented census sizes within natural populations of *A. decemloba* (Neel & Somers 2001). Although *A. decemloba* is an annual, we chose 12,000 generations because populations likely have a seed bank, which increases the generation time. For the anthropogenic hypothesis, all populations were occupied and simulations were run for 5,000 generations. To model recent anthropogenic influences on habitat, during the final 200 generations the dispersal rate decreased to 1.0^{-4}, which represented decreasing the likelihood of migration among existing populations. There was also a linear reduction in population size such that the smallest populations had a carrying

capacity of 45 individuals. Based on exploratory analyses, the initial 4,800 generations

ensured that mutation-drift equilibrium had been established.

RESULTS

Population level polymorphism and differentiation

The proportion of polymorphic loci within each population ranged from 0.10 within

MDNY to 1.0 in both GCGA and BCGA; the mean proportion of polymorphic loci was

0.75 (Table 4.1). The number of alleles, *A*, harbored within each population averaged

88.8; levels of allelic richness when corrected for differences in sample size, A_R, were

low and averaged 2.52. The southern populations in South Carolina and Georgia showed

the highest levels of allelic diversity and MDNY possessed only 27 and 1.11 alleles for *A*

and A_R, respectively (Table 4.1). On average each population possessed an appreciable

number of private alleles ($\overline{A_P} = 11.8$), but the four most southern populations are

primarily responsible for this value given that more northern populations had values

closer to 4; the BVMA population had only one private allele and MDNY harbored no

private alleles

Observed levels of heterozygosity were consistently lower than expected

heterozygosity with some populations consisting of individuals that were nearly

homozygous across all loci. For example, H_o = 0.007, 0.039, 0.047, and 0.064 for

MDNY, SHNY, SDMD, and BVMA, respectively; estimates of H_e for all those

populations was an order of magnitude greater (Table 4.1). In contrast, southern

populations had H_o values closer to 0.5 and had no large discrepancy between H_o and H_e.

As expected based on the levels of heterozygosity, the inbreeding coefficient, R_{IS}, was

extremely high for both MDNY (R_{IS} = 0.997) and SHNY (R_{IS} = 0.831). The populations

in South Carolina and Georgia had much lower R_{IS} values (Table 4.1).

Estimates of R_{ST} were similar to the levels of differentiation based on D_{EST} (averages

across all pairs of populations were 0.628 and 0.664, respectively). Both measures

suggest a strong level of population structure. However, the two methods differ as to

which populations are most similar and different from one another. R_{ST} was smallest

between LCGA and GCGA (R_{ST} = 0.116) but D_{EST} was smallest between BCGA and

GCGA (D_{EST} = 0.128). R_{st} was largest between PCCT and CCSC (R_{ST} = 0.925) and the

D_{EST} was greatest between MDNY and BVMA (D_{EST} = 0.844) (Table 4.2).

Based on the morphological traits, squared Mahalanobis distances ranged from 15.59

between PCCT and EPRI to 255.03 PCCT and CCSC; the mean value was 96.02 (Table

4.3). However, CCSC is quite different from all populations. Tests for significant

differences in Squared Mahalanobis distances were highly significant ($P \ll 0.05$) among

all populations.

The amount of phenotypic variance within populations was significantly less than

that among populations for 62% of the 91 pairwise comparisons across all 44

morphological traits. Variance for height, pedicel, pedicel to bract ratio, and corolla

throat width were significantly greater among than within populations for over 70

pairwise comparisons (i.e., >75%; P <0.05). Calyx diameter width showed the least

amount of differentiation between populations in that only 24 of the 91 (26%)

comparisons were significant (P <0.05). After averaging all pairwise variance

components, the average pairwise estimate of P_{ST} was 0.651; P_{ST} was lowest between

BVMA and SMNY (P_{ST} = 0.412) and greatest between LCGA and EPRI (P_{ST} = 0.871)

(Table 4.2).

Phylogeographic results

We found a significant negative relationship between latitude and A_R and H_e

(P<<0.05) (Fig. 4.2). No populations showed evidence of a statistically significant

excess of heterozygosity, which would be present had the populations recently

experienced a population bottleneck (Luikart et al. 1998). The results from

GENECLASS v2.0 showed no statistical support for any of the samples being recent

immigrants into the population from which they were sampled.

Based on the Mantel test, there was a positive and statistically significant correlation

(r^2 = 0.151, P < <0.01) between pairwise R_{ST} values and geographic distances (Fig. 4.3);

the relationship between P_{ST} and geographic distances was also significant but stronger

than that observed based on genetic differentiation (r^2 = 0.416, P <<0.01; Fig. 4.4).

However, the spatial autocorrelation tests indicate that the overall correlation of both P_{ST}

and R_{ST} with geographic distance is predominantly due to relatively geographically close

populations. Pairwise R_{ST} values were only significant for populations less than 250km

or between 250-500km apart; pairwise P_{ST} values were only significantly correlated with

geographic distances for populations less than 250km apart (P <0.05). P_{ST} and R_{ST} were

also significantly positively correlated with one another (r^2 = 0.114; P <0.01; Fig. 4.3).

After removing the degree to which P_{ST} and R_{ST} were correlated (i.e., the degree of

differentiation based on putatively neutral loci), the correlation between the residuals of

P_{ST} and geographic distance was still positive and statistically significant (r^2 = 0.420; P

<< 0.01).

Our observed measure of discordance between the neighbor-joining tree based on D_a distances and the hypothetical cladogram representing the leading edge hypothesis was 21 (i.e., 4 deletion and 17 insertion events were necessary for the two cladograms to be identical; Fig. 4.1). The mean level of discordance between the topology based on D_a distances and 10,000 randomly generated topologies, which can be interpreted as 10,000 randomly generated phylogeographic hypotheses, was 47.96 and the minimum value was 25 and the maximum was 73; there is statistically significant support that the topology based on D_a distances is concordant with a topology reflecting our phylogeographic hypothesis in that no random topologies had discordance values less than our observed value ($P < 0.001$). The level of discordance between the topology based on squared Mahalanobis distances and the hypothetical topology was also significant in that only 10 of the 10,000 random topologies had discordance values less than our observed value of 38 additional events ($\bar{x} = 56.78$, min = 31 and max = 87; $P < 0.01$).

Simulations

The simulation of the leading edge hypothesis showed the expected decrease in allelic diversity and observed heterozygosity as the distance from the initial occupied patch (e.g., refugial population) increased (Fig. 4.4), which was most similar to our observed results. The observed heterozygosities of the initial patch and neighboring patches were greatest than the other patches but smaller than those we observed among the southern populations, ~0.5 and 0.6, respectively; the farthest populations from the initial occupied patch had levels of heterozygosity similar to those found in the northern populations of *A. decemloba* (i.e., < 0.2; Fig. 4.4). Levels of allelic diversity under the leading edge scenario were also similar to observed values in *A. decemloba* (Fig. 4.4).

The results of the simulations of the anthropogenic hypothesis were also similar to our observed results. A population bottleneck for the final 200 generations of the simulation resulted in decreased H_o and allelic diversity in the 'northern' patches (Fig. 4.4). Based on the parameters we used, this implies that 200 generations of increased isolation and bottlenecking of populations that were previously large and founded from a large number of individuals is a sufficient amount of time for a pattern indicative of a range expansion to arise.

The results of the recolonization simulation showed the expected pattern of little difference in allelic diversity and heterozygosity among border patches (Fig. 4.4). Although not substantially lower, the projected reduction in those genetic measures among populations in the middle of the distribution was observed (Fig. 4.4). Consequently, the patterns resulting from this simulation are quite different than those we observed among *A. decemloba* populations.

DISCUSSION

Historical range expansions and contractions as a result of glacial cycles during the Pleistocene have likely had a profound effect on the geographic distribution and patterns of differentiation among conspecific populations (Hewitt 2000). Consequently, in addition to contemporaneous events, the process by which populations were established may be a prevailing explanation for extant patterns of relatedness. Based on the analysis of 20 microsatellite loci and morphological traits, we found that the distribution and patterns of relatedness among of 14 populations of *Agalinis decemloba* closely matched the expectations under a unidirectional range expansion (i.e., northern emigration events out of southern refugial populations). The observed patterns may have also arisen as a

result of recent anthropogenic events that caused the extirpation of populations and population genetic bottlenecks. We found that it is highly unlikely that *A. decemloba* populations existed in northern latitudes during the Wisconsin glaciations (i.e., north-south recolonization hypothesis; Soltis et al. 1997).

In accordance with the leading-edge phylogeographic hypothesis, we found that allelic diversity and heterozygosity within populations of *A. decemloba* decreased with increasing latitude (Fig. 4.2). The neighbor-joining trees inferred with D_a distances and phenetic similarity were statistically compatible with the cladogram depicting the colonization route expected under the leading edge hypothesis (Fig. 4.1). There was evidence that morphological variation, in addition to genetic diversity, decreased with increasing latitude (i.e., a negative relationship between the average standard error across all morphological traits within each population and latitude ($r^2 = 0.163$)); however, this relationship was not significant ($P = 0.152$).

Within eastern North America, the southeast has been characterized as having a high number of endemic species and, consequently, likely represents a refugium for species during periods when more northern regions were inhabitable (Estill & Cruzan 2001; Sorrie & Weakley 2001). In addition to our results, other studies support this hypothesis. For example, allozyme variation in northern populations of *Asclepias exaltata* (Asclepiadaceae) was significantly less than that found in the southern putatively refugial populations in southern Appalachian mountains (Broyles 1998). A comparative study of 11 species of *Polygonella* (Polygonaceae) found the unexpected pattern of higher levels of allozyme diversity within the more narrowly distributed southern species relative to the species in the north with larger geographic ranges (Lewis & Crawford 1995). As a

means of resolving this paradox, the authors state that northern species were recently

established as a result of dispersal out of and subsequent reproductive isolation from

southern species located in Florida; establishment from only a few individuals explains

the unexpected pattern of lower allelic diversity within the more widespread *Polygonella*

species (Lewis & Crawford 1995). However, northern populations of the herbaceous

perennial *Trillium grandiflorum* (Melanthiaceae) did not show a significant decrease in

allozyme diversity relative southern populations (Griffin & Barrett 2004). The authors

do not suggest that northern populations existed during glaciations but that larger

effective population sizes in northern populations maintained high levels of diversity

(Griffin & Barrett 2004).

Alternatively, studies of species in eastern North America have found that northern

populations may have persisted throughout the last glacial maximum in putatively

glaciated regions. For example, northern populations of both the red maple (*Acer*

rubrum) and beech (*Fagus granifolia*) tree species in eastern North America do not have

substantially reduced haplotype diversity relative to southern populations (McLachlan et

al. 2005). Two distinct northern and southern cpDNA haplotypes are also found within

Liriodendron tulipifera leading to the conclusion that two geographically isolated refugial

populations existed during the glacial cycles of the Pleistocene (Sewell et al. 1996).

However, there is also evidence that tree species (e.g., spruce *Picea* spp.) were forced out

of northern regions during the last glacial advance and subsequently migrated north

following deglaciation events (Davis & Shaw 2001).

Recent anthropogenic effects

Another alternative phylogeographic hypothesis we evaluated invokes more contemporaneous events (i.e., the anthropogenic hypotheses) and begins with the assumption that historically a greater number of populations existed and that populations were larger than that of today. Fragmentation or extirpation events resulting from anthropogenic activity over the past 200-300 years would have increased the isolation and reduced population sizes causing a corresponding increase in population differentiation and decrease in allelic diversity. The putative extirpation of multiple populations (e.g., the type locality from Washington DC and Pennsylvania populations of *Agalinis decemloba*; loss of the *A. acuta* type locality from Martha's Vineyard, MA; and the fact that 52 populations were known historically but only 12 natural occurrences existed upon listing *A. acuta* (Suckling 2006)) provides strong evidence that fewer northern populations exist today than were present during the early 20th century. Additionally, the height of deforestation of many parts of the northeastern United States occurred during the early to late 1800s where approximately 80% of land was unforrested (Foster 1992; Motzkin et al. 1999). Estimates of the degree of loss and degradation of habitat show that in addition to California, the Southern, Midwestern, and Northeastern regions all have been extensively impacted (Noss et al. 1995); over the past few decades the southern United States also has had the highest rate of conversion of rural lands to urban uses (Alig et al. 2004).

Our own experience trying to locate populations provides evidence that, in addition to the northeast, habitat loss and fragmentation is frequent in southeast. For example, across multiple years of sampling, we were not able to locate 22 of 30 historically

documented populations visited in Virginia, North Carolina, South Carolina and Georgia.

Seasonal fluctuations in flowering time and a severe drought during 2007 suggest that

populations may still exist and could explain why we could not locate some populations.

We also found that many sites had either been converted to commercial or residential

properties, been subjected to disturbances associated with construction for the widening

of a road, or the habitat had transitioned to a later successional stage, all of which would

facilitate the extirpation of those populations. However, based on personal observations,

northern populations in New England are more frequently found in small areas of suitable

habitat that likely restricts population sizes relative to more southern populations. As a

result, there is evidence that populations have been extirpated throughout the species

range but smaller population sizes are more frequent in the northern populations relative

to those in the south.

Despite the evidence that anthropogenic activities have impacted *A. decemloba*,

many empirical studies have found that there has been an insufficient amount of time

since anthropogenically induced changes to the landscape occurred for the manifestation

of the expected genetic consequences. A study of the endangered grassland species,

Globularia bisnagarica, using AFLP data found little evidence that the increased

fragmentation and reduction of Belgian grasslands since the late 16th century has

significantly eroded genetic similarity or allelic diversity likely (Honnay et al. 2007).

Based on variation at inter-simple sequence repeat (ISSR) markers, a study of *Viola*

pubescens found that increased urbanization over the last century had not substantially

reduced genetic diversity (Culley et al. 2007). The timescale over which

anthropogenicaly induced deforestation occurred within Madagascar (i.e., over the past

200-300 years), was also found to be insufficient to explain the levels of diversity and connectivity observed among populations of the tree species *Dalbergia monticola* (Andrianoelina et al. 2009). However, a generation time of approximately 50 years for *Dalbergia monticola* makes it less likely to exhibit patterns associated with relatively recent fragmentation and bottlenecking events. A phylogeographic investigation of *Trillium reliquum* also found that recent habitat fragmentation due to anthropogenic development would have been too recent, and that populations would therefore still maintain a signature associated with shared evolutionary history (Gonzales & Hamrick 2005).

In contrast, the results of our simulations show that relatively recent anthropogenic events can in fact have a significant effect on population genetic patterns. Simulated populations subjected to a reduction in size for 200 generations had low levels of allelic diversity and heterozygosity that closely resembled the leading-edge hypothesis and our observed patterns (Fig. 4.4). However, the magnitude of change between the southern and northern populations of *A. decemloba* in observed heterozygosity is substantially greater than observed among the different simulations. A potential explanation for this discrepancy is that southern *A. decemloba* populations, in addition to being larger, also outcross more frequently than northern populations; we did not vary selfing rate within the simulations. Assuming that northern populations of *A. decemloba* are disproportionately affected by human activities, results from the simulation of the anthropogenic hypothesis show that recent events can produce a pattern indicative of a unidirectional range expansion. Consequently, it is likely that the magnitude of differences we observed may have primarily resulted from the establishment of

populations, but those differences have also been facilitated by evolutionary processes acting after establishment (e.g., Broyles 1998; e.g., Griffin & Barrett 2004).

Dispersal dynamics

An important component to the range expansion hypothesis is determining the mechanism responsible for the establishment of northern populations. Reid's paradox states that a species' assumed dispersal potential is often insufficient to account for the rapid range expansion following deglaciation and/or geographic distances found among plant populations in general (Clark et al. 1998; Reid 1899). Since Reid's study on the distribution of Oaks in Britain (Reid 1899) many other studies have also documented the existence of a paradox. For example, a study investigating *Daviesia triflora* (Papilionaceae) found that the primary dispersal mechanism of ants could not account for source-sink dynamics observed among populations in that the average dispersal distance of ants was two to three orders of magnitude too short (He et al. 2009). Another study on the ant-dispersed *Asarum canadense* (Aristolochiacae) also found that the distribution of populations could not be explained by the putative primary dispersal mechanism (Cain et al. 1998). As such, those two studies suggest that long distance dispersal events through either unknown or less likely mechanisms (e.g., rare wind or animal mediated dispersal) are necessary to resolve the observed paradox. Simulation studies show that Reid's paradox can also be resolved by increasing the probability of long distance dispersal events through modeling a dispersal kernel with longer tails (Ibrahim et al. 1996). Dispersal potential has also been modeled, and empirically validated, to be a dynamic process in that the availability of suitable habitat associated with, for example, the

receding glaciers could drive the evolution of increased dispersal distance (e.g., spatial selection; Phillips et al. 2008).

Seeds of *A. decemloba* are primarily gravity dispersed and thought to stay within a 3.5-meter radius of the parental plant, which also reflects a minimum potential annual dispersal distance (NatureServe 2007). Since the recession of the glaciers, the species could have only have dispersed tens of kilometers; thus, we face a paradox given the geographic range among populations relative to the dispersal potential of *A. decemloba*. Wind, which is more plausible for taller trees than low-growing woodland or grassland species (Pakeman 2001), is highly unlikely to have served as a long-distance dispersal mechanisms for *A. decemloba* seeds. Endozoochory, the ingestion of seeds and subsequent dispersal after passing through the gut of an animal, provides an additional mechanism that could facilitate the long distance dispersal of low-growing plant species. It is documented that rabbits can explain the recolonization of disturbed grassland habitat (Pakeman et al. 1999); deer have also been shown to disperse seeds of *Trillium grandiflorum* > 3km in the northeastern United States (Vellend et al. 2003). Given that grazing by small animals such as meadow voles *(Microtus pennsylvanica)* or rabbits (e.g., *Sylvilagus* spp.) has been observed among populations of *A. decemloba* (NatureServe 2007; Neel 2002), endozoochory represents a probable mechanism through which post-glacial long distance dispersal events could have occurred.

Molecular and phenotypic evolution

Statements about natural selection are increasingly being made based on the magnitude of differences observed between phenotypic traits (P_{ST}, which is analogous to Q_{ST}) and neutral genetic markers (F_{ST}) assayed from wild populations (Gay et al. 2009;

Raeymaekers et al. 2007; Saether et al. 2007; Storz 2002). However, due to the inability

to differentiate between differences in phenotypic traits that are due to additive genetic

variance versus environmentally induced phenotypic plasticity (e.g., Pujol et al. 2008),

we agree that caution must be exercised in making statements about natural selection

when morphological traits are assayed among natural populations. Our average estimates

of phenotypic and genetic differentiation were similar (i.e., 0.628 and 0.615, respectively;

Table 4.2), which could be interpreted to mean that neutral processes are primarily

responsible for divergence at phenotypic traits (e.g., Hendry 2002). However, rather than

focusing on the magnitude of differences, we discuss the spatial autocorrelation of the

phenotypic and genetic differences and the correlation between them as a means to make

statements about the relative roles of genetic drift and natural selection.

Differentiation based on both microsatellite and morphological variation showed a

pattern indicative of isolation by distance (Fig. 4.3). Given the putative neutrality of the

microsatellite loci, this pattern is likely the result of restricted gene flow and the influence

of genetic drift, which have facilitated divergence (Slatkin 1993). However, the

coefficient of determination is relatively low between differentiation at neutral genetic

markers and geographic distance ($r^2 = 0.151$) and the correlation is only significant over a

short distance (e.g., <500km) relative to the distribution of the populations; therefore,

genetic drift is likely the dominant force and populations are not in migration-drift

equilibrium (e.g., Hutchison & Templeton 1999). The strength of the isolation by

distance pattern observed with phenotypic traits was much stronger ($r^2 = 0.416$), which

could be the result of either neutral processes (e.g., Merila & Crnokrak 2001; e.g.,

Mitchell-Olds & Schmitt 2006) or an adaptive response as a result of a latitudinal

gradient (e.g., Claussen et al. 1940). Examples of the latter include the legume *Chamaecrista fasciculata*, in which reciprocal transplant experiments of populations spanning a large latitudinal gradient in the central United States clearly showed that populations had adapted to their local environments (Etterson 2004a); populations at lower latitudes were also found to have lower heritabilities and, therefore, may be more susceptible to extirpation due to the inability to adapt to changing environments (Etterson 2004b). Latitudinal clines associated with adaptive differences have also been found among European populations of *Arabidopsis thaliana* and changes in flowering time have been identified as key drivers of diversification (Mitchell-Olds & Schmitt 2006).

The partial Mantel test, which was employed as a means to differentiate between adaptive and neutral explanations for phenotypic differentiation, indicated that the strong correlation between P_{ST} and geographic distance remained after accounting for the degree of differentiation expected due to genetic drift (e.g., that observed based on R_{ST}). Such a result has been interpreted as evidence for natural selection due to spatial differences in that a latitudinal cline causes local adaptation and subsequent greater phenotypic differentiation than that expected as a result of neutral processes (Storz 2002). However, although the correlation between P_{ST} and R_{ST} may be significant, that relationship is relatively week ($r^2 = 0.114$; Fig. 4.3) such that genetic differentiation is not a good indicator of phenotypic differentiation. As a result there is evidence that the two are either not being controlled by the same evolutionary force or are influenced differently by the same force. Because we did not control for environmental differences we are not able to distinguish between natural selection or phenotypic plasticity in explaining patterns of phenotypic differentiation. Unfortunately, due to low germination rates in a common

garden environment, the additional studies necessary to adequately address the evolutionary forces responsible for the phenotypic pattern we observed among *A. decemloba* populations may not be possible.

Conclusions

This study is one of only a few that explicitly considers the influence of historical geologic and recent anthropogenic events on shaping extant patterns of diversity of herbaceous species with ranges spanning both glaciated and unglaciated regions of eastern North America (e.g., Griffin & Barrett 2004). Although this investigation was enriched by analyzing morphological data, difficulties in differentiating between neutral processes and natural selection as explanations for the phenotypic differences we observed further illustrates the problems with sampling phenotypic variation in the wild (e.g., Pujol et al. 2008). Given the degree of concordance between the expected patterns associated with the leading edge hypothesis and those observed through the analysis of molecular and phenotypic variation it appears that northward colonization events following deglaciation during the Pleistocene played a predominant role in shaping patterns of differentiation among extant populations of *Agalinis decemloba*. In addition to historical processes, anthropogenically induced extirpations and decreases in population sizes may have contributed to the observed patterns and levels of genetic diversity, which suggests that recent human activities may be significantly altering the evolutionary dynamics of extant taxa.

Table 4.1. Population genetic characteristics of species and populations based on 20 microsatellite loci across 14 *Agalinis* populations. Collection locality; latitude and longitude; sample sizes, N and in parentheses the number of samples for which morphological data was acquired; proportion of polymorphic loci, P; number of alleles, A, N_A; number of private alleles, A_P; allelic richness as calculated in FSTAT using a rarefaction method, A_R; observed (H_o) and expected (H_E) heterozygosity; and, fixation index, R_{IS}.

Population	Collection locality	Latitude	Longitude	N (morphology)	P	A	A_R	A_P	H_o	H_e	R_{IS}
A. decemloba ssp. *decemloba*											
AACPEMA	Barnstable Co., MA	41.687	-70.456	29 (14)	0.65	45	1.67	4	0.072	0.227	0.372
AACBVMA	Barnstable Co., MA	41.582	-70.527	29 (27)	0.7	45	1.58	1	0.064	0.191	0.528
AACPCCT	Windham Co., CT	41.687	-71.923	29 (30)	0.7	52	1.82	9	0.167	0.293	0.730
AACEPRI	Washington Co., RI	41.528	-71.577	24 (29)	0.85	70	2.30	5	0.165	0.414	0.630
AACMDNY	Suffolk Co., NY	41.052	-71.939	30 (14)	0.1	27	1.11	0	0.007	0.029	0.997
AACSHNY	Suffolk Co., NY	41.039	-71.927	24 (10)	0.45	31	1.21	4	0.039	0.062	0.831
AACHPNY	Nassau Co., NY	40.734	-73.586	31 (28)	0.65	47	1.85	5	0.137	0.295	0.669
AACSDMD	Baltimore Co., MD	39.417	-76.830	30 (30)	0.6	46	1.79	11	0.045	0.291	0.642
ADEL1VA	Lunenberg Co. VA	36.972	-78.358	24 (8)	0.95	80	2.55	9	0.168	0.490	0.569
ADEWCNC	Wake Co. NC	35.877	-78.766	33 (29)	0.9	87	2.42	7	0.237	0.398	0.525
A. decemloba ssp. *tenella*											
ATECCSC	Colleton Co. SC	32.934	-80.623	12 (12)	0.95	108	3.65	13	0.518	0.648	0.305
ATELCGA	Lowndes Co. GA	30.796	-83.381	30 (30)	0.95	230	4.56	50	0.579	0.709	0.184
ATEGCGA	Grady Co., GA	30.751	-84.086	30 (30)	1	207	4.54	25	0.523	0.739	0.247
ATEBCGA	Brooks Co, GA	30.713	-83.490	31 (6)	1	168	4.16	22	0.557	0.708	0.182
Mean across populations				27.57 (21.21)	0.75	88. 8	2.52	11.8	0.234	0.392	0.529

Table 4.2. Measures of pairwise differentiation among 14 populations of *A. decemloba*. Populations are arranged in order of decreasing latitude; above and below the diagonal are P_{ST} and R_{ST} values, respectively.

	Population													
	PCMA	BVMA	PCCT	EPRI	MDNY	SHNY	HPNY	SDMD	L1VA	WCNC	CCSC	LCGA	GCGA	BCGA
PCMA		0.439	0.555	0.590	0.505	0.451	0.612	0.616	0.521	0.634	0.704	0.801	0.777	0.715
BVMA	0.397		0.635	0.518	0.564	0.411	0.630	0.629	0.488	0.736	0.740	0.829	0.821	0.681
PCCT	0.526	0.541		0.459	0.658	0.528	0.683	0.778	0.565	0.783	0.742	0.834	0.796	0.649
EPRI	0.519	0.547	0.660		0.624	0.552	0.571	0.725	0.521	0.785	0.810	0.871	0.823	0.689
MDNY	0.648	0.679	0.869	0.454		0.587	0.558	0.549	0.608	0.582	0.663	0.723	0.733	0.638
SHNY	0.602	0.638	0.713	0.538	0.516		0.581	0.623	0.441	0.599	0.645	0.759	0.766	0.661
HPNY	0.502	0.565	0.514	0.509	0.540	0.525		0.623	0.663	0.754	0.686	0.825	0.774	0.677
SDMD	0.839	0.842	0.831	0.736	0.894	0.848	0.754		0.654	0.584	0.670	0.707	0.766	0.603
L1VA	0.372	0.220	0.283	0.596	0.637	0.454	0.367	0.620		0.635	0.687	0.748	0.727	0.669
WCNC	0.573	0.684	0.657	0.236	0.439	0.555	0.371	0.734	0.503		0.713	0.668	0.719	0.589
CCSC	0.893	0.915	0.925	0.596	0.888	0.846	0.812	0.897	0.796	0.584		0.721	0.719	0.518
LCGA	0.804	0.840	0.831	0.398	0.677	0.703	0.722	0.833	0.740	0.410	0.340		0.497	0.444
GCGA	0.795	0.845	0.837	0.376	0.703	0.728	0.704	0.835	0.748	0.353	0.262	0.116		0.429
BCGA	0.830	0.874	0.872	0.501	0.799	0.802	0.750	0.857	0.779	0.458	0.304	0.274	0.352	

Figure 4.1. Location of the 14 populations analyzed, A; fictional cladogram corresponding to the phylogeographic hypothesis being tested, B; the neighbor-joining tree based on D_a distances calculated from 20 microsatellite loci (numbers represent percentage of 1000 bootstrap replicates a given clade was present and branches are in D_a units), C; and a UPGMA phenogram based on squared Mahalanobis distances among populations.
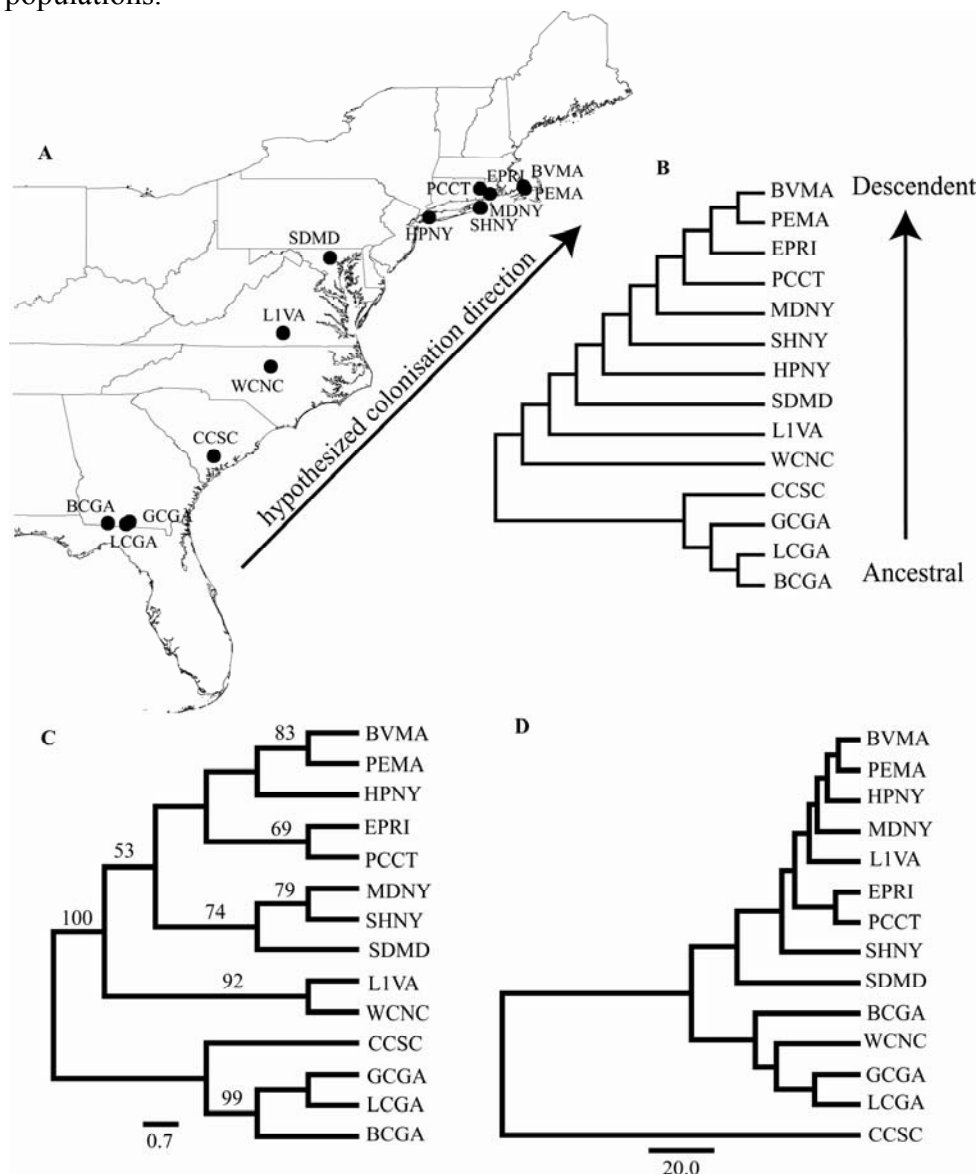
Figure 4.2. Regressions of genetic diversity on latitude for 14 populations of *A. decemloba*. (A) allelic richness, A_R; (B) Hardy-Weinberg expected heterozygosity, H_e.

A.

B.
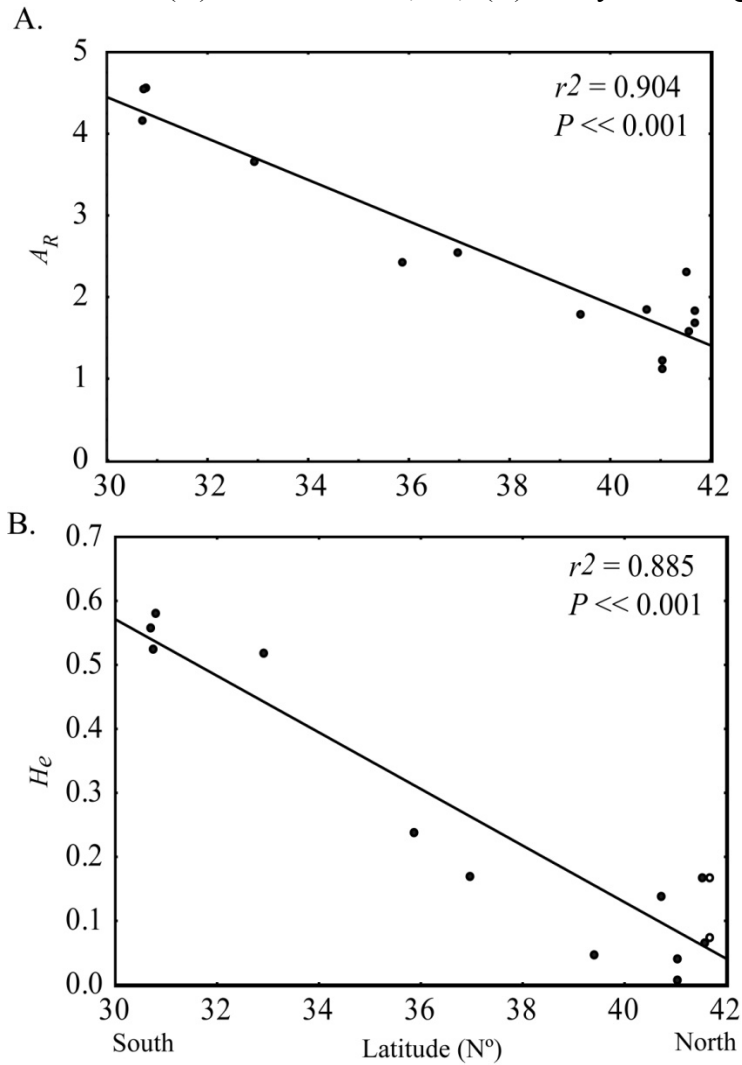
Figure 4.3. (A) Patterns of isolation by distance based on genetic (R_{ST}) and phenotypic (P_{ST}) differentiation and the correlation between those two measures of differentiation (B).

A.
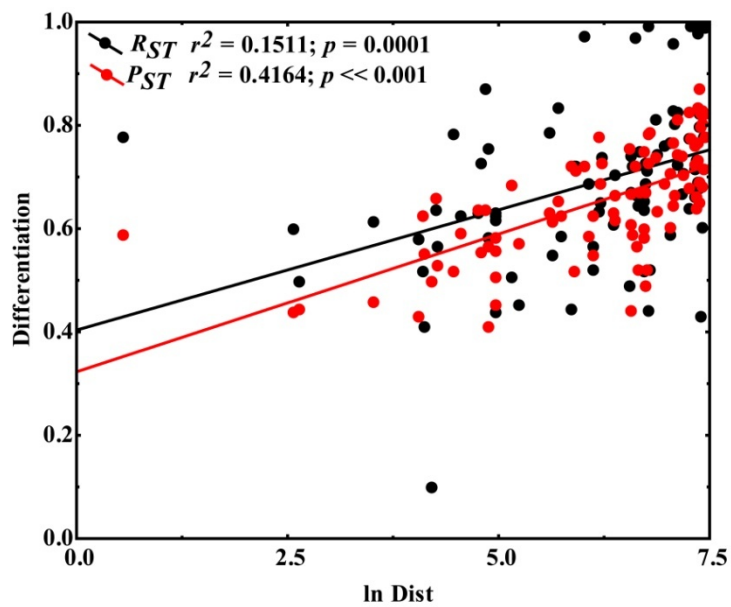
B.
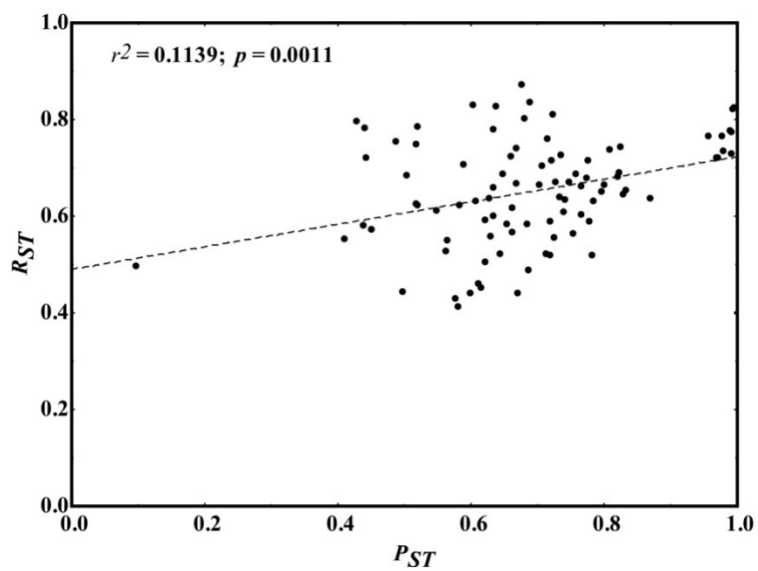
Figure 4.4. The observed results and those from each of the three alternative phylogeographic hypothesis. Populations are ordered based on increasing latitude for the observed data or, in the case of simulations, with increasing distance from putative southern refugia. (A) Observed heterozygosity, H_o; and (B) allelic richness.

A.
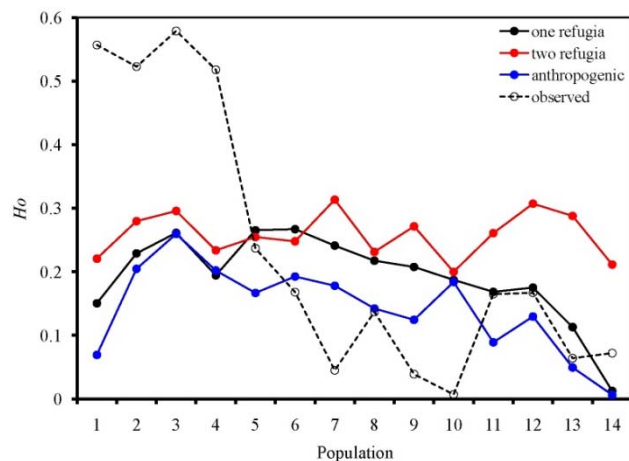
B.
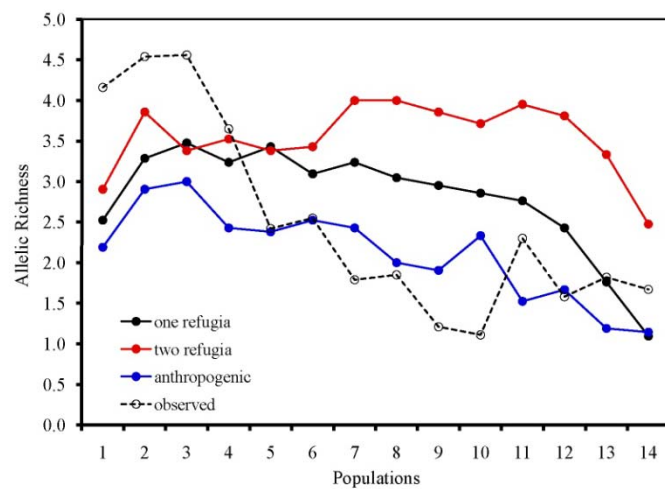
CHAPTER 5: ASSESSING THE CONSERVATION IMPLICATIONS OF POPULATION GENETIC ANALYSES IN THE SELF-COMPATIBLE ANNUAL *AGALINIS DECEMLOBA* SSP. *DECEMLOBA* (OROBANCHACEAE)

ABSTRACT

The levels of diversity within and differentiation among populations are in large part determined by a species' life-history characteristics. For a self-compatible annual taxon with gravity dispersed seeds, the patterns of genetic diversity that are expected based on those life-history characteristics (e.g., low levels of allelic diversity among isolated populations) may also be indicative of an increased risk of extinction due to genetic factors. Consequently, the expected correlation between population genetic parameters and fitness may not always hold true for those species which, if true, would have important conservation implications. In this study, we conducted population genetic analyses based on 21 microsatellite loci to determine the extent to which genetic factors pose an extinction risk for the *Agalinis decemloba* ssp. *decemloba* (Orobanchaceae), which includes individuals that have been historically ascribed to the federally endangered species *Agalinis acuta*. Of the 13 populations sampled, 6 exhibited extremely low levels of observed heterozygosity (< 0.09); the mean estimate of the population level inbreeding coefficient (R_{IS}) was 0.615. Selfing rates estimated with a Bayesian clustering method were also high and maximum-likelihood estimates of the pairwise relationships within populations revealed most individuals to be full-sibs. Populations are also highly differentiated from one another ($D_{EST} = 0.573$; $\Theta = 0.565$). Although the patterns we observed are expected based on the taxon's life history characteristics (i.e., self-compatible annual with gravity dispersed seeds), a conservative

approach is to assume that those populations with the highest levels of homozygosity and inbreeding are likely suffering a reduction in fitness due to genetic factors. However, the lack of gene flow and low levels of allelic diversity among all populations signify that a major conservation priority should be to protect all occurrences from the extinction risks associated with demographic and environmental stochasticity. Empirical detection of seed augmentation and creation of de novo populations from various seed sources are also documented.

INTRODUCTION

Based on the correlation between population genetic parameters and fitness, the analysis of molecular variation among individuals (e.g., differences in allele length at microsatellite loci) can provide an indirect measure of whether a species is facing an increased risk of extinction due to genetic factors (Frankham et al. 2002). Such information is of particular use to conservation managers since it can be incorporated to increase the efficacy of management actions (e.g., Haig 1998). However, the proposed utility of population genetic analyses for conservation is also controversial (e.g., Spielman et al. 2004); it has been argued that demographic and environmental stochasticity are likely to be of greater importance because they can drive a species to extinction before the negative effects associated with genetic factors can impact a species (Lande 1988). The interpretation of population genetic analyses for the conservation of a species is also complicated by the correlation between life-history characteristics and specific patterns of genetic diversity that are often associated with a reduction in fitness (e.g., populations of self-compatible annual plant taxa often exhibit low allelic diversity and strong population differentiation; Nybom 2004). Furthermore, the strength of the expected positive relationship between genetic diversity, population size, and fitness within plants likely varies depending on a species' life-history characteristics (i.e., fitness and genetic variation were independent of one another in self-compatible species; Leimu et al. 2006).

The risks that species may face because of genetic factors can be broadly summarized as short-term declines in fitness due to inbreeding depression and the long-term erosion of genetic diversity as a result of genetic drift and consecutive bottlenecks

(Amos & Balmford 2001). Inbreeding is of particular interest because when high there is the possibility that a species' progeny may be experiencing a reduction in fitness due to the increased probability that recessively deleterious alleles will be present as homozygous genotypes (Charlesworth & Charlesworth 1999; Frankham 1995). Molecular markers can also be used to estimate the effective population size which, when low, genetic drift may be facilitating the loss of allelic diversity through increased rates of fixation(Amos & Balmford 2001), promote the fixation of selectively disadvantageous alleles (i.e., drift load; Whitlock 2000), and increase population differentiation (Kimura 1983). Low values of N_e and the lack of genetic diversity associated with such populations may also inhibit a species' ability to withstand environmental fluctuations (Reed & Frankham 2003). Given that habitat loss is the greatest threat to the persistence of most species (Wilcove et al. 1998), populations may frequently experience a reduction in populations size which is expected to increase inbreeding depression and genetic load (Frankham et al. 2002; Ingvarsson 2001; Petit et al. 1998; Young et al. 1996). Consequently, the ability to detect population bottlenecks through the analysis of molecular data (Luikart et al. 1998) may be a particular useful tool for conservation biologists.

Despite the expectations regarding the relationship between genetic diversity and fitness, the patterns of genetic variation observed within a species are partially determined by its life history characteristics (i.e., breeding system, dispersal potential and lifespan). Outcrossing plant species often have higher levels of genetic diversity and a smaller amount of population differentiation than selfing species; populations of species that have gravity dispersed seeds are more differentiated from one another than species

with wind dispersed seeds; annual species show higher levels of population differentiation than long-lived perennials (e.g., Hamrick & Godt 1996; Hamrick et al. 1979; Nyborn & Bartish 2000). Based on these generalities, within self-compatible annual taxa with gravity dispersed seeds the patterns of genetic diversity may also be similar to those that are often associated with an increased extinction risk as a result of genetic factors. Selfing rates are also known to vary depending on pollinator availability and population size , which may have negative fitness consequences (e.g., Herlihy & Eckert 2002) or represent an evolved reproductive assurance mechanism that does not (Jain 1976; Lloyd 1979) (Kalisz et al. 2004). Consequently, for conservation managers it may be particularly difficult to determine whether low levels of allelic variation and high levels of inbreeding and population differentiation within self-compatible taxa confers a reduction in fitness as a result of anthropogenic effects or reflects a stable situation that is best explained by life history characteristics.

In this study, we investigate the conservation implications of population genetic analyses for *Agalinis decemloba* ssp. *decemloba* (Orobanchaceae) based on the analysis of 21 microsatellite loci. Recent analyses evaluating the evolutionary distinctiveness of *Agalinis acuta* Pennell have shown that it does not appear to be distinct from *Agalinis decemloba* Greene (Pennell) (Pettengill & Neel 2008) such that it has been proposed the two be synonymized as subspecies *Agalinis decemloba* ssp. *decemloba* (Pettengill and Neel in prep). Another subspecies, *A. decemloba* ssp. *tenella*, is represented by what has historically been known as *Agalinis tenella* Pennell. *Agalinis acuta* was listed under the Endangered Species Act (ESA; 1973) in 1987 due to conversion of its grassland habitat to agricultural, residential, industrial, and commercial development; destruction of habitat

was also cited as the most serious threat to the persistence of the species (U. S. Fish and Wildlife Service 1988). Vegetation succession to closed-canopy forest has also caused habitat loss and is a continuing threat at all remaining sites (U. S. Fish and Wildlife Service 1988). As such, remaining populations of this species persist only at sites that are consistently managed to reduce cover of competing vegetation through mowing or prescribed burning (e.g., regular mowing or prescribed burning maintains an unforested sandplain grassland community at the two remaining naturally occurring populations on Cape Cod, Massachusetts, USA).

MATERIALS AND METHODS

Study system

    Agalinis decemloba ssp. *decemloba* is a fall blooming annual native to eastern North America where it is found on the coastal plain in eastern Massachusetts; Rhode Island; Connecticut; and New York; the piedmont in Maryland; southern Virginia; and central North Carolina. Individuals are found in habitats characterized by open vegetation such as sandplain grassland and in openings in pine–oak forest on nutrient poor soils such as glacial outwash sands or serpentine. Large population fluctuations have been documented. Censuses in the two naturally occurring populations on Cape Cod between 1980 and 2000 ranged from 1 to 4253 plants at one site and from 0 to 3674 plants at the other (Neel & Somers 2001). Mechanisms responsible for fluctuations include weather conditions (e.g., lowest population sizes followed Hurricane Bob in 1991) and mowing regimes (e.g., populations increased dramatically after altering the mowing regime at one site) (Neel 2002; Neel & Somers 2001). In addition to being an annual, *A. decemloba* ssp. *decemloba* lacks any obvious mechanism for long distance dispersal and seeds are

primarily gravity dispersed. Bumble bees, *Bombus* spp., are likely the primary pollinator, but members of the Syrphidae (Order Diptera) have also frequently been observed visiting flowers (Neel 2002). The species is highly self-compatible species with 97% of selfed flowers setting fruit; however, fruits resulting from selfed flowers exhibited a 17-20% reduction in seed set compared with open pollinated flowers (Neel 2002). Geitonogamy (i.e., the transfer of pollen to different flowers on the same individual) and autonomous selfing late in anthesis as corollas dehisce (i.e., 'corolla dragging;' Dole 1990) are likely mechanisms through which selfing occurs.

Population sampling, DNA extraction and microsatellites

Tissue was collected from 328 individuals representing 13 populations during the fall blooming seasons of 2005-2007 (Fig. 5.1). Where possible, we sampled 30 individuals per population, but for many populations the number of samples is less than 30, which reflects the total number of individuals we could locate (Table 5.1). Two populations, WANY and EPRI, are known to have been anthropogenically created or modified with seed from different populations (M. Jordan, C. Raithel, and S. Ruhren pers. comm.). At SDMD, we sampled to represent the extent of variation within that relatively large area such that we collected individuals from potentially spatially isolated patches. L1VA individuals were collected from different sides of a road along a powerline corridor. Other populations consisted of individuals within one relatively continuous patch. We also attempted to obtain samples from an additional 20 populations that were known based on historical records (e.g., herbarium specimens and Natural Heritage records) or on information provided by local botanists. No individuals could be located at these additional sites due to either unsuitable habitat (e.g., commercial development, recently

mowed, or habitat was in a later successional stage), severe drought, or that our sampling efforts did not coincide with flowering times that vary annually.

Total genomic DNA from all samples was isolated from fresh or frozen (-80°C) leaves and flower buds by grinding 50-100 mg of tissue to powder in liquid nitrogen with a mortar and pestle, and then using GenElute Plant Genomic DNA Kits (Sigma) or QIAGEN DNEasy Kits (QIAGEN) following manufacturer instructions. Extractions were also carried out on a QIAGEN BioSprint 96 robotic workstation using QIAGEN's BioSprint 96 DNA Plant Kit.

All samples were assayed for variation at 21 microsatellite loci that were developed for *Agalinis acuta*; the development of which is described in Appendix B (Pettengill et al. 2009).

Genetic diversity and N_e

Allelic richness (A_R) using rarefaction to account for differences in sample size was calculated using FSTAT v.2.9.3.2 (Goudet 1995). FSTAT was also used to test for deviations from Hardy-Weinberg equilibrium (HWE) and linkage equilibrium (LE). The program GENEPOP v4.0 (Raymond & Rousset 1995) was used to estimate inbreeding within populations as measured by R_{IS} (Rousset 1996). We calculated the observed (H_o) and expected (H_e) heterozygosity at the locus, population, and subspecies level using the program ARLEQUIN v3.1 (Excoffier et al. 2005).

Single sample estimates of N_e based on the degree of linkage disequilibrium (Waples 2006) were calculated using the program LDNE (Waples & Do 2008). This method is based on Burrows' Δ estimate of linkage disequilibrium that does not depend on the assumption of random mating making it particularly useful in systems where selfing and

consanguineous matings are likely (Waples 2006). Because alleles with extreme frequencies (i.e., close to 0 or 1) can result in a biased but also more precise estimate, we estimated N_e by excluding alleles with frequencies below two critical thresholds (P_{crit} = 0.1 and 0.01). We present the 95% confidence intervals based on the jackknife procedure in which new estimates are calculated after removing in turn each of the 210 pairwise locus combination. The jackknife method is less biased than the parametric method that assumes independence among the pairwise comparisons and results in too narrow of confidence intervals (Waples 2006; Waples & Do 2008). We also evaluated coalescent-based estimates of effective population size scaled by the mutation rate (i.e., Θ from LAMARC; Kuhner 2006) but found results to be unreliable. Estimates of Θ consistently approached the upper bound of the prior for populations with higher levels of heterozygosity and allelic diversity (e.g., when the upper bound on Θ was 30, estimates of Θ were ~28) . The large range of intrapopulation diversity and heterozygosity among the populations is likely responsible for the unreliable estimates in that invariant populations are known to cause problems for coalescent based methods (Kuhner 2006).

Given the small population sizes observed upon sampling (e.g., the targeted 30 individuals could not be located and sampled from six populations), we were interested in whether any of the populations had experienced a recent bottleneck. We used the program BOTTLENECK v1.2.02 (Piry et al. 1999) to test for recent bottlenecks with significance based on 1000 permutations. We present the results from the two-phased mutational model, which is recommended for microsatellite loci and the one-tailed (for a heterozygote excess) Wilcoxon-sign rank test (α = 0.05) that is the most powerful of three tests offered by the program (Piry et al. 1999).

Population differentiation and migration analyses

We used the program MARK v3.1 (Ritland 2006) to estimate the relationships of

individuals within and among populations as the probability that pairs of individuals were

full-sibs, parent-offspring, half-sibs, first cousins, or unrelated. MARK uses a maximum-

likelihood method to assign pairs of individuals to a specific relationship class (described

in Mousseau et al. 1998). The likelihood of a relationship is equal to the observed level

of relatedness, *r*, which is calculated as the product across the 21 microsatellite loci of the

probabilities of identity by descent of alleles from two individuals, given the probability

of identity by descent expected of that relationship. The results are presented as the

probabilities that two individuals fall into each of the relationship classes; the class that

has the highest probability represents the relationship that best describes the two

individuals.

We calculated Jost's D_{EST} (Jost 2008) using the program SMOGD v1.2.3 (*Software*

for the Measurement of Genetic Diversity) (Crawford accepted) to provide an estimate of

population structure. Jost's D_{EST} is argued to be a more accurate measure of population

differentiation than those used previously (e.g., G_{ST} and F_{ST}) in that it does not suggest

low levels of differentiation among populations that share few or no alleles in common or

underestimate the degree of differentiation when within population heterozygosity is high

(Heller & Siegismund 2009; Jost 2008). We also estimate the more traditional measure

of population differentiation, F_{ST} (i.e., Θ of Weir & Cockerham 1984) using GENEPOP

to facilitate comparisons with other studies and with D_{EST}.

To evaluate the differences among populations in allelic composition, we conducted

an inter-class principal components analysis (Dolédec & Chessel 1987) using the

program *adegenet* (Jombart 2008) written in R (R Development Core Team 2008). This method maximizes the variance among populations while also displaying the relationship of individuals (i.e., multilocus genotypes) in ordination space. Analyses were based on the variance-covariance matrix and the significance of the difference among populations was based on 999 permutations using the randtest.between function. For each population, 95% inertia ellipses are shown that depict the amount variance among individuals.

We used the model based method implemented in InStruct (Gao et al. 2007) to estimate the degree of genetic structure among samples. We chose InStruct because, in addition to quantifying the degree of nonrandom mating as a result of restricted gene flow, it also incorporates a measure of selfing or inbreeding within clusters (Gao et al. 2007). Therefore, the model implemented in InStruct is well suited for organisms such as *A. decemloba* ssp. *decemloba* in which both selfing and restricted dispersal are likely to cause nonrandom mating. Furthermore, clustering methods that do not account for mating among close relatives (e.g., STRUCTURE; Pritchard et al. 2000) can result in erroneous signatures of population structure (Anderson & Dunham 2008; Gao et al. 2007). We ran InStruct treating the number of clusters as a random variable to be estimated by the MCMC algorithm with a maximum value of 19. A run consisted of two independent chains of 20^5 generations with 10^5 generations serving as burnin. We considered chains at a given value of K (i.e., the number of populations) converged if the Gelman-Rubin statistic (an ANOVA based measure comparing within-chain and among-chain variance where when equal indicates stationarity; Gelman & Rubin 1992) was <1.1 (Gao et al. 2007). The value of K with the lowest Deviance Information Criteria was chosen as the best fitting model.

We used the program GENECLASS v2.0 (Piry et al. 2004) to determine whether any

of the sampled individuals represented recent immigrants into the population from which

they were sampled using the Bayesian method of Rannala and Mountain (1997) and the

Λ statistic, which is the recommend measure of whether an individual is a migrant

(Paetkau et al. 2004). Individuals are identified as migrants based on the degree of

genotypic disequilibrium they exhibit relative to the other members of the population

from which they were sampled. The population within which an immigrant has the

maximum level of LE is identified as the most likely source population. Significance

was assessed through creating populations of the same size as the original populations

using the resampling algorithm of Paetkau et al. (2004), which was carried out until

10000 simulated assignment criteria had been estimated; α was 0.01 (i.e., the null

hypothesis that an individual germinated in the population from which it was sampled is

rejected if the observed Λ for that individual is greater than 99% of the distribution

associated with the 10000 Λs from the simulated data).

RESULTS

Genetic diversity and N_e

Among the 328 samples, 718 alleles were observed across all 21 microsatellites, the

proportion of which were polymorphic was 0.95 (Table 5.1). H_o was low ($\bar{x} = 0.018$)

and was an order of magnitude lower than H_e ($\bar{x} = 0.634$) (Table 5.1). The mean

proportion of polymorphic loci within all 13 populations was 0.61 and the average

number of alleles per population was 56; on average 9.23 of these alleles were private

(Table 5.1). The 30 samples from Montauk Downs, NY (MDNY) were polymorphic at

only a tenth of the loci and had only 28 alleles, 2 of which were exclusive to that

population (Table 5.1). This population also had the lowest H_o and H_e, 0.006 and 0.028, respectively, and the highest R_{IS} (0.997). The two most southern populations had the greatest portion of polymorphic loci (L1VA, 0.95; WCNC, 0.81) and the largest number of alleles. Of the northern populations, RCRI had the greatest portion of polymorphic loci (0.71) and had a high H_o (0.239) (Table 5.1).

99 out of 2730 possible comparisons of loci within populations were out of HWE and all populations showed a significant departure from HWE. After a bonferroni correction for multiple tests, 2.7% (75 of 2730) of the locus-by-locus comparisons significantly deviated from LE. However, 84% of these significant deviations were exclusive to two populations; SDMD had 22 and L1VA had 41 pair-wise locus deviations from linkage equilibrium.

Although N_e varied depending on the critical threshold associated with the minimum allele frequency, the averages under each threshold were extremely low; average N_e when $P_{crit} = 0.01$ was 21.6 and when $P_{crit} = 0.1$ N_e averaged 12.6 (Table 5.2). Those averages exclude two instances where estimates of N_e are unrealistic (i.e., a negative estimate for SNMA and the exceedingly large estimate of $N_e = 1 \times 10^8$ for MDNY; Table 5.2). The upper bounds to the confidence intervals for six and four populations under $P_{crit} = 0.01$ and $P_{crit} = 0.1$, respectively, are infinity. Rather than actually representing truly large effective population sizes, infinite upper bounds can also result when there are very low values of heterozygosity within populations or insufficient power to estimate N_e (Waples pers. comm.). Exploratory analyses using the parametric method of calculating confidence intervals produced the expected narrower intervals (i.e., the upper bound was infinity for 2 rather than 4 confidence intervals), which illustrate the bias that arises when

independence among pairwise locus comparisons is assumed (Waples 2006; Waples & Do 2008). The larger estimates under a lower allele frequency were also observed by (Waples 2006), and can be explained by the fact that more alleles present at a locus will likely reduce the degree of linkage disequilibrium thus increasing the estimate of N_e compared to when such alleles are excluded. Focusing on $P_{crit} = 0.01$, which may be a more appropriate threshold for hypervariable loci such microsatellites, the largest estimate of N_e was 132.5 and the smallest was 1.5 that were observed in WCNC and SDMD, respectively. In addition to WCNC, BVMA and PCCT were the only other populations with estimates of $N_e > 15$ (Table 5.2).

Only two populations, SHNY and WCNC, showed evidence of having experienced a recent bottleneck under the two-phased mutational model and using the one-tailed Wilcoxon sign-rank test for heterozygote excess.

Population structure and migration

The pairwise relationships of the vast majority of individuals within populations were best described as being full-sibs and individuals from different populations were best described as unrelated (Fig. 5.2). There were very few intermediate levels of relatedness in that we did not find a substantial number of pairs of individuals being classified as parent-offspring, half-sibs, or first cousins. Exceptions are the individuals from L1VA that do exhibit intermediate levels of relatedness, which may be an artifact of collecting individuals across a potential barrier to gene flow in that different locally breeding neighborhoods were sampled. Individuals within WANY also show intermediate relationships, which likely reflects how that population was created from seed from at least three different populations (Fig. 5.2).

The levels of differentiation among populations estimated with Jost's D_{EST} and Θ averaged 0.573 and 0.565, respectively. Pairwise population estimates of D_{EST} and Θ were both lowest between RCRI and EPRI ($D_{EST} = 0.086$, $\Theta = 0.0716$), which is to be expected given that the latter was created from seeds from the former population (Table 5.4). However, the two measures differed in the pair of populations that were depicted to be the most differentiated. D_{EST} was greatest between WCNC and BVMA ($D_{EST} = 0.828$), but between those two populations Θ was 0.650. More surprisingly was that based on Θ, SHNY and MDNY showed the largest degree of differentiation ($\Theta = 0.878$), but between those two populations D_{EST} was 0.333 (Table 5.4). Although differences exist, both measures indicate a high degree of population differentiation..

A global test for the significance of the degree of differentiation among populations based on the inter-class principal components analysis was highly significant in that our observed level of differentiation among populations was greater than all values from 999 permuted datasets ($P < 0.0001$). The first two axes from the inter-class principal components analysis explained 37% of the variance among populations. The first principal components axis primarily explained differences among the populations in the northern portion of the species range; the second principal components axis explained the variance due to the latitudinal cline associated with the populations (Fig. 5.3). The distribution of multilocus genotypes from individuals sampled from WANY are quite spread out relative to the mean and hence the 95% inertia ellipse is large. Most other populations have small 95% inertia ellipses and individual multilocus genotypes are a short distance from the within population centroid (i.e., mean; Fig. 5.3).

The value of *K* with the lowest Deviance Information Criteria from the InStruct was 12 (logL= -6269.513; DIC = 12539.026) (Fig. 5.4). The Gelman-Rubin statistic was 0.999 indicating that the independent chains associated with each analysis had converged. Interestingly, the L1VA population is divided into two separate clusters, and the EPRI and RCRI populations are found in the same cluster with individuals from EPRI also showing a similarity to individuals from PCCT. The majority of individuals from WANY are found within the cluster also occupied by HPNY but with admixture indicating a similarity to individuals from MDNY (Fig. 5.4). Estimates of the selfing rate were high ($\bar{S} = 0.730$; Table 5.2), which can be interpreted as the proportion of offspring produced via selfing within a cluster and indicates that in addition to restricted gene flow genetic, mating among close relatives also contributes to the degree of nonrandom mating (Fig. 5.4).

The results from GENECLASS v2.0 identified only four samples as being recent immigrants into the population from which they were sampled. Three individuals from EPRI were classified as recent migrants from RCRI and one individual from EPRI was identified as being a recent migrant from RCRI.

DISCUSSION

Populations of the self-compatible taxon, *Agalinis decemloba* ssp. *decemloba*, exhibited a wide range in the levels of genetic diversity, inbreeding, and effective population size such that the extent to which genetic issues are of conservation concern likely depends on the population being considered. However, there is also the possibility that extremely low levels of heterozygosity and high levels of inbreeding could be explained by the taxon's life history characteristics and, therefore, may not represent a

conservation concern. To help differentiate between levels of diversity that are an artifact of life-history characteristics and those that are emblematic of an extinction risk due to extrinsic factors (e.g., population bottlenecks and habitat fragmentation as a result of anthropogenic factors; Michalski & Durka 2007), below we discuss our results in light of other experiments involving taxa with similar life history characteristics and studies investigating the correlation between fitness and genetic parameters among self-compatible species.

Genetic diversity, N_e*, and population structure*

All populations showed a significant deviation from HWE but, like many plants, *A. decemloba* ssp. *decemloba* exhibits a degree of self-compatibility and the sessile nature of individuals can result in small local breeding neighborhoods, both of which violate the assumption of random mating associated with HWE (Clegg 1980). The Wahlund effect (Wahlund 1928) may also explain departures from HWE, particularly within L1VA and SDMD because individuals from these populations may have been sampled across a barrier to gene flow. Regardless of the explanation for the observed departures from HWE, the deficit of heterozygosity and corresponding high levels of homozygosity may have vital consequences for the fitness of populations.

Important for the conservation of *A. decemloba* ssp. *decemloba* is that the low levels of H_o and A_R we observed may be indicative of an increased extinction risk due to the inability of individuals to adapt to changes in the environment (Frankham et al. 2002). However, the expected relationship between genetic variation and fitness depends strongly on the mating-system of a species. Fitness and genetic variation have been shown to be positively correlated within outcrossing species but independent of each

other within self-compatible taxa (Leimu et al. 2006). Based on previous empirical

research, we know that *A. decemloba* ssp. *decemloba* is highly self-compatible in that

97% of selfed flowers set fruit (e.g., Neel 2002). Comparisons with other studies provide

support for the conclusion that some populations of the *A. decemloba* ssp. *decemloba*

have levels of H_o that are characteristic of predominantly selfing species. From a

summary of investigations of plant species using microsatellite loci, mixed-mating

species had an average H_o of 0.51; selfing species had an average H_o of 0.05 and (Nybom

2004). Among the 13 populations we sampled H_o ranged from 0.006 to 0.239, which is

suggestive of a predominantly selfing population and moderately outcrossed population,

respectively. Although the evidence for self-compatibility and the lack of correlation

between genetic variation and fitness within such taxa (Leimu et al. 2006) make it

difficult to determine whether low genetic diversity is causing a reduction in fitness, a

conservative conclusion is that a lack of genetic diversity decreases the ability of

populations to withstand environmental fluctuations. The 6 populations with $H_o < 0.100$

and $A_R < 2.0$ would be the most at risk (Table 5.2).

Unlike the estimates of genetic diversity, populations of *A. decemloba* ssp.

decemloba are more differentiated from one another and have N_e estimates that are

substantially lower than those found in many species with similar life history

characteristics. For example, the average of the geometric mean of census sizes from

different conspecific populations for the 26 self-compatible species assayed in Leimu et

al. (2006) was 437.3. If we assume a N_e/N ratio of 0.27, which was the mean value from

the analysis of 26 studies involving species of conservation concern (Palstra & Ruzzante

2008), then the average N_e value from Leimu et al. (2006) is 118.1. The mean estimate of

N_e, under the same N_e/N ratio of 0.27, for the 22 plant species that are both rare and self-compatible from Leimu et al. (2006) is 126.9. In stark contrast, estimates of N_e for the majority of populations of *A. decemloba* spp. *decemloba* are < 10 and drastically lower than either that documented in Leimu et al. (2006). Consequently, the influence of drift in reducing allelic diversity or fixing deleterious alleles is a conservation concern within *A. decemloba* spp. *decemloba*. The level of differentiation among populations we observed ($\overline{\Theta} = 0.56$) was greater than the mean Θ of 0.40 observed across other studies involving annual species (Nybom 2004). This degree of isolation among populations suggests that populations are not benefitting from the infusion of novel alleles that can reduce the level of inbreeding depression (Richards 2000) and increase the ability of populations to adapt to changes in the environment (Lande 1988; but see Lenormand (2002) for a discussion of 'migration load').

Our results imply that only SHNY and WCNC may have recently experienced a recent population bottleneck and, therefore, suffer from the increased extinction risk associated with a reduction in size (Holsinger 2000; Young et al. 1996). The other populations may have undergone a more historical bottleneck (i.e., greater than a dozen generations in the past; Luikart et al. 1998), but have persisted long enough to reach mutation-drift equilibrium. Although fluctuations in census sizes have been documented for some populations (i.e., BVMA and PEMA; Neel & Somers 2001), the existence of a seed bank may also explain why those populations do not exhibit the genetic signature associated with population bottlenecks (e.g., Nunney 2002). If other populations did in fact experience a more historical population bottleneck (e.g., upon being founded), their persistence may suggest that they are resilient and purged the problems associated with

such demographics events. Studies illustrating that self-compatible taxa are less likely

than outcrossing species to suffer the negative consequences (i.e., substantial decrease in

H_e, P, A, and increase in F_{IS}) associated with a reduction in population size and habitat

fragmentation (Aguilar et al. 2008; Honnay et al. 2007) also support the idea that *A.*

decemloba ssp. *decemloba* may be resilient to the effects associated with a reduction in

population size.

Inbreeding depression, self-compatibility, and reproductive assurance

Although low levels of genetic diversity may inhibit the ability of individuals to

withstand environmental variation (Reed & Frankham 2003), the degree of inbreeding

within populations has been stated to be a greater threat to the persistence of populations

(Amos & Balmford 2001). A high level of self-compatibility has been observed for what

was known as *A. acuta* (i.e., 97% of self-pollinated *A. acuta* flowers set fruit; Neel 2002),

but our results are the first to document that populations are also highly inbred (mean

population R_{IS} = 0.607, Table 5.1). In particular, three populations have R_{IS} values close

to unity (e.g., SHNY, R_{IS} = 0.831; WANY, R_{IS} = 0.898; and MDNY, R_{IS} = 0.997).

However, given the low levels of allelic diversity within populations, our estimates of

inbreeding may be an artifact of outcrossing among individuals with similar genotypes,

which could also explain the results from MARK and InStruct that indicate most

individuals within populations are full-sibs and there is high rate of selfing within

clusters, respectively.

There is a large body of literature describing how the high levels of inbreeding that

we observed within many populations of *A. decemloba* ssp. *decemloba* may in fact not

result in a substantial decrease in fitness. Through successive generations of selfing or

inbreeding the level of inbreeding depression can be reduced along with purging of the genetic load (e.g., Crnokrak & Barrett 2002). This scenario is supported by empirical research documenting that inbreeding depression is negatively correlated with selfing rate (i.e., predominantly selfing species exhibited a 43% reduction in inbreeding depression relative to outcrossing species (Husband & Schemske 1996)); a lack of a correlation between selfing and inbreeding depression may also be more prominent in annual rather than perennial plant species (Byers & Waller 1999). Based on simulation studies, the likelihood of this scenario is greatest when recessive genes associated with negative fitness are lethal, rather than detrimental, as homozygotes. This is because a high degree of inbreeding (e.g., full-sib mating) provides a mechanism through which the genetic load of a population can be purged; high fecundity of individuals that are not killed can help reduce the risk of extirpation during the purging process (Hedrick 1994; also see Holsinger 1988).

Studies estimating fitness consequences associated with different levels of inbreeding within congeneric taxa also provide evidence that inbreeding may not result in a decrease in fitness within populations of *A. decemloba* ssp. *decemloba*. For example, a study involving two populations of *Agalinis skinneriana* found no difference in fruit set within each population among different crossing treatments (e.g., selfed and xenogamous matings) (Dieringer 1999). Between two populations of *Agalinis auriculata*, self-pollinated individuals also did not show a reduction in fruit set, seed set, or seed mass relative to outcrossed individuals where seed germination and availability of host plants, because the species is hemiparasitic, were identified as more likely causes for a decrease in population size than inbreeding depression (Mulvaney et al. 2004).

The high levels of inbreeding within populations of *A. decemloba* ssp. *decemloba*

may also reflect an evolved response to assure reproduction when population sizes are

small and pollinator availability is low (e.g., Jain 1976). Autogamous pollination was

also suggested to be a reproductive assurance mechanism in small populations of *A.*

skinneriana within which there was a significantly greater potential selfing rate (99%)

relative to the larger population (85%) (Dieringer 1999). Other empirical studies show

that this shift can result from natural fluctuations in pollinator visitation or due to

anthropogenic induced fragmentation that may cause decline in pollinator visitation rates

(Aguilar et al. 2008). In terms of conservation, some have argued that such a shift

indicates that populations may not suffer the negative fitness consequences often

associated with low levels of genetic diversity and high inbreeding. For example, among

populations of the annual *Collinsia verna* (Scrophulariaceae) pollinator failure strongly

influenced the degree of selfing within populations; inbreeding depression on early acting

traits within selfed individuals was actually lower than that observed among outcrossed

individuals (Kalisz et al. 2004). Kalisz et al. (2004) also state that years of complete

pollinator failure (i.e., when no pollinators were present) have likely occurred and that

under such conditions selfing would have been extremely high, which would have

reduced the genetic load. Based on nine microsatellite loci there were strong differences

among selfing ($H_o = 0.02$–0.06) and outcrossed ($H_o = 0.13$–0.31) populations of

Arabidopsis lyrata (Brassicaceae) and the authors suggest that selfed populations may

have purged deleterious recessive mutations (Mable & Adam 2007). However, the

degree of seed discounting within selfing individuals of the perennial *Aquilegia*

canadensis (Ranunculaceae) from 10 populations was shown to negatively impact fitness

relative to outcrossed individuals (Herlihy & Eckert 2002). The authors state that the reproductive assurance provided by selfing within *Aquilegia canadensis* will be advantageous if there are mechanisms that delay autogamous mating and, thus, maximize the potential for outcrossing (Herlihy & Eckert 2002). The dimorphism of the two pairs of stamens within *A. decemloba* ssp. *decemloba* (i.e., a ventral and dorsal pair relative to the stigmatic surface) may represent an evolved trait that facilitates outcrossing when pollinators are present but ensures autogamous selfing upon corolla senescence.

Conservation implications

Agalinis acuta (which we have advocated be subsumed under *Agalinis decemloba* ssp. *decemloba*) received federal protection on public lands under the ESA due to the threat of extinction posed by a decrease in suitable habitat. Because of this decrease in habitat, most extant populations are small and dramatic fluctuations in population sizes have also been documented; our results from the analysis of 21 microsatellite loci are concordant with the expectations associated with such a history (i.e., low levels of heterozygosity and allelic richness within and strong differentiation among populations). However, the observed pattern of genetic diversity is common within self-compatible annual taxa (Hamrick & Godt 1989; Nybom 2004) and the expected correlation of our observed levels of genetic variation and fitness is not always true of self-compatible taxa (Leimu et al. 2006). As for the levels of inbreeding, there are empirical and simulation studies that provide evidence to support the claim that *Agalinis decemloba* ssp. *decemloba* may have purged deleterious recessive alleles through high rates of selfing and not suffer a reduction in fitness due to inbreeding depression (e.g., Crnokrak & Barrett 2002). However, the magnitude of inbreeding estimates within three populations

of *Agalinis decemloba* ssp. *decemloba* (e.g., MDNY, SHNY, and WANY; Table 5.2) warrants additional studies to investigate whether such populations do suffer an increased risk of extinction.

Due to the low levels of allelic diversity and small estimates of N_e, populations of *Agalinis decemloba* ssp. *decemloba* may not be able to adapt to environmental changes (e.g., Reed & Frankham 2003; e.g., Stebbins 1957). Additionally, populations are strongly isolated and harbor unique genetic diversity such that the priority for conservation should be to ensure the persistence of all extant populations. Given these conservation concerns, conservation managers should strive to increase the size of these populations and help buffer against any stochastic events that could result in extirpation. These management actions would facilitate achieving two of the three criteria stated in the recovery plan for *A. acuta* (i.e., 20 stable and wild occurrences and protection of at least 15 of these occurrences; U.S. Fish and Wildlife Service 1989b). The third criterion was to determine whether plants could be propagated from seed. In addition to the fact that populations have been established from seed from different populations (e.g., EPRI from RCRI), our results show that molecular techniques can detect and confirm the source population (e.g., Figs. 5.3 & 5.4). The successful creation of populations from multiple seed sources also implies that conservation managers of *A. decemloba* ssp. *decemloba* may not have to be particularly concerned about a decrease in fitness associated with outbreeding depression (Lynch 1991); experiments are needed to validate this assumption. Assuming that *A. decemloba* ssp. *decemloba* receives the federal protection afforded *A. acuta*, if the number of populations can be increased and they are

protected against habitat destruction, then *A. decemloba* ssp. *decemloba* could be among the few taxa to be delisted as a direct result of management actions (Noecker 1998).

Conclusions

We suggest that a conservative approach is to assume that the populations of *A. decemloba* ssp. *decemloba* with extremely high inbreeding coefficients are experiencing an increased extinction risk due to inbreeding (e.g., MDNY, SHNY, and WANY). However, as a result of strong population differentiation and low values of allelic diversity and N_e within populations, a conservation priority should be to guard against the risk posed by habitat destruction and environmental stochasticity. Additional meta-analyses that differentiate taxa based on mating system and empirical studies that directly measure fitness differences among populations of highly selfing plant species will help to clarify whether, in general, demographic, and environmental stochasticity rather than genetic factors are of primary concern in ensuring the persistence of such species.

Table 5.1. Characteristics of 21 microsatellite across 328 *A. decemloba* ssp. *decemloba* individuals representing 13 putative populations. (*A*, number alleles, H_o, observed heterozygosity; H_e, expected heterozygosity; R_{IS} inbreeding coefficient; D_{EST}, Jost's unbiased estimator of population differentiation; and percent amplification success)

Locus	Range Size (bp)	A	H_o	H_e	R_{IS}	D_{EST}	Success (%)
Agac.M1	242-310	29	0.182	0.930	0.805	0.933	97.26
Agac.M6	272-360	21	0.146	0.920	0.841	0.904	98.17
Agac.M33	203-219	10	0.063	0.582	0.891	0.488	96.34
Agac.M11	200-232	9	0.077	0.580	0.867	0.484	99.09
Agac.M14	185-244	21	0.137	0.861	0.841	0.851	97.87
Agac.M28	271-291	6	0.031	0.282	0.890	0.194	98.48
Agac.M42	271-293	9	0.115	0.801	0.857	0.797	98.17
Agac.M46	212-286	21	0.216	0.866	0.751	0.831	97.56
Agac.ca 11	148-174	12	0.125	0.682	0.817	0.512	100.00
Agac.ca20	404-436	14	0.136	0.686	0.802	0.639	98.78
Agac.ca26	218-252	8	0.169	0.760	0.778	0.714	97.56
Agac.aag46	457-484	6	0.108	0.597	0.820	0.529	96.34
Agac.ca10	158-170	8	0.146	0.379	0.614	0.183	97.87
Agac.aag29	335-343	4	0.028	0.202	0.861	0.156	97.87
Agac.ca48	280-300	6	0.070	0.429	0.837	0.374	100.00
Agac.ca 21	144-160	6	0.070	0.415	0.831	0.342	100.00
Agac.taca12	430-470	10	0.155	0.792	0.804	0.731	94.21
Agac.taca45	185-375	33	0.298	0.934	0.681	0.929	99.39
Agac.taca04	311-427	14	0.108	0.714	0.849	0.585	99.09
Agac.ca45	292-294	2	0.000	0.164	1.000	0.122	98.17
Agac.ca33	214-286	13	0.104	0.719	0.855	0.668	99.70
Average	**n/a**	**12.47**	**0.118**	**0.633**	**0.823**	**0.570**	**98.24**

Table 5.2. Population genetic characteristics at the subspecies and population level based on 21 microsatellite loci across 13 populations of *A. decemloba* ssp. *decemloba*. Sample sizes, n; proportion of polymorphic loci, P; number of alleles, A; number of private alleles, A_P; allelic richness as calculated in FSTAT using a rarefaction method, A_R; observed (H_o) and expected (H_e) heterozygosity; fixation index, R_{IS}; and, the selfing rate inferred with InStruct, S.

Population	Collection Locality	n	Amplification Success (%)	P	A	A_P	A_R	H_o	H_e	R_{IS}	S
A. decemloba ssp. *decemloba*		**328**	**98.24**	**0.95**	**718**	**n/a**	**12.476**	**0.018**	**0.634**	**n/a**	**n/a**
BVMA	Bay View, Barnstable Co., MA	29	98.03	0.57	46	5	1.665	0.061	0.182	0.528	0.802
SNMA	Scrubby Neck, Dukes Co., MA	16	99.70	0.62	44	5	1.745	0.087	0.224	0.233	0.780
PEMA	Percival Cemetery, Barnstable Co., MA	29	98.85	0.62	46	8	1.755	0.069	0.216	0.372	0.813
PCCT	Plainfield Cemetery, Windham Co., CT	29	99.01	0.67	55	5	1.993	0.177	0.302	0.717	0.546
EPRI	Eppley Wildlife Sanctuary, Washington Co., RI	24	94.64	0.67	74	9	2.563	0.169	0.416	0.629	0.622
RCRI	Richmond Cemetery, Washington Co., RI	20	99.05	0.71	60	3	2.502	0.239	0.405	0.195	0.622
HPNY	Hempstead Plains, Nassau Co., NY	31	99.08	0.67	51	4	2.030	0.151	0.312	0.663	0.645
MDNY	Montauk Downs, Suffolk Co., NY	30	99.37	0.10	28	2	1.148	0.006	0.028	0.997	0.943
SHNY	Shadmoor State Park, Suffolk Co., NY	24	98.02	0.43	32	6	1.264	0.037	0.059	0.831	0.841
WANY	Warhol Preserve, Suffolk Co., NY	9	94.71	0.48	58	5	2.609	0.124	0.421	0.898	0.794
SDMD	Soldiers Delight, Baltimore Co., MD	30	99.37	0.62	49	10	1.926	0.046	0.302	0.645	0.861
L1VA	Lunenberg Co. VA	24	96.10	0.95	83	30	2.849	0.168	0.496	0.577	0.606
WCNC	William B. Umstead State Park, Wake Co. NC	33	99.40	0.81	92	28	2.874	0.226	0.415	0.530	0.611
Mean across populations		**25.2**	**98.11**	**0.61**	**56**	**9.23**	**2.630**	**0.125**	**0.290**	**0.607**	**0.730**

Table 5.3. Estimates of N_e based on linkage disequilibrium; P_{crit} represents the threshold allele frequency below which alleles are not included in the estimation procedure. Numbers in parentheses are 95% confidence intervals based on the parametric method described in WAPLES 2006.

| Population | N_e | | | |
	$P_{crit} = 0.01$		$P_{crit} = 0.1$	
BVMA	29.1	(7.7 - ∞)	6	(1.8 - 23.1)
PCMA	3.3	(2.3 - 8)	1.2	(0.9 - 1.7)
SNMA	- 84*	(11 - ∞)	18.7	(1.6 - ∞)
PCCT	47.4	(18.3 - ∞)	28.9	(8 - ∞)
EPRI	10.3	(7.5 - 14.3)	11.6	(6 - 24.9)
RCRI	6.7	(4 - 9.9)	3.9	(2.6 - 7.1)
HPNY	13.3	(5.5 - 33)	3.9	(2.1 - 13)
MDNY	2	(-9.8 - ∞)	1×10^8*	(1×10^8 - ∞)
SHNY	4.6	(0.8 - ∞)	0.5	(0.1 - 1.6)
WANY	6.4	(2.8 - 15.3)	2.4	(1.6 - 5)
SDMD	1.5	(1.1 - 1.9)	0.7	(0.5 - 0.8)
L1VA	2.2	(1.9 - 2.5)	0.7	(0.6 - 0.7)
WCNC	132.5	(51 - ∞)	72.6	(18.1 - ∞)
Average	21.6		12.6	

* Not included in calculating the average estimate of N_e.

Table 5.4. Measures of pairwise differentiation among 13 populations of *A. decemloba* ssp. *decemloba;* Θ (Weir & Cockerham 1984) is above and D_{EST} (Jost 2008) is below the diagonal.

	BVMA	SNMA	PCMA	PCCT	EPRI	RCRI	HPNY	MDNY	SHNY	WANY	SDMD	L1VA	WCNC
BVMA		0.706	0.642	0.604	0.557	0.551	0.619	0.845	0.805	0.583	0.647	0.559	0.650
SNMA	0.635		0.619	0.588	0.526	0.532	0.598	0.857	0.793	0.543	0.547	0.511	0.559
PCMA	0.487	0.520		0.587	0.537	0.554	0.560	0.826	0.781	0.511	0.606	0.534	0.617
PCCT	0.543	0.588	0.548		0.382	0.390	0.537	0.712	0.653	0.417	0.499	0.424	0.524
EPRI	0.567	0.589	0.542	0.410		0.072	0.470	0.716	0.646	0.388	0.487	0.401	0.486
RCRI	0.526	0.581	0.552	0.380	0.086		0.463	0.715	0.659	0.389	0.498	0.402	0.489
HPNY	0.585	0.612	0.526	0.587	0.582	0.545		0.711	0.664	0.186	0.516	0.442	0.445
MDNY	0.647	0.687	0.665	0.514	0.653	0.591	0.561		0.878	0.732	0.695	0.674	0.691
SHNY	0.607	0.592	0.621	0.495	0.589	0.587	0.548	0.333		0.671	0.628	0.607	0.639
WANY	0.522	0.572	0.472	0.451	0.551	0.521	0.190	0.451	0.462		0.420	0.359	0.408
SDMD	0.634	0.517	0.612	0.503	0.579	0.579	0.573	0.506	0.469	0.506		0.484	0.526
L1VA	0.675	0.720	0.687	0.587	0.679	0.674	0.580	0.699	0.656	0.586	0.703		0.334
WCNC	0.828	0.715	0.777	0.685	0.730	0.716	0.517	0.697	0.657	0.548	0.684	0.514	

Figure 5.1. Locations of the 13 *A. decemloba* ssp. *decemloba* populations included in this study. See Table 5.2 for population codes.

1. PEMA
2. BVMA
3. SNMA
4. PCCT
5. EPRI
6. RCRI
7. SHNY
8. WANY
9. MDNY
10. HPNY
11. SDMD
12. L1VA
13. WCNC
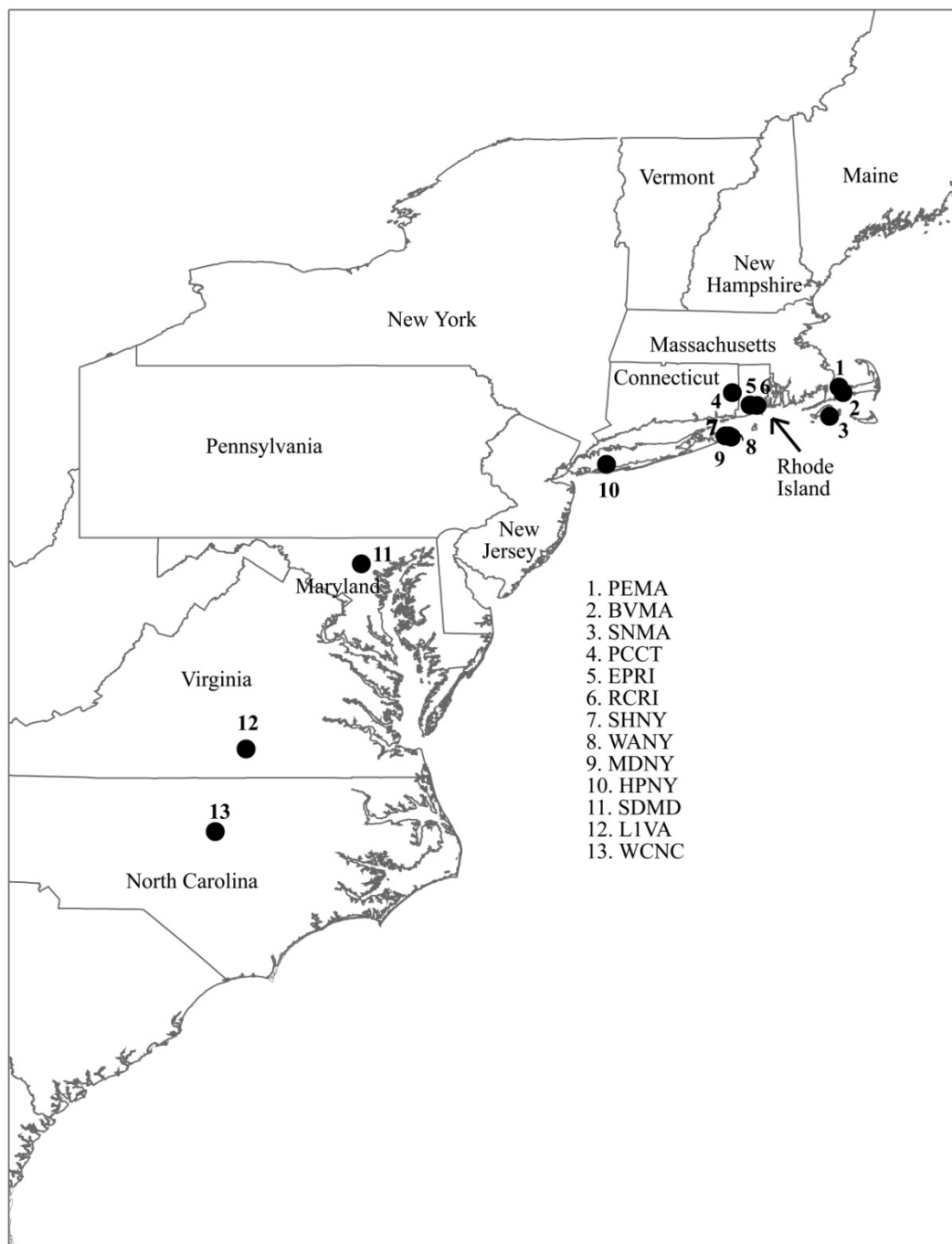
Figure 5.2. The most probable relationships inferred using the program MARK (Ritland 2006) of pairs of individuals sampled within a population (2a) and the distribution of the relationships of intra - and inter-population pairs of individuals (2b).

2a)

.
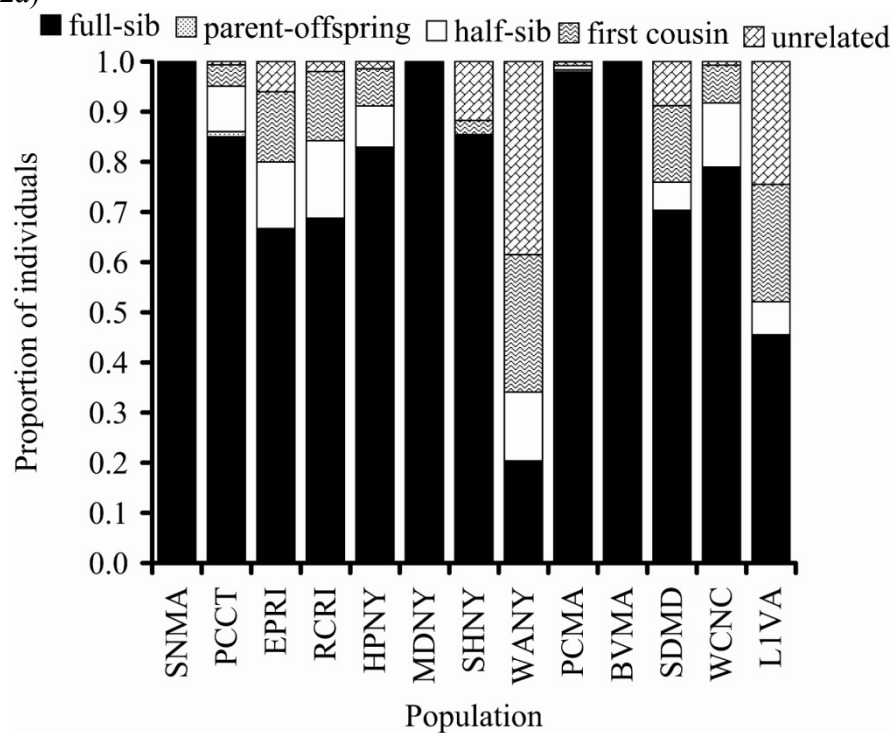
2b)
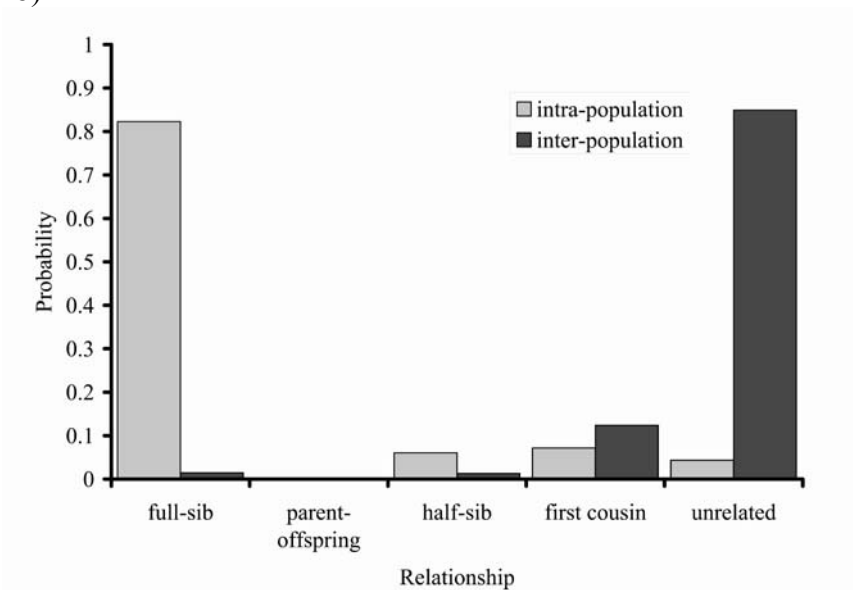
Figure 5.3. Inter-class principal components analysis based on 21 microsatellite loci from 13 populations of *A. decemloba* ssp. *decemloba*. Points represent individuals (i.e., a multilocus genotype) and lines indicate the distance of an individual to the mean of the population from which they were sampled. Circles depict the 95% inertia ellipses and the histogram in the upper right corner shows the relativity of the eigenvalues of each principal component. Individuals from the same population have the same color and population codes are within rectangles.
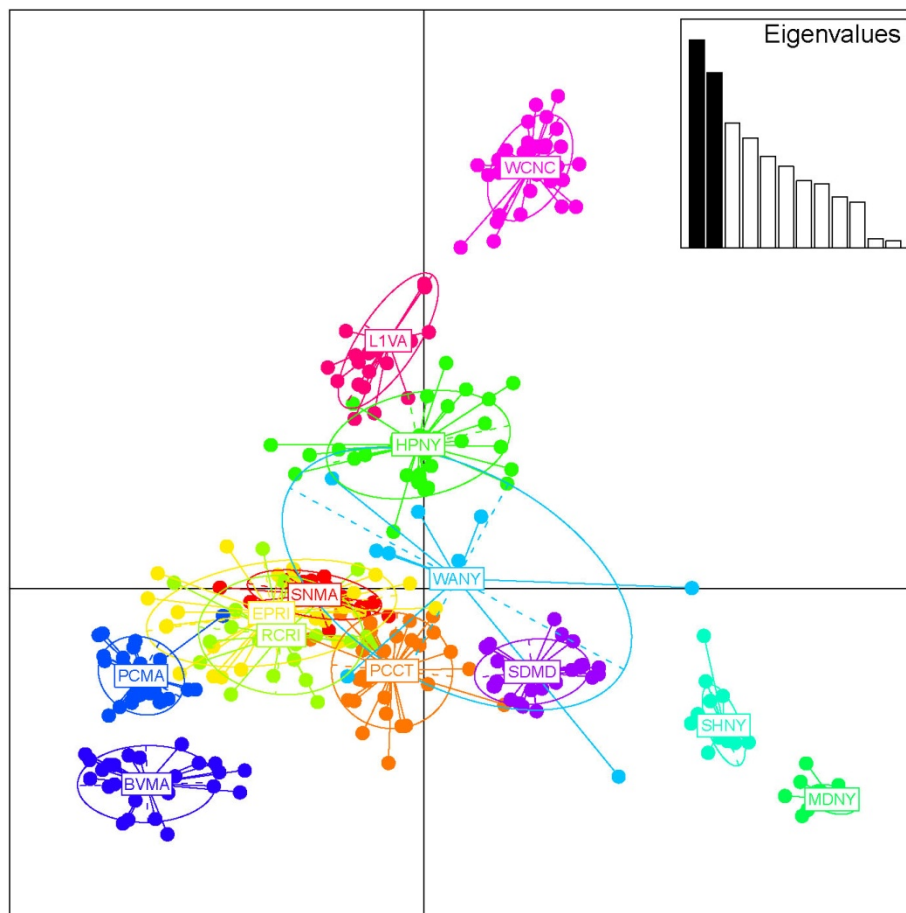
Figure 5.4. Distruct (Rosenberg 2004) diagram of the assignment of individuals to clusters based on the program InStruct where $K = 12$ are the assignments when K was treated as a random variable. Numbers to the right of the figures represent selfing rates within each population and can be interpreted as the proportion of offspring within each cluster that were produced through selfing.
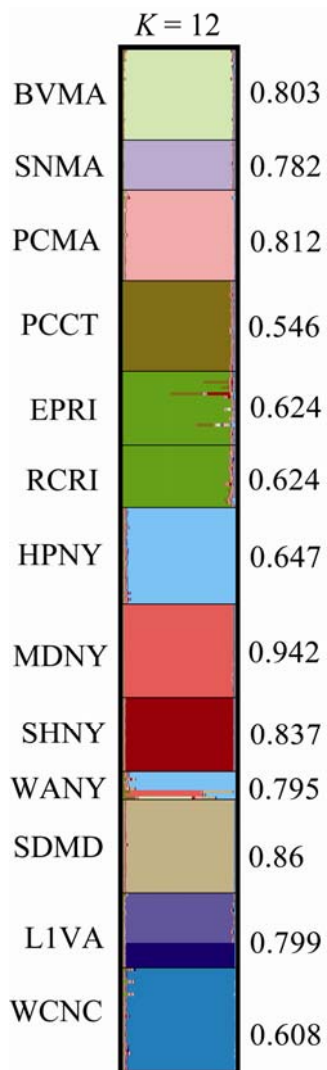
APPENDIX A

General population locations and Genbank accession numbers for loci sampled from North American Agalinis species examined in Chapter 1 (Pettengill and Neel 2008). Section and subsection classifications follow J.M. Canne-Hilliker. Genbank accession numbers for those sequences with "N & C (2004)" can be found in Neel and Cummings (2004).

| Taxon	Sampled Location	*mat*K	*rbc*L	*ndh*F	*rps*2	*trn*T -*trn*F	*trn*H-*psb*A	*rpo*B	ITS
Section Erectae									
A. acuta 125CT	Windham Co., CT				EU827965	EU828128	EU827882	EU828046	EU827810
A. acuta 139RI	Washington Co., RI				EU827966	EU828129	EU827883	EU828047	EU827811
A. acuta 13PCMA	Sandwich, Barnstable Co. MA				EU827967	EU828130	EU827884	EU828048	EU827812
A. acuta 1BVMA	Waquoit Bay, Barnstable Co. MA	N & C (2004)	N & C (2004)	N & C (2004)	EU827968	EU828131	EU827885	EU828049	EU827813
A. acuta 211HPNY	Nassau Co., NY				EU827969	EU828132	EU827886	EU828050	EU827814
A. acuta 229MDNY	Montauk Downs, Suffolk Co. NY				EU827970	EU828133	EU827887	EU828051	EU827815
A. acuta 265SMNY	Shadmoor,Suffolk Co., NY				EU827971	EU828134	EU827888	EU828052	EU827816
A. acuta 33SNMA	Dukes Co., MA				EU827972	EU828135	EU827889	EU828053	EU827817
A. acuta 51MD	Baltimore Co., MD				EU827973	EU828136	EU827890	EU828054	EU827818
A. aphylla 3FL	Liberty Co., FL	N & C (2004)	N & C (2004)	N & C (2004)	EU827974	EU828137	EU827891	EU828055	EU827819
A. aphylla 4AL	Geneva Co., AL				EU827975	EU828138	EU827892	EU828056	EU827820
A. decemloba 6VA	Lunenberg Co., VA	EU828211	EU828220		EU827982	EU828145	EU827899	EU828063	
A. decemloba 9NC	Randolph Co., NC				EU827983	EU828146	EU827900	EU828064	EU827826
A. gattingeri 45MO	Crawford Co., MO		EU828224		EU827993	EU828156	EU827910	EU828074	EU827836
A. gattingeri 8MO	Hickory Co., MO				EU827994	EU828157	EU827911	EU828075	EU827835
A. gattingeri 1LA	Webster Parish, LA				EU827992	EU828155	EU827909	EU828073	EU827823
A. obtusifolia 13AL	Geneva Co., AL				EU828008	EU828171	EU827925	EU828089	EU827849
A. obtusifolia 14AL	Mobile Co., AL			EU828233		EU828172	EU827926	EU828089	EU827851
A. obtusifolia 20FL	Liberty Co., FL	N & C (2004)	N & C (2004)	N & C (2004)	EU828010	EU828174	EU827928	EU828092	EU827852
A. obtusifolia 6AL	Mobile Co., AL				EU828011	EU828175	EU827929	EU828093	
A. obtusifolia 8AL	Geneva Co., AL			EU828234		EU828176	EU827930	EU828094	
A. oligophylla 12AL	Tyler Co., TX				EU828012	EU828177	EU827931	EU828095	
A. oligophylla 1AL	Mobile Co., AL	EU828216	EU828235		EU828013	EU828178	EU827932	EU828096	
A. oligophylla 5AL	Mobile Co., AL				EU828014	EU828179	EU827933	EU828097	

Taxon	Sampled Location	*mat*K	*rbc*L	*ndh*F	*rps*2	*trn*T -*trn*F	*trn*H-*psb*A	*rpo*B	ITS
A. oligophylla 8TX	Vernon Parish, LA				EU828015	EU828180	EU827934	EU828098	
A. skinneriana 106MD	Prince Georges Co., MD		EU828239		EU828028	EU828193	EU827947	EU828110	EU827864
A. skinneriana 78MD	Dorchester Co., MD				EU828029	EU828194	EU827948	EU828111	EU827865
A. skinneriana 90MO	Vernon Co., MO		EU828240		EU828030	EU828195	EU827949	EU828112	EU827866
A. tenella 1GA	Ware Co., GA	EU828215	EU828241		EU828009	EU828173	EU827927	EU828091	EU827850
A. tenella 11GA	Lowndes Co., GA				EU828032	EU828197	EU827951	EU828114	EU827868
A. tenella 13GA	Grady Co., GA				EU828033	EU828198	EU827952	EU828115	EU827869
A. tenella 3SC	Colleton Co., SC				EU828034	EU828199	EU827953	EU828116	EU827870
A. tenella 4GA	Ware Co., GA	N & C (2004)	N & C (2004)		EU828035	EU828200	EU827954	EU828117	EU827871
A. tenella 9GA	Lanier Co., GA				EU828036	EU828201	EU827955	EU828118	EU827872
A. viridis 2LA	Natchitoches Parish, LA	EU828218	EU828242		EU828040	EU828205	EU827959	EU828122	EU827876
A. viridis 9IL	DeSoto Parish, LA				EU828041	EU828206	EU827960	EU828123	EU827877
Section Heterophyllae									
A. auriculata 1IA	Story Co., IA				EU827976	EU828139	EU827893	EU828057	EU827821
A. auriculata 7IL	Will Co., IL	N & C (2004)	N & C (2004)		EU827977	EU828140	EU827894	EU828058	EU827822
A. calycina	Pecos Co., TX		EU828219		EU827978	EU828141	EU827895	EU828059	
A. heterophylla 5TX	Cameron Co., TX				EU827997	EU828160	EU827914	EU828078	
A. heterophylla 8TX	Stephens Co., TX				EU827998	EU828161	EU827915	EU828079	EU827839
A. heterophylla TX	Grimes Co., TX	N & C (2004)	N & C (2004)	N & C (2004)	EU827979	EU828142	EU827896	EU828060	
Section Linifoliae									
A. linifolia 2FL	Liberty Co., FL	N & C (2004)	N & C (2004)	N & C (2004)	EU828003	EU828166	EU827920	EU828084	EU827844
A. linifolia 4GA	Cinch Co., GA		EU828231		EU828004	EU828167	EU827921	EU828085	EU827845
Section Purpureae									
Subsection Pedunculares									
A. edwardsiana 1TX	Stephens Co., TX	EU828212	EU828221		EU827986	EU828149	EU827903	EU828067	EU827829
A. homalantha 1TX	Tyler Co., TX		EU828227		EU827999	EU828162	EU827916	EU828080	EU827840
A. homalantha 2TX	Jasper Co., TX		EU828228		EU828000	EU828163	EU827917	EU828081	EU827841
A. pulchella 3GA	Grady Co., GA	N & C (2004)	N & C (2004)		EU828020	EU828185	EU827939	EU828102	EU827857
A. pulchella 4FL	Florida		EU828237		EU828021	EU828186	EU827940	EU828103	
A. strictifolia 4	Stephens Co., TX				EU828031	EU828196	EU827950	EU828113	EU827867
A. strictifolia TX	Cameron Co., TX				EU827981	EU828144	EU827898	EU828062	EU827825
A. navasotensis 1TX	Tyler Co. TX		EU828232		EU828006	EU828169	EU827923	EU828087	EU827847

Taxon	Sampled Location	*mat*K	*rbc*L	*ndh*F	*rps*2	*trn*T -*trn*F	*trn*H-*psb*A	*rpo*B	ITS
A. navasotensis 5TX	Grimes Co., TX				EU828007	EU828170	EU827924	EU828088	EU827848
Subsection Purpureae									
A. fasciculata 1LA	Grimes Co., TX	EU828213	EU828222		EU827987	EU828150	EU827904	EU828068	EU827830
A. fasciculata 2GA	Long Co., GA	N & C (2004)	N & C (2004)		EU827988	EU828151	EU827905	EU828069	EU827831
A. fasciculata 4LA	Caddo Parish, LA				EU827989	EU828152	EU827906	EU828070	EU827832
A. harperi 13FL	Liberty Co., FL		EU828225		EU827995	EU828158	EU827912	EU828076	EU827837
A. harperi 14NC	Brunswick Co., NC		EU828226		EU827996	EU828159	EU827913	EU828077	EU827838
A.. maritima TX	Cameron Co., Texas				EU827980	EU828143	EU827897	EU828061	EU827824
A. maritima 2MA	Barnstable CO. MA				EU828005	EU828168	EU827922	EU828086	EU827846
A. paupercula 4MA	Barnstable CO. MA				EU828016	EU828181	EU827935	EU828099	EU827853
A. paupercula 7NY	Shadmoor, Suffolk Co. NY				EU828017	EU828182	EU827936	EU828100	EU827854
A. purpurea 101VA	Fauquier Co, VA				EU828022	EU828187	EU827941	EU828104	EU827858
A. purpurea 1AL	Mobile Co., AL	EU828217	EU828238		EU828023	EU828188	EU827942	EU828105	EU827859
A. purpurea 64MD	Dorchester Co., MD				EU828024	EU828189	EU827943 EU827941	EU828106	EU827860
A. purpurea 6SC	Harry Co. SC				EU828025	EU828190	4	EU828107	EU827861
A. tenuifolia 2VA	Prince Edward Co., VA				EU828038	EU828203	EU827957	EU828120	EU827874
A. tenuifolia 5IA	Story Co., Iowa	N & C (2004)	N & C (2004)	N & C (2004)	EU828039	EU828204	EU827958	EU828121	EU827875
A. tenuifolia 10LA	Caddo Parish, LA				EU828037	EU828202	EU827956	EU828119	EU827873
Subsection Setaceae									
A. laxa 3SC	Colleton Co., SC	EU828214	EU828229		EU828001	EU828164	EU827918	EU828082	EU827842
A. laxa 4GA	Long Co., GA		EU828230		EU828002	EU828165	EU827919	EU828083	EU827843
A. plukenettii 2FL	Washington Co., FL	N & C (2004)	N & C (2004)		EU828018	EU828183	EU827937		EU827855
A. plukenettii 4GA	Georgia		EU828236		EU828019	EU828184	EU827938	EU828101	EU827856
A. setacea 3VA	Prince Edward Co., VA	N & C (2004)	N & C (2004)		EU828026	EU828191	EU827945	EU828108	EU827862
A. setacea 7MD	Wicomico Co., MD				EU828027	EU828192	EU827946	EU828109	EU827863
Section Tenuifolieae									
A. filicaulis 5FL	Grady Co., GA		EU828223		EU827991	EU828154	EU827908	EU828072	EU827833
A. filicaulis 1AL	Mobile Co., AL				EU827990	EU828153	EU827907	EU828071	EU827834
A. divaricata 3FL	Liberty Co., FL				EU827985	EU828147	EU827901	EU828065	EU827827
A. divaricata 5FL	Washington Co., FL	N & C (2004)			EU827985	EU828148	EU827902	EU828066	EU827828
Outgroup Species									

Taxon	Sampled Location	*mat*K	*rbc*L	*ndh*F	*rps*2	*trn*T -*trn*F	*trn*H-*psb*A	*rpo*B	ITS
Aureolaria pectinata	Liberty Co., Florida				EU828042	EU828206	EU827961	EU828124	EU827878
Aureolaria pedicularia	Prince Edward Co., VA	N & C (2004)	N & C (2004)	N & C (2004)	EU828043	EU828208	EU827962	EU828125	EU827879
Brachystigma wrightii	Cochise Co., AZ	N & C (2004)	N & C (2004)	N & C (2004)	EU828044	EU828209	EU827963	EU828126	EU827880
Dasistoma macrophylla	Ames, Iowa	N & C (2004)	N & C (2004)	N & C (2004)	EU828045	EU828210	EU827964	EU828127	EU827881

APPENDIX B

Below is the Molecular Ecology Resources article published by Pettengill et al. (2009). It

describes the development and characteristics (e.g., primer sequences and repeat motifs)

of the 21 microsatellites that were used in Chapters 2, 4, and 5.

Characterization of 21 microsatellite loci within *Agalinis acuta* (Orobanchaceae)

and cross-species amplification among closely related taxa

ABSTRACT

We report the isolation and characterization of 21 microsatellites from the federally listed

endangered plant species *Agalinis acuta* (Orobanchaceae). Within *A. acuta*, these loci

show moderate levels of allelic variation (averaging 2.61 alleles per locus) and low levels

of heterozygosity (average observed heterozygosity = 0.177). Because of taxonomic

ambiguity surrounding this listed species, these microsatellites were also tested for cross-

species amplification in five additional congeneric species. In addition to being useful

for evaluating the evolutionary distinctiveness of *A. acuta*, these microsatellites can also

provide information relevant to conservation management strategies by characterizing

genetic diversity within *A. acuta*.

Agalinis acuta Pennell (Family Orobanchaceae) is a fall-blooming annual plant native to eastern North America. Recent phylogenetic evidence has suggested that *A. acuta* may not be distinct from *A. tenella* or *A. decemloba* (Neel & Cummings 2004; Pettengill & Neel 2008). The latter two species have also been synonymized with *A. obtusifolia* (Kartesz 1994), which further complicates the issue of resolving the taxonomic status of *A. acuta*. The microsatellites described here were developed to provide molecular markers suitable for elucidating the relationships among these four closely related species.

Total genomic DNA from putative *A. acuta* samples and the additional species that were evaluated for cross-species amplification success (*A. decemloba*, *A. tenella*, *A. obtusifolia*, *A. setacea*, and *A. skinneriana*) was isolated from fresh or frozen (-80°C) leaves and flower buds by grinding 50-100 mg of tissue to powder in liquid nitrogen with a mortar and pestle, and then using GenElute Plant Genomic DNA Kits (Sigma) or QIAGEN DNEasy Kits (QIAGEN) following manufacturer instructions. Extractions of additional samples were also carried out on a QIAGEN BioSprint 96 robotic workstation using QIAGEN's BioSprint 96 DNA Plant Kit.

Using genomic DNA pooled from two *A. acuta* individuals from Soldiers Delight Natural Environment Area, Baltimore Co., MD, four microsatellite libraries corresponding to four motifs (CA-, AAG-, CAG-, and TACA-) were created by Genetic Identification Services Inc., Chatsworth, CA, USA (GIS; http://www.genetic-id-services.com/) using proprietary magnetic bead capture technology. A total of 244 clones were sequenced (100 by GIS and a subsequent 144 at the University of Maryland College Park (UMCP)). Clones were sequenced on an ABI PRISM® 377 DNA Sequencer using

the DYEnamic™ ET Terminator Cycle Sequencing Kit (Amersham Biosciences) (GIS

protocol) or on an ABI 3730xl DNA Analyzer with a BigDye® Terminator v3.1 Cycle

Sequencing Kit (UMCP). Ninety-five sequences ($n = 48$, GIS; $n = 47$, UMCP) contained

an acceptable number of repeats and sufficient flanking region within which primers

could be designed (48 CA-, 32 AAG-, 15 TACA-). Primers were designed for these

candidate loci using DesignerPCR version 1.03 (Research Genetics, Inc. Huntsville, AL)

(GIS) or PRIMER3 (Rozen & Skaletsky 2000) (UMCP).

Amplification of each candidate microsatellite locus was tested in two individuals

representing two *A. acuta* populations. Upon successful amplification in *A. acuta*, primer

pairs were further tested for cross amplification in at least two individuals from each of

the other putative species (*A. decemloba*, *A. tenella*, *A. obtusifolia*, *A. setacea*, and *A.

skinneriana*). Polymerase chain reactions (PCR) were conducted using a PTC-200

Thermal Cycler (MJ Research, Watertown, MA, USA). Reaction volumes were 10μl and

included 1μl of genomic DNA (2 – 5ng), 1X PCR buffer (Tris·Cl, KCl, $(NH4)_2SO_4$, 15

mM $MgCl_2$, stabilizers (QIAGEN); pH 8.7), 1X Q-solution (QIAGEN), 1 mM $MgCl_2$,

0.25 mM dNTPs, 0.075uM of each primer, and 0.25U of TopTaq DNA polymerase

(QIAGEN). The amplification cycle for all loci was the following touchdown program:

94C, 3min; 30 cycles of (94C, 40sec; 63C, 40sec [-0.3C/cycle]; 72C, 30sec); 5 cycles of

(94C, 40sec; 55C, 40sec; 72C, 30sec); 72C, 10min. Amplification success was

determined through agarose gel electrophoresis and subsequent ethidium bromide

staining.

Loci for which we obtained a product within the expected repeat size range were

further tested in four individuals from each of two *A. acuta* populations to determine if

they yielded polymorphic products that could be reliably scored. Loci that appeared

monomorphic in *A. acuta* were evaluated further for polymorphism across the other taxa.

For fragment analysis we used the same reaction mixture and amplification program

described above, except for the substitution of the fluorescently labeled forward primer

(6-FAM, VIC, or NED from Applied Biosystems). PCR products were electrophoresed

with an ABI 3730xl DNA Analyzer and using GeneScan™ -500 ROX™ Size Standard

(Applied Biosystems). Allele sizes were initially estimated using GENEMAPPER

version 3.7 (Applied Biosystems), but all electropherograms were examined manually

before assigning final genotypes. Loci that could not be scored reliably were no longer

pursued.

The screening process yielded 15 microsatellite loci that were polymorphic in 29 *A.*

acuta samples from Windham Co., CT (Table 1). An additional six loci were

polymorphic when evaluated across all putative species. We successfully genotyped 99%

of the *A. acuta* and 96% of the *A. decemloba* samples (Tables 1 & 2). Amplification

success was slightly lower in *A. tenella* (92%) and substantially lower among the *A.*

obtusifolia (55%) and *A. setacea* (62%) samples (Table 2). All loci failed to amplify

across the majority of individuals in *A. skinneriana*. The lack of any product from the

PCR indicates that failure to acquire genotypes was due to alteration of priming sites.

The computer program ARLEQUIN (Excoffier et al. 2005) was used to calculate

number of alleles, observed (H_o) and expected (H_e) heterozygosity (Tables 1 & 2) and to

test for departures from Hardy-Weinberg equilibrium (HWE) and linkage equilibrium

(LE) among loci within the *A. acuta* population. The *A. acuta* samples had fewer alleles

than the other species (Tables 1 & 2). Certain loci could not be amplified in *A.*

obtusifolia and *A. setacea* (Table 2). After a Bonferroni correction for multiple comparisons (Rice 1988), significant deviations from HWE and LE were observed for 5 loci (Table 1) and 2 pairwise locus combinations (Agac.ca20 - Agac.taca12 and Agac.ca26 - Agac.taca12), respectively. This number of departures from HWE is not surprising given the degree of self-compatibility that has been observed within *A. acuta* (Neel 2002) and the fact that many populations exist as small isolated groups of individuals.

In addition to the fact that the majority of the microsatellites described here cross amplify within *A. decemloba*, *A. tenella*, and *A. obtusifulia*, preliminary analyses have further confirmed their utility in elucidating the evolutionary distinctiveness of *A. acuta* (e.g., presence of private alleles within each of the putative species and monospecific groupings based on genetic distances). The difference in the degree of allelic diversity at the microsatellite loci among these species suggests that the loci will also be useful in additional population genetic and comparative phylogeography studies.

Table 1. Characteristics of 21 microsatellite loci developed for *A. acuta*: locus name, repeat motif, primer sequences, GenBank Accession no., number of alleles, range of PCR product, observed heterozygosity, and expected heterozygosity. Loci that were monomorphic among the *A. acuta* samples are listed because they were polymorphic in the other putative species (Table 2).

Locus	Repeat Motif in Clone	Primer Sequences (5'-3')±	GenBank Accession no.	% success	No. of Alleles	Size (bp)	H_o	H_E
Agac.M1	CA[25]	For – *6FAM* - TTGATGTAGCACCACAAGTCAC Rev – CATGACCAAGTAATTGCAGTCA	FJ754652	100	3	283-302	0.379	0.399
Agac.M6*	CA[22]CT[21]	For – *6FAM* - GGTGATCGACGTAAATGTGA Rev – CGTCAGTTGTACGAAGCAGT	FJ754649	100	4	354-360	0.138	0.407
Agac.M33*	CA[7]T[1]CA[6]	For – *6FAM* - GGAGCTAACGCCTAACAGTAG Rev – ACAAAATGTGTTCTTGAGAGGT	FJ754648	89.66	5	205-219	0.115	0.590
Agac.M11	CA[27]	For – *6FAM* - GGAGCAGAAGTTGGGATTC Rev – AAATGCCTTACGGATGACC	FJ754647	100	3	208-212	0.103	0.220
Agac.M14	CA[31]	For – *NED* - CCTCGACAAGGTAACAAGAA Rev – AAGCTGCTAACGATGACAAC	FJ754646	96.55	5	197-221	0.357	0.485
Agac.M28	CA[13]	For – *NED* - TAAACCAGCCCGCTAAGC Rev – ATCCTGCCGCCAAAATAC	FJ754650	96.55	1	285-285	0.0	n/a
Agac.M42	CA[25]	For – *6FAM* - GTACCCTTCAAGTTTGACCTAA Rev – GCGGTTTTTGGAAATAGAG	FJ754651	100	2	283-285	0.172	0.216
Agac.M46*	CATA[19]	For – *6FAM* - TCGGCAAACTCCAGTGAC Rev – TTGAGCCCATCCTCTGTG	FJ754666	96.55	3	230-254	0.250	0.544
Agac.ca11*	CA[15]	For – *VIC* - GCTTCCTCTTTCCACCTGAGTA Rev – GCCAATGAAAGTCGGTAAGTTG	FJ754653	100	3	150-160	0.001	0.455
Agac.ca20	CA[16]	For – *VIC* - AATTGAGCAGGAATCAAGTCAT Rev – CTGTTTTACGAGAACTGCCTGA	FJ754654	100	2	406-410	0.414	0.503
Agac.ca26	CA[4]CGCACA A[1]TA[3]CA[12]	For – *NED* - AAATGAAACAGTGACCAGGGAA Rev – GCAATCCGAAAAAGATGAGAGC	FJ754655	100	3	248-252	0.379	0.493
Agac.aag46	AAG[10]	For – *VIC* - GTGACGATAAGTCGGTCAATCA Rev – CACAGTCTTACCATGCGAACTA	FJ754656	100	1	475-475	0.0	n/a
Agac.ca10	CA[12]	For – *6FAM* - GCCCTTACTCTCACATTTGCTA Rev – GGTTTGTCGATTGAACCTCTCT	FJ754657	100	3	160-165	0.276	0.272
Agac.aag29	AAG[10]	For – *NED* - TTGACGAAGTAAAGGACATCGG Rev – TCACTATCTCAGACACCGTCAT	FJ754658	100	1	339-339	0.0	n/a
Agac.ca48	CA[12]	For – *6FAM* - CCAGATGCACAGACTCCATAAA Rev – CGTGGGATCAGGTAGATACGTT	FJ754659	100	1	296-296	0.0	n/a

Locus	Repeat	Primer sequence (5'–3')	GenBank			Size		
Agac.ca 21	CA[10]	For – *NED* - TTTGTTGCGTTGAAATCCTCAC Rev – GACTAGACTCCAAGCTCGATCA	FJ754660	100	1	156-156	0.0	n/a
Agac.taca12	TACA[11]	For – *NED* - CGAGATCGAAGGAACAACTTCA Rev – CGATCACAAAGCACGAACTAAC	FJ754661	100	3	442-454	0.379	0.482
Agac.taca45	TACA[35]	For – *6FAM* - TTCCATAATGCCCCCATCAAAT Rev – AGCGACCAAAGTGTATTTTCCT	FJ754662	100	4	329-341	0.552	0.635
Agac.taca04	TACA[21]	For – *6FAM* - CTCACTCCATACAAGGATGCTC Rev – CGTTTCGGTCCGGTTCTC	FJ754663	100	2	331-335	0.103	0.100
Agac.ca45	CA[7]	For – *NED* - CAGTGCTCGTGTTGTATTTTGG Rev – GATTTTGTCAACTTGCTCCACC	FJ754664	100	1	294-294	0.0	n/a
Agac.ca33*	CA[24]	For – *VIC* - GGCTAGTTTGTCCACCATCATA Rev – ACTTAGTAGCATCGTTTGAGCC	FJ754665	100	4	214-284	0.103	0.540

* = Loci that exhibited significant deviations from HWE

[±] = *6FAM*, *NED*, and *VIC* are the fluorescent dyes used to label the forward primer for genotyping

Table 2. Characteristics of 21 microsatellite loci in *A. decemloba, A. tenella, A. obtusifolia,* and *A. setacea.* Numbers in parentheses represent the number of individuals evaluated followed by geographic location.

| | Cross Amplification Success | | | | | | | | | | | |
| | *A. decemloba* (33) (Wake Co. NC) | | | *A. tenella* (30) (Lowndes Co. GA) | | | *A. obtusifolia* (35) (Liberty Co. FL) | | | *A. setacea* (31) (Lunenberg Co. VA) | | |
Locus	Range (bp)	No.of Alleles	% Success	Range (bp)	No.of Alleles	% Success	Range (bp)	No. of Alleles	% Success	Range (bp)	No. of Alleles	% Success
Agac.M1	244-274	6	100	218-296	15	90.00	204-292	9	37.14	200-208	3	100
Agac.M6	346-356	7	100	272-342	13	93.33	n/a	0	0.00	226-246	2	100
Agac.M33	209-209	1	100	203-223	10	96.67	133-215	5	14.29	205-223	9	90.32
Agac.M11	200-223	5	100	168-242	17	90.00	162-166	3	82.86	174-175	2	100
Agac.M14	209-239	6	96.97	195-297	21	90.00	147-313	9	94.29	225-234	4	100
Agac.M28	273-285	3	100	243-293	6	83.33	239-288	10	74.29	241-276	6	61.29
Agac.M42	271-293	6	100	263-295	15	100	271-365	13	51.43	225-245	2	100
Agac.M46	238-266	3	100	200-270	17	96.67	226-310	14	82.86	n/a	0	0.0
Agac.ca 11	148-168	5	100	152-198	18	100	158-196	11	97.14	132-160	7	100
Agac.ca20	408-436	8	100	390-478	23	96.67	440-514	13	71.43	436-436	1	100
Agac.ca26	238-250	2	93.94	250-254	3	100	216-252	12	71.43	238-250	3	16.13
Agac.aag46	457-478	3	100	460-511	10	100	481-494	5	97.14	n/a	0	0.0
Agac.ca10	157-164	3	100	154-160	2	96.67	156-164	5	97.14	152-164	3	41.94
Agac.aag29	339-343	2	81.82	339-351	4	100	n/a	0	0.00	n/a	0	0.0
Agac.ca48	280-300	6	100	284-296	3	83.33	292-294	2	2.86	292-296	3	93.55
Agac.ca 21	150-156	3	100	152-173	9	96.67	144-146	2	94.29	157-157	3	22.58
Agac.taca12	450-470	5	42.42	434-482	6	43.33	n/a	0	0.00	n/a	0	0.0
Agac.taca45	202-337	11	100	178-261	15	96.67	n/a	0	0.00	191-196	3	100
Agac.taca04	311-365	4	93.94	295-381	16	100	335-459	30	94.29	300-348	3	96.77
Agac.ca45	294-294	1	100	292-292	1	86.67	n/a	0	0.00	n/a	0	0.0
Agac.ca33	230-232	2	100	216-246	12	100	214-248	14	91.43	238-248	3	100

APPENDIX C

Data sheet used to record morphological characteristics.

Location: _____
Plant Number: _____ Date: _____ Observers: _____
_____

Plant Characteristics			
Height	Width	# Major Branches	Inflorescence Type
	1: 2:		
Stem Shape:		Stem Surface:	

Leaf Characteristics			
Length		Widest Width	Shape
1: 2: 3: 4: 5:		1: 2: 3: 4: 5:	

Flower Characteristics			
Flower Color	# Flowers in Bloom	# Flowers/Node	Pedicel Length
			1: 2: 3: 4: 5:
Bract Length: 1: 5:	2:	3:	4:

Corolla Characteristics			
Adaxial Tube Length	Abaxial Tube Length	Tube Diameter at Midpoint	Throat Diameter
		H: W:	H: W:
Upper Lobe	Lower Lobe:	Upper Lobe	Lower Lobe Margin
L: W:	L: W:	Margin:	Margin:
Distance from tube to Lobe Tip:		Inner Surface:	Inner Surface:
Style Length: Stigma Length:		Outer Surface	Outer Surface

Stamen Characteristics			
Dorsal Anther Size	D. Filament Length	Ventral Anther Size	V. Filament Length
L: W:		L: W:	

Calyx Characteristics			
	Adaxial Tube Length	Abaxial Tube Length	Tube Diameter at Midpoint
			H: W:
	Upper Lobe		Lower Lobe:
Flower	L: W:		L: W:
Fruit	L: W:		L: W:

BIBLIOGRAPHY

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

Akaike, H. 1974. New look at statistical model identification. IEEE Transactions on Automatic Control **19**:716-723.

Alig, R. J., J. D. Kline, and M. Lichtenstein. 2004. Urbanization on the US landscape: looking ahead in the 21st century. Landscape and Urban Planning **69**:219-234.

Alvarez, I., and J. F. Wendel. 2003. Ribosomal ITS sequences and plant phylogenetic inference. Molecular Phylogenetics and Evolution **29**:417-434.

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

Anderson, E. C., and K. K. Dunham. 2008. The influence of family groups on inferences made with the program Structure. Molecular Ecology Resources **8**:1219-1229.

Andreasen, K. 2005. Implications of molecular systematic analyses on the conservation of rare and threatened taxa: Contrasting examples from Malvaceae. Conservation Genetics **6**:399-412.

Andrianoelina, O., B. Favreau, L. Ramamonjisoa, and J. M. Bouvet. 2009. Small effect of fragmentation on the genetic diversity of *Dalbergia monticola,* an endangered tree species of the eastern forest of Madagascar, detected by chloroplast and nuclear microsatellites. Annals of Botany **104**.

Armstrong, K. F., and S. L. Bar. 2005. DNA barcodes for biosecurity: invasive species identification. Philosophical Transactions of the Royal Society B-Biological Sciences **360**:2373-2373.

Avise, J. C., J. Arnold, R. M. Ball, E. Bermingham, T. Lamb, J. E. Neigel, C. A. Reeb, and N. C. Saunders. 1987. Intraspecific phylogeography - the mitochondrial-DNA bridge between population genetics and systematics. Annual Review of Ecology and Systematics **18**:489-522.

Bacon, C. D., and C. D. Bailey. 2006. Taxonomy and conservation: A case study from *Chamaedorea alternans*. Annals of Botany **98**:755-763.

Baldwin, B. G. 2007. Adaptive radiation of shrubby tarweeds (*Deinandra*) in the California Islands parallels diversification of the Hawaiian silversword alliance (Compositae-Madiinae). American Journal of Botany **94**:237-248.

Barbara, T., C. Palma-Silva, G. M. Paggi, F. Bered, M. F. Fay, and C. Lexer. 2007. Cross-species transfer of nuclear microsatellite markers: potential and limitations. Molecular Ecology **16**:3759-3767.

Barraclough, T. G., and S. Nee. 2001. Phylogenetics and speciation. Trends in Ecology and Evolution **16**:391-399.

Baum, D. A. 1998. Individuality and the existence of species through time. Systematic Biology **47**:641-653.

Baum, D. A. 2009. Species as ranked taxa. Systematic Biology **58**:74-86.

Baum, D. A., and K. L. Shaw. 1995. Genealogical perspectives on the species problem. Pages 289-303. Monographs in Systematic Botany: Experimental and molecular approaches to plant biosystematics. Missouri Botanical Garden, St. Louis.

Baumel, A., M. L. Ainouche, R. J. Bayer, A. K. Ainouche, and M. T. Misset. 2002. Molecular phylogeny of hybridizing species from the genus *Spartina* Schreb. (Poaceae). Molecular Phylogenetics and Evolution **22**:303-314.

Bazinet, A. L., and M. P. Cummings. in press. The Lattice Project: a grid research and production environment combining multiple grid computing models in M. H. W. Weber, editor. Distributed & Grid Computing – Science Made Transparent for Everyone. Principles, Applications and Supporting Communities. Tectum Publishing House, Marburg.

Bazinet, A. L., D. S. Myers, J. Fuetsch, and M. P. Cummings. 2007. Grid Services Base Library: A high-level, procedural application programming interface for writing Globus-based Grid services. Future Generation Computer Systems-The International Journal of Grid Computing Theory Methods and Applications **23**:517-522.

Beardsley, P. M., S. E. Schoenig, J. B. Whittall, and R. G. Olmstead. 2004. Patterns of evolution in western North American *Mimulus* (Phrymaceae). American Journal of Botany **91**:474-489.

Bennett, J. R., and S. Mathews. 2006. Phylogeny of the parasitic plant family Orobanchaceae inferred from phytochrome A. American Journal of Botany **93**:1039-1051.

Bicknell, E. P. 1915. *Agalinis acuta*. Bulletin of the Torrey Botanical Club **42**:338.

Blaxter, M. L. 2004. The promise of a DNA taxonomy. **359**:669-679.

Broughton, R. E., and R. G. Harrison. 2003. Nuclear gene genealogies reveal historical, demographic and selective factors associated with speciation in field crickets. Genetics **163**:1389-1401.

Brown, M. L., and R. G. Brown 1984. Herbaceous plants of Maryland. Port City Press, Baltimore, Maryland.

Broyles, S. B. 1998. Postglacial migration and the loss of allozyme variation in northern populations of *Asclepias exaltata* (Asclepiadaceae). American Journal of Botany **85**:1091-1097.

Brumback, W. E., and S. L. Kelley. 1990. *Agalinis acuta* cultivation and seed bank research. Page 21. New England Wildflower Society, Framingham, Massachusetts.

Brunsfeld, S. J., D. E. Soltis, and P. S. Soltis. 1991. Patterns of Genetic-Variation in *Salix* Section Longifoliae (Salicaceae). American Journal of Botany **78**:855-869.

Burnham, K. D., and D. R. Anderson 1998. Model selection and inference: a practical information-theoretic approach. Springer-Verlag, New York, New York, USA.

Burns, J. M., D. H. Janzen, M. Hajibabaei, W. Hallwachs, and P. D. N. Hebert. 2008. DNA barcodes and cryptic species of skipper butterflies in the genus *Perichares* in Area de Conservacion Guanacaste, Costa Rica. Proceedings of the National Academy of Sciences of the United States of America **105**:6350-6355.

Byers, D. L., and D. M. Waller. 1999. Do plant populations purge their genetic load? Effects of population size and mating history on inbreeding depression. Annual Review of Ecology and Systematics **30**:479-513.

Cain, M. L., H. Damman, and A. Muir. 1998. Seed dispersal and the Holocene migration of woodland herbs. Ecological Monographs **68**:325-347.

Canne-Hilliker, J. M. 1987. Patterns of floral development in *Agalinis* and allies (Scrophulariaceae). II. Floral development of *Agalinis densiflora*. American Journal of Botany **74**:1419-1430.

Canne-Hilliker, J. M. 1988. *Agalinis* (Scrophulariaceae) in Peru and Bolivia. Brittonia **40**:433-440.

Canne-Hilliker, J. M., and M. Dubrule. 1993. A new species of *Agalinis* (Scrophulariaceae) from Grimes County, Texas. SIDA **15**:425-440.

Canne-Hilliker, J. M., and C. M. Kampny. 1991. Taxonomic significance of leaf and stem anatomy of *Agalinis* (Scrophulariaceae) from the U.S.A. and Canada. Canadian Journal of Botany **69**:1935-1950.

Canne, J. M. 1979. A light and scanning electron microscope study of seed morphology in *Agalinis* (Scrophulariaceae) and its taxonomic significance. Systematic Botany **4**:281-296.

Canne, J. M. 1980. Seed surface features in *Aureolaria, Brachystigma, Tomanthera*, and certain South American *Agalinis* (Scrophulariaceae). Systematic Botany **5**:241-252.

Canne, J. M. 1981. Chromosome counts in *Agalinis* and related taxa (Scrophulariaceae). **59**:1111-1116.

Canne, J. M. 1982. The taxonomic significance of seedling morphology in *Agalinis* (Scrophulariaceae). Canadian Journal of Botany **61**:1868-1874.

Canne, J. M. 1983. The taxonomic significance of seedling morphology in *Agalinis* (Scrophulariaceae). Canadian Journal of Botany **62**:454-456.

Canne, J. M. 1984. Chromosome numbers and the taxonomy of North American *Agalinis* (Scrophulariaceae). Canadian Journal of Botany **62**:454-456.

Carroll, S. P., and C. W. Fox, editors. 2008. Conservation Biology Evolution in Action. Oxford University Press, Oxford.

Carstens, B. C., and L. L. Knowles. 2007. Estimating species phylogeny from gene-tree probabilities despite incomplete lineage sorting: An example from Melanoplus grasshoppers. Systematic Biology **56**:400-411.

Case, M. A., H. T. Mlodozeniec, L. E. Wallace, and T. W. Weldy. 1998. Conservation genetics and taxonomic status of the rare Kentucky lady's slipper: *Cypripedium kentuckiense* (Orchidaceae). American Journal of Botany **85**:1779-1786.

CBOL Plant Working Group. 2009. A DNA barcode for land plants. Proceedings of the National Academy of Sciences of the United States of America **106**:12794-12797.

Channell, R., and M. V. Lomolino. 2000. Dynamic biogeography and conservation of endangered species. Nature **403**:84-86.

Chapuis, M. P., and A. Estoup. 2007. Microsatellite null alleles and estimation of population differentiation. Molecular Biology and Evolution **24**:621-631.

Charlesworth, B., and D. Charlesworth. 1999. The genetic basis of inbreeding depression. Genetical Research **74**:329-340.

Chase, M. W., N. Salamin, M. Wilkinson, J. M. Dunwell, R. P. Kesanakurthi, N. Haidar, and V. Savolainen. 2005. Land plants and DNA barcodes: short-term and long-term goals. Philosophical Transactions of the Royal Society B-Biological Sciences **360**:1889-1895.

Clark, J. S., C. Fastie, G. Hurtt, S. T. Jackson, C. Johnson, G. A. King, M. Lewis, J. Lynch, S. Pacala, C. Prentice, E. W. Schupp, T. Webb, and P. Wyckoff. 1998. Reid's paradox of rapid plant migration - Dispersal theory and interpretation of paleoecological records. Bioscience **48**:13-24.

Claussen, J., D. D. Keck, and W. M. Hiesey 1940. Experimental studies on the nature of species. I. The effect of varied environment on western North American plants. Carnegie Institution of Washington, Washington, D.C.

Clegg, M. T. 1980. Measuring plant mating systems. BioScience **30**:814-818.

Comes, H. P., and R. J. Abbott. 2001. Molecular phylogeography, reticulation, and lineage sorting in Mediterranean *Senecio* sect. *Senecio* (Asteraceae). Evolution **55**:1943-1962.

Comes, H. P., and J. W. Kadereit. 1998. The effect of Quaternary climatic changes on plant distribution and evolution. Trends in Plant Science **3**:432-438.

Coyne, J. A., and H. A. Orr 2004. Speciation. Sinauer Associates, Inc., Sunderland, Massachusetts.

Cracraft, J. 1989. Speciation and its ontology: The empirical consequences of alternative species concepts for understanding patterns and processes of differentiation. Pages 28-59 in D. Otte, and J. A. Endler, editors. Speciation and Its Consequences. Sinauer Associates, Sunderland, MA.

Crandall, K. A., O. R. P. Bininda-Emonds, G. M. Mace, and R. K. Wayne. 2000. Considering evolutionary processes in conservation biology. Trends in Ecology and Evolution **15**:290-295.

Crawford, N. G. accepted. SMOGD: Software for the Measurement of Genetic Diversity. Molecular Ecology Resources.

Crawford, N. G. submitted. SMOGD: Software for the Measurement of Genetic Diversity. Molecular Ecology Resources.

Crisp, M. D., and G. T. Chandler. 1996. Paraphyletic species. Telopea **6**:813-844.

Crnokrak, P., and S. C. H. Barrett. 2002. Purging the genetic load: A review of the experimental evidence. Evolution **56**:2347-2358.

Crutzen, P. J. 2002. Geology of mankind. Nature **415**:23-23.

Cuénoud, P., V. Savolainen, L. W. Chatrou, M. Powell, R. J. Grayer, and M. W. Chase. 2002. Molecular phylogenetics of Caryophyllales based on nuclear 18S rDNA and plastid *rbc*L, *atp*B, and *mat*K DNA sequences. American Journal of Botany **89**:132-144.

Culley, T. M., S. J. Sbita, and A. Wick. 2007. Population genetic effects of urban habitat fragmentation in the perennial herb *Viola pubescens* (Violaceae) using ISSR markers. Annals of Botany **100**:91-100.

Cummings, M. P., M. C. Neel, and K. L. Shaw. 2008. A genealogical approach to quantifying lineage divergence. Evolution **62**:2411-2422.

Cwynar, L. C., and G. M. Macdonald. 1987. Geographical variation of lodgepole pine in relation to population history. American Naturalist **129**:463-469.

Darling, J. A., and M. J. Blum. 2007. DNA-based methods for monitoring invasive species: a review and prospectus. Biological Invasions **9**:751-765.

Davis, M. B., and R. G. Shaw. 2001. Range shifts and adaptive responses to Quaternary climate change. Science **292**:673-679.

Dawnay, N., R. Ogden, R. McEwing, G. R. Carvalho, and R. S. Thorpe. 2007. Validation of the barcoding gene COI for use in forensic genetic species identification. Forensic Science International **173**:1-6.

de Oliveira, L., F. de Sousa, T. Wendt, G. K. Brown, D. E. Tuthill, and T. M. Evans. 2007. Monophyly and phylogenetic relationships in *Lymania* (Bromeliaceae :

Bromelioideae) based on morphology and chloroplast DNA sequences. Systematic Botany **32**:264-270.

de Queiroz, K. 2007. Species concepts and species delimitation. Systematic Biology **56**:879 - 886.

de Queiroz, K., and M. J. Donoghue. 1988. Phylogenetic systematics and the species problem. Cladistics **4**:317-338.

dePamphilis, C. W., N. D. Young, and A. D. Wolfe. 1997. Evolution of plastid gene *rps2* in a lineage of hemiparasitic and holoparasitic plants: Many losses of photosynthesis and complex patterns of rate variation. Proceedings of the National Academy of Sciences of the United States of America **94**:7367-7372.

DeSalle, R., M. G. Egan, and M. Siddall. 2005. The unholy trinity: taxonomy, species delimitation and DNA barcoding. Philosophical Transactions of the Royal Society B-Biological Sciences **360**:1905-1916.

Dieringer, G. 1991. Variation in individual flowering time and reproductive success of *Agalinis strictifolia* (Scrophulariaceae). American Journal of Botany **78**:497-503.

Dieringer, G. 1999. Reproductive biology of *Agalinis skinneriana* (Scrophulariaceae), a threatened species. Journal of the Torrey Botanical Society **126**:289-295.

Dobzhansky, T. 1937. Genetics and the origin of species. Columbia University Press, New York, New York.

Dole, J. A. 1990. Role of corolla abscission in delayed self-pollination of *Mimulus guttatus* (Scrophulariaceae). American Journal of Botany **77**:1505-1507.

Dolédec, S., and D. Chessel. 1987. Rythmes saisonniers et composantes stationnelles en milieu aquatique I- Description d'un plan d'observations complet par projection de variables. Acta Oecologica-Oecologia Generalis **8**:403–426.

Donoghue, M. J. 1985. A critique of the biological species concept and recommendations for a phylogenetic alternative. Bryologist **88**:172-181.

Dueck, L. A., and K. M. Cameron. 2008. Molecular evidence on the species status and phylogenetic relationships of *Spiranthes parksii*, an endangered orchid from Texas. Conservation Genetics **9**:1617-1631.

Eckert, C. G., K. E. Samis, and S. C. Lougheed. 2008. Genetic variation across species' geographical ranges: the central-marginal hypothesis and beyond. Molecular Ecology **17**:1170-1188.

Edgar, R. C. 2004a. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics **5**:1-19.

Edgar, R. C. 2004b. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research **32**:1792-1797.

Edwards, C. E., D. Lefkowitz, D. E. Soltis, and P. S. Soltis. 2008a. Phylogeny of *Conradina* and related southeastern scrub mints (Lamiaceae) based on *Gap*C gene sequences. International Journal of Plant Sciences **169**:579-594.

Edwards, C. E., D. E. Soltis, and P. S. Soltis. 2008b. Using patterns of genetic structure based on microsatellite loci to test hypotheses of current hybridization, ancient hybridization and incomplete lineage sorting in *Conradina* (Lamiaceae). Molecular Ecology **17**:5157-5174.

Edwards, D., A. Horn, D. Taylor, V. Savolain, and J. A. Hawkins. 2008c. DNA barcoding of a large genus, *Aspalathus* L. (Fabaceae). Taxon **57**:1317-1327.

Elias-Gutierrez, M., and M. Valdez-Moreno. 2008. A now cryptic species of *Leberis* Smirnov, 1989 (Crustacea, Cladocera, Chydoridae) from the Mexican semi-desert region, highlighted by DNA barcoding. Hidrobiologica **18**:63-74.

Ellstrand, N. C., and D. R. Elam. 1993. Population genetic consequences of small population size: implications for plant conservation. Annual Review of Ecology and Systematics **24**:217-242.

Erickson, D. L., J. Spouge, A. Resch, L. A. Weigt, and W. J. Kress. 2008. DNA barcoding in land plants: developing standards to quantify and maximize success. Taxon **57**:1304-1316.

Estill, J. C., and M. B. Cruzan. 2001. Phytogeography of rare plant species endemic to the southeastern United States. Castanea **66**:3-23.

Etterson, J. R. 2004a. Evolutionary potential of *Chamaecrista fasciculata* in relation to climate change. 1. Clinal patterns of selection along an environmental gradient in the great plains. Evolution **58**:1446-1458.

Etterson, J. R. 2004b. Evolutionary potential of *Chamaecrista fasciculata* in relation to climate change. II. Genetic architecture of three populations reciprocally planted along an environmental gradient in the great plains. Evolution **58**:1459-1471.

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

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

Faith, D. P., and A. M. Baker. 2006. Phylogenetic diversity (PD) and biodiversity conservation: some bioinformatics challenges. Evolutionary Bioinformatics Online **2**:70-77.

Fallon, S. M. 2007. Genetic data and the listing of species under the U.S. Endangered Species Act. Conservation Biology **21**:1186-1195.

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

Fazekas, A. J., K. S. Burgess, P. R. Kesanakurti, S. W. Graham, S. G. Newmaster, B. C. Husband, D. M. Percy, M. Hajibabaei, and S. C. H. Barrett. 2008. Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. PLoS ONE **3**.

Fazekas, A. J., P. R. Kesanakurti, K. S. Burgess, D. M. Percy, S. W. Graham, S. C. H. Barrett, S. G. Newmaster, M. Hajibabaei, and B. C. Husband. 2009. Are plant species inherently harder to discriminate than animal species using DNA barcoding markers? Molecular Ecology Resources **9**:130-139.

Ferguson, J. W. H. 2002. On the use of genetic divergence for identifying species. Biological Journal of the Linnean Society **75**:509-516.

Foster, D. R. 1992. Land-use history (1730-1990) and vegetation dynamics in central New England, USA. Journal of Ecology **80**:753-772.

Foster, D. R., G. Motzkin, D. Bernardos, and J. Cardoza. 2002. Wildlife dynamics in the changing New England landscape. Journal of Biogeography **29**:1337-1357.

Frankham, R. 1995. Conservation genetics. Annual Review of Genetics **29**:305-327.

Frankham, R., J. D. Ballou, and D. A. Briscoe 2002. Introduction to Conservation Genetics. Cambridge University Press, Cambridge, U.K.

Funk, D. J., and K. E. Omland. 2003. Species-level paraphyly and polyphyly: frequency, causes and consequences with insights from animal mitochondrial DNA. Annual Review of Ecology, Evolution, and Systematics **34**:397-423.

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

Gay, L., G. Neubauer, M. Zagalska-Neubauer, J. M. Pons, D. A. Bell, and P. A. Crochet. 2009. Speciation with gene flow in the large white-headed gulls: does selection counterbalance introgression? Heredity **102**:133-146.

Gelman, A., and D. B. Rubin. 1992. Inference from iterative simulation using multiple sequences (with discussion). Statistical Science **7**:457–511.

Gerlach, J. D., B. S. Bushman, J. K. McKay, and H. Meimberg. 2009. Taxonomic confusion permits the unchecked invasion of vernal pools in California by low mannagrass (*Glyceria declinata*). Invasive Plant Science and Management **2**:92-97.

Goldstein, P. Z., and R. DeSalle. 2000. Phylogenetic species, nested hierarchies, and character fixation. Cladistics-the International Journal of the Willi Hennig Society **16**:364-384.

Goldstein, P. Z., R. DeSalle, G. Amato, and A. P. Vogler. 2000. Conservation genetics at the species boundary. Conservation Biology **14**:120-131.

Gonzales, E., and J. L. Hamrick. 2005. Distribution of genetic diversity among disjunct populations of the rare forest understory herb, *Trillium reliquum*. Heredity **95**:306-314.

Goudet, J. 1995. FSTAT (Version 1.2): A computer program to calculate F-statistics. Journal of Heredity **86**:485-486.

Gould, S. J., and R. F. Johnston. 1972. Geographic variation. Annual Review of Ecology and Systematics **3**:457-498.

Greene, E. L. 1899. Scrophulariaceae *Gerardia decemloba*. Pittonia **4**:51.

Griffin, S. R., and S. C. H. Barrett. 2004. Post-glacial history of *Trillium grandiflorum* (Melanthiaceae) in eastern North America: inferences from phylogeography. Am. J. Bot. **91**:465-473.

Griffith, B., J. M. Scott, J. W. Carpenter, and C. Reed. 1989. Translocation as a species conservation tool - status and strategy. Science **245**:477-480.

Gu, X., Y. X. Fu, and W. H. Li. 1995. Maximum-likelihood-estimation of the heterogeneity of substitution rate among nucleotide sites. Molecular Biology and Evolution **12**:546-557.

Haig, S. 1998. Molecular contributions to conservation. Ecology **79**:413-425.

Haig, S. M., E. A. Beever, S. M. Chambers, H. M. Draheim, B. D. Dugger, S. Dunham, E. Elliott-Smith, J. B. Fontaine, D. C. Kesler, B. J. Knaus, I. F. Lopes, P. Loschl, T. D. Mullins, and L. M. Sheffield. 2006. Taxonomic considerations in listing subspecies under the US Endangered Species Act. Conservation Biology **20**:1584-1594.

Hall, B., G. Motzkin, D. R. Foster, M. Syfert, and J. Burk. 2002. Three hundred years of forest and land-use change in Massachusetts, USA. Journal of Biogeography **29**:1319-1335.

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

Hamrick, J. L., and M. J. W. Godt. 1989. Allozyme diversity in plant species. Pages 43-63 in A. H. D. Brown, M. T. Clegg, A. L. Kahler, and B. S. Weir, editors. Plant population genetics; breeding and genetic resources. Sinauer, Sunderland, Massachusetts, USA.

Hamrick, J. L., and M. J. W. Godt. 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.

Hamrick, J. L., Y. B. Linhart, and J. B. Mitton. 1979. Relationship between life history characteristics and electrophoretically detectable genetic variation in plants. Annual Review of Ecology and Systematics **10**:173-200.

He, T. H., B. B. Lamont, S. L. Krauss, N. J. Enright, B. P. Miller, and A. D. Gove. 2009. Ants cannot account for interpopulation dispersal of the arillate pea *Daviesia triflora*. New Phytologist **181**:725-733.

Hebert, P. D. N., A. Cywinska, S. L. Ball, and J. R. DeWaard. 2003. Biological identifications through DNA barcodes. Proceedings of the Royal Society of London Series B-Biological Sciences **270**:313-321.

Hedrick, P. W. 1994. Purging inbreeding depression and the probability of extinction: full-sib mating. Heredity **73**:363-372.

Hedrick, P. W. 1999. Perspective: Highly variable loci and their interpretation in evolution and conservation. Evolution **53**:313-318.

Heller, R., and H. R. Siegismund. 2009. Relationship between three measures of genetic differentiation G_{ST}, D_{EST}, and G'_{ST}: How wrong have we been? Molecular Ecology **18**:2080-2083.

Hendry, A. P. 2002. $Q_{ST} >= \neq < F_{ST}$? Trends in Ecology & Evolution **17**:502-502.

Herlihy, C. R., and C. G. Eckert. 2002. Genetic cost of reproductive assurance in a self-fertilizing plant. Nature **416**:320-323.

Hewitt, G. 2000. The genetic legacy of the Quaternary ice ages. Nature **405**:907-913.

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

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

Hey, J., and J. Wakeley. 1997. A coalescent estimator of the population recombination rate. Genetics **145**:833-846.

Holder, M. T., J. A. Anderson, and A. K. Holloway. 2001. Difficulties in detecting hybridization. Systematic Biology **50**:978-982.

Hollingsworth, M. L., A. A. Clark, L. L. Forrest, J. Richardson, R. T. Pennington, D. G. Long, R. Cowan, M. W. Chase, M. Gaudeul, and P. M. Hollingsworth. 2009. Selecting barcoding loci for plants: evaluation of seven candidate loci with species-level sampling in three divergent groups of land plants. Molecular Ecology Resources **9**:439-457.

Holsinger, K. E. 1988. Inbreeding depression doesn't matter: the genetic basis of mating-system evolution. Evolution **42**:1235-1244.

Holsinger, K. E. 2000. Demography and extinction in small populations in A. Young, and G. Clarke, editors. Genetics, Demography, and the Viability of Populations. Cambridge University Press, Cambridge, England.

Honnay, O., D. Adriaens, E. Coart, H. Jacquemyn, and I. Roldan-Ruiz. 2007. Genetic diversity within and between remnant populations of the endangered calcareous grassland plant *Globularia bisnagarica* L. Conservation Genetics **8**:293-303.

Hudson, R. R., and J. A. Coyne. 2002. Mathematical consequences of the genealogical species concept. Evolution **56**:1557-1565.

Husband, B. C., and D. W. Schemske. 1996. Evolution of the magnitude and timing of inbreeding depression in plants. Evolution **50**:54-70.

Hutchison, D. W., and A. R. Templeton. 1999. Correlation of pairwise genetic and geographic distance measures: Inferring the relative influences of gene flow and drift on the distribution of genetic variability. Evolution **53**:1898-1914.

Ibrahim, K. M., R. A. Nichols, and G. M. Hewitt. 1996. Spatial patterns of genetic variation generated by different forms of dispersal during range expansion. Heredity **77**:282-291.

Ingvarsson, P. K. 2001. Restoration of genetic variation lost - the genetic rescue hypothesis. Trends in Ecology & Evolution **16**:62-63.

Isaac, N. J. B., J. Mallet, and G. M. Mace. 2004. Taxonomic inflation: its influence on macroecology and conservation. Trends in Ecology & Evolution **19**:464-469.

Jain, S. K. 1976. Evolution of inbreeding in plants. Annual Review of Ecology and Systematics **7**:469-495.

James, F. C., and C. E. McCulloch. 1990. Multivariate analysis in ecology and systematics: Panacea or Pandora's box? Annual Review of Ecology and Systematics **21**:129-166.

Jombart, T. 2008. *adegenet*: a R package for the multivariate analysis of genetic markers. Bioinformatics **24**:1403-1405.

Jost, L. 2008. G_{ST} and its relatives do not measure differentiation. Molecular Ecology **17**:4015-4026.

Kalisz, S., D. W. Vogler, and K. M. Hanley. 2004. Context-dependent autonomous self-fertilization yields reproductive assurance and mixed mating. Nature **430**:884-887.

Kampny, C. M., and J. M. Canne-Hilliker. 1987. Patterns of floral development in *Agalinis* and allies (Scrophulariaceae) I. Floral development of *Agalinis fasciculata* and *A. tenuifolia*). Canadian Journal of Botany **65**:2255-2262.

Kaplan, N. L., T. Darden, and R. R. Hudson. 1988. The coalescent process in models with selection. Genetics **120**:819-829.

Kartesz, J. T. 1994. A synonymized checklist of the vascular flora of the United States, Canada, and Greenland. Timber Press, Portland, OR.

Kartesz, J. T. 1999. A synonymized checklist and atlas with biological attributes for the vascular flora of the United States, Canada, and Greenland in J. T. Kartesz, and C. A. Meacham, editors. Synthesis of the North American Flora, Version 1.0. North Carolina Botanical Garden, Chapel Hill, NC.

Kimball, R. T., D. J. Crawford, and E. B. Smith. 2003. Evolutionary processes in the genus *Coreocarpus*: Insights from molecular phylogenetics. Evolution **57**:52-61.

Kimura, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide-sequences. Journal of Molecular Evolution **16**:111-120.

Kimura, M. 1983. The neutral theory of molecular evolution. Cambridge University Press, Cambridge.

King, T. L., J. F. Switzer, C. L. Morrison, M. S. Eackles, C. C. Young, B. A. Lubinski, and P. Cryan. 2006. Comprehensive genetic analyses reveal evolutionary distinction of a mouse (*Zapus hudsonius preblei*) proposed for delisting from the US Endangered Species Act. Molecular Ecology **15**:4331-4359.

Knowles, L. L., and B. C. Carstens. 2007. Delimiting species without monophyletic gene trees. Systematic Biology **56**:887-896.

Kress, J. W., and D. L. Erickson. 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**:e508.

Kress, W. J., K. J. Wurdack, E. A. Zimmer, L. A. Weigt, and D. H. Janzen. 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.

Kuhner, M. K. 2006. LAMARC 2.0: Maximum likelihood and Bayesian estimation of population parameters. Bioinformatics **22**:768-770.

Kumar, S., K. Tamura, and M. Nei. 2004. MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. Briefings in Bioinformatics **5**:150-163.

Lahaye, R., M. Van der Bank, D. Bogarin, J. Warner, F. Pupulin, G. Gigot, O. Maurin, S. Duthoit, T. G. Barraclough, and V. Savolainen. 2008. DNA barcoding the floras of biodiversity hotspots. Proceedings of the National Academy of Sciences of the United States of America **105**:2923-2928.

Lande, R. 1988. Genetics and demography in biological conservation. Science **241**:1455-1460.

Langella, O. 2002. Populations v1.2.28. http://bioinformatics.org/~tryphon/populations/.

Leache, A. D., M. S. Koo, C. L. Spencer, T. J. Papenfuss, R. N. Fisher, and J. A. McGuire. 2009. Quantifying ecological, morphological, and genetic variation to delimit species in the coast horned lizard species complex (Phrynosoma). Proceedings of the National Academy of Sciences of the United States of America **106**:12418-12423.

Legendre, P., and A. Vaudor. 1991. Le Progiciel R: analyse multidimensionnelle, analyse spatiale. Page Version MacIntosh. Université de Montréal, Montréal, Québec.

Leimu, R., P. Mutikainen, J. Koricheva, and M. Fischer. 2006. How general are positive relationships between plant population size, fitness and genetic variation? Journal of Ecology **94**:942-952.

Lenormand, T. 2002. Gene flow and the limits to natural selection. Trends in Ecology & Evolution **17**:183-189.

Lesica, P., and F. W. Allendorf. 1995. When are peripheral populations valuable for conservation? Conservation Biology **9**:753-760.

Levin, R. A., W. L. Wagner, P. C. Hoch, W. J. Hahn, A. Rodriguez, D. A. Baum, L. Katinas, E. A. Zimmer, and K. J. Sytsma. 2004. Paraphyly in tribe Onagreae:

Insights into phylogenetic relationships of Onagraceae based on nuclear and chloroplast sequence data. Systematic Botany **29**:147-164.

Levins, R. 1969. Some demographic and genetic consequences of environmental heterogeneity for biological control. Bulletin of the Entomological Society of America **15**:237-240.

Lewis, P. O. 1998. A genetic algorithm for maximum-likelihood phylogeny inference using nucleotide sequence data. Molecular Biology and Evolution **15**:277-283.

Lewis, P. O., and D. J. Crawford. 1995. Pleistocene refugium endemics exhibit greater allozymic diversity than widespread congeners in the genus *Polygonella* (Polygonaceae). American Journal of Botany **82**:141-149.

Little, D. P., and D. W. Stevenson. 2007. A comparison of algorithms for the identification of specimens using DNA barcodes: examples from gymnosperms. Cladistics **23**:1-21.

Lloyd, D. G. 1979. Some reproductive factors affecting the selection of self-fertilization in plants. American Naturalist **113**:67-79.

Lomolino, M. V., B. R. Riddle, and J. H. Brown 2006. Biogeography. Sinauer Associates, Sunderland.

Luikart, G., F. W. Allendorf, J. M. Cornuet, and W. B. Sherwin. 1998. Distortion of allele frequency distributions provides a test for recent population bottlenecks. Journal of Heredity **89**:238-247.

Luikart, G., and J. M. Cornuet. 1998. Empirical evaluation of a test for identifying recently bottlenecked populations from allele frequency data. Conservation Biology **12**:228-237.

Lynch, M. 1991. The genetic interpretation of inbreeding depression and outbreeding depression. Evolution **45**:622-629.

Lynch, M., and B. Walsh 1998. Genetics and Analysis of Quantitative Traits. Sinauer Associates, Sunderland.

Mable, B. K., and A. Adam. 2007. Patterns of genetic diversity in outcrossing and selfing populations of *Arabidopsis lyrata*. Molecular Ecology **16**:3565-3580.

Maddison, W., and D. Maddison. 2001. MacClade 4.06 software. Sinauer Associates, Sunderland.

Maddison, W. P. 1997. Gene trees in species trees. Systematic Biology **46**:523-536.

Maddison, W. P., and L. L. Knowles. 2006. Inferring phylogeny despite incomplete lineage sorting. Systematic Biology **55**:21-30.

Mallett, J. 1995. A species definition for the Modern Synthesis. Trends in Ecology and Evolution **10**:294-299.

Marshall, J. C., E. Arevalo, E. Benavides, J. L. Sites, and J. W. Sites. 2006. Delimiting species: Comparing methods for mendelian characters using lizards of the *Sceloporus grammicus* (Squamata : Phrynosomatidae) complex. Evolution **60**:1050-1065.

Mayr, E. 1942. Systematics and the Origin of Species. Columbia University Press, New York.

McKay, J. K., and R. G. Latta. 2002. Adaptive population divergence: markers, QTL and traits. Trends in Ecology and Evolution **17**:285-291.

McLachlan, J. S., J. S. Clark, and P. S. Manos. 2005. Molecular indicators of tree migration capacity under rapid climate change. Ecology **86**:2088-2098.

Meier, R., K. Shiyang, G. Vaidya, and P. K. L. Ng. 2006. DNA barcoding and taxonomy in diptera: A tale of high intraspecific variability and low identification success. Systematic Biology **55**:715-728.

Merila, J., and P. Crnokrak. 2001. Comparison of genetic differentiation at marker loci and quantitative traits. Journal of Evolutionary Biology **14**:892-903.

Meyer, C. P., and G. Paulay. 2005. DNA barcoding: error rates based on comprehensive sampling. **3**:2229-2238.

Michalski, S. G., and W. Durka. 2007. High selfing and high inbreeding depression in peripheral populations of *Juncus atratus*. Molecular Ecology **16**:4715-4727.

Missouri Botanical Garden. 2007. The VAST (VAScular Tropicos) Database. Missouri Botanical Garden.

Mitchell-Olds, T., and J. Schmitt. 2006. Genetic mechanisms and evolutionary significance of natural variation in *Arabidopsis*. Nature **441**:947-952.

Moritz, C. 1994. Defining evolutionary significant units for conservation. Trends in Ecology and Evolution **9**:373-375.

Moritz, C. 1995. Uses of molecular phylogenies for conservation. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences **349**:113-118.

Morrison, D. A. 2009. Why would phylogeneticists ignore computerized sequence alignment? Systematic Biology **58**:150-158 168.

Motzkin, G., W. A. Patterson III., and D. R. Foster. 1999. A historical perspective on pitch pine-scrub oak communities in the Connecticut Valley of Massachusetts. Ecosystems **2**:255-273.

Mousseau, T. A., K. Ritland, and D. D. Heath. 1998. A novel method for estimating heritability using molecular markers. Heredity **80**:218-224.

Mulvaney, C. R., B. Molano-Flores, and D. W. Whitman. 2004. The reproductive biology of *Agalinis auriculata* (Michx.) Raf. (Orobanchaceae), a threatened North American prairie inhabitant. International Journal of Plant Sciences **165**:605-614.

Murray, T. E., U. Fitzpatrick, M. J. F. Brown, and R. J. Paxton. 2008. Cryptic species diversity in a widespread bumble bee complex revealed using mitochondrial DNA RFLPs Conservation Genetics **9**:653-666.

Mymudes, M. S., and D. H. Les. 1993. Morphological and Genetic-Variability in *Plantago cordata* (Plantaginaceae), a Threatened Aquatic Plant. American Journal of Botany **80**:351-359.

Nason, J. D., J. L. Hamrick, and T. H. Fleming. 2002. Historical vicariance and postglacial colonization effects on the evolution of genetic structure in *Lophocereus*, a Sonoran Desert columnar cactus. Evolution **56**:2214-2226.

NatureServe. 2007. NatureServe Explorer: An online encyclopedia of life [web application]. Version 6.3. NatureServe, Arlington, Virginia. Available http://www.natureserve.org/explorer.

Neel, M. C. 2002. Conservation implications of the reproductive ecology of *Agalinis acuta* (Scrophulariaceae). American Journal of Botany **89**:972-980.

Neel, M. C., and M. P. Cummings. 2004. Section-level relationships of North American *Agalinis* (Orobanchaceae) based on DNA sequence analysis of three chloroplast gene regions. BMC Evolutionary Biology **4**:15.

Neel, M. C., and P. Somers. 2001. Summary Report on six experiments examining establishment and maintenance of *Agalinis acuta* populations.

Neff, N. A., and L. F. Marcus 1980. A survey of multivariate methods for systematics. Privately Published, New York, New York.

Nei, M., F. Tajima, and Y. Tateno. 1983. Accuracy of estimated phylogenetic trees from molecular data. Journal of Molecular Evolution **19**:153–170.

Neuenschwander, S., F. Hospital, F. Guillaume, and J. Goudet. 2008. quantiNemo: an individual-based program to simulate quantitative traits with explicit genetic architecture in a dynamic metapopulation. Bioinformatics **24**:1552-1553.

Newmaster, S. G., A. J. Fazekas, R. A. D. Steeves, and J. Janovec. 2008. Testing candidate plant barcode regions in the Myristicaceae. Molecular Ecology Resources **8**:480-490.

Newmaster, S. G., and S. Ragupathy. 2009. Testing plant barcoding in a sister species complex of pantropical *Acacia* (Mimosoideae, Fabaceae). Molecular Ecology Resources **9**:172-180.

Nicole, F., F. Tellier, A. Vivat, and I. Till-Bottraud. 2007. Conservation unit status inferred for plants by combining interspecific crosses and AFLP. Conservation Genetics **8**:1273-1285.

Nixon, K. C., and Q. D. Wheeler. 1990. An amplification of the phylogenetic species concept. Cladistics **6**:211-223.

Noecker, R. J. 1998. Endangered species list revisions: a summary of delisting and downlisting. Congressional Research Service, Washington, D.C.

Noss, R. F., E. T. LaRoe, and J. M. Scott 1995. Endangered Ecosystems of the United States: a Preliminary Assessment of Loss and Degradation. Biological Report 28, U.S. Department of the Interior, National Biological Service, Washington, D.C., USA.

Nunney, L. 2002. The effective size of annual plant populations: The interaction of a seed bank with fluctuating population size in maintaining genetic variation. American Naturalist **160**:195-204.

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

Nyborn, H., and I. V. Bartish. 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.

Olfelt, J. P., G. R. Furnier, and J. J. Luby. 2001. What data determine whether a plant taxon is distinct enough to merit legal protection? A case study of *Sedum integrifolium* (Crassulaceae). American Journal of Botany **88**:401-410.

Olmstead, R. G., C. W. DePamphilis, A. D. Wolfe, N. D. Young, W. J. Elisons, and P. A. Reeves. 2001. Disintegration of the Scrophulariaceae. American Journal of Botany **88**:348-361.

Olmstead, R. G., and P. A. Reeves. 1995. Evidence for the polyphyly of the Scrophulariaceae based on chloroplast *rbc*L and *ndh*F sequences. Annals of the Missouri Botanical Garden **82**:176-193.

Ota, T. 1993. DISPAN: Genetic distance and phylogenetic analysis. Pennsylvania State University, University Park. http://iubio.bio.indiana.edu/soft/molbio/ibmpc/.

Paetkau, D. 1999. Using genetics to identify intraspecific conservation units: a critique of current methods. Conservation Biology **13**:1507-1509.

Paetkau, D., R. Slade, M. Burden, and A. Estoup. 2004. Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power. Molecular Ecology **13**:55-65.

Paetkau, D., L. P. Waits, P. L. Clarkson, L. Craighead, and C. Strobeck. 1997. An empirical evaluation of genetic distance statistics using microsatellite data from bear (Ursidae) populations. Genetics **147**:1943-1957.

Page, R. D. M. 1998. GeneTree: comparing gene and species phylogenies using reconciled trees. Bioinformatics **14**:819-820.

Page, R. D. M., and E. C. Holmes 1998. Molecular Evolution: A Phylogenetic Approach. Blackwell Sciences, Oxford.

Pakeman, R. J. 2001. Plant migration rates and seed dispersal mechanisms. Journal of Biogeography **28**:795-800.

Pakeman, R. J., J. Engelen, and J. P. Attwood. 1999. Rabbit endozoochroy and seedbank build-up in an acidic grassland. Plant Ecology **145**:83-90.

Palmer, J. D., editor. 1985. Evolution of chloroplast and mitochondrial DNA in plants and algae, Plenum, New York, pp 131-240.

Palmer, J. D., D. E. Soltis, and M. W. Chase. 2004. The plant tree of life: An overview and some points of view. **91**:1437-1445.

Palstra, F. P., and D. E. Ruzzante. 2008. Genetic estimates of contemporary effective population size: what can they tell us about the importance of genetic stochasticity for wild population persistence? Molecular Ecology **17**:3428-3447.

Peakall, R., and P. E. Smouse. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes **6**:288-295.

Pennell, F. W. 1913a. Studies in the Agalinanae, a subtribe of the Rhinanthaceae. I. Nomenclature of the Neartic genera. Bulletin of the Torrey Botanical Club **40**:119-130.

Pennell, F. W. 1913b. Studies in the Agalinanae, a subtribe of the Rhinanthaceae. II. Species of the Atlantic Coastal Plain. Bulletin of the Torrey Botanical Club **40**:401-439.

Pennell, F. W. 1928. *Agalinis* and allies in North America I. Proceedings of the Academy of Natural Sciences of Philadelphia **80**:339-449.

Pennell, F. W. 1929. *Agalinis* and allies in North America II. Proceedings of the Academy of Natural Sciences of Philadelphia **81**:111-249.

Pennell, F. W. 1935. The Scrophulariaceae of eastern temperate North America (41 *Gerardia*). Academy of Natural Sciences of Philadelphia Monographs **1**:419-476.

Petit, R. J., I. Aguinagalde, J. L. de Beaulieu, C. Bittkau, S. Brewer, R. Cheddadi, R. Ennos, S. Fineschi, D. Grivet, M. Lascoux, A. Mohanty, G. M. Muller-Starck, B. Demesure-Musch, A. Palme, J. P. Martin, S. Rendell, and G. G. Vendramin. 2003. Glacial refugia: Hotspots but not melting pots of genetic diversity. Science **300**:1563-1565.

Petit, R. J., S. Brewer, S. Bordacs, K. Burg, R. Cheddadi, E. Coart, J. Cottrell, U. M. Csaikl, B. van Dam, J. D. Deans, S. Espinel, S. Fineschi, R. Finkeldey, I. Glaz, P. G. Goicoechea, J. S. Jensen, A. O. Konig, A. J. Lowe, S. F. Madsen, G. Matyas, R. C. Munro, F. Popescu, D. Slade, H. Tabbener, S. G. M. de Vries, B. Ziegenhagen, J. L.

de Beaulieu, and A. Kremer. 2002. Identification of refugia and post-glacial colonisation routes of European white oaks based on chloroplast DNA and fossil pollen evidence. Forest Ecology and Management **156**:49-74.

Petit, R. J., A. El Mousadik, and O. Pons. 1998. Identifying populations for conservation on the basis of genetic markers. Conservation Biology **12**:844-855.

Pettengill, J. B., R. K. Burnett, and M. C. Neel. 2009. Characterization of 21 microsatellites within *Agalinis acuta* (Orobanchaceae) and cross-species amplification among closely related taxa. Molecular Ecology Resources.

Pettengill, J. B., and M. C. Neel. 2008. Phylogenetic patterns and conservation among North American members of the genus *Agalinis* (Orobanchaceae). BMC Evolutionary Biology **8**:264.

Phillips, B. L., G. P. Brown, J. M. J. Travis, and R. Shine. 2008. Reid's paradox revisited: The evolution of dispersal kernels during range expansion. American Naturalist **172**:S34-S48.

Piry, S., A. Alapetite, J. M. Cornuet, D. Paetkau, L. Baudouin, and A. Estoup. 2004. GENECLASS2: A software for genetic assignment and first-generation migrant detection. Journal of Heredity **95**:536-539.

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

Posada, D., and K. A. Crandall. 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics **14**:817-818.

Posada, D., and K. A. Crandall. 2001. Intraspecific gene genealogies: trees grafting into networks. Trends in Ecology & Evolution **16**:37-45.

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

Pujol, B., A. J. Wilson, R. I. C. Ross, and J. R. Pannell. 2008. Are Q_{ST}-F_{ST} comparisons for natural populations meaningful? Molecular Ecology **17**:4782-4785.

R Development Core Team 2008. *R*: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna.

Raeymaekers, J. A. M., J. K. J. Van Houdt, M. H. D. Larmuseau, S. Geldof, and F. A. M. Volckaert. 2007. Divergent selection as revealed by P_{ST} and QTL-based F_{ST} in three-spined stickleback (*Gasterosteus aculeatus*) populations along a coastal-inland gradient. Molecular Ecology **16**:891-905.

Rafinesque, C. S. 1836. New Flora and Botany of North America

Ragupathy, S., S. G. Newmaster, M. Murugesan, and V. Balasubramaniam. 2009. DNA barcoding discriminates a new cryptic grass species revealed in an ethnobotany study by the hill tribes of the Western Ghats in southern India. Molecular Ecology Resources **9**:164-171.

Rannala, B., and J. L. Mountain. 1997. Detecting immigration by using multilocus genotypes. Proceedings of the National Academy of Sciences, USA **94**:9197-9201.

Ratnasingham, S., and P. D. N. Hebert. 2007. BOLD: The Barcode of Life Data System ([www.barcodinglife.org)](www.barcodinglife.org). Molecular Ecology Notes **7**:355-364.

Raymond, M., and F. Rousset. 1995. Genepop (Version-1.2) - population genetics software for exact tests and ecumenicism. Journal of Heredity **86**:248-249.

Redding, D. W., and A. O. Mooers. 2006. Incorporating evolutionary measures into conservation prioritization. Conservation Biology **20**:1670-1678.

Reed, D. H., and R. Frankham. 2003. Correlation between fitness and genetic diversity. Conservation Biology **17**:230-237.

Reid, C. 1899. The Origin of the British Flora. Dulau, London.

Rice, W. R. 1988. Analyzing tables of statistical tests. Evolution **43**:223-225.

Richards, C. M. 2000. Inbreeding depression and genetic rescue in a plant metapopulation. American Naturalist **155**:383-394.

Rieseberg, L. H., and L. Brouillet. 1994. Are many plant species paraphyletic? Taxon **43**:21-32.

Ritland, K. 2006. MARK - Genetic Marker Analysis Program. University of British Columbia, Vancouver, BC, Canada. http://genetics.forestry.ubc.ca/ritland/programs.html.

Rosenberg, N. A. 2003. The shapes of neutral gene genealogies in two species: probabilities of monophyly, paraphyly, and polyphyly in a coalescent model. Evolution **57**:1465-1477.

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

Ross, H. A., S. Murugan, and W. L. S. Li. 2008. Testing the reliability of genetic methods of species identification via simulation. Systematic Biology **57**:216-230.

Rousset, F. 1996. Equilibrium values of measures of population subdivision for stepwise mutation processes. Genetics **142**:1357-1362.

Rozas, J., J. C. Sanchez-DelBarrio, X. Messeguer, and R. Rozas. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics **19**:2496-2497.

Rozen, S., and H. Skaletsky. 2000. Primer3 on the WWW for general users and for biologist programmers in S. Krawetz, and S. Misener, editors. Bioinformatics methods and protocols. Humana Press, Totowa.

Rubinoff, D. 2006. Utility of mitochondrial DNA barcodes in species conservation. **20**:1026-1033.

Russell, E. W. B., R. B. Davis, R. S. Anderson, T. E. Rhodes, and D. S. Anderson. 1993. Recent centuries of vegetational change in the glaciated north-eastern United States. Journal of Ecology **81**:647-664.

Ryder, O. A. 1986. Species conservation and systematics - the dilemma of subspecies. Trends in Ecology & Evolution **1**:9-10.

Saether, S. A., P. Fiske, J. A. Kalas, A. Kuresoo, L. Luigujoe, S. B. Piertney, T. Sahlman, and J. Hoglund. 2007. Inferring local adaptation from Q_{ST}-F_{ST} comparisons: neutral genetic and quantitative trait variation in European populations of great snipe. Journal of Evolutionary Biology **20**:1563-1576.

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

Sanderson, M. J., and M. J. Donoghue. 1989. Patterns of variation in levels of homoplasy. Evolution **43**:1781-1795.

Sarkar, I. N., P. J. Planet, T. E. Bael, S. E. Stanley, M. Siddall, R. DeSalle, and D. H. Figurski. 2002a. Characteristic attributes in cancer microarrays. Journal of Biomedical Informatics **35**:111-122.

Sarkar, I. N., P. J. Planet, and R. Desalle. 2008. CAOS software for use in character-based DNA barcoding. Molecular Ecology Resources **8**:1256-1259.

Sarkar, I. N., J. W. Thornton, P. J. Planet, D. H. Figurski, B. Schierwater, and R. DeSalle. 2002b. An automated phylogenetic key for classifying homeoboxes. Molecular Phylogenetics and Evolution **24**:388-399.

Sass, C., D. P. Little, D. W. Stevenson, and C. D. Specht. 2007. DNA barcoding in the Cycadales: testing the ootential of oroposed barcoding markers for species identification of cycads. PLoS ONE **2**:e1154.

Savolainen, V., R. S. Cowan, A. P. Vogler, G. K. Roderick, and R. Lane. 2005. Towards writing the encyclopaedia of life: an introduction to DNA barcoding. Philosophical Transactions of the Royal Society B-Biological Sciences **360**:1805-1811.

Schaal, B. A., D. A. Hayworth, K. M. Olsen, J. T. Rauscher, and W. A. Smith. 1998. Phylogeographic studies in plants: problems and prospects. Molecular Ecology **7**:465-474.

Schneeweiss, G. M., A. Colwell, J. M. Park, C. G. Jang, and T. F. Stuessy. 2004. Phylogeny of holoparasitic *Orobanche* (Orobanchaceae) inferred from nuclear ITS sequences. Molecular Phylogenetics and Evolution **30**:465-478.

Sewell, M. M., C. R. Parks, and M. W. Chase. 1996. Intraspecific chloroplast DNA variation and biogeography of North American *Liriodendron* L (Magnoliaceae). Evolution **50**:1147-1154.

Shaw, J., E. B. Lickey, J. T. Beck, S. B. Farmer, W. S. Liu, J. Miller, K. C. Siripun, C. T. Winder, E. E. Schilling, and R. L. Small. 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. L. 2002. Conflict between mitochondrial and nuclear DNA phylogenies of a recent species radiation: what mitochondrial DNA 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.

Shimodaira, H. 2002. An approximately unbiased test of phylogenetic tree selection. Systematic Biology **51**:492-508.

Shimodaira, H., and M. Hasegawa. 2001. CONSEL: For assessing the confidence of phylogenetic tree selection. Bioinformatics **17**:1246-1247.

Sites, J. W., Jr., and J. C. Marshall. 2004. Empirical criteria for delimiting species. Annual Review of Ecology, Evolution, and Systematics **35**:199-229.

Slatkin, M. 1993. Isolation by distance in equilibrium and non-equilibrium populations. Evolution **47**:264-279.

Small, R. L., J. A. Ryburn, R. C. Cronn, T. Seelanan, and J. F. Wendel. 1998. The tortoise and the hare: Choosing between noncoding plastome and nuclear ADH sequences for phylogeny reconstruction in a recently diverged plant group. American Journal of Botany **85**:1301-1315.

Smouse, P. E., J. C. Long, and R. R. Sokal. 1986. Multiple regression and correlation extensions of the mantel test of matrix correspondence. Systematic Zoology **35**:627-632.

Smouse, P. E., and R. Peakall. 1999. Spatial autocorrelation analysis of individual multiallele and multilocus genetic structure. Heredity **82**:561-573.

Snider, J. A. 1969. A variation study of *Agalinis decemloba* (Greene) Pennell and *Agalinis obtusifolia* Raf. in the Carolinas. Unpublished Master of Arts thesis, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina.

Sokal, R. R., and T. J. Crovello. 1970. The biological species concept: a critical evaluation. The American Naturalist **104**:127-153.

Sokal, R. R., and C. D. Michener. 1958. A statistical method for evaluating systematic relationships

University of Kansas Science Bulletin **38**.

Soltis, D. E., M. A. Gitzendanner, D. D. Strenge, and P. S. Soltis. 1997. Chloroplast DNA intraspecific phylogeography of plants from the Pacific Northwest of North America. Plant Systematics and Evolution **206**:353-373.

Soltis, D. E., and R. K. Kuzoff. 1995. Discordance between nuclear and chloroplast phylogenies in the *Heuchera* group (Saxifragaceae). Evolution **49**:727-742.

Soltis, E. D. S., P. S. 2000. Contributions of plant molecular systematics to studies of molecular evolution. Plant Molecular Biology **42**:45-75.

Soltis, P. S., and M. A. Gitzendanner. 1999. Molecular systematics and the conservation of rare species. Conservation Biology **13**:471-483.

Sorrie, B. A., and A. S. Weakley. 2001. Coastal plain vascular plant endemics: Phytogeographic patterns. Castanea **66**:50–82.

Spielman, D., B. W. Brook, and R. Frankham. 2004. 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.

Spitze, K. 1993. Population structure in *Daphnia obtusa* - Quantitative genetic and allozymic variation. Genetics **135**:367-374.

Starr, J. R., R. F. C. Naczi, and B. N. Chouinard. 2009. Plant DNA barcodes and species resolution in sedges (*Carex*, Cyperaceae). Molecular Ecology Resources **9**:151-163.

Statsoft, I. 2004. STATISTICA (data analysis software system) Version 6. Statsoft, Tulsa, Oklahoma, USA.

Stebbins, G. L. 1950. Variation and Evolution in Plants. Columbia University Press, New York, New York.

Stebbins, G. L. 1957. Self fertilization and population variability in the higher plants. American Naturalist **91**:337-354.

Stewart, H. M., and J. M. Canne-Hilliker. 1998. Floral development of *Agalinis neoscotica*, *Agalinis paupercula* var. *borealis*, and *Agalinis purpurea* (Scrophulariaceae): Implications for taxonomy and mating system. International Journal of Plant Sciences **159**:418-439.

Stewart, H. M., S. C. Stewart, and J. M. Canne-Hilliker. 1996. Mixed mating system in *Agalinis neoscotica* (Scrophulariaceae) with bud pollination and delayed pollen germination. International Journal of Plant Sciences **157**:501-508.

Storz, J. F. 2002. Contrasting patterns of divergence in quantitative traits and neutral DNA markers: analysis of clinal variation. Molecular Ecology **11**:2537-2551.

Streisfeld, M. A., and J. R. Kohn. 2005. Contrasting patterns of floral and molecular variation across a cline in *Mimulus aurantiacus*. Evolution **59**:2548-2559.

Suckling, K. 2006. Measuring the Success of the Endangered Species Act: Recovery Trends in the Northeastern United States. Center for Biological Diversity, Tucson.

Swofford, D. L. 2003. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4.0. Sinauer Associates, Sunderland.

Syring, J., K. Farrell, R. Businsky, R. Cronn, and A. Liston. 2007. Widespread genealogical nonmonophyly in species of *Pinus* subgenus *Strobus*. Systematic Biology **56**:163-181.

Taberlet, P., E. Coissac, F. Pompanon, L. Gielly, C. Miquel, A. Valentini, T. Vermat, G. Corthier, C. Brochmann, and E. Willerslev. 2007. Power and limitations of the chloroplast *trn*L (UAA) intron for plant DNA barcoding. Nucleic Acids Research **35**:-.

Taberlet, P., L. Gielly, G. Pautou, and J. Bouvet. 1991. Universal primers for amplification of 3 noncoding regions of chloroplast DNA. Plant Molecular Biology **17**:1105-1109.

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

Takahata, N. 1989. Gene genealogy in three related populations: consistency probability between gene and population trees. Genetics **122**:957-966.

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

Tamura, K., J. Dudley, M. Nei, and S. Kumar. 2007. MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. Molecular Biology and Evolution **24**:1596-1599.

Titus, K., J. A. Mosher, and B. K. Williams. 1984. Chance-corrected classification for use in discriminant analysis: ecological applications. American Midland Naturalist **107**:1-7.

Trewick, S. A. 2008. DNA Barcoding is not enough: mismatch of taxonomy and genealogy in New Zealand grasshoppers (Orthoptera : Acrididae). Cladistics **24**:240-254.

Tsitrone, A., M. Kirkpatrick, and D. A. Levin. 2003. A model for chloroplast capture. Evolution **57**:1776-1782.

Turesson, G. 1922. The species and the variety as ecological units. Hereditas **3**:100-113.

U. S. Fish and Wildlife Service. 1988. Determination of *Agalinis acuta* to be an endangered species. Federal Register **53**:34701-34705.

U. S. Fish and Wildlife Service. 1996. Endangered and Threatened Wildlife and Plants: Reclassification of *Erigeron maguirei* (Maguire daisy) from Endangered to Threatened. Federal Register **61**.

U.S. Fish and Wildlife Service. 1989a. Final rule: Delisting of *Echinocereus engelmannii* var. *purpureus* (Purple-Spined Hedgehog Cactus). Federal Register **54** 48749-48751.

U.S. Fish and Wildlife Service. 1989b. Sandplain gerardia (*Agalinis acuta*) recovery plan, Newton, Massachusetts.

U.S. Fish and Wildlife Service. 1999. Final Rule To Remove the Plant *Echinocereus lloydii* (Lloyd's Hedgehog Cactus) From the Federal List of Endangered and Threatened Plants. Federal Register **64**.

U.S.C. 1973. Endangered Species Act §1531-§1544.

USDA NRCS. 2004. The PLANTS Database. National Plant Data Center.

USDA NRCS. 2008. The PLANTS Database. National Plant Data Center. http://plants.usda.gov.

Van Valen, L. 1976. Ecological species, multispecies, and oaks. Taxon **25**:233-239.

Vellend, M., J. A. Myers, S. Gardescu, and P. L. Marks. 2003. Dispersal of *Trillium* seeds by deer: Implications for long-distance migration of forest herbs. Ecology **84**:1067-1072.

Vogler, A. 2006. Will DNA barcoding advance efforts to conserve biodiversity more efficiently than traditional taxonomic methods? Reply. Frontiers in Ecology and the Environment **5**:272-273.

Wahlund, S. 1928. Zusammensetzung von Population und Korrelationserscheinung vom Standpunkt der Vererbungslehre aus betrachtet. Hereditas **11**:65–106.

Walters, C. 2005. MSc thesis: Genetic relationships among Spiranthes parksii and congeneric species. Texas A&M University, College Station.

Waples, R. S. 2006. A bias correction for estimates of effective population size based on linkage disequilibrium at unlinked gene loci. Conservation Genetics **7**:167-184.

Waples, R. S., and C. Do. 2008. LDNE: a program for estimating effective population size from data on linkage disequilibrium. Molecular Ecology Resources **8**:753-756.

Weakley, A. S. 2008. Flora of the Carolinas, Virginia, Georgia, northern Florida, and surrounding areas. University of North Carolina at Chapel Hill, UNC Herbarium, North Carolina Botanical Garden.

Weir, B. S., and C. C. Cockerham. 1984. Estimating F-statistics for the analysis of population structure. Evolution **38**:1358-1370.

White, T. J., T. Bruns, S. Lee, and J. Taylor. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. Pages 315-332 in I. M., D. Gelfand , J. Sninsky, and T. White, editors. PCR Protocols: A Guide to Methods and Applications. Academic Press, San Diego.

Whitlock, M. C. 2000. Fixation of new alleles and the extinction of small populations: Drift load, beneficial alleles, and sexual selection. Evolution **54**:1855-1861.

Wiens, J. J. 2007. Species delimitation: new approaches for discovering diversity. Systematic Biology **56** 875-878.

Wiens, J. J., and M. R. Servedio. 2000. Species delimitation in systematics: inferring diagnostic differences between species. Proceedings of the Royal Society of London Series B-Biological Sciences **267**:631-636.

Wilcove, D. S., D. Rothstein, J. Dubow, A. Phillips, and E. Losos. 1998. Quantifying threats to imperiled species in the United States. BioScience **48**:607-615.

Will, K. W., B. D. Mishler, and Q. D. Wheeler. 2005. The perils of DNA barcoding and the need for integrative taxonomy. Systematic Biology **54**:844-851.

Wood, T. E., and T. Nakazato. 2009. Investigating species boundaries in the Giliopsis group of *Ipomopsis* (Polemoniaceae): Strong discordance among molecular and morphological markers. American Journal of Botany **96**:853-861.

Wu, C. I. 2001. The genic view of the process of speciation. Journal of Evolutionary Biology **14**:851-865.

Yang, F. S., and X. Q. Wang. 2007. Extensive length variation in the cpDNA *trn*T-*trn*F region of hemiparasitic *Pedicularis* and its phylogenetic implications. Plant Systematics and Evolution **264**:251-264.

Yang, F. S., X. Q. Wang, and D. Y. Hong. 2003. Unexpected high divergence in nrDNA ITS and extensive parallelism in floral morphology of *Pedicularis* (Orobanchaceae). Plant Systematics and Evolution **240**:91-105.

Yang, Z. B. 1994. Estimating the pattern of nucleotide substitution. Journal of Molecular Evolution **39**:105-111.

Yassin, A., P. Capy, L. Madi-Ravazzi, D. Ogereau, and J. R. David. 2008. DNA barcode discovers two cryptic species and two geographical radiations in the invasive drosophilid *Zaprionus indianus*. Molecular Ecology Resources **8**:491-501.

Yeh, F. C., R. Yang, and T. Boyle. 1997. POPGENE. A Microsoft Window-based freeware for population genetic analysis. Version 1.21.

Yi, T., J. Wen, A. Golan-Goldhirsh, and D. E. Parfitt. 2008. Phylogenetics and reticulate evolution in *Pistacia* (Anacardiaceae). American Journal of Botany **95**:241-251.

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

Zwickl, D. J. 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. The University of Texas at Austin.