To the Graduate Council:

I am submitting herewith a dissertation written by Joey Thomas Shaw entitled "Systematics of the North American plums (*Prunus* subgenus *Prunus* section *Prunocerasus*; Rosaceae)." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Botany.

Randall L. Small
Major Professor

We have read this dissertation
and recommend its acceptance:

Edward E. Schilling

Joseph Williams

Sally P. Horn

Accepted for the Council:

Ann Mayhew
Vice Chancellor and
Dean of Graduate Studies

(Original Signatures are on file with official student records)

Systematics of the North American plums (*Prunus* subgenus *Prunus* section *Prunocerasus*; Rosaceae)

A Dissertation Presented for the Doctor of Philosophy Degree

The University of Tennessee, Knoxville

Joey Thomas Shaw

May 2005

ACKNOWLEDGEMENTS

I want to thank the Department of Botany at the University of Tennessee for allowing me the opportunity, not only to obtain my graduate degrees, but to reach beyond my degrees and achieve as much I could make time for. I sincerely thank my major professor Randy Small for his guidance, time, money, advice, and more importantly for his friendship—his direction through the course of this study has been like that of an older brother. I also want to thank the other members of my committee: Ed Schilling, Sally Horn, and Joe Williams for their helpful guidance and advice during my term as a graduate student. B. Eugene Wofford also served as a mentor and friend throughout the course of this research. Additionally, I want to thank my labmates: Ed Lickey, John Beck, Charles Winder, Wusheng Liu, and Larry Estes for their friendships and many helpful and educating discussions. To the secretaries Eunice Turner and Eileen Coltharp, I owe thanks for too many things to list.

Emily, Aidan, and the rest of my family gave me tremendous support during the seven years that I have been gone.

Lastly, need to acknowledge several sources of funding that allowed me both the time and the money to do this project: National Science Foundation Dissertation Improvement Grant (DEB-0407948); Yates Fellowship, the University of Tennessee; Alexander Hollaender Fellowship, the University of Tennessee; Graduate Student Research Award, Botanical Society of America; Science Alliance, University of Tennessee; Sharp Fund, the University of Tennessee, Hesler Fund, the University of Tennessee; Dennis-Breedlove Fund for Field Research, the University of Tennessee; and Core Student Award, Southern Appalachian Botanical Society.

ABSTRACT

The North American plums (*Prunus* subgenus *Prunus* section *Prunocerasus*; Rosaceae) are a closely related group with approximately 17 commonly recognized species and lesser taxa. They are infamous for their very poor development of reproductively isolating barriers and most are interfertile in many combinations. This interfertility blurs nearly all morphologically-based taxonomic boundaries. Even still, geographically related morphological variation exists and when intermediates are ignored several taxa may be seen as being reasonably different from one another with respect to both morphology and ecology. Additionally, the ranges of most of the North American plum taxa overlap with the ranges of several others. The only exception is *P. subcordata*, which is the only species in the section whose range is west of the Rocky Mountains. The aim of this dissertation research was to infer a phylogeny for the group in an attempt to understand their complicated evolutionary history. Emphasis was placed on using molecular tools (e.g., PCR and DNA sequencing) to tease apart their intricate relationships. The importance of this course of study includes but is not limited to: (1) better understanding a difficult group of taxa for academic reasons, (2) accumulating data to better understand these taxa so that conservation efforts can be better focused (e.g., *P. geniculata* is Federally Endangered), (3) better understanding the relationships among a group of plants with economic importance, (4) testing the limits of tools currently used in plant molecular systematics—the North American plums posit a problem at the boundary between molecular phylogenetics and population genetics, and (5) adding to a body of knowledge surrounding longstanding biogeographic questions of North America, namely the Atlantic and Gulf Coastal Plain disjunction to the Great Lakes region and the eastern North America-western North America disjunction.

This research dissertation is the accumulation of information from four original research papers and a National Science Foundation Dissertation Improvement Grant. All papers have either been published, submitted, or will be submitted to the *American Journal of Botany*.

The results of this coarse of study showed that (1) predictable rate heterogeneity exists among noncoding cpDNA regions and several rarely used regions provide more mutations to phylogenetic investigations than the most commonly used regions (Part 3). (2) The North American plums are monophyletic (Part 4). (3) Most North American plum taxa are para- and polyphyletic with respect to their chloroplasts—more than one of the three primary chloro-haplotypes was observed in 12 of 17 of the North American plum taxa (Part 5). (4) Most North American plum taxa are not monophyletic with respect to their nuclear encoded *s6pdh* genes—three primary haplotypes are shared among most taxa (Part 6). Total evidence provided by this investigation strongly supports the hypothesis of the North American plums are a syngameon, or hybridizing species group.

TABLE OF CONTENTS

Part 1. OVERVIEW OF *PRUNUS* SUBGENUS *PRUNUS* SECTION *PRUNOCERASUS* (ROSACEAE); PROBLEMS AND PERSPECTIVES 1
Introduction 2
Goals of the Project 5
Literature Cited 7
Appendix 11
Part 2. THE TORTOISE AND THE HARE II: RELATIVE UTILITY OF 21 NONCODING CHLOROPLAST DNA SEQUENCES FOR PHYLOGENETIC ANALYSIS 14
Abstract 15
Introduction 15
Materials and Methods 17
Results 25
Discussion 27
Literature Cited 40
Appendix 53
Part 3. ADDRESSING THE "HARDEST PUZZLE IN AMERICAN POMOLOGY:" PHYLOGENY OF *PRUNUS* SECTION *PRUNOCERASUS* (ROSACEAE) BASED ON SEVEN NONCODING CHLOROPLAST DNA REGIONS 67
Abstract 68
Introduction 68
Materials and Methods 72
Results 74
Discussion 76
Literature Cited 83
Appendix 88
Part 4. CHLOROPLAST DNA PHYLOGENY AND PHYLOGEOGRAPHY OF THE NORTH AMERICAN PLUMS (*PRUNUS* SUBGENUS *PRUNUS* SECTION *PRUNOCERASUS*; ROSACEAE) 96
Abstract 97
Introduction 97
Materials and Methods 100
Results 101
Discussion 102
Literature Cited 108
Appendix 115
Part 5. CONCORDANT PHYLOGEOGRAPHIC PATTERNS OF THE CHLOROPLAST AND NUCLEAR GENOMES IN THE NORTH AMERICAN PLUM SYNGAMEON (*PRUNUS* SUBGENUS *PRUNUS* SECTION *PRUNOCERASUS*; ROSACEAE) 143
Abstract 144
Introduction 144
Materials and Methods 147
Results 149
Discussion 151
Literature Cited 156

Appendix 161
Part 6. GENERAL CONCLUSIONS AND FUTURE DIRECTIONS 187

VITA 191

LIST OF TABLES

Part 1. Table 1. Commonly recognized taxa within *Prunocerasus* 12

Part 2. Table 1. Sources of cpDNA sequences employed in this study 54

Part 3. Table 1. List of taxa used in this investigation along with current classification scheme, source and voucher numbers, status of species of conservation concern, and GenBank accession numbers 89

Part 4. Table 1. List of taxa used in this investigation: Chloroplast DNA phylogeny and phylogeography of the North American plums (*Prunus* subgenus *Prunus* section *Prunocerasus*; Rosaceae). 116

Part 5. Table 1. List of taxa used in this investigation and nDNA *s6pdh* and cpDNA haplotypes 162

LIST OF FIGURES

Part 1.

Figure 1. Phylogenetic relationships of *Prunocerasus* shown with five classically used morphological characters 13

Part 2.

Figure 1. Phylogenetic studies using cpDNA *matK* and noncoding regions published in *American Journal of Botany*, *Systematic Botany*, *Molecular Phylogenetics and Evolution*, and *Plant Systematics and Evolution* from 1995 through 2002 .. 58

Figure 2. Simplified phylogenetic representation, modified from APG II (2003), of the 10 lineages used in the survey of 21 noncoding cpDNA regions 59

Figure 3. Scaled map of the 21 noncoding cpDNA regions surveyed in this investigation (based on the *Nicotiana* chloroplast genome [Wakasugi et al., 1998]) 60

Figure 4. Representation of the PIC (potentially informative character) values among the noncoding regions (*z*-axis) across taxonomic groups (*x*-axis) 62

Figure 5. The average PIC (potentially informative character) value of the six lineages with complete data sets for each region 63

Figure 6. Scatter plot showing the relationship of region length and its PIC value for each lineage 64

Figure 7. The average PIC (potentially informative character) value of six lineages with complete data sets for each region compared to easily coamplifiable regions. 65

Figure 8. Comparison of the predictability of a three-species survey with complete data sets used to generate phylogenetic hypotheses for *Prunus* (Shaw and Small, 2004) and *Hibiscus* 66

Part 3.

Figure 1. Strict consensus of 25 171 equally parsimonious trees resulting from maximum parsimony analysis using the combined data from the *rpL16* and *trnG* introns as well as the *trnS-trnG* and *trnH-psbA* intergenic spacers (3270 aligned nucleotide positions, 134 are parsimony informative; tree length = 422 steps; consistency index [CI] = 0.92; retention index [RI] = 0.94) 91

Figure 2. Bayesian analysis tree for *Prunus* using the combined data from the *rpL16* and *trnG* introns as well as the *trnS-trnG* and *trnH-psbA* intergenic spacers (3270 aligned nucleotide positions, 134 are parsimony informative) 93

Figure 3. Phylogenetic relationships of *Prunocerasus* shown with the classifications of Waugh (1899, 1901) and Wight (1915) in addition to five morphological characters classically used to delimit taxonomic boundaries within the section. 95

Part 4.

Figure 1. A maximum parsimony cladogram from seven cpDNA regions, 4375 aligned nucleotide positions (modified from Shaw and Small, 2004)..... 122

Figure 2a-c. Strict consensus of 3 equally parsimonious trees resulting from maximum parsimony analysis using the *rpL16* intron (797 aligned bp from the 3´ end of the intron)..... 123

Figure 3. TCS gene genealogy of *Prunus* subgenus *Prunus* section *Prunocerasus* from the *rpL16* intron. Haplotypes observed in this study are represented by squares. 127

Figure 4(a) to (k). Geographical locations of haplotypes observed throughout the ranges of the 18 taxa of this study..... 128

Figure 5. Cladogram of *Prunus* subgenus *Prunus* section *Prunocerasus* based on the most frequent haplotype observed in each taxon 139

Figure 6(a) to (b). The distribution of the six major haplotypes (A, B, C, S, T, and pU) observed in this study 140

Figure 7. Proposed migration route of the North American plums through North America..... 142

Part 5.

Figure 1. Strict consensus of 653 equally parsimonious trees resulting from maximum parsimony analysis using the nDNA *s6pdh* intron one (955 aligned nucleotide positions, 85 are parsimony informative; tree length = 165 steps; consistency index [CI] = 0.72; retention index [RI] = 0.93)..... 164

Figure 2. Bayesian analysis tree using the nDNA *s6pdh* intron one (955 aligned nucleotide positions, 85 are parsimony informative) 166

Figure 3. TCS gene genealogy of *Prunus* subgenus *Prunus* section *Prunocerasus* from the *s6pdh* intron one 168

Figure 4. (a) to (k). Geographical locations of haplotypes observed throughout the ranges of the 18 taxa of this study..... 170

Figure 5. (a) to (d). The distribution of the three primary (□, □, and □) and two possible recombinant (- and 29) haplotypes observed in this study 181

Figure 6. Superposition of the ranges of all North American plum taxa showing that the ranges of the majority of the named taxa overlap in a region from northeast Texas through Oklahoma and Arkansas to Missouri and west Tennessee 185

Figure 7. Distribution of the primary chloroplast haplotypes 186

Part 1

OVERVIEW OF *PRUNUS* SUBGENUS *PRUNUS* SECTION *PRUNOCERASUS* (ROSACEAE); PROBLEMS AND PERSPECTIVES

INTRODUCTION

Overview of* Prunus *and Rosaceae*—Prunus* L. (Rosaceae) is a morphologically diverse genus of about 200 species (Rehder, 1940; Robertson, 1974) of trees and shrubs, many of which are economically important. Several species are used for lumber (e.g., black cherry), ornamentals (e.g., flowering cherries), and medicinally (*Prunus africanum* is used to treat prostatic hyperplasia). Most importantly, many *Prunus* species are valuable sources of edible fruits and/or seeds, including species of plums/prunes, peaches, apricots, cherries, and almonds. According to the USDA statistics of fruits and tree nuts, the value of plums, peaches, cherries, and almonds for 2003 in the United States was approximately 2.54 billion dollars; the USA is not the world leader in production of many of these crops. In addition to these domesticates, the genus includes numerous wild species of ethnobotanical importance because of their uses for food, timber, and medicine.

The majority of *Prunus* species are native to the Northern Hemisphere, where the genus is widely distributed in both the Old and New Worlds, while approximately 75 species have tropical and subtropical distributions (Rehder, 1940; Willis, 1973; Mabberley, 1997).

Prunus has most often been classified, along with several other genera whose members also have base chromosome number x=8 and drupaceous fruits, in Amygdaloideae, one of the four traditionally recognized subfamilies of Rosaceae (e.g., Schulze-Menz, 1964). The other members of Amygdaloideae include *Maddenia*, with 4-5 Asian species, *Pygeum*, with about 40 tropical Asian and one African species, *Prinsepia*, with three to four Asian species, and the monotypic *Oemleria* from western North America. In the most recent subfamilial classification of Rosaceae, Takhtajan (1997) recognized twelve subfamilies; his Amygdaloideae includes the four genera just mentioned, *Prunus* s. s. and five segregate genera, and *Exochorda*. The last genus, comprising one to five species native to central Asia to Korea, stands out within the subfamily because its fruit is a capsule rather than a drupe, but several lines of evidence including chromosome number (Goldblatt, 1976), wood anatomy (Zhang, 1992), and a chloroplast DNA (cpDNA) *rbcL* phylogeny (Morgan et al., 1994) suggested placing *Exochorda* within Amygdaloideae.

Through the last 300 years, the taxonomy within and surrounding *Prunus* has been controversial. *Prunus* has been variously lumped and split by different workers. For example, Tournefort (1700) recognized six distinct genera within *Prunus s.l.*, while Linnaeus (1754) recognized four genera. Bentham and Hooker (1865) were the first to lump the six genera of Tournefort under *Prunus*, which they subdivided into seven sections. In his recent treatment of Rosaceae, Takhtajan (1997) recognized six separate genera, but the most widely accepted classification of *Prunus* is Rehder's (1940) broad interpretation, in which he divided the genus into five subgenera: *Amygdalus* (peaches and almonds), *Cerasus* (cherries), *Laurocerasus* (laurel-cherries), *Padus* (bird-cherries), and *Prunus* (plums and apricots). He further split the subgenera into sections that mostly correspond to Old World and New World groups. Finally, while most western investigators have synonymized *Pygeum* to *Prunus* subgenus *Laurocerasus* (Kalkman 1965), some eastern workers (e.g., Yü et al., 1986; Hô, 2000) still recognize *Pygeum* as a distinct genus.

Recent molecular phylogenetic studies of relationships across Rosaceae (e.g., Morgan et al., 1994; Potter et al., 2002; Evans and Campbell, 2002; Evans et al., 2002) and within *Prunus* (Lee and Wen, 2001; Bortiri et al., 2001, 2002; Shaw and Small, 2004 [Part 3 of this dissertation]) have challenged both the circumscriptions of Amygdaloideae and infrageneric classifications of *Prunus*. Morgan et al.'s (1994) *rbcL* study provided moderate support for a sister relationship between *Prunus* and a strongly supported clade including *Exochorda*, *Oemleria*, and *Prinsepia*, reflected in Takhtajan's (1997) circumscription of Amygdaloideae. Several more recent family-level analyses, however, while supporting the monophyly of both *Prunus* and the clade including the other three genera, do *not* support the sister relationship between those two groups. Instead, *Prunus* variously appears by itself within an array of "spiraeoid" lineages, as sister to one of those, or as sister to Maloideae s. l., but never with strong support (Potter, 2003). Furthermore, neither *Pygeum* nor *Maddenia* has been sampled in any of these family-level analyses; however, Lee and Wen's (2001) phylogenetic analysis of *Prunus* based on ITS data indicated that *Maddenia* is nested within *Prunus*.

The nuclear ITS study of 38 *Prunus* species by Lee and Wen (2001) revealed two major clades within *Prunus*, *Amygdalus*-*Prunus* (the plum / apricot / peach clade) and *Cerasus*-*Laurocerasus*-*Padus* (the cherry clade). None of the individual subgenera was supported as monophyletic and accessions of subgenus *Cerasus* section *Microcerasus* were embedded in the plum / apricot / peach clade. These relationships were also supported in other studies that used somewhat different taxon sampling and additional nuclear / chloroplast DNA loci (Bortiri et al., 2001; Shaw and Small, 2004 [Part 3 of this dissertation]). The consensus of these studies is that a broad interpretation of *Prunus* is warranted because: 1) many of the previously recognized subgenera and sections are not monophyletic; and 2) genetic variability across the genus is relatively low. Because of limited phylogenetic resolution provided by the markers chosen coupled with the limitations on the number of species sampled, none of these studies has unequivocally supported or refuted classical groups based on morphology. Therefore, there is still no well-resolved phylogenetic framework on which infrageneric revision can be based.

It is evident from the foregoing section that *Prunus* is in need of further taxonomic and phylogenetic study and that the genus provides opportunities to address numerous compelling evolutionary questions. Recent molecular phylogenetic analyses have elucidated various aspects of *Prunus* systematics, but because these studies differed in the specific issues addressed and the taxa sampled, and because a significant number of species has not been included in any of them, many questions about infrageneric taxon delimitation, character evolution, and historical biogeography remain unresolved. The primary focus of this dissertation is to study the phylogenetic relationships of one group within *Prunus*, the North American plums of subgenus *Prunus* section *Prunocerasus*.

***Overview of* Prunus *subgenus* Prunus *section* Prunocerasus—**Although many *Prunus* species are economically significant, the Eurasian plums of section *Prunus* are those most heavily used in agriculture. But, they have proven difficult to grow in many parts of North America because they are not well adapted to drought, frost, and particularly in the eastern and southern states, fungal, bacterial, and viral diseases (Fogle,

1978). Section *Prunocerasus* taxa are currently not cultivated on a large scale, although they have been in the past (Waugh, 1899; 1901: *P. americana*, *P. hortulana*, *P. angustifolia*, *P. subcordata*), and more recently, *P. maritima* (R. Uva, pers. comm.; Dunn, 2002; Karp, 2003) and *P. angustifolia* (Reid and Gast, 1993) are under study in agricultural trials. Because nearly all section *Prunocerasus* species are known to naturally hybridize, not only with themselves but also with section *Prunus*, the potential for better-adapted agricultural and horticultural varieties may be unlimited (Wight, 1915). Over 2800 varieties of European plums have been cultivated (Fogle, 1978) and since there are several more North American plum species than there are European plum species, the potential for cultivated varieties is greater (Hedrick et al., 1911). Thus, understanding the classification and evolutionary history of section *Prunocerasus* is of economic importance as well as biological interest.

Members of section *Prunocerasus* are adapted to wide climatic, edaphic, and geographic ranges from the sand pine scrub of central Florida (*P. geniculata*) to the salty coastal dunes of New England (*P. maritima*) north to damp alluvial stream valleys and limestone ridges of Canada (*P. nigra*). *Prunus americana* has the greatest range and is common to perennially damp soils throughout the eastern U.S. While section *Prunocerasus* is generally an eastern North American taxon that is absent throughout the Rocky Mountain forests, Palouse Prairies, and the Intermountain and warm deserts of the western U.S., *P. subcordata* is a western disjunct species that prefers rich moist soils along the western slope of the Sierra Nevada Mountains of California.

Although some species of North American plums are relatively common and widespread (e.g. *P. americana*), several others are of conservation concern. Of greatest concern is *P. geniculata* that has a global distribution of 4 counties in central Florida and is Federally Endangered. *Prunus maritima* var. *gravesii* is presumed extirpated and was only known from a single population in Connecticut. Seven other species in section *Prunocerasus* are legally protected at the state level (see Table 1, all tables and figures located in the Appendix following the Literature Cited).

Waugh (1899, 1901) was among the first investigators of the North American plums and he noted their extensive hybridization and intergradation. He divided them into four informally named series, each of which he described as a "hybrid swarm" of species that graded into each other where their ranges overlapped. For example, he wrote that within the Hortulana Series *P. hortulana* var. *mineri* Bailey grades into *P. americana* to the north and into *P. hortulana* var. *waylandii* Bailey to the south and then also into *P. munsoniana* and eventually *P. angustifolia* (Waugh, 1899). He also thought that *P. hortulana* and its relatives were of hybrid origin (*P. americana* x *P. angustifolia*). Wight (1915) provides the most complete treatment of *Prunocerasus* and he informally recognized six groups (see Table 1), which are morphologically coherent and largely in agreement with Waugh (1899, 1901).

Morphological taxonomy has been difficult within section *Prunocerasus* because species boundaries are blurred by interspecific similarities and intraspecific variation, and likely by interspecific hybridization (Waugh 1899, 1901; Hedrick et al., 1911; Wight, 1915; Rehder, 1940). However, near the centers of their respective ranges, many species are morphologically relatively easily separable from the rest. There are 17 commonly recognized taxa in section *Prunocerasus* (see Table 1), several of which grade into other

taxa. This has created the taxonomic confusion. For example, Fernald (1950 p. 877) described *P. americana* var. *americana* as "passing insensibly into *P. americana* var. *lanata*" (which was described as *P. lanata* Sudw.) whereas Gleason and Cronquist (1991) placed *P. americana* var. *lanata* as a synonym of *P. mexicana* (characterized by densely pubescent leaves). Robertson (1974 p. 659) noted that these two species need more study "to ascertain their distinctiveness and distribution." In another example, *P. umbellata* has been subdivided into *P. injucunda*, which has pubescent leaves and twigs, and *P. mitis* which has glabrous twigs (Radford et al., 1968); both of these putative taxa differ from *P. umbellata* which is completely glabrous. Sargent (1902) recognized *P. tarda*, but Wight (1915 p. 54) noted "*P. umbellata* ssp. *injucunda* merges imperceptibly into *P. umbellata* ssp. *tarda*." Other examples of taxonomic disagreement abound and excessive 'splitting' within this group was noted as early as 1915 (Wight, 1915). Because fluid morphological characters and blurred geographic species boundaries have resulted in ambiguous and conflicting species delimitation and classification, many section *Prunocerasus* species have been variously accepted, ranked, lumped, and split. This has created a great deal of taxonomic confusion in addition to bewildering nomenclatural synonymy. Clearly, morphological data alone cannot define species boundaries and relationships within section *Prunocerasus*.

The earliest work to employ molecular tools (isozymes) suggested that section *Prunocerasus* is polyphyletic and experienced two immigration events into North America (Mowrey and Werner, 1990). More recently, Lee and Wen (2001) included four *Prunocerasus* species in an ITS analysis of *Prunus*, but their data had weak resolution in this part of the tree. Bortiri et al. (2001) combined data from the cpDNA *trnL-trnF* spacer and ITS sequences to address phylogenetic relationships within *Prunus* and although their data support monophyly of some eastern North American section *Prunocerasus* representatives, *P. subcordata*, of northwest North America, was not resolved from species outside section *Prunocerasus*. Finally, Bortiri et al. (2002), in a second broad analysis of *Prunus*, using the single copy nuclear gene *s6pdh*, showed that *P. subcordata* is sister to other sampled section *Prunocerasus* species. While these studies conditionally support a close relationship among section *Prunocerasus* species, none unequivocally support or refute its monophyly or amply sample to infer infrasectional relationships.

GOALS OF THE PROJECT

Given the lack of understanding of this economically important group of plants, the goals of this dissertation research are:

1. To determine which noncoding chloroplast regions are the most suitable for low-level phylogenetic study. This study is relevant to the North American plums because they are very closely related to one another. To accomplish this we identified 21 noncoding chloroplast DNA regions that had been used in published phylogenetic studies. We used sequences of these 21 regions from 10 seed plant lineages spanning all seed plants to identify which noncoding chloroplast regions potentially offer the greatest number of mutations to phylogenetic studies.

2. To rigorously test the monophyly of section *Prunocerasus*. To accomplish this, we employed a multi-region cpDNA data set including all currently recognized section *Prunocerasus* species as well as representative species of all other sections of *Prunus* subgenus *Prunus* and all other subgenera of *Prunus*.
3. To test taxon monophyly among the North American plums using chloroplast DNA. Because hybridization is known to be prevalent in the group—and therefore species monophyly cannot be assumed—we sampled cpDNA sequences for multiple individuals per taxon to show if (or which of) the North American plum taxa are monophyletic.
4. To infer a nuclear DNA-based phylogeny of the North American plums using a parallel sampling scheme to the cpDNA-based phylogeny. This study is an integral component because phylogenies derived from different genomes are necessary to both corroborate phylogenetic hypotheses and identify reticulation events.

This course of investigation will allow us to assess species homogeneity (or the lack thereof) and detect hybridization and introgression. An understanding of the evolutionary history of the North American plums can then be used to (1) better understand their complicated taxonomy, (2) clarify species boundaries (or lack thereof), (3) evaluate the relationships and status of commonly recognized intraspecific taxa, and (4) gain insight into the phylogeographic history of the North American plums.

Species Concepts and Taxon Identification—Species circumscription in section *Prunocerasus* has long been problematic owing to the variation within species coupled with the continuous (rather than discrete) characters that separate species. A morphological species concept has been the primary working concept in *Prunus* taxonomy and will be employed in this study in identification of specimens for phylogenetic analysis. A primary goal of this study was to use phylogenetic analysis to determine whether species circumscriptions based on a morphological species concept are monophyletic (and thus also phylogenetic species). Identification of taxa followed published descriptions and keys and included the commonly recognized taxa in the following works: Small, 1933; Rehder, 1940, Fernald, 1950, Steyermark, 1963, Radford et al., 1968; Correll and Johnston; 1970, Godfrey, 1988; Wunderlin, 1988; Gleason and Cronquist, 1991.

Multiple Accession Sampling Scheme—Some species of section *Prunocerasus* are easily distinguishable and geographically or ecologically isolated from other taxa, are unlikely to have experienced interspecific gene flow, and thus will not be heavily sampled. For example, *P. subcordata* is easily distinguished, not only morphologically, but it is widely geographically separated. Also, two taxa within the Beach plum clade will not require a multiple accession sampling scheme because *P. geniculata* and *P. maritima* var. *gravesii* have severely restricted ranges and are easily morphologically identifiable from the most widespread member of that clade, *P. maritima* var. *maritima*. However, taxa of both the American and Chickasaw clades (Fig. 1) are morphologically variable and their ranges overlap; therefore these taxa require multiple sampling of each

of the taxa to clarify relationships. Taxon ranges were determined by Little (1971, 1976, 1977) and USDA national PLANTS database (USDA, NRCS, 2002).

***Biogeographic Analysis*—**The species of section *Prunocerasus* present two primary biogeographic questions. The first concerns their origin relative to Eurasian species and the second concerns the nature of the eastern North American (majority of species) – western North American (*P. subcordata*) disjunction. A biogeographical hypothesis of section *Prunocerasus* will provide yet another opportunity to evaluate general biogeographic relationships among Eurasian, western North American, and eastern North American taxa (e.g., Xiang, 1998a, 1998b; Wen, 1999; Davis et al., 2002). This is a geographic distribution that has been the subject of several phylogenetically-based biogeographic analyses that have yet to yield a consensus pattern of relationships. Although a phylogenetic hypothesis is necessary to address these questions, more data are needed to resolve the sister taxon to section *Prunocerasus*. Additionally, because the sister taxon to *Prunus* is unknown (Bortiri et al. 2001) we cannot root a biogeographic hypothesis. Therefore, although we speculate on possible migration routes for the North American plums in the following chapters, and a phylogenetic hypothesis for section *Prunocerasus* will certainly provide a backbone upon which to base a biogeographic hypothesis, this inquiry needs further study.

LITERATURE CITED

BENTHAM, G., AND J. D. HOOKER. 1865. Genera plantarum, vol. 1. London: Reeve and Co.

BORTIRI, E., S. H. OH, F. Y. GAO, AND D. POTTER. 2002. The phylogenetic utility of nucleotide sequences of sorbitol 6-phosphate dehydrogenase in *Prunus* (Rosaceae). *American Journal of Botany* 89: 1697-1708.

BORTIRI, E., S.-H. OH, J. JIANG, S. BAGGETT, A. GRANGER, C. WEEKS, M. BUCKINGHAM, D. POTTER, AND D. PARFITT. 2001. Phylogeny and systematics of *Prunus* (Rosaceae) as determined by sequence analysis of ITS and the chloroplast *trn*L-*trn*F spacer DNA. *Systematic Botany* 26: 797-807.

CORRELL, D. S., AND M. C. JOHNSTON. 1970. Manual of the vascular plants of Texas. Texas Research Foundation, Renner, Texas, USA.

DUNN, K. L. 2002. Growing ripe for development: Beach plums as a regional crop. *Farming: The Journal of Northeast Agriculture.* September, pp. 19, 30-32. (www.farmingmagazine.com).

EVANS R. C., C. S. CAMPBELL, D. POTTER, D. MORGAN, T. ERIKSSON, L. ALICE, S.-H. OH, P. E. BORTIRI, F. GAO, J. SMEDMARK, AND M. ARSENAULT. 2002. A Rosaceae phylogeny. Botany2002 Meetings, Madison, WI. Abstract: http://www.botany2002.org/sympos13/abstracts/9.shtml

EVANS, R. C., AND C. S. CAMPBELL. 2002. The origin of the apple subfamily (Maloideae; Rosaceae) is clarified by DNA sequence data from duplicated GBSSI genes. *American Journal of Botany* 89: 1478-1484.

FERNALD, M. L. 1950. Gray's Manual of Botany, 8th ed. American Book Company, New York, New York, USA.

FOGLE, H. W. 1978. Plum improvement in the United States. *Acta Horticulturae* 74: 35-40.

GLEASON, H. A., AND A. CRONQUIST. 1991. Manual of vascular plants of northeastern United States and adjacent Canada, 2nd ed. New York Botanical Garden, Bronx, New York, USA.

GODFREY, R. K. 1988. Trees, shrubs, and woody vines of northeastern Florida and adjacent Georgia and Alabama. The University of Georgia Press, Athens, Georgia, USA.

GOLDBLATT, P. 1976. Cytotaxonomic studies in the tribe Quillajeae (Rosaceae). *Annals of the Missouri Botanical Garden* 63: 200-206.

HEDRICK, U. P., W. H. ALDERMAN, M. J. DORSEY, O. M. TAYLOR, AND R. WELLINGTON. 1911. The plums of New York. J. B. Lyon Company state printers, Albany, New York, USA.

HO, P.-H. 2000. Cây co Viêt Nam --- an illustrated flora of Vietnam. Ho Chi Minh City: Nha Xuat Ban Tre.

KALKMAN, C. 1965. The Old World species of *Prunus* subgen. *Laurocerasus* including those formerly referred to *Pygeum*. *Blumea* 13: 1-115.

KARP, D. 2003. The beach plums: hard to tame, but growers don't give up. The New York Times. available at http://www.nytimes.com/2003/09/03/dining/03PLUM.html.

LEE, S.-T., AND J. WEN. 2001. A phylogenetic analysis of *Prunus* and the Amygdaloideae (Rosaceae) based on ITS sequences of nuclear ribosomal DNA. *American Journal of Botany* 88: 150-160.

LINNAEUS, C. 1754. Genera plantarum, 5th ed. Stockholm, Sweden

LITTLE, E.L. JR. 1971. Atlas of United States Trees. Vol. 1. Conifers and important hardwoods. US Dept. of Agric. Misc. Publ. 1146.

LITTLE, E.L. JR. 1976. Atlas of the United States Trees, Vol. 3. minor western hardwoods. US Dept. Agric. Misc. Publ. 1314.

LITTLE, E.L. JR. 1977. Atlas of the United States Trees, Vol. 4, minor eastern hardwoods. US Dept. Agric. Misc. Publ. 1342.

MABBERLEY, D. J. 1997. The plant-book: a portable dictionary of the vascular plants utilizing Kubitzki's The families and genera of vascular plants (1990-), Cronquist's An integrated system of classification of flowering plants (1981), and current botanical literature arranged largely on the principles of editions 1-6 (1896/97-1931) of Willis's A dictionary of the flowering plants and ferns. New York: Cambridge University Press.

MORGAN, D. R., D. E. SOLTIS, AND K. R. ROBERTSON. 1994. Systematic and evolutionary implications of *rbc*L sequence variation in Rosaceae. *American Journal of Botany* 81: 890-903.

MOWREY, B. D., AND D. J. WERNER. 1990. Phylogenetic relationships among species of *Prunus* as inferred by isozyme markers. *Theoretical and Applied Genetics* 80: 129-133.

POTTER, D. 2003. Molecular phylogenetic studies in Rosaceae. Pp. 319-351 in: Sharma, A. K., A. Sharma, eds., Plant Genome: Biodiversity and Evolution, Vol. I, Pt. A: Phanerogams. Enfield, New Hampshire: Science Publishers, Inc.

POTTER, D., F. GAO, P. E. BORTIRI, S.-H. OH, AND S. BAGGETT. 2002. Phylogenetic relationships in Rosaceae inferred from chloroplast *mat*K and *trnL-trnF* nucleotide sequence data. *Plant Systematics and Evolution* 231: 77-89.

RADFORD, A. E., H. E. AHLES, AND C. R. BELL. 1968. Manual of the Vascular Flora of the Carolinas. The University of North Carolina Press, Chapel Hill, North Carolina, USA.

REHDER, A. 1940. Manual of cultivated trees and shrubs hardy in North America, exclusive of the subtropical and warmer temperate regions, 2d, rev. and enl. edition. The Macmillan Company, New York, New York, USA.

REID, W., AND K. L. B. GAST. 1993. The potential for domestication and utilization of native plums in Kansas. *In* J. Janick, and J. E. Simon [eds.], New crops, 520-523. Wiley, New York, New York, USA.

ROBERTSON, K. R. 1974. The genera of Rosaceae in the southeastern United States. *Journal of the Arnold Arboretum* 55: 303-332, 344-401, 611-662.

SARGENT, C. S. 1902. Texas, Louisiana, and Arkansas [*P. tarda*]. *Botanical Gazette* 33: 108.

SCHULZE-MENZ, G. K. 1964. Rosaceae. Pp. 209-218 in: Melchior, H., ed., Engler's Syllabus der Pflanzenfamilien II (12th ed.) Berlin: Gebrüder Borntraeger.

SHAW, J., AND R. L. SMALL. 2004. Addressing the "hardest puzzle in American pomology:" Phylogeny of *Prunus* section *Prunocerasus* (Rosaceae) based on seven noncoding chloroplast DNA regions. *American Journal of Botany* 91: 985-996.

SMALL, J. K. 1933. Manual of the southeastern flora. Published by the author, The Science Press Printing Company, Lancaster, Pennsylvania, USA.

STEYERMARK, J. A. 1963. Flora of Missouri. Iowa State University Press, Ames, Iowa, USA.

TAKHTAJAN, A. L. 1997. Diversity and classification of flowering plants. New York: Columbia University Press.

TOURNEFORT, J. P. DE. 1700. Institutiones Rei Herbariae, ed. 1. Paris, France.

USDA, NRCS. 2002. The PLANTS Database, Version 3.5 (http://plants.usda.gov). National Plant Data Center, Baton Rouge, Louisiana, USA.

WAUGH, F. A. 1899. The pollination of plums. *Vermont Agricultural Experimentation Station, Annual Report* 12: 189-209.

WAUGH, F. A. 1901. Plums and plum culture. Orange Judd, New York, New York, USA.

WEN, J. 1999. Evolution of Eastern Asian and Eastern North American disjunct distributions in flowering plants. Annual Review of Ecology and Systematics 30: 421-455.

WIGHT, W. F. 1913. North American species of the genus *Amygdalus*. Dudley Memorial Volume. Leland Stanford Junior University Publications, University Series, California, USA.

WIGHT, W. F. 1915. Native American species of *Prunus*. *Bulletin of the U.S. Department of Agriculture* 179: 1-75.

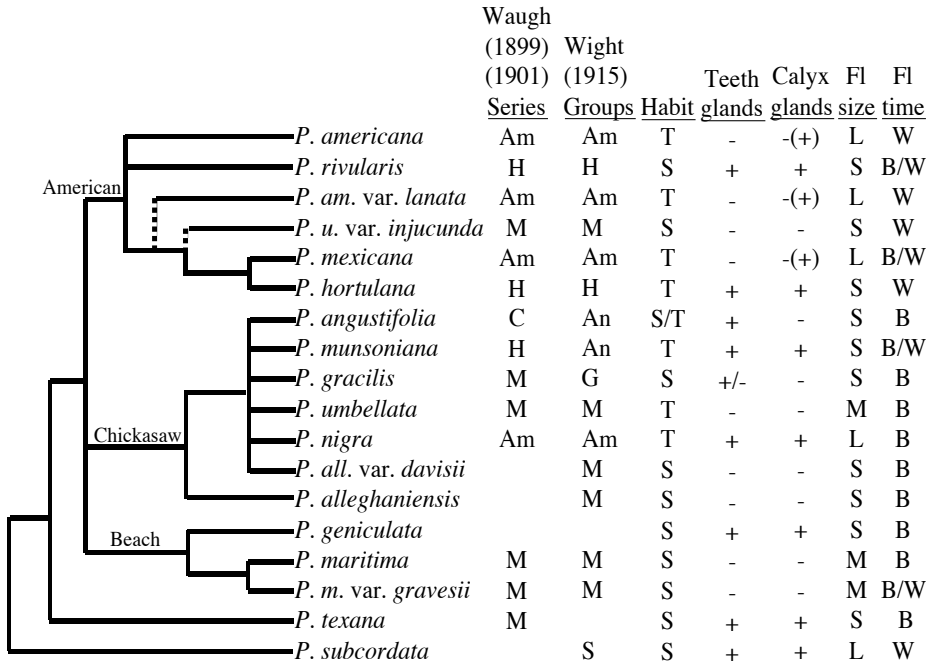
WILLIS, J. C. 1973. A dictionary of the flowering plants and ferns, 8th ed., revised by H. K. Airy Shaw. New York: Cambridge University Press.

WUNDERLIN, R. P. 1988. Guide to the vascular plants of Florida. University of Florida Press, Gainesville, Florida, USA.
XIANG, Q.Y., D.E. SOLTIS, AND P.S. SOLTIS. 1998b. The Eastern Asian and Western North American floristic disjunction: congruent phylogenetic patterns in seven diverse genera. Molecular Phylogenetics and Evolution 10: 178-190.
XIANG, Q.Y., D.J. CRAWFORD, A.D. WOLF, Y.C. TANG, AND C. DEPAMPHILIS. 1998a. Origin and biogeography of *Aesculus* L. (Hippocastanaceae): a molecular phylogenetic perspective. Evolution 52: 988-997.
YÜ, T. T., L. T. LU, T. C. KU, C. L. LI AND S. X. CHEN. 1986. Rosaceae (3), Amygdaloideae. Flora Reipublicae Popularis Sinicae, vol. 38. Beijing: Science Press.
ZHANG S-Y. 1992. Systematic wood anatomy of the Rosaceae. *Blumea* 37: 81-158.

APPENDIX; PART 1

Table 1. Commonly recognized taxa within *Prunocerasus*. Modified classification scheme of Wight (1915) and status of species of conservation concern. State listings are from USDA (2002): E=Endangered and T = Threatened. *P. geniculata* is Federally Endangered.

	Prunocerasus	Species	Cons. concern	Source and Voucher
Wight's (1915) Groups		*P. geniculata* R.M. Harper	FE	JSh898; TENN
	Americana	*P. americana* Marshall	T-NH, VT	JSh038; TENN
		P. a. Marshall var. *lanata* Sudw.		J. Beck 4955; TENN
		P. mexicana S. Watson	PE-OH	JSh919-072; TENN
		P. nigra Ait.	E-IA, PE-OH	JSh979-125; TENN
	Angustifolia	*P. angustifolia* Marshall	E-NJ	JSh785; TENN
		P. munsoniana Wight & Hedrick		JSh810; TENN
	Gracilis	*P. gracilis* Engelm. & A. Gray		JSh936-088; TENN
	Hortulana	*P. hortulana* L.H. Bailey		JSh821-017; TENN
		P. rivularis Scheele		Endquist 3372; BRIT
	Maritima	*P. alleghaniensis* Porter	E-NJ, T-MD, PA	JSh837-030; TENN
		P.a. var. *davisii* (Wight.) Sarg.		G. Schmidt; TENN
		P. maritima Marshall	E-ME, MD, PA	JSh877-045; TENN
		P. m. var. *gravesii* (Small) G.J. Anderson	PE-CT	Conn Greenhouse; TENN
		P. umbellata Elliott		JSh774-003; TENN
		P. u. Elliott var. *injucunda* (Small) Sarg.		JSh958-108; TENN
	Subcordata	*P. subcordata* Benth.		Syring, J. 1; TENN



	Waugh (1899) (1901) Series	Wight (1915) Groups	Habit	Teeth glands	Calyx glands	Fl size	Fl time
P. americana	Am	Am	T	-	-(+)	L	W
P. rivularis	H	H	S	+	+	S	B/W
P. am. var. *lanata*	Am	Am	T	-	-(+)	L	W
P. u. var. *injucunda*	M	M	S	-	-	S	W
P. mexicana	Am	Am	T	-	-(+)	L	B/W
P. hortulana	H	H	T	+	+	S	W
P. angustifolia	C	An	S/T	+	-	S	B
P. munsoniana	H	An	T	+	+	S	B/W
P. gracilis	M	G	S	+/-	-	S	B
P. umbellata	M	M	T	-	-	M	B
P. nigra	Am	Am	T	+	+	L	B
P. all. var. *davisii*		M	S	-	-	S	B
P. alleghaniensis		M	S	-	-	S	B
P. geniculata			S	+	+	S	B
P. maritima	M	M	S	-	-	M	B
P. m. var. *gravesii*	M	M	S	-	-	M	B/W
P. texana	M		S	+	+	S	B
P. subcordata		S	S	+	+	L	W

Fig. 1. Phylogenetic relationships of section *Prunocerasus* shown with five classically used morphological characters. Habit: T=tree, S=shrub; Teeth glands: + = glands present on each tooth of the leaf blade, - = glands absent on each tooth of the leaf blade; Calyx glands: + = glands on margin of calyx lobes, - = glands absent on margin of calyx lobes; Flowers large (L) = > 15 mm wide, Flowers small (S) = < 15 mm wide, Flowers midsized (M) = ~ 15 mm wide; Time of flowering (Fl): W = flowers emerging with the leaves, B = flowers present before leaves emerge.

Part 2

THE TORTOISE AND THE HARE II: RELATIVE UTILITY OF 21 NONCODING CHLOROPLAST DNA SEQUENCES FOR PHYLOGENETIC ANALYSIS

This chapter is a lightly revised version of a paper by the same name published in the *American Journal of Botany* in 2005 by Joey Shaw, Edgar B. Lickey, John T. Beck, Susan B. Farmer, Wusheng Liu, Jermey Miller, Kunsiri C. Siripun, Charles T. Winder, Edward E. Schilling, and Randall L. Small:

Shaw, J., E. Lickey, J. Beck, S. Farmer, W. Liu, J. Miller, K. C. Siripun, C. Winder, E. E. Schilling, and R. Small. 2005. The tortoise and the hare II: relative utility of 21 non-coding chloroplast DNA sequences for phylogenetic analysis. *American Journal of Botany* 92: 142-166.

My use of "we" in this chapter refers to my co-authors and myself. My primary contributions to this paper include (1) development of the core ideas underlying the project, (2) initiation of the proposal to the Hesler Fund for the money that supported this project, (3) a portion of the lab work not limited to the sequences for *Prunus* and, most importantly, coordination of the laboratory efforts of my co-authors, and (4) most of the writing.

ABSTRACT

Chloroplast DNA sequences are a primary source of data for plant molecular systematic studies. A few key papers have provided the molecular systematics community with universal primer pairs for noncoding regions that have dominated the field, namely *trnL-trnF* and *trnK/matK*. These two regions have provided adequate information to resolve species relationships in some taxa, but often provide little resolution at low taxonomic levels. To obtain better phylogenetic resolution, sequence data from these regions are often coupled with other sequence data. Choosing an appropriate cpDNA region for phylogenetic investigation is difficult because of the scarcity of information about the tempo of evolutionary rates among different noncoding cpDNA regions. The focus of this investigation was to determine whether there is any predictable rate heterogeneity among 21 noncoding cpDNA regions identified as phylogenetically useful at low levels. To test for rate heterogeneity among the different cpDNA regions, we used three species from each of ten groups representing eight major phylogenetic lineages of phanerogams. The results of this study clearly show that a survey using as few as three representative taxa can be predictive of the amount of phylogenetic information offered by a cpDNA region and that rate heterogeneity exists among noncoding cpDNA regions.

INTRODUCTION

Chloroplast DNA sequences are the primary source of data for inferring plant phylogenies, rivaled only perhaps by nuclear ribosomal ITS sequences in recent years (Baldwin, 1992; Baldwin et al., 1995; Álvarez and Wendel, 2003). Early in the plant molecular systematics era chloroplast DNA (cpDNA) was surveyed through restriction site polymorphism studies (see Olmstead and Palmer [1994] for a review of cpDNA studies through the early 1990s). As DNA sequencing technology became available, comparative studies of cpDNA gene sequences began to accumulate sparked by the observations of Ritland and Clegg (1987) and Zurawski and Clegg (1987). A landmark

publication, the angiosperm *rbcL* study of Chase et al. (1993), set the stage for the increased use of cpDNA sequences for phylogenetic studies. Most early publications employed sequences of *rbcL* and were focused on suprageneric taxonomic questions (e.g., Chase et al., 1993). Subsequent workers began to explore additional gene sequences such as *ndhF* (Olmstead and Sweere, 1994; Olmstead and Reeves, 1995; Clark et al., 1995; Kim and Jansen, 1995), *atpB* (Hoot et al., 1995; Jensen et al., 1995; Wolf, 1997), and *matK* (Johnson and Soltis, 1994; Steele and Vilgalys, 1994). Simultaneously, noncoding regions of the chloroplast were being explored for lower level taxonomic studies under the assumption that noncoding regions should be under less functional constraint than coding regions and should provide greater levels of variation for phylogenetic analyses (Gielly and Taberlet, 1994). Among the first regions to be exploited were the *trnT-trnL-trnL-trnF* region (Taberlet et al., 1991), the *atpB-rbcL* intergenic spacer (Golenberg et al., 1993; Ehrendorfer et al., 1994; Hodges and Arnold, 1994; Manen et al., 1994), and the noncoding intron portions of the *trnK*/*matK* region (Johnson and Soltis, 1994; Steele and Vilgalys, 1994). Following these pioneering studies, the use of noncoding cpDNA regions has continually increased and is now routinely employed for studies of phylogeny at intergeneric and interspecific levels. Even though many noncoding regions have been explored by different workers (e.g., Taberlet et al., 1991; Johnson and Soltis, 1994; Demesure et al., 1995, Dumolin-Lapegue et al., 1997; Sang et al., 1997; Small et al., 1998) many investigators continue to use a limited number of regions. A survey of papers published from 1995 through 2002 in *American Journal of Botany*, *Systematic Botany*, *Molecular Phylogenetics and Evolution*, and *Plant Systematics and Evolution* illustrates that the number of investigations employing noncoding cpDNA is rapidly increasing (Fig. 1; all tables and figures are located in the Appendix following the Literature Cited). However, of 445 studies, 342 (77%) used some portion of either *trnK-matK-trnK*, the *trnL* intron, and/or the *trnL-trnF* spacer. Two other relatively popular regions are the *rpS16* and *rpL16* introns. Studies that employed *rpS16*, *rpL16*, *trnK-matK-trnK*, or *trnL-trnL-trnF* (either alone or in combination with other regions) account for approximately 84% of all noncoding cpDNA-based phylogenetic investigations since 1995 and approximately 83% of the studies in 2002. This illustrates that, although the number of phylogenetic investigations using noncoding cpDNA is increasing every year, so too is the continued reliance on a few regions. Figure 1 also shows the slow increase in the use of other noncoding cpDNA regions, such as the *trnH-psbA* and *trnS-trnG* intergenic spacers, which have nearly always been added to supplement data collected from *trnL-trnL-trnF* or *trnK-matK-trnK*. It is important to note this apparent reliance on a few regions is in spite of the fact that in comparative studies, the phylogenetic utility of *trnL-trnL-trnF* and *trnK*/*matK* is often limited with respect to other regions (Sang et al., 1997; Small et al., 1998; see below for others).

As the majority of current phylogenetic investigations are focused at shallower phylogenetic levels, regions like the *trnL* intron, the *trnL-trnF* spacer, and the *trnK* intron/*matK* gene have provided satisfactory information in some groups (Bellstedt et al., 2001; Ge et al., 2002), but often yield poor resolution in other groups, at least in some clades (Bell and Patterson, 2000; Cuénoud et al., 2000; Hardig et al., 2000; Goldblatt et al., 2002; Klak et al., 2003; Muellner et al., 2003; Samuel et al., 2003). To obtain

additional data and provide better phylogenetic resolution, sequences from these popular regions are often coupled with other sequence data, cpDNA or otherwise (Sang et al., 1997; Wang et al., 1999; Hardig et al., 2000; Kusumi et al., 2000; Azuma et al., 2001; Bortiri et al., 2001; Soltis et al., 2001; Bayer et al., 2002; Cronn et al., 2002; Hartmann et al., 2002; Mast and Givnish, 2002; Nyffeler, 2002; Schönenberger and Conti, 2003; Yamane et al., 2003), because additional data are often required to generate a phylogenetic hypothesis with acceptable resolution.

It has been clearly shown that the phylogenetic utility of different noncoding cpDNA regions within a given taxonomic group can vary tremendously (Sang et al., 1997; Small et al., 1998; Xu et al., 2000; Hartmann et al., 2002; Mast and Givnish, 2002; Cronn et al., 2002; Hamilton et al., 2003; Perret et al., 2003; Sakai et al., 2003), but choosing an appropriate cpDNA region for phylogenetic investigation is often difficult because of the paucity of information about the relative tempo of evolution among different noncoding cpDNA regions. Gielly and Taberlet (1994, p. 774) wrote: "it is not easy, for many reasons, to establish a rule for the choice of a particular region of the chloroplast genome for resolving phylogenies." While many authors have compared relative rates of evolution among a few noncoding regions (Sang et al., 1997; Small et al., 1998; Wang et al., 1999; Kusumi et al., 2000; Xu et al., 2000; Soltis et al., 2001; Cronn et al., 2002; Mast and Givnish, 2002; Hamilton et al., 2003; Perret et al., 2003; Sakai et al., 2003; Yamane et al., 2003), these studies are all of a relatively narrow phylogenetic context and there is no consensus as to variability in evolutionary rates among noncoding cpDNA regions across a broad phylogenetic range. To our knowledge, the only work that has attempted to compare levels of variation among several different noncoding cpDNA regions across a wide range of lineages is Aoki et al. (2003). However, their results are equivocal because of insufficient data. Therefore, for most investigators, choosing the appropriate region for phylogenetic investigation at a particular taxonomic level is often guesswork.

We present a comparison of 21 noncoding cpDNA regions sampled across all of the major lineages of phanerogams sensu APG II (2003) (Fig. 2). Sequence divergence and, more importantly, the amount of information offered to phylogenetic investigations by the various noncoding cpDNA regions is compared across lineages to assess the phylogenetic utility of each. In this investigation, we determine whether there is any predictable rate heterogeneity among different noncoding chloroplast regions that have been employed in the field of molecular systematics. We will also provide a discussion of the often used noncoding cpDNA regions and present a general protocol for selecting potential noncoding cpDNA regions useful to systematic investigations.

MATERIALS AND METHODS

***Taxonomic sampling*—**Species and lineages sampled in this study are listed in Table 1. Sampling focused on representing all major phanerogam lineages sensu APG II (2003) (Fig. 2, Table 1) in addition to representing different habits and life strategies (e.g., woody perennials, herbaceous perennials, and herbaceous annuals). Three fairly closely related species were chosen to represent each of 10 lineages. Earlier workers have shown that analysis of very closely related species, or even accessions of the same species, is likely to yield little or no information (e.g., Aoki et al., 2003) which would

limit a comparison of different noncoding cpDNA regions. We therefore chose three species within each lineage that we knew from other studies, or our own unpublished data, were from separate but closely related clades. For each lineage two species were chosen to represent ingroup taxa of different clades, while the third was chosen as a closely related outgroup taxon (O.G.). Voucher information and GenBank accession numbers are listed in Table 1.

***cpDNA sampling*—**After extensive literature review, 21 noncoding cpDNA regions were identified that had been previously used in interspecific or intraspecific phylogenetic investigations (Fig. 3). Some additional regions were added because they flank these previously utilized regions. All regions surveyed in this study occur in the Large Single Copy (LSC) region of the chloroplast genome. Listed as they occur on the Wakasugi et al. (1998) *Nicotiana* cpDNA map starting at the junction of Inverted Repeat A, they include: *trnH-psbA*; *psbA*-3´*trnK*; 3´*trnK-matK*; *matK*-5´*trnK*; *rpS*16 intron; *trnS-trnG*; *trnG* intron; *rpoB-trnC*; *trnC-ycf6*; *ycf6-psbM*; *psbM-trnD*; *trnD-trnT*; *trnS-trnfM*; *trnS-rpS4*; *rpS4-trnT*; *trnT-trnL*; *trnL* intron; *trnL-trnF*; 5´*rpS12-rpL20*; *psbB-psbH*; and *rpL*16 intron. Based on the Wakasugi et al. (1998) *Nicotiana* chloroplast map, these 21 regions comprise 14321 bp (35%) of the 40732 bp of the noncoding LSC.

Because the main focus of this investigation was to highlight cpDNA regions that may be the most beneficial to low-level systematic studies, coding cpDNA regions were excluded because they tend to provide fewer variable characters than their noncoding counterparts. Although the *rbcL* gene has arguably been the largest contributor to our phylogenetic understanding of chloroplast-containing life forms and has even been suggested to be nearly as informative as some noncoding regions of the cpDNA molecule (Chase et al., 2000), it was not included here. This is because *rbcL* is "sometimes too conserved to clarify relationships between closely related genera" (Gielly and Taberlet, 1994, p. 769) and other studies have shown it to provide fewer variable characters than several different noncoding regions (e.g., Renner, 1999; Richardson et al., 2000; Asmussen and Chase, 2001; Stefanovic et al., 2002; Salazar et al., 2003).

The *atpB-rbcL* spacer, perhaps one of the first intergenic spacers to be widely used, was excluded from our analysis because it is apparently of little infrageneric phylogenetic utility. It has consistently provided fewer variable characters compared to the entire *trnK* intron (Azuma et al., 2001), *trnH-psbA* (Azuma et al., 2001; Schönenberger and Conti, 2003; Hamilton et al., 2003), 5´*rpS12-rpL20* (Hamilton et al., 2003), *rpL16* (Renner, 1999; Schönenberger and Conti, 2003), *rpS16* (Schönenberger and Conti, 2003), or *trnL-trnL-trnF* (Mayer et al., 2003).

Another well-characterized region found in the literature but excluded from this study is the *rpoC1* intron. The *rpoC1* intron was excluded here because it was shown to be less informative in cotton (*Gossypium* sp.) than *atpB-rbcL*, *trnL-trnF*, *ndhA*, and *rpL16* (Small et al., 1998) and it yielded fewer characters than *rpL16*, *rpS16*, and *matK* in a study of the Apiaceae subfamily Apioideae (Downie et al., 2001). Although this region appears to show appropriate levels of variation for studies above the family level, it was noted as being "largely inappropriate to infer phylogeny among closely related taxa" (Downie et al., 1996, p. 14).

For the sake of clarity, we wish to point out that it is important to use specific terminology to describe a region of interest. For example, authors have used "*trnL-trnF*" to mean either the *trnL* intron plus *trnL-trnF* spacer or just the *trnL-trnF* spacer. To be precise we will use, for example, "*trnL-trnF*" to indicate the intergenic spacer alone, but "*trnL-trnL-trnF*" to indicate the intron plus the intergenic spacer. In addition, because there are multiple tRNA genes in the chloroplast genome that encode tRNAs for the same amino acid, it is desirable to denote the specific tRNA gene by the addition of the anti-codon as a superscript. For example, one of the regions we found to be highly variable is the $trnS^{GCU}$-$trnG^{UUC}$ intergenic spacer, which is different than the $trnS^{UGA}$-$trnG^{GCC}$ intergenic spacer that lies within the $trnS^{UGA}$-$trnfM^{CAU}$ region (Fig 3).

***Molecular techniques*—**Because the genes surrounding noncoding regions are highly conserved across seed plants (and especially within angiosperms), many polymerase chain reaction (PCR) primers for amplification and sequencing could be used across the diverse taxonomic groups of this study. Nearly all of the primer regions used here were published in other studies. However, alignment of GenBank sequences from a wide array of phanerogam lineages was used to determine the universality of the previously published primers, modify problematic primers, and aid in the construction of new primers. In some cases, we designed new primers for regions not previously surveyed, or to help sequence through difficult regions (e.g., polynucleotide runs). Unless otherwise noted, all of the primers listed below and in Fig. 3 were successfully used for both amplification and sequencing reactions in all taxonomic groups.

DNA was extracted from leaf tissue using either the DNeasy Plant Mini Kit (Qiagen, Valencia, California, USA) or the CTAB method (Doyle and Doyle, 1987). PCR was performed using either Eppendorf or MJ Research thermal cyclers in 20-50 μL volumes with the following reaction components: 1 μL template DNA (~ 10-100 ng), 1X buffer (PanVera / TaKaRa, Madison, Wisconsin, USA or Promega, Madison, Wisconsin, USA), 200 μmol/L each dNTP, 3.0 mmol/L $MgCl_2$, 0.1 μmol/L each primer, and 1.25 units *Taq* (PanVera / TaKaRa or Promega). Some reactions included bovine serum albumin with a final concentration of 0.2 μg/μL to improve amplification of difficult templates. In a few cases, 10 μmol/L tetramethyl ammonium chloride (TMACl) was included in the PCR solution because it is reported to reduce problems associated with long polynucleotide runs (Oxelman et al., 1997). However, we did not perform a comparative study to determine whether or not its presence actually improved our sequences. PCR amplification protocols and reaction conditions were continuously optimized throughout this investigation for all regions across all lineages. Material and methodological information and primer sequences specific to each of the different noncoding cpDNA regions are described below. All primer sequences are written in standard 5' to 3' orientation and their relative positions and orientations are illustrated in Fig. 3. A key to the shorthand for the following PCR parameters is as follows: initial denaturing step (temperature, time); number of repetitions of the amplification cycle [#X (denaturing temperature, time; primer annealing temperature, time; chain extension temperature, time)]; final extension step (temperature, time). All reactions ended with a final 4°C hold step.

PCR products were purified prior to sequencing with either the QIAquick PCR Purification Kit (Qiagen, Valencia, California, USA) or ExoSAP-IT (USB, Cleveland, Ohio, USA). All DNA sequencing was performed with the ABI Prism BigDye Terminator Cycle Sequencing Ready Reaction Kit, v. 2.0 or 3.1 (Perkin-Elmer / Applied Biosystems, Foster City, California, USA), using the thermal cycle parameters 80°C, 5 min; 30X (96°C, 10 s; 50°C, 5 s; 60°C, 4 min). The products were electrophoresed and detected on an ABI Prism 3100 automated sequencer (University of Tennessee Molecular Biology Resource Facility). All sequences have been deposited in GenBank, and accession numbers are provided in Table 1.

trnH^{GUG}-*psbA*—The PCR parameters for this region were 80°C, 5 min; 35X(94°C, 30 s; 50-56°C, 30 s; 72°C, 1 min); 72°C, 10 min with primers **trnH**^{GUG} (CGC GCA TGG TGG ATT CAC AAT CC) (Tate and Simpson, 2003) and **psbA** (GTT ATG CAT GAA CGT AAT GCT C) (Sang et al., 1997). This region amplified and sequenced easily for all lineages. Because the average length of this region is relatively short (~500 bp), only the *trnH* primer was used in sequencing in most cases.

psbA-3´*trnK*^{UUU}-[*matK*]-5´*trnK*^{UUU}—These regions were the most problematic of any in this investigation. A variety of previously published and newly designed primers were required to amplify and sequence these regions, and very few completely universal primers were identified. We included only the noncoding portions of this region: *psbA*-3´*trnK* spacer, 3´*trnK*-*matK* intron, and *matK*-5´*trnK* intron. The *matK* gene was excluded primarily because it is a coding region, but also because of the inefficiency in designing the many primers that would be necessary to obtain this region for all lineages. In many cases, after amplifying the entire *trnK*-*matK*-*trnK* fragment, we were unable to sequence the PCR product with either the amplification or internal primers. However, if the region was PCR amplified in smaller sections using internal primers we were able to sequence these amplicons using the same primers that had previously failed. This phenomenon was observed independently in the laboratories of both E. E. Schilling and R. L. Small, as well as by J. Panero (University of Texas, personal communication) and R. Rapp (Iowa State University, personal communication) who suggested that dimethylsulfoxide might help during sequencing. Different primer combinations were often required for different taxa. The gymnosperm lineage is not represented in this data set because gymnosperm-specific primers were not obtained (Kusumi et al., 2000). The primers used in this study include: **psbA5´R** (AAC CAT CCA ATG TAA AGA CGG TTT), **ALS-11F** (ATC TTT CGC ATT ATT ATA G) (M. Nepokroeff, University of South Dakota, personal communication), **matKAR** (CTG TTG ATA CAT TCG A) (Kazempour Osaloo et al., 1999), **matKM** (TCG ACT TTC TGG GCT ATC) (Tate and Simpson, 2003), **matK1** (AAC TAG TCG GAT GGA GTA G) (Johnson and Soltis, 1994), **matK5** (TGT CAT AAC CTG CAT TTT CC) (Panero and Crozier, 2003), **matK5´R** (GCA TAA ATA TAY TCC YGA AAR ATA AGT GG), **matK6** (TGG GTT GCT AAC TCA ATG G) (Johnson and Soltis, 1994), **matK8F** (TCG ACT TTC TTG TGC TAG AAC TTT) (Steele and Vilgalys, 1994), **matK5PSIF** (CTA TGG CTC CAA TTC TGG T), **matK5PSIR** (CCG CAT CAG GCA CTA ATC TA).

Hibiscus and *Minuartia* protocol: Amplification of the *matK*-5´*trnK* spacer used the matK6 and matK5´R primers with the PCR parameters 80°C, 5 min; 35X(95°C, 1 min; 50°C, 1 min with a ramp of 0.3°C/s; 65°C, 5 min); 65°C, 5 min. This spacer was sequenced with the matK6 primer. The *psbA-trnK-matK* spacers were amplified using the matKM (*Hibiscus*) or ALS-11F (*Minuartia*) and psbA5´R primers using the parameters 80°C, 5 min; 30X(94°C, 30 s; 50°C, 30 s; 72°C, 2 min); 72°C, 5 min. This region was sequenced using the psbA5´R primer.

Magnolia, *Prunus*, and *Gratiola* protocol: Amplification of the *matK*-5´*trnK* spacer used the matK6 and matK5 primers with the parameters 80°C, 5 min; 30-35X(94°C, 1 min; 50°C, 1 min; 72°C, 1.5 min); 72°C, 5 min. Amplification of the *psbA*-3´*trnK-matK* spacers was done using the matK8F and psbA5´R primers with the same PCR protocol.

Trillium-Pseudotrillium protocol: Amplification of the *matK*-5´*trnK* spacer used the matK6 and matKAR primers with the parameters 80°C, 5 min; 30-35X(94°C, 1 min; 50°C, 1 min; 72°C, 2 min); 72°C, 5 min. Amplification of the *psbA*-3´*trnK-matK* spacers used the matK8F and psbA5´R primers with the same PCR parameters. Because of two poly-A/T runs, matK5PSIF and matK5PSIR were used for internal sequencing.

Solanum, *Carphephorus-Trilisa*, *Eupatorium* protocol: Amplification of the *matK*-5´*trnK* spacer used the matK6 and matK5 primers with the parameters 80°C, 5 min; 35X(95°C, 1 min; 50°C, 1 min; 65°C, 5 min); 65°C, 5 min. Both primers were also used for sequencing reactions. The *psbA*-3´*trnK-matK* spacers were amplified with the psbA5´R and ALS-11F for *Solanum americanum* and *S. physalifolium*, matKM for *S. ptychanthum,* and matK8F for *Eupatorium* and *Carphephorus-Trilisa* with the above parameters. All were sequenced using only the psbA5´R primer.

rpS16—This region was amplified using the parameters 80°C, 5 min; 35X(94°C, 30 s; 50-55°C, 30 s; 72°C, 1 min); 72°C, 5 min, with primers **rpS16F** (AAA CGA TGT GGT ARA AAG CAA C) and **rpS16R** (AAC ATC WAT TGC AAS GAT TCG ATA), which are modified from Oxelman et al. (1997). Both primers were also used in sequencing reactions. This region amplified and sequenced easily for all angiosperm taxa and two of the three gymnosperm representatives with minimal troubleshooting. Despite trying several different PCR programs, annealing temperatures, and $MgCl_2$ concentrations, we were unable to amplify this region for *Cryptomeria japonica*.

trnS^{GCU}-*trnG*^{UUC}-*trnG*^{UUC}—For this region, three different protocols were used and in most cases the *trnS-trnG* spacer and the *trnG* intron were amplified as one fragment. For most taxa protocol 1was successful. Both protocols 1 and 2 used the primers **trnS**$^{\mathbf{GCU}}$ (AGA TAG GGA TTC GAA CCC TCG GT) and **3´trnG**$^{\mathbf{UUC}}$ (GTA GCG GGA ATC GAA CCC GCA TC). Additional primers **5´trnG2G** (GCG GGT ATA GTT TAG TGG TAA AA) (toward *trnG*) and **5´trnG2S** (TTT TAC CAC TAA ACT ATA CCC GC) (toward *trnS*) were sometimes used to amplify only the *trnG* intron, and for sequencing longer fragments and templates with a difficult poly-A repeat.

Protocol 1: This is a two-step PCR protocol with primer annealing and chain extension occurring at the same temperature, using the parameters 80°C, 5 min;

30X(95°C, 1 min; 66°C, 4 min); 66°C, 10 min. A final $MgCl_2$ concentration of 1.5 mmol/L (rather than 3.0 mmol/L) was used.

Protocol 2: This protocol was used when amplification with protocol 1 was problematic. The parameters are 80°C, 5 min; 35X(95°C, 1 min; 50°C, 1 min with a ramp of 0.3°C/s; 65°C, 5 min); 65°C, 10 min. This protocol always coamplifies the *trnS*^{UGA} and *trnG*^{GGC} part of the *trnS*^{UGA}-*trnfM*^{CAU} spacer. The result of this protocol yields two equal-intensity, but well-separated bands in a test gel, the larger of which was always the target *trnS*^{GCU}-*trnG*^{UUC}. The desired fragment was excised from the gel and cleaned with a QIAquick Gel Extraction Kit. Because of the sequence similarity of these two different *trnS* and *trnG* genes, primer design was difficult and the protocols needed to be very specific to amplify only the correct region.

Protocol 3: Independent inversions in monocots (Hiratsuka et al., 1989) and Asteraceae (Jansen and Palmer, 1987) interrupt the *trnS*^{UGA}-*trnG*^{GGC} spacer preventing amplification. However, using the 3´trnG and 5´trnG2G primers, we successfully amplified and sequenced the *trnG* intron for *Trillium*-*Pseudotrillium*, *Carphephorus*-*Trilisa*, and *Eupatorium*. The amplification parameters for the *trnG* intron are 80°C, 5 min; 35X(95°C, 1 min; 50°C, 1 min with a ramp of 0.3°C/s; 65°C, 5 min); 65°C, 5 min.

rpoB-*trnC*^{GCA}—This region amplified easily for most angiosperm taxa using primers **trnC^{GCA}R** (CAC CCR GAT TYG AAC TGG GG) and **rpoB** (CKA CAA AAY CCY TCR AAT TG), modified from Ohsako and Ohnishi (2000). The PCR parameters for this region are 80°C, 5 min; 30-35X(96°C, 1 min; 50-57°C, 2 min; 72°C, 3 min); 72°C, 5 min. For unknown reasons, we were unable to amplify this region for *Taxodium*, *Glyptostrobus*, or *Cryptomeria*.

trnC^{GCA}-*ycf6*-*psbM*-*trnD*^{GUC}—Two different, but equally successful, protocols were used to amplify this region. For *Gratiola*, *Hibiscus*, *Magnolia*, *Minuartia*, *Prunus*, and *Taxodium*, we amplified the entire approximately 3-kb *trnC* to *trnD* fragment. For *Carphephorus*-*Trilisa*, *Eupatorium*, *Solanum*, and *Trillium*-*Pseudotrillium*, we amplified the fragments *trnC*-*psbM* and *ycf6*-*trnD*. Both protocols used the same PCR parameters, which were 80°C, 5 min; 35X(94°C, 1 min; 50-55°C, 1 min; 72°C, 3.5 min); 72°C, 5 min. PCR and sequencing primers included **trnC^{GCA}F** (CCA GTT CRA ATC YGG GTG) (modified from Demesure et al., 1995), **ycf6R** (GCC CAA GCR AGA CTT ACT ATA TCC AT), **ycf6F** (ATG GAT ATA GTA AGT CTY GCT TGG GC), **psbMR** (ATG GAA GTA AAT ATT CTY GCA TTT ATT GCT), **psbMF** (AGC AAT AAA TGC RAG AAT ATT TAC TTC CAT), **Taxodium-psbMF2** (CTT TTG TTC GGG TGA GAA AGG), and **trnD^{GUC}R** (GGG ATT GTA GYT CAA TTG GT) (modified from Demesure et al., 1995). This region required only moderate troubleshooting. After trying several different PCR modifications, we were unable to obtain the *psbM*-*trnD* segment for *Carphephorus*-*Trilisa*. In nearly all surveyed lineages, a poly-A/T run exists between *psbM* and *trnD*, but created sequencing difficulties in only a few cases.

trnD^{GUC}-*trnT*^{GGU}—This spacer amplified easily for most taxa using Demesure et al. (1995) primers **trnD^{GUC}F** (ACC AAT TGA ACT ACA ATC CC) and **trnT^{GGU}** (CTA CCA CTG AGT TAA AAG GG). The PCR parameters for this region are 80°C, 5 min;

30X(94°C, 45 s; 52-58°C, 30 s; 72°C, 1 min); 72°C, 5 min. Internal sequencing primers **trnE^{UUC}** (AGG ACA TCT CTC TTT CAA GGA G) and **trnY^{GUA}** (CCG AGC TGG ATT TGA ACC A) were created because of poly-A/T repeats that were difficult to sequence and the atypically large size of the region in a few taxa. A large inversion in the Asteraceae, excluding the Barnadesieae (Jansen and Palmer, 1987), interrupts the *trnD-trnT* spacer precluding its use. This region also appears to be absent in the *Pinus* chloroplast genome (Wakasugi et al., 1994), which may explain why we were unable to amplify this region for *Taxodium*, *Glyptostrobus*, or *Cryptomeria*.

trnS^{UGA}*-trnfM*^{CAU}—The amplification parameters for this region are 80°C, 5 min; 30X(94°C, 30 s; 55°C, 30 s; 72°C, 2 min); 72°C, 5 min, using Demesure et al. (1995) primers **trnS^{UGA}** (GAG AGA GAG GGA TTC GAA CC) and **trnfM^{CAU}** (CAT AAC CTT GAG GTC ACG GG). This region amplified and sequenced easily for most taxa with minimal troubleshooting.

As explained in the *trnS*^{GCU}*-trnG*^{UUC}*-trnG*^{UUC} region above, *trnG*^{GCC} occurs between *trnS*^{UGA}*-trnfM*^{CAU}. Because there is so little difference between the sequences of these *trnS* and *trnG* genes, the two independent *trnS-trnG* regions will coamplify under certain amplification parameters. However, a seemingly counterintuitive advantage to such sequence similarity is that primer **3´trnG^{UUC}** (and possibly primers **5´trnG2G** and **5´trnG2S**) can be used as an internal sequencing primer for the *trnS*^{UGA}*-trnfM*^{CAU} region.

trnS^{GGA}*-rpS4-trnT*^{UGU}*-trnL*^{UAA}*-trnL*^{UAA}*-trnF*^{GAA}—Because of an initial lack of communication, we PCR amplified several of the taxa using different primer combinations, all of which worked well. However, for all of the lineages of angiosperm taxa, this region was easily amplified in two fragments. The first, *trnS*-5´*trnL*, was amplified using primers **trnS^{GGA}** (TTA CCG AGG GTT CGA ATC CCT C) and **5´trnL^{UAA}R (TabB)** (TCT ACC GAT TTC GCC ATA TC) (Taberlet et al., 1991) with the parameters 96°C, 5 min; 35X(96°C, 1 min; 50-55°C, 2 min; 72°C, 2.5 min); 72°C, 5 min. The second fragment, *trnL5´-trnF*, was amplified using primers **trnL5´^{UAA}F (TabC)** (CGA AAT CGG TAG ACG CTA CG) (Taberlet et al., 1991) and **trnF^{GAA} (TabF)** (ATT TGA ACT GGT GAC ACG AG) (Taberlet et al., 1991) with the parameters 80°C, 5 min; 35X(94°C, 1 min; 50°C, 1 min; 72°C, 2 min); 72°C, 5 min. Several internal sequencing primers were used and included **rpS4R2** (CTG TNA GWC CRT AAT GAA AAC G), **trnT ^{UGU}R** (AGG TTA GAG CAT CGC ATT TG), **trnT ^{UGU}F (TabA)** (CAT TAC AAA TGC GAT GCT CT) (Taberlet et al., 1991), **trnT ^{UGU}2F** (CAA ATG CGA TGC TCT AAC CT) (trnA2 of Cronn et al., 2002), **3´trnL^{UAA}R (TabD)** (GGG GAT AGA GGG ACT TGA AC) (Taberlet et al., 1991), and **3´trnL^{UAA}R (TabE)**)GGT TCA AGT CCC TCT ATC CC) (Taberlet et al., 1991).

5´rpS12-rpL20—This region amplified and sequenced easily for almost all taxa using primers **5´rpS12** (ATT AGA AAN RCA AGA CAG CCA AT) and **rpL20** (CGY YAY CGA GCT ATA TAT CC), both modified from Hamilton (1999a). Amplification parameters were 96°C, 5 min; 35X(96°C, 1 min; 50-55°C, 1 min; 72°C, 1 min); 72°C, 5 min. Although amplification of this region was successful for *Trillium ovatum*,

sequencing reactions using either primer failed repeatedly, even for several different accessions of this species.

psbB-psbH—This region amplified and sequenced easily for all taxa using primers **psbB** (TCC AAA AAN KKG GAG ATC CAA C) and **psbH** (TCA AYR GTY TGT GTA GCC AT), both modified from Hamilton (1999a). Amplification parameters were 80°C, 5 min; 35X(94°C, 30 s; 57-60°C, 30 s; 72°C, 1 min); 72°C, 5 min.

rpL16—This region amplified and sequenced easily for all taxa with minimal troubleshooting using primers **rpL16F71** (GCT ATG CTT AGT GTG TGA CTC GTT G) and **rpL16R1516** (CCC TTC ATT CTT CCT CTA TGT TG) (Small et al., 1998). Amplification parameters were 80°C, 5 min; 35X(95°C, 1 min; 50°C, 1 min with a ramp of 0.3°C/s; 65°C, 5 min); 65°C, 4 min.

***cpDNA compilation and analysis*—**Sequencher 3.0 (Gene Codes Corp., 1998) was used to compile contiguous sequences (contigs) of each accession from electropherograms generated on the automated sequencer. Positions of coding and noncoding (gene, exon, and intron) borders were determined by comparison with either *Arabidopsis* (NC 000932), *Lotus* (NC 001874), or *Nicotiana* (NC 002694) entire cpDNA sequences in GenBank. Terminal coding regions and, in a few rare cases, unreadable ends of the PCR amplicons were excluded from the contigs. Small coding regions within some of the noncoding regions (e.g., $trnE^{UUC}$ and $trnY^{GUA}$ within the $trnD^{GUC}$-$trnT^{GGU}$ spacer) were not excluded from the contigs. Sequences of each of the three-species groups were aligned using Clustal X (Thompson et al., 2001) and manually corrected using McClade v. 4.0 to produce an alignment with the fewest number of changes (indels or nucleotide substitutions). All polymorphic sites found in the three-species groups were rechecked against the original electropherograms. Alignments are available upon request from J. Shaw, E. B. Lickey, or R. L. Small.

The number of nucleotide substitutions, indels, and inversions (hereafter referred to collectively as Potentially Informative Characters or PICs) between the two ingroup species and between either ingroup species and the outgroup species were tallied for each noncoding cpDNA region in each of the lineages. Because indels have been shown to be prevalent and often phylogenetically informative (Golenberg et al., 1993; Morton and Clegg, 1993; Gielly and Taberlet, 1994), they were scored in this study, as were inversions. Indels, any nucleotide substitutions within the indels, and inversions were scored as independent, single characters. We then estimated the proportion of observed mutational events for each noncoding cpDNA region using a modified version of the formula used in O'Donnell (1992) and Gielly and Taberlet (1994). The proportion of mutational events (or % variability) = $[(NS + ID + IV) / L] \times 100$, where NS = the number of nucleotide substitutions, ID = the number of indels, IV = the number of inversions, and L = the total sequence length.

***Assessment of a correlation between variability and length*—**To assess whether or not the length of the different noncoding cpDNA regions accounts for the number of PICs observed within a particular region, we used a simple regression analysis. Because

of the variation in phylogenetic distance between species in the different lineages we could not combine all lineages in a single regression. Instead, we performed 10 separate regressions (one per lineage) and calculated r^2 for each to determine how much of the variation seen in the PIC values is explained by the length of the region.

***Cost / benefit analysis of coamplifiable noncoding cpDNA regions*—**In the above analyses, each noncoding region was treated individually. However, several adjacent, shorter, noncoding cpDNA regions may be coamplified as a single contiguous unit. We surveyed several cpDNA region combinations to assess the potential phylogenetic utility of coamplifiable regions from a cost/benefit perspective. For example, the *trnL* intron and *trnL-trnF* spacer are often coamplified, and most of the time these two regions are sequenced with the same two primers that were used in PCR (TabC and TabF). From a cost/benefit perspective, it is beneficial to amplify and sequence both of these regions together instead of separately by maximizing the number of characters obtained per two sequencing reactions. Our sequencing reactions always yielded easily readable sequence data of 800 bp from a single-primer sequencing reaction. We therefore limited what we categorize as "coamplifiable" regions to those whose total length average is < approximately 1500 bp and can be sequenced entirely with two sequencing reactions. These coamplifiable regions include *psbA-3´trnK-matK*, *trnS-trnG-trnG*, *trnC-ycf6-psbM*, *ycf6-psbM-trnD*, *rps4-trnT-trnL*, and *trnL-trnL-trnF*.

***Assessment of the predictive value of a three-species sample study*—**Our inferences from these data rely on the assumption that a sample of three species is predictive of the overall levels of variation that will be found in an entire data set. To test the predictive power of a three-species survey we compared the number of PICs among the three species with the respective complete data sets of 18 taxa of *Prunus* section *Prunocerasus* (Shaw and Small, 2004) and nine taxa of *Hibiscus* section *Furcaria* (R. L. Small et al., unpublished data), each with a single outgroup. The comparison of the *Prunus* data sets was made with introns *trnL*, *trnG*, *rpS16*, and *rpL16* and intergenic spacers *trnL-trnF*, *trnH-psbA*, and *trnS-trnG*, and the comparison of the *Hibiscus* data sets was made with introns *rpS16*, *rpL16*, and *trnG* and intergenic spacers *trnD-trnT*, *rpoB-trnC*, *trnH-psbA*, and *trnS-trnG*. Regression lines were calculated and their slopes were compared on a scatterplot for each data set comparison.

RESULTS

Assessment of the noncoding cpDNA regions surveyed across phanerogam lineages—A few regions were excluded from analysis because they were missing in some taxa or their lineages due to inversions such as those observed in the *trnS-trnG* spacer in *Trillium-Pseudotrillium*, *Eupatorium*, and *Carphephorus-Trilisa*, the *rpoB-trnC* spacer in *Eupatorium* and *Carphephorus-Trilisa*, and the *trnD-trnT* spacer in *Taxodium*, Eupatorium, and Carphephorus-Trilisa. Some regions were excluded because of their inconsistency or inability to amplify such as the *rpoB-trnC* region in *Taxodium* and the *rpS16* intron in *Cryptomeria*. Others had to be excluded because, for unknown reasons, they could not be sequenced, such as the *psbM-trnD* region in *Carphephorus-Trilisa* and the *5´rpS12-rpL20* region in *Trillium ovatum*. After the exclusion of the problematic

regions, taxa, and most coding regions, 133504 bp from 21 noncoding cpDNA regions from 10 phanerogamic lineages were sequenced. Of that, we observed 2968 nucleotide substitutions, 1260 indels, and six inversions for a total of 4234 PICs. Nucleotide substitutions account for 70.1% of the variable characters, while indels and inversions account for 29.8% and 0.14%, respectively. No obvious differences were observed in the amount of variability or number of PICs between intergenic spacers and introns.

We did not apply statistical analyses to these data because of potentially different rates of evolution among the different lineages, the incongruent phylogenetic distances between the species in each lineage, and the exclusion of some regions because of structural rearrangement of the cpDNA molecule or PCR amplification or sequencing difficulties. Thus, the following discussion is based on our qualitative interpretation of the results, which are compiled in Table 2 (not shown here but available in the original published version of this paper) and simplified in Fig. 4.

Figure 4 illustrates the disparity in the PICs offered by different noncoding cpDNA regions and that this trend exists across all phylogenetic lineages. Of the 21 separate noncoding cpDNA regions surveyed in this investigation, several consistently provided more PICs per our cost/benefit criterion of two sequencing reactions than other regions across all lineages. Specifically, five intergenic spacers (*trnD-trnT*, *trnS-trnG*, *rpoB-trnC*, *ycf6-psbM*, and *trnS-trnfM*) provided more PICs than the other surveyed regions. The PIC value for each region was averaged across the six lineages for which there are complete data sets to illustrate more clearly the general trend (Fig. 5).

Assessment of a correlation between PICs and length—A scatterplot showing the relationship of the PIC values with respect to the length of the region is shown in Fig. 6. For each lineage, a regression line was drawn (not shown) and coefficients of determination were calculated. Coefficients of determination ranged from 22% in *Eupatorium* to 83% in *Taxodium*, which means that length of the region explains 83% of the variation in PIC value observed in *Taxodium*, while it only explains 22% of the variation in PIC value observed in *Eupatorium*. However, there is an apparent outlier within the *Taxodium* data set (circled in the upper right corner of Fig. 6), the removal of which drops the r^2 value from 0.83 to 0.54. Although Fig. 6 reveals the intuitively obvious conclusion that the length of a region accounts for a proportion of the PICs, this figure shows that length does not explain all, or even a majority in many cases, of the variability within a particular region. Many examples can be found where very different PIC values are found within regions that are nearly the same length.

Cost / benefit analysis of coamplifiable noncoding cpDNA regions—Several regions may be coamplified, sequenced, and successfully contiged with the same two PCR primers, and from a cost perspective, they are equal to amplifying and sequencing a portion of each alone. These combined regions include *psbA-3´trnK-matK*, *trnS-trnG-trnG*, *trnC-ycf6-psbM*, *ycf6-psbM-trnD*, *rps4-trnT-trnL*, and *trnL-trnL-trnF*. The results of the comparison between the coamplifiable regions are shown in Fig. 7. This figure shows that the *trnS-trnG* spacer combined with the *trnG* intron (*trnS-trnG-trnG*) potentially provides the greatest number of PICs compared to all other regions that could be amplified and sequenced with two primers.

***Assessment of the predictive value of a three-species sample study*—**Analysis of the predictive value of a three-species survey of a particular cpDNA region indicates that as the number of PICs in a three-species survey increases, so too will the actual number of variable characters in a complete data set generated by that region (Fig. 8). A regression analysis of each of the *Prunus* and *Hibiscus* data sets reveals remarkably similar slopes. It is apparent from the linear relationship that a preliminary survey of three species is highly predictive of the amount of information that a noncoding cpDNA region might offer to a phylogenetic investigation.

DISCUSSION

Since 1995, the number of molecular systematic investigations that employ noncoding cpDNA sequence data has increased every year (Fig. 1). However, most of these studies (about 77% from 1995 to 2002) have used some portion of the *trnL-trnL-trnF* or *trnK-matK-trnK* regions and very few investigators have sampled from the myriad other noncoding regions of the cpDNA molecule. Because of this, little is known about the relative rates of evolution among the different noncoding cpDNA regions, and most investigators continue to rely on these two very popular regions.

The initial goal of this investigation was to provide a comparison of noncoding cpDNA regions to see if there are any that reliably yield a greater number of variable characters (PICs) at low taxonomic levels, and thus would be of greater value to systematic studies than the often used *trnL-trnL-trnF* or *trnK-matK-trnK* regions. To do so we used three-species surveys representing most of the major phylogenetic lineages of phanerogams (sensu APG II, 2003). To test the predictive power of a three-species survey we compared the surveys of seven regions in *Prunus* and eight regions in *Hibiscus* with their respective complete data sets (Fig. 8). Figure 8 shows that as the number of PICs in survey of three species increases, so will the actual number of variable characters in a complete data set generated from those regions. Therefore, a survey of three species is highly predictive of the amount of information that a noncoding cpDNA region might offer to a phylogenetic investigation and is an effective means of comparison between different noncoding cpDNA regions.

Most investigators, when comparing different DNA regions, have used either of two metrics that are not wholly separate. One tallies the number of variable characters including nucleotide substitutions, indels, and inversions (PICs), while the other calculates the percent variability, or percent divergence of a region, by dividing the total number of variable characters by the total length of the region. It is necessary to emphasize that, from the viewpoint of systematists, the total number of variable characters offered by a region is more important than the percent variability. A highly variable but extremely short region may not provide a sufficient number of variable characters with which to generate a resolved phylogeny. As systematists, we are interested in obtaining the greatest number of variable characters per sequencing reaction, arguably the costliest portion of sequence acquisition, where current techniques and equipment allow for 600-800 bp of easily readable nucleotides per reaction. Therefore, it would be ideal to use cpDNA regions that combine high variability in fragments of approximately 700-1500 bp that can easily be sequenced with one or two primers, ideally

the original amplification primers. To show that the number of PICs offered to systematic studies is not due solely to total length of a region, we regressed PICs on length of the region for all of the regions surveyed in this study (Fig. 6). It is apparent that while the length of the region accounts for some proportion of the PIC value, there is a large amount of unaccountable variation in this trend. Within *Prunus*, for example, regions that are between 261 and 307 bp contain between 2 and 14 PICs, while regions that are from 709 to 783 bp contain between 2 and 34 PICs, with the largest region not accounting for the greatest PIC value.

Our results clearly show that a disparity exists in the information offered to phylogenetic investigations by different noncoding cpDNA regions. Additionally, we show that the most widely used noncoding cpDNA regions in infrageneric systematic investigations, namely the *trnL-trnL-trnF* and *trnK-matK* intron regions, consistently provide fewer PICs than several other choices, such as *trnS-trnG-trnG*, *trnC-ycf6-psbM*, *trnD-trnT*, *trnT-trnL*, and *rpoB-trnC*.

***Discussion of each of the regions*—**Below is a summary of each of the 21 different noncoding cpDNA regions that we have surveyed in this study including a brief history of their utility in previous studies and an assessment of their utility based on the results of this study. Because there is no intuitively straightforward way to rank each of the regions, we have divided the regions into three tiers based on their overall qualitative usefulness (Fig. 5). Tier 1 contains five regions that on average consistently provide the greatest number of PICs across all phylogenetic lineages. Tier 2 includes the next five regions that may provide some useful information, but they may be less than optimal in providing the number of characters needed for a well-resolved phylogenetic study. Tier 3 comprises those regions that consistently provide the fewest PICs across all lineages and are therefore not recommended for low-level studies because better noncoding cpDNA choices exist. Ranking these regions in three tiers offers information relevant to studies focused on very low taxonomic levels where researchers might opt to choose one or more regions that likely contain the highest number of PICs. In addition, this ranking scheme is also useful in providing information to researchers who may wish to couple quickly evolving regions with more slowly evolving Tier 2 or Tier 3 regions, which might allow for resolution within the clade of interest in addition to confidence alignment with an outgroup (Asmussen and Chase, 2001).

trnH^{GUG}*-psbA (Tier 3)*—Inquiry into the *trnH-psbA* intergenic spacer began with Aldrich et al. (1988) who showed that indels were prevalent in this region, even between closely related species. An early study that showed this region to be of value to systematics is Sang et al. (1997) who noted that it was highly variable compared to *matK* and *trnL-trnF*. The utility of *trnH-psbA* was also shown by Hamilton (1999b) who used it for an intraspecific study within *Corythophora* (Lecythidaceae). Subsequent to these two studies, several investigators have used this region to study closely related genera and species (Azuma et al., 1999; Chandler et al., 2001; Mast and Givnish, 2002; Fukuda et al., 2003; Miller et al., 2003; Tate and Simpson, 2003). It has also been used in an intraspecific investigation (Holdregger and Abbott, 2003). At higher levels, *trnH-psbA* has proven to be largely unalignable (Laurales: Renner, 1999; Saxifragaceae: Soltis et al.,

2001; Lecythidaceae: Hamilton et al., 2003). In a study of the relative rates of nucleotide and indel evolution, Hamilton et al. (2003) showed *trnH-psbA* to be more divergent, based on percent variability, than *trnS-trnG*, *psbB-psbH*, *atpB-rbcL*, *trnL-trnF*, and *5′rpS12-rpL20*. Although studies have shown that *trnH-psbA* contains a very high percentage of variable characters (Azuma et al., 2001; Hamilton et al., 2003), this spacer is usually coupled with other regions because it is comparatively short and may not yield enough characters with which to build a well-resolved phylogeny.

The average length of *trnH-psbA* is 465 bp, and it ranges from 198 to 1077 bp. Based on our data, and data of the previous workers listed above, the 1077-bp length found in *Trillium-Pseudotrillium* is atypical. Although this spacer is the second-most variable on a percent basis, we include it in Tier 3 because its relatively short length provides few overall characters. However, it amplified and sequenced easily across all lineages and can be sequenced with only one primer in most taxa. It is also worth noting that the ends of this spacer, roughly 75 bp from either gene, are relatively conserved compared to the middle portion of this spacer, which is highly indel prone (Aldrich et al., 1988), and contains several poly-A/T runs. Most of the numerous observed indels were relatively short, but a 132-bp indel was observed among the *Hibiscus* accessions. Among more distantly related taxa, this indel-prone middle region may generate a relatively high amount of homoplasy due to apparent indel "hot spots" with numerous, repeating, and overlapping indels.

psbA-3′*trnK*^{UUU}-[*matK*]-5′*trnK*^{UUU} *(Tier 3 + Tier 3 + Tier 3)*—The *matK* gene region (*trnK-matK-trnK*) or some portion of it was first employed in intrafamilial phylogenetic studies by Steele and Vilgalys (1994) and Johnson and Soltis (1994). Since then, this region has been a primary tool in phylogenetic investigations below the family level, but it has also been suggested as an effective tool above the familial level (Hilu and Liang, 1997; Hilu et al., 2003). The frequency of infrageneric phylogenetic use of this region is second only to *trnL-trnL-trnF*, representing 22 vs. 55%, respectively, of studies in 2002 (Fig. 1). Several studies have used the entire *trnK-matK-trnK* region (e.g., Johnson and Soltis, 1994; Sang et al., 1997; Hardig et al., 2000; Miller and Bayer, 2001), while most have carved out various portions depending on variable primer success and availability. Additionally, some investigators have used the intergenic spacer between *psbA* and 3′*trnK* (Winkworth et al., 2002; Pedersen and Hadenäs, 2003). In some studies the 3′*trnK* intron to some 3′ portion of *matK* was used (Wang et al., 1999; Schultheis, 2001; Winkworth et al., 2002; Hufford et al., 2003; Salazar et al., 2003). Others have used some 5′ portion of *matK* to 5′*trnK* (Plunkett et al., 1996; Ohsako and Ohnishi, 2000, 2001; Chandler et al., 2001), and still others have used part of the *matK* gene only (Kajita et al., 1998; Bayer et al., 2002; Cuénoud et al., 2002; Ge et al., 2002; Samuel et al., 2003). In many of the abovementioned investigations, several sequencing primers were required in addition to the PCR primers to piece together sequences for the entire desired region. Also, truly universal primers cannot be designed due to the variability of the gene across broad phylogenetic lineages, and often primers have to be made that are specific to different groups (e.g., Wang et al., 1999; Hardig et al., 2000; Hu et al., 2000; Miller and Bayer, 2001; Mort et al., 2001; Pridgeon et al., 2001; Bayer et al., 2002; Hilu et al., 2003). Therefore, in terms of cost, the *matK* region is relatively expensive because it

often involves several sequencing reactions from multiple unique primers. Although *matK* is putatively the most variable coding region found within cpDNA (Neuhaus and Link, 1987; Olmstead and Palmer, 1994), it was excluded from this study primarily because it is a coding region and not part of our focus. Furthermore, the gene's large size would require the development of several internal sequencing primers, and with few strategically placed conserved regions, the number of primers for specific lineages becomes too cumbersome for the scope of this investigation. Therefore, we only included both ends of the *trnK* intron in addition to the *psbA*-3´*trnK* intergenic spacer.

Although the above discussion may read as to denigrate the *psbA*-3´*trnK*-[*matK*]-5´*trnK* region, this was not our intent. Because of the plethora of data already available, this region is valuable with respect to its potential in comparative studies (e.g., the placement of taxa whose phylogenetic positions are ambiguous).

The *psbA*-3´*trnK* intergenic spacer is usually shorter than either portion of the *trnK* intron, averaging 268 bp in length and ranging from 212 to 430 bp. The 5´ end of the *trnK* intron is consistently larger, with an average of 747 bp and a range of 704-860 bp, than the 3´ end of the *trnK* intron, which averages 314 bp and ranges from 257 to 533 bp. While the 3´*trnK* portion of the intron is more variable on a percent basis than the 5´ portion, the 5´*trnK* portion consistently provides more PICs, as was reported in *Acacia* by Miller and Bayer (2001). Compared to other noncoding cpDNA regions surveyed in this study, both the *trnK* intron and *psbA*-3´*trnK* spacer provide relatively few variable characters and are ranked in Tier 3. Even combining the two halves of the intron yields an average PIC value below that of several other regions (Fig. 7). It is our opinion that the entire *psbA-trnK-matK-trnK* region is less suitable for infrageneric phylogenetic investigation than several other choices because it is very large, less informative than other regions, lacks sufficiently conserved coding regions where "universal" primers can be anchored, and was inexplicably problematic during sequencing.

rpS16 (Tier 2)—The ribosomal protein 16 small subunit gene (*rpS16*) contains a group II intron that was first used in a phylogenetic context by Oxelman et al. (1997). Since this initial investigation the *rpS16* intron has been used to successfully resolve relationships among genera in Rubiaceae subfamily Rubioideae (Andersson and Rova, 1999), Arecaceae subfamily Calamoideae (Baker et al., 2000), Arecaceae (Asmussen and Chase, 2001), Fabaceae tribe Glycininae (Lee and Hymowitz, 2001), Marantaceae (Andersson and Chase, 2001), Apiaceae subfamily Apioideae (Downie and Katz-Downie, 1999), and Colchicaceae (Vinnersten and Reeves, 2003). However, within each of these studies, infrageneric resolution was weak. Other studies have also shown that *rpS16* is usually not variable enough to resolve infrageneric relationships (Baker et al., 2000; Edwards and Gadek, 2001; Wanntorp et al., 2001; Popp and Oxelman, 2001; Aagesen and Sanso, 2003; Ingram and Doyle, 2003).

Originally, the *rps16* intron was suggested to be a valuable tool for investigation at the family level and below (Oxelman et al., 1997), but the accumulated literature in addition to our data suggests it will often not provide enough characters to resolve relationships below generic levels. The intron averages 846 bp in length and ranges from 784 to 946 bp. The *rpS16* intron is typically more informative than the *trnL-trnL-trnF* region (Fig. 7), but it frequently contains fewer PICs than other choices and is therefore

included in Tier 2. A poly-A/T run in most lineages (especially *Prunus* and *Hibiscus*) at the 3′ end of the intron may be problematic in sequencing from that direction. This region cannot be used in some taxa because all or some part of the *rpS16* gene is absent from some members of Linaceae, Malpighiaceae, Passifloraceae, Salicaceae, Polygalaceae, Turneraceae, Violaceae (see Downie and Palmer, 1992), Connaraceae, Eucommiaceae, Fagaceae, Krameriaceae, Fabaceae (see Doyle et al., 1995), *Marchantia polymorpha* (Ohyama et al., 1986), *Pinus thunbergii* (Tsudzuki et al., 1992), *Pisum sativum* (Nagano et al., 1991) and *Epifagus virginiana* (Wolfe et al., 1992).

trnS^{GCU}*-trnG*^{UUC}*-trnG*^{UUC} *(Tier 1 + Tier 2)*—Hamilton (1999a, b) designed primers for the intergenic spacer between *trnS* and *trnG* (*trnS-trnG*) to study population dynamics within a tropical tree species in *Corythophora* (Lecythidaceae) and subsequently published them along with the suggested amplification protocol. Xu et al. (2000) designed nearly the same primers for this spacer for use in *Glycine* (Fabaceae). Subsequent studies have shown this region to be highly variable. Olson (2002a) showed that the *trnS-trnG* spacer sequences are largely unalignable between genera in the Caricaceae-Moringaceae clade, and Xu et al. (2000) showed the *trnS-trnG* spacer to be among the most informative of nine noncoding cpDNA regions within two closely related subgenera of *Glycine*. In another study within *Glycine*, Sakai et al. (2003) showed the *trnS-trnG* spacer to contain many more PICs than *atpB-rbcL*, *rpS11-rpL36*, and *rpS*3-*rpL16*. The *trnS-trnG* spacer was reported to show intraspecific variation in *Moringa* (Moringaceae) by Olson (2002b) and *Corythophora* (Lecythidaceae) by Hamilton (1999a, b). Perret et al. (2003) showed the *trnS-trnG* spacer to provide more PICs than *rpL16*, *trnL* intron, *trnL-trnF* spacer, *trnT-trnL* spacer, and *atpB-rbcL* spacer in the tribe Sinningieae (Gesneriaceae) and Hamilton et al. (2003) showed it to be more informative than *psbB-psbH*, *atpB-rbcL*, *trnL-trnF*, and *5′rpS12-rpL20* in *Corythophora* (Lecythidaceae). However, Shönenberger and Conti (2003) showed the *trnS-trnG* spacer to contain fewer PICs than the *rpS16* and *rpL16* introns, but having more PICs than the *trnH-psbA* spacer, *atpB-rbcL* spacer, and part of the *matK* exon. Gaskin and Schaal (2003) showed that *trnS-trnG* is five times more variable than the *trnL-trnF* spacer and contained more variable characters than nuclear ribosomal ITS in *Tamarix*. Lastly, Pacak and Szweykowska-Kulinska (2000) designed primers to study the group II *trnG* intron which was not included in the abovementioned studies. They found that this intron provided several nucleotide substitutions in a group where the *trnL* intron was invariant. Pedersen and Hedenäs (2003) also used the *trnG* intron and showed that, although it did not contain as many variable characters as *rpL16,* it provided nearly twice as many as *trnL-trnF*.

Because Hamilton's (1999a) "G" primer was designed in the 5′ *trnG* exon and the *trnG* intron was shown to be relatively variable, we designed a 3′ exon *trnG* primer that would allow the *trnS-trnG* intergenic spacer and the *trnG* intron to be coamplified. Additionally, we created internal sequencing primers that are located in the internal 5′ *trnG* exon near the position of Hamilton's (1999a) *trnG* primer.

The *trnG* intron averages 763 bp in length and ranges from 697 to 1008 bp, while the *trnS-trnG* spacer averages 763 bp and ranges from 619 to 1035 bp. Alone the *trnS-trnG* spacer ranks in Tier 1, while the *trnG* intron ranks in Tier 2. Additionally, the *trnS-*

trnG spacer not only provides the greatest number of PICs, but also has the highest percent variability with an average of 4.74%. However, when combined as a coamplifiable unit, *trnS-trnG-trnG* averages approximately 1500 bp and provides the greatest number of PICs per two (very rarely three) sequencing reactions compared to any other regions, single or combined, surveyed in this study (Figs. 5 and 7). All lineage representatives included in this study have a poly-A/T run in the *trnG* intron near the 3´*trnG* end, which is usually not long enough to affect PCR or sequencing. Sometimes, as is the case in *Taxodium*, this poly-A/T run is over 30 bp and prohibits sequencing from that direction. Sequencing with internal primers usually alleviates this problem. Approximately the same number of indels were found in both the *trnS-trnG* spacer and the *trnG* intron, and large indels were noted in *Hibiscus* (73 bp) and in the complete data set of *Prunus* (358 bp) (Shaw and Small, 2004).

Because of independent structural rearrangements in both monocots (Hiratsuka et al., 1989) and Asteraceae, excluding Barnadesieae (Jansen and Palmer, 1987), the *trnS-trnG* spacer does not exist in these taxa. However, the *trnG* intron can be used in both of these groups, but there may be better choices for such studies (Fig. 5).

rpoB-trnC^{GCA} *(Tier 1)*—The *rpoB-trnC*^{GCA} region was first used by Ohsako and Ohnishi (2000; 2001) in their study of intra- and interspecific relationships in *Fagopyrum* (Polygonaceae). They showed this spacer to contain enough variable characters to distinguish between closely related species in addition to showing some intraspecific variation. Studying only intraspecific relationships in *Fagopyrum cymosum*, Yamane et al. (2003) found 15 informative characters. The *rpoB-trnC* spacer has also been used in an intergeneric study of subtribe Clematidinae (Ranunculaceae) where O. Miikeda et al. (Tokyo Metropolitan Mizuho-nogei High School, personal communication) found 66 potentially informative characters.

The average length of this region is 1174 bp with a range of 914-1309 bp, comparable with that found in the studies listed above. The center region of this intergenic spacer contains several relatively small (< 8 bp) poly-A/T strings, and the *rpoB-trnC* spacer contains several relatively large indels (47 bp in *Prunus*, 58 bp in *Hibiscus*, 81 bp in *Gratiola*). This spacer, ranked in Tier 1, is among the most informative regions with respect to the number of PICs offered for infrageneric investigations (Figs. 5 and 7).

trnC^{GCA}*-ycf6-psbM-trnD*^{GUC} *(Tier 3 + Tier 2 + Tier 2)*—The *trnC*^{GCA}*-trnD*^{GUC} region is approximately 3200 bp long in *Nicotiana* (Wakasugi et al., 1998) and includes the genes *ycf6* and *psbM* which are 90 and 105 bp long, respectively. This region was first identified as a potential region for phylogenetic study by Demesure et al. (1995) who reported a length of 3000 bp in *Quercus* (Fagaceae). Demesure et al. (1996) subsequently used the region in a PCR-RFLP phylogeographic study in *Fagus* (Fagaceae). In their PCR-RFLP study of the interspecific relationships in *Allium* (Alliaceae), Mes et al. (1997) also used the *trnC-trnD* region. Hartmann et al. (2002) compared sequences of a 1149-bp portion of this region among eleven cactus species across five genera to determine the origin of *Lophocereus* (Cactaceae), but only found nine informative characters. Sequences of this region have been used to elucidate

infrageneric relationships in *Humulus* (Cannabaceae) (A. Murakami, Kirin Brewery Company, Ltd., unpublished data) and *Panax* (Araliaceae) (Lee and Wen, 2004). In both studies, internal primers were designed to completely sequence this region which is about 2600 bp in *Humulus* and up to 3000 bp in *Panax*. Within the aligned data set for *Panax*, Lee and Wen (2004) report 71 informative characters plus 20 informative indels.

Because this entire region is large enough to be cumbersome and each of the three intergenic spacers greatly varies in length, we have analyzed the three intergenic spacers separately. The average length of the entire region is 2480 bp with a range of 1726-3460 bp. The *trnC-ycf6* intergenic spacer averages 690 bp with a range of 246-1071 bp, the *ycf6-psbM* spacer averages 825 bp with a range of 406-1283 bp, and the *psbM-trnD* spacer averages 965 bp and ranges from 506 to 1801 bp. All three of these regions appear to be prone to large indels. We observed indels of 232 bp in *Trillium-Pseudotrillium*, and 64 and 89 bp in *Gratiola* in the *trnC-ycf6* spacer, indels of 107 bp in *Carphephorus-Trilisa / Eupatorium*, 86 bp in *Trillium-Pseudotrillium*, and 371 bp in *Gratiola* in the *ycf6-psbM* spacer, and indels of 40 bp in *Minuartia*, 142 bp in *Eupatorium*, 360 and 45 bp in *Taxodium-Glyptostrobus-Cryptomeria* in the *psbM-trnD* spacer. Of the three regions, *ycf6-psbM* and *psbM-trnD* rank in Tier 2, while *trnC-ycf6* ranks at the top of Tier 3. Of these three regions, the *ycf6-psbM* intergenic spacer provides the greatest number of PICs. The value of *ycf6-psbM* may actually be an underestimate because many potential characters may have been hidden in the large portions of missing data caused by large indels among the three-species sequence data sets. When combined with either *trnC-ycf6* or *psbM-trnD*, the *ycf6-psbM* spacer is the second-most variable coamplifiable region behind *trnS-trnG-trnG*. When the entire *trnC-ycf6-psbM-trnD* region is compared to *trnS-trnG-trnG*, it provides a greater number of PICs on average, 78 vs. 64 respectively (Fig. 7), but this much larger fragment requires at least two additional sequencing reactions. Therefore, there are likely better combinations of regions to obtain a greater number of characters.

trnD^{GUC}*-trnT*^{GGU} *(Tier 1)*—Using the aligned sequences of *Oryza*, *Nicotiana*, and *Marchantia*, Demesure et al. (1995) developed a primer pair anchored within the *trnD* and *trnT* genes to amplify the noncoding intergenic spacer and the embedded *trnY*^{GUA} and *trnE*^{UUC} genes, which are 84 and 73 bp, respectively. Friesen et al. (2000) used sequence data from this region to investigate the phylogenetic relationships among some *Allium* species and the monotypic *Milula* (Alliaceae), in which the resulting parsimony trees were well resolved and comparable to those generated with nuclear ITS sequences. The *trnD-trnT* region provided sufficient characters to separate populations in an intraspecific study of *Cunninghamia konishii* (Cupressaceae) and the very closely related *Cunninghamia lanceolata* (Lu et al., 2001). In a phylogenetic study of the arecoid lineage of palms (Arecaceae), Hahn (2002) showed that this region provided more variable characters than either the *trnQ-rpS16* intergenic spacer or the *atpB* and *rbcL* genes combined. To trace wild parentage and potential hybridization among cultivated rootstocks of *Juglans* (Juglandaceae), Potter et al. (2002) sequenced the *trnD-trnT* region, along with *trnT-trnL* and *trnL-trnF*, and ITS. Although they could not obtain the middle portion of the *trnD-trnT* spacer because of a large poly-A/T, this region still provided more variable characters than *trnT-trnL* and *trnL-trnF* and provided the same number of

variable characters as those two regions put together, while providing only one fewer character than ITS. That the *trnD-trnT* spacer might provide an equivalent number of variable characters as ITS was also suggested by Feliner et al. (2002). In a phylogenetic study of *Brassica* and *Raphanus* (Brassicaceae), Yang et al. (2002) showed that the ~1150 bp *trnD-trnT* region evolves 1.1 times faster than the ~1641-bp *trnT-trnL-trnL-trnF* and that both regions provided nearly the same number of nucleotide substitutions, 345 and 346, respectively.

The *trnD-trnT* intergenic spacer averages 1066 bp, ranging from 578 to 1403 bp, and provides the greatest number of PICs compared to all of the uncombined regions we surveyed (Fig. 5). This region amplified and sequenced easily for all of the lineages except for the two representatives of the Asteraceae clade where this spacer is interrupted by the same inversion involving the *trnS-trnG* spacer (Jansen and Palmer, 1987). As an aside, we attempted to amplify the resulting *trnD*^{GUC}*-trnS*^{GCU} and *trnT*^{GGU}*-trnG*^{UUC} using the primers above for our Asteraceae representatives, but these attempts were not entirely successful and provided no useful information (data not shown). Within the *trnD-trnT* region, several relatively large indels were noted (246 bp in *Minuartia*, 42 bp in *Trillium*, 87 bp in *Hibiscus*) as well as poly-A/T runs and poly-AT repeats. Because of these runs and repeats, we designed (but never used beyond making sure that they work), universal internal sequencing primers embedded in the *trnE* and *trnY* genes.

trnS^{UGA}*-trnfM*^{CAU} *(Tier 1)*—A universal primer pair for the amplification of the *trnS-trnfM* intergenic spacer was developed by Demesure et al. (1995), using aligned sequences from *Oryza*, *Nicotiana*, and *Marchantia*. Subsequent studies used the PCR-RFLP method to investigate geographically structured intraspecific variation (e.g., El Mousadik et al., 1996; Stehlik, 2002; Stehlik et al., 2002). Zuber and Widmer (2000) used sequences of the *trnS-trnfM* region to assess genetic variation within and among host-specific subspecies of *Viscum album* (Viscaceae). They showed that the *trnS-trnfM* spacer provided more nucleotide substitutions than *trnL-trnL-trnF*, *trnH-trnK*, or ITS, 8 vs. 2, 7, 5, respectively. Chassot et al. (2001) showed that *trnS-ycf9*, approximately one-third of the *trnS-trnfM* region (Fig. 3), provided 76 informative characters, whereas *trnL-trnF* provided 83 and the *trnL* intron provided 59. Although Hartmann et al. (2002) investigated the phylogenetic origins of *Lophocereus* (Cactaceae) using *trnS-trnfM* (and others), they only sequenced approximately 375 bp of the ~ 1.5 kb spacer; therefore a comparison of the relative utility within this study could not be made.

The average length of the *trnS-trnfM* region is 1119 bp with a range of 856-1804 bp. Embedded in this region are coding regions *trnG*^{GCC}, *ycf9*, and *psbZ*. Although it contains three genes, our data, in addition to the previously mentioned studies, indicate this region exhibits a relatively high PIC value. Within *Minuartia* and *Prunus*, two large indels of 43 and 141 bp, respectively, were observed. Several of the taxa surveyed here showed poly-A/T runs, most of which were not long enough to affect sequencing. However, an approximately 250-bp portion of the *Hibiscus* data set was excluded from our analyses because it consisted of nearly all A's and T's and could not be confidently aligned.

trnS^{GGA}*-rpS4 (Tier 3)*—Cranfill (2001) suggested that the *trnS-rpS4* spacer might be of phylogenetic utility and described it as useful for investigations below the family level because it evolves at a rate similar to ITS. Subsequently, Smith and Cranfill (2002) used this region in intrafamilial reconstruction of the thelypteroid ferns, although it was combined with the *rpS4* gene and the *trnL-trnF* spacer and comparative analyses among these regions were not discussed. Hennequin et al. (2002) coupled this spacer with *rbcL* and the *rpS4* genes in an investigation of *Hymenophyllum* (Hymenophyllaceae) and showed that it provided more variable characters than either. However, the utility of this spacer with respect to Hennequin et al. (2002) should be taken with caution because it was compared only to coding regions.

The *trnS-rpS4* intergenic spacer averages 273 bp and ranges from 209 to 314 bp. This Tier 3 intergenic spacer yielded the lowest PIC value of any region surveyed in this study and is therefore not recommended as a systematic tool for infrageneric studies.

rpS4-trnT^{UGU} *(Tier 3)*—Saltonstall (2001) amplified the *rps4-trnT* intergenic spacer along with several other intergenic spacers to test the "universality" of the primers in one member of each of the six major subfamilies of Poaceae. She then used sequence data from *rpS4-trnT*, along with several other regions, to assess the amount of intraspecific polymorphism within each of these regions in *Phragmites australis*. Because only numbers of haplotypes were reported, it is difficult to make inferences as to the comparative utility of regions within her study. However, she did report that this spacer provided more haplotypes than *trnH-psbA*, *trnT-trnE*, *rpoB-trnC*, *trnL*, *trnL-trnF* and some lesser known intergenic spacers. Before the start of this study, we (J. Shaw and R. L. Small) failed in several attempts to obtain the *trnT-trnL* intergenic spacer in *Prunus* and *Hibiscus* because of the problematic nature of the TabA primer (see *trnT*^{UGU}-*trnL*^{UAA}-*trnL*^{UAA}-*trnF*^{GAA} discussion below). Therefore, we designed a primer embedded in the *rpS4* gene that allowed us to obtain the *trnT-trnL* spacer in addition to the *rpS4-trnT* spacer.

Although Saltonstall (2001) reported that the *rpS4-trnT* intergenic spacer is 750-950 bp in grasses, we found it averages 402 bp with a range of 345 to 785 bp. This region shows a relatively low PIC value (see Fig. 5), and is ranked in Tier 3.

trnT^{UGU}*-trnL*^{UAA}*-trnL*^{UAA}*-trnF*^{GAA} *(Tier 1 + Tier 3 + Tier 3)*—One of the first sets of universal PCR primers for noncoding cpDNA was published by Taberlet et al. (1991). These primer sets span a region comprising three tRNA genes—*trnT*^{UGU}, *trnL*^{UAA}, and *trnF*^{GAA}. The noncoding portions of the region include a Group I intron that interrupts the *trnL* gene, as well as the intergenic spacers between *trnT-trnL* and *trnL-trnF*. Taberlet et al. (1991) described primer sequences situated in conserved regions of the tRNA genes for amplifying each of these regions and demonstrated amplification in land plants ranging from bryophytes to pteridophytes, gymnosperms, and angiosperms. Because of the near-universal nature of the primers and their early publication, these regions have become the most widely used noncoding cpDNA sequences in plant systematics. As of December 2003, Web of Science lists 579 citations of the Taberlet et al. (1991) paper. Usually these regions are employed in studies of closely related species or genera, but a recent study by Borsch et al. (2003) used the entire region to evaluate

relationships among basal angiosperms. Renner (1999) used the *trnT-trnL* and *trnL-trnF* spacers, along with other coding and noncoding regions, in an analysis of the order Laurales. Bremer et al. (2002) also employed the entire region, along with other coding and noncoding regions, in a phylogenetic analysis of Asterids, as did Stech et al. (2003) who analyzed *trnL* intron and *trnL-trnF* spacer sequences in a broad survey of land plants and algae.

The *trnT-trnL* intergenic spacer has been the least used of the Taberlet et al. (1991) regions due to difficulties with PCR amplification in many plant groups (personal observation and personal communication from colleagues). This difficulty apparently stems from the *trnT* primer, Taberlet et al. (1991) primer "A", but a new PCR-amplification primer designed by Cronn et al. (2002) works in all of the taxa surveyed in this study (data not shown). Studies that have used the *trnT-trnL* spacer often report that it provides greater variation than other surveyed regions, including the *trnL* intron and the *trnL-trnF* spacer (e.g., Böhle et al., 1994; Small et al., 1998; Cronn et al., 2002; Downie et al., 2002). This spacer exhibits a wide range of sizes in different plant groups from ~ 400–1500 bp and often includes large A/T rich regions that may be difficult to align among divergent sequences.

The *trnL* intron is the only chloroplast Group I intron (Palmer, 1991). It has a specific secondary structure and several highly conserved regions that are found among all Group I introns (Westhof and Michel, 1996; Stech et al., 2003). This intron ranges in size from as small as ~ 250 bp in pteridophytes and bryophytes (Stech et al., 2003) to over 1400 bp in some angiosperms (e.g., *Disa* [Orchidaceae], Bellstedt et al. [2001]). The *trnL-trnF* spacer is generally shorter than the *trnL* intron, ranging from less than 100 bp in mosses and liverworts (Stech et al., 2003) up to ~ 500 bp in seed plants.

Sequences of the *trnL* intron and *trnL-trnF* spacer have been employed in numerous studies, oftentimes together because they can be coamplified using the "C" and "F" primers of Taberlet et al. (1991). In studies where both the *trnL* intron and the *trnL-trnF* spacer have been sequenced for a common set of taxa, the number of parsimony-informative characters in the *trnL-trnF* spacer is often greater than or equal to the *trnL* intron, despite the fact that the *trnL* intron is usually larger than the *trnL-trnF* spacer. For example, in *Lepidium* (Brassicaceae) the aligned length of the *trnL* intron was 519 bp vs. 350 bp in the *trnL-trnF* spacer, yet only 46 parsimony-informative sites were detected in the *trnL* intron vs. 52 in the *trnL-trnF* spacer (Mummenhoff et al., 2001). Similarly, in *Disa* (Orchidaceae) the *trnL* intron was 1412 bp aligned vs. 359 bp for the *trnL-trnF* spacer, yet 72 parsimony-informative substitutions were found in each region, despite the fact that the *trnL* intron is almost four times longer (Bellstedt et al., 2001). Lastly, Yang et al. (2002) showed that the rate of nucleotide substitution in the *trnL* intron is about 33% of that within the *trnT-trnL* or *trnL-trnF* spacers. Presumably, this observation is due to greater functional constraints on the *trnL* intron that must assume a correct secondary structure for proper removal.

A comprehensive list of studies using the *trnT-trnL-trnL-trnF* regions is spatially impossible and the majority of the papers cited in this work contain at least a portion of this region. However, the following list provides some representative papers and emphasizes those studies that have employed combinations of *trnT-trnL* and *trnL-trnF* spacer and/or *trnL* intron sequences or comparisons of these sequences with other coding

or noncoding regions. These studies include: Gielly and Taberlet (1994); Sang et al. (1997); Small et al. (1998); Mort et al. (2001); Bremer et al. (2002); Cronn et al. (2002); Goldblatt et al. (2002); Hartmann et al. (2002); Mast and Givnish (2002); Borsch et al. (2003); Fukuda et al. (2003); Jobson et al. (2003); Miller et al. (2003); Salazar et al. (2003); Simpson et al. (2003); and Stech et al. (2003).

Our data support the previous findings that the *trnT-trnL* spacer is much more variable than the *trnL-trnF* spacer and that the *trnL* intron is the least variable of these three regions. The *trnT-trnL* spacer averages 752 bp, ranging from 527 to 1023 bp, and exhibits a few large indels (174 bp in *Minuartia*, 46 bp in *Hibiscus*, 61 bp in *Gratiola*, 63 bp in *Trillium*). The *trnL* intron averages 499 bp, ranging from 395 to 602 bp, whereas the *trnL-trnF* spacer averages 362 bp and ranges from 207 to 474 bp. Separately, the *trnL* intron and the *trnL-trnF* spacer are Tier 3 regions, while the *trnT-trnL* spacer is a Tier 1 region. When the *trnL* intron and the *trnL-trnF* spacer (*trnL-trnL-trnF*) are combined and compared to the other combined regions, this region still ranks behind several others (Fig. 7). However, when the *trnT-trnL* spacer is coamplified with either the *rps4-trnT* spacer or the *trnL* intron (Fig. 7), this combined region is among the most variable of the combined regions, behind only *trnS-trnG-trnG*, *trnC-ycf6-psbM*, or *ycf6-psbM-trnD*.

5´rpS12-rpL20 (Tier 3)—Hamilton (1999a, b) designed primers for the intergenic spacer between *5´rps12* and *rpL20* to study population dynamics within a tropical tree species in the genus *Corythophora* (Lecythidaceae) and subsequently published them along with the suggested amplification protocol.

The *5´rps12-rpL20* intergenic spacer averages 783 bp and ranges from 715 to 866 bp. This is comparable to the length of approximately 880 bp in *Corythophora alta* reported by Hamilton (1999b). Contrary to Hamilton (1999a), this Tier 3 region consistently shows a relatively low PIC value and does not appear suitable for low level investigations.

psbB-psbH (Tier 3)—Hamilton (1999a, b) designed primers for the intergenic spacer between *psbB* and *psbH* to study population dynamics within a tropical tree species in *Corythophora* (Lecythidaceae) and subsequently published them along with the suggested amplification protocol. Xu et al. (2000) showed this region to be less informative than *trnH-psbA*, *trnS-trnG*, and *trnT-trnL*, but more informative than *atpB-rbcL* and *ndhD-ndhE*. Schütze et al. (2003) also used this region in addition to the *atpB-rbcL* spacer in an investigation of the Suaedoideae (Chenopodiaceae) and showed that it was only slightly more than half as informative as *atpB-rbcL*. One reason for the lack of variable characters within *psbB-psbH* is that two genes, *psbT* (~100 bp) and *psbN* (~130 bp), comprise nearly half of the approximately 527-bp spacer between *psbB* and *psbH*.

The *psbB-psbH* intergenic region averages 527 bp, ranging from 250 to 604 bp, and contains approximately 230 bp of coding sequence, which is relatively invariable across all lineages. In reference to the number of PICs offered to investigators, this Tier 3 region ranks toward the bottom and is therefore not a suitable region for low-level investigation, contrary to Hamilton (1999a).

rpL16 (Tier 2)—Posno et al. (1986) first demonstrated homology between the chloroplast *rpL*16 region from *Spirodela* (Lemnaceae) and the ribosomal protein L16 of *E. coli*, and observed the presence of a Group II intron that split the coding region into shorter and longer exons. Comparisons of whole chloroplast sequences confirmed its presence and pointed to the *rpL16* region as having high sequence divergence in flowering plants (Wolfe et al., 1987). Jordan et al. (1996) published the first attempt to use *rpL16* intron sequence data for phylogenetic studies, but reported relatively little variation in Lemnaceae. Application of data from the *rpL16* region has been primarily for phylogenetic analysis at the infrageneric and familial levels (Kelchner and Clark, 1997; Baum et al., 1998; Schnabel and Wendel, 1998; Small et al., 1998; Seelanan et al., 1999; Applequist and Wallace, 2000; Downie et al., 2000; Zhang, 2000; Shaw, 2000; Baumel et al., 2001; Mast et al., 2001; Butterworth et al., 2002; Cronn et al., 2002; Les et al., 2002; Mast and Givnish, 2002; Pfeil et al., 2002; Pires and Sytsma, 2002; Kimball et al., 2003; Perret et al., 2003). There have been some reports of variability within species (e.g., Seelanan et al., 1999; Les et al., 2002), but only a few studies have reported using it specifically for examination of intraspecific variation (Xu et al., 2000; Kimura et al., 2003).

The *rpL16* intron averages 1002 bp, ranging from 811 to 1208 bp, and is especially indel prone in the D3 bulge region (Baum et al., 1998; Kelchner, 2000; Kelchner, 2002; Pfeil et al., 2002). Relative to the other regions surveyed here, this intron ranks in Tier 2. Finally, the *rpL16* intron is absent in at least some Geraniaceae, Goodeniaceae, and Plumbaginaceae (Campagna and Downie, 1998), precluding its use in some groups.

***Indels vs. nucleotide substitutions*—**A number of authors have addressed the issue of the relative frequencies of nucleotide substitutions and indels in noncoding cpDNA sequences. Clegg et al. (1994) noted that indels may occur more frequently than nucleotide substitutions. Golenberg et al. (1993) and Gielly and Taberlet (1994) suggested that indels occur with nearly the same frequency as nucleotide substitutions. On the contrary, our results agree more with Small et al. (1998). We found that nucleotide substitutions account for 70.1% of the PIC value, while indels account for only 29.8% and inversions only 0.14% of the 35% of the noncoding LSC region we surveyed. As systematic investigations move more toward lower levels and any variable characters become important, indels, which may be homoplasious at deeper levels (Golenberg et al., 1993), are of great utility for infrageneric studies.

Several of the regions were rich in strings of mononucleotide repeats and/or small tandem repeat units that are likely the result of slipped-strand mispairing (Levinson and Gutman, 1987). Polynucleotide (A/T) repeats and/or small tandem repeats (AT) were especially noted in the *trnH-psbA*, *psbA*-3′*trnK*, *matK*-5′*trnK*, *trnS-trnfM*, *trnS-trnG*, *trnD-trnT*, *trnT-trnL* spacers and in the *rpS16* and *trnG* introns. Length variation, because of relatively large indels, was noted in several regions, described above.

***Implications of this study*—**The results of our study do not point to a "holy grail" of noncoding cpDNA regions that can be universally used for low-level systematic studies. However, our results do highlight several noncoding cpDNA regions that are

better suited for low-level investigation than many commonly employed choices. The Tier 1 intergenic spacers that provided the greatest numbers of PICs in order of most to least include *trnD-trnT*, *rpoB-trnC*, *trnS-trnG*, *trnS-trnfM*, and *trnT-trnL*. Several other regions, designated as Tier 2, may also be useful to investigators if one or more of the Tier 1 regions cannot be obtained or simply to add more data including *rpS16*, *rpL16*, *ycf6-psbM*, *trnG*, and *psbM-trnD*. Regions that consistently provided the fewest PICs include *trnC-ycf6*, *5´rpS12-rpL20*, *trnH-psbA*, *matK*-5´*trnK*, *rpS4-trnT*, *trnL-trnF*, *trnL*, 3´*trnK-matK*, *psbB-psbH*, *psbA*-3´*trnK*, and *trnS-rpS4*. Because several of the more variable regions are often adjacent to other surveyed regions, they are easily coamplifiable and may be amplified and sequenced with little to no additional cost. As we have shown, we can amplify a combined *trnS-trnG-trnG* fragment (Tier 1 + Tier 2 region) to yield a fragment that exhibits the greatest number of PICs per two sequencing reactions as compared to other noncoding cpDNA choices. Other combinable regions that yield relatively high PIC values include *ycf6-psbM-trnD*, *trnC-ycf6-psbM*, and *rpS4-trnT-trnL* (Fig. 7). Because these combined regions include small internal coding regions, internal primers already exist and may be used in cases where sequencing is difficult. The five shorter, uncombined Tier 1 regions may provide a more cost-efficient alternative if your sequencing reactions yield confidently useable reads of < 700 bp.

An additional important finding of this work is that a preliminary three-species survey can be used to determine the relative utility of a given region prior to implementing a full scale sequencing project. Such a full-scale assault, blindly choosing a reportedly useful region which may yield little information, may be a risky and costly venture in terms of time and resources. While the regions we identify as Tier 1 are consistently among the most variable among all lineages we tested, there is still variation within and among lineages with respect to phylogenetic utility of the regions. A pilot study employing a small number (e.g., three in this study) of taxa can quickly identify which particular Tier 1 region may likely be the most informative within a particular species group.

In summary, the data we present indicate that there is indeed phylogenetically significant and predictable rate heterogeneity among noncoding cpDNA regions. While considerable variation exists among lineages, several noncoding cpDNA regions are identified that consistently provide greater levels of sequence variation compared to other regions that consistently yield low levels of variation, such as the commonly employed *trnL-trnL-trnF* and *trnK*/*matK*. More phylogenetically informative variation appears to be present in the chloroplast genome than previously inferred, based on the accumulated evidence from a small number of apparently more slowly evolving noncoding regions. In addition, we show the importance and applicability of performing pilot studies to identify appropriate regions for further study. A small survey with as few as three species can be predictive of the overall levels of variation likely to be found in a larger scale study. The application of the top tier regions we have identified for future infrageneric studies and the continued exploration of noncoding regions of the chloroplast genome for variable markers are warranted.

Relative to the cpDNA genome, comparatively few noncoding cpDNA regions were surveyed in this study, and unsampled cpDNA regions may be found that yield a greater number of PICs than any of the top tier regions of this study. Therefore, we are

currently in the midst of a companion study where we are adding data from other noncoding cpDNA regions to this data set in continuation of our search for the "holy grail."

LITERATURE CITED

AAGESEN, L., AND A. M. SANSO. 2003. The phylogeny of the Alstroemeriaceae, based on morphology, *rps16* intron, and *rbcL* sequence data. *Systematic Botany* 28: 47-69.

ALDRICH, J., B. W. CHERNEY, E. MERLIN, AND L. CHRISTOPHERSON. 1988. The role of insertions/deletions in the evolution of the intergenic region between *psbA* and *trnH* in the chloroplast genome. *Current Genetics* 14: 137-146.

ÁLVAREZ, I., AND J. F. WENDEL. 2003. Ribosomal ITS sequences and plant phylogenetic inference. *Molecular Phylogenetics and Evolution* 29: 417-434.

ANDERSSON, C. L., AND J. H. E. ROVA. 1999. The *rpS16* intron and the phylogeny of the Rubioideae (Rubiaceae). *Plant Systematics and Evolution* 214: 161-186.

ANDERSSON, C. L., AND M. W. CHASE. 2001. Phylogeny and classification of Marantaceae. *Botanical Journal of the Linnean Society* 135: 275-287.

AOKI, K., T. SUZUKI, AND N. MURAKAMI. 2003. Intraspecific sequence variation of chloroplast DNA among the component species of evergreen broad-leaved forests in Japan. *Journal of Plant Research* 116: 337-344.

APG II. 2003. An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG II. *Botanical Journal of the Linnean Society* 141: 399-436.

APPLEQUIST, W. L., AND R. S. WALLACE. 2000. Phylogeny of the Madagascan endemic family Didiereaceae. *Plant Systematics and Evolution* 221: 157-166.

ASMUSSEN, C. B., AND M. W. CHASE. 2001. Coding and noncoding plastid DNA in palm systematics. *American Journal of Botany* 88: 1103-1117.

AZUMA, H., J. G. GARCÍA-FRANCO, V. RICO-GRAY, AND L. B. THIEN. 2001. Molecular phylogeny of the Magnoliaceae: the biogeography of tropical and temperate disjunctions. *American Journal of Botany* 88: 2275-2285.

AZUMA, H., L. B. THIEN, AND S. KAWANO. 1999. Molecular phylogeny of *Magnolia* (Magnoliaceae) inferred from cpDNA sequences and evolutionary divergence of the floral scents. *Journal of Plant Research* 112: 291-306.

BAKER, W. J., T. A. HEDDERSON, AND J. DRANSFIELD. 2000. Molecular phylogenetics of subfamily Calamoideae (Palmae) based on nrDNA ITS and cpDNA *rpS16* intron sequence data. *Molecular Phylogenetics and Evolution* 14: 195-217.

BALDWIN, B. G. 1992. Phylogenetic utility of the internal transcribed spacers of nuclear ribosomal DNA in plants: an example from the Compositae. *Molecular Phylogenetics and Evolution* 1: 3-16.

BALDWIN, B. G., M. J. SANDERSON, J. M. PORTER, M. F. WOJCIECHOWSKI, C. S. CAMPBELL, AND M. J. DONOGHUE. 1995. The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. *Annals of the Missouri Botanical Garden* 82: 247-277.

BAUM, D. A., R. L. SMALL, AND J. F. WENDEL. 1998. Biogeography and floral evolution of baobabs (*Adansonia*, Bombacaceae) as inferred from multiple data sets. *Systematic Biology* 47: 181-207.

BAUMEL, A., M. L. AINOUCHE, AND J. E. LEVASSEUR. 2001. Molecular investigations in populations of *Spartina anglica* C. E. Hubbard (Poaceae) invading coastal Brittany (France). *Molecular Ecology* 10: 1689-1701.

BAYER, R. J., D. G. GREBER, AND N. H. BAGNALL. 2002. Phylogeny of Australian Gnaphalieae (Asteraceae) based on chloroplast and nuclear sequences, the *trnL* intron, *trnL-trnF* intergenic spacer, *matK*, and ETS. *Systematic Botany* 27: 801-814.

BELL, C. D., AND R. W. PATTERSON. 2000. Molecular phylogeny and biogeography of *Linanthus* (Polemoniaceae). *American Journal of Botany* 87: 1857-1870.

BELLSTEDT, D. U., H. P. LINDER, AND E. H. HARLEY. 2001. Phylogenetic relationships in *Disa* based on noncoding *trnL-trnF* chloroplast sequences: evidence of numerous repeat regions. *American Journal of Botany* 88: 2088-2100.

BÖHLE, U. –R., H. HILGER, R. CERFF, AND W. F. MARTIN. 1994. Noncoding chloroplast DNA for plant molecular systematics at the infrageneric level. *In* B. Schierwater, B. Streit, G. P. Wagner, and R. DeSalle [eds.], Molecular ecology and evolution: approaches and applications, 391-403. Birkhäuser Verlag, Basel, Germany.

BORSCH, T., K. W. HILU, D. QUANDT, V. WILDE, C. NEINHUIS, AND W. BARTHLOTT. 2003. Noncoding plastid *trnT-trnF* sequences reveal a well resolved phylogeny of basal angiosperms. *Journal of Evolutionary Biology* 16: 558-576.

BORTIRI, E., S. H. OH, J. G. JIANG, S. BAGGETT, A. GRANGER, C. WEEKS, M. BUCKINGHAM, D. POTTER, AND D. E. PARFITT. 2001. Phylogeny and systematics of *Prunus* (Rosaceae) as determined by sequence analysis of ITS and the chloroplast *trnL-trnF* spacer DNA. *Systematic Botany* 26: 797-807.

BREMER, B., K. BREMER, N. HEIDARI, P. ERIXON, R. G. OLMSTEAD, A. A. ANDERBERG, M. KÄLLERSJÖ, AND E. BARKHORDARIAN. 2002. Phylogenetics of asterids based on 3 coding and 3 noncoding chloroplast DNA markers and the utility of noncoding DNA at higher taxonomic levels. *Molecular Phylogenetics and Evolution* 24: 274-301.

BUTTERWORTH, C. A., J. H. COTA-SANCHEZ, AND R. S. WALLACE. 2002. Molecular systematics of the tribe Cacteae (Cactaceae: Cactoideae): a phylogeny based on *rpl16* intron sequence variation. *Systematic Botany* 27: 257-270.

CAMPAGNA, M. L., AND S. R. DOWNIE. 1998. The intron in chloroplast gene *rpL16* is missing from the flowering plant families Geraniaceae, Goodeniaceae, and Plumbaginaceae. *Transactions of the Illinois State Academy of Science* 91: 1-11.

CHANDLER, G. T., R. J. BAYER, AND M. D. CRISP. 2001. A molecular phylogeny of the endemic Australian genus *Gastrolobium* (Fabaceae: Mirbelieae) and allied genera using chloroplast and nuclear markers. *American Journal of Botany* 88: 1675-1687.

CHASE, M. W., A. Y. DE BRUIJN, A. V. COX, G. REEVES, P. J. RUDALL, M. A. T. JOHNSON, AND L. E. EGUIARTE. 2000. Phylogenetics of Asphodelaceae (Asparagales): an analysis of plastid *rbcL* and *trnL-F* DNA sequences. *Annals of Botany* 86: 935-951.

CHASE, M. W., ET AL. 1993. Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbcL*. *Annals of the Missouri Botanical Garden* 80: 528-580.

CHASSOT, P., S. NEMOMISSA, Y. –M. YUAN, AND P. KÜPFER. 2001. High paraphyly of *Swertia* L. (Gentianaceae) in the *Gentianella*-lineage as revealed by nuclear and chloroplast DNA sequence variation. *Plant Systematics and Evolution* 229: 1-21.

CLARK, L. G., W. ZHANG, AND J. F. WENDEL. 1995. A phylogeny of the grass family (Poaceae) based on *ndhF* sequence data. *Systematic Botany* 20: 436-460.

CLEGG, M. T., B. S. GAUT, G. H. LEARN, JR., AND B. R. MORTON. 1994. Rates and patterns of chloroplast DNA evolution. *Proceedings of the National Academy of Sciences, USA* 91: 6795-6801.

CRANFILL, R. B. 2001. Phylogenetic utility of plastid ribosomal protein S4 (*rpS4*) in land plants. *American Journal of Botany* 87: 121 (Abstract).

CRONN, R. C., R. L. SMALL, T. HASELKORN, AND J. F. WENDEL. 2002. Rapid diversification of the cotton genus (*Gossypium*: Malvaceae) revealed by analysis of sixteen nuclear and chloroplast genes. *American Journal of Botany* 89: 707-725.

CUÉNOUD, P., M. A. DEL PERO MARTINEZ, P. A. LOIZEAU, R. SPICHIGER, S. ANDREWS, AND J. F. MANEN. 2000. Molecular phylogeny and biogeography of the genus *Ilex* L. (Aquifoliaceae). *Annals of Botany* 85: 111-122.

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 *rbcL*, *atpB*, and *matK* DNA sequences. *American Journal of Botany* 89: 132-144.

DEMESURE, B., B. COMPS, AND R. PETIT. 1996. Chloroplast DNA phylogeography of the common beech (*Fagus sylvatica* L.) in Europe. *Evolution* 50: 2515-2520.

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

DOWNIE, S. R., R. L. HARTMAN, S.-F. JIE, D. S. KATZ-DOWNIE. 2002. Polyphyly of the spring-parsleys (*Cymopterus*): molecular and morphological evidence suggests complex relationships among the perennial endemic genera of western North American Apiaceae. *Canadian Journal of Botany* 80: 1295-1324.

DOWNIE, S. R., AND D. S. KATZ-DOWNIE. 1999. Phylogenetic analysis of chloroplast *rps16* intron sequences reveals relationships within the woody southern African Apiaceae subfamily Apioideae. *Canadian Journal of Botany* 77: 1120-1135.

DOWNIE, S. R. , D. S. KATZ-DOWNIE, AND K. J. CHO. 1996. Phylogenetic analysis of Apiaceae subfamily Apioideae using nucleotide sequences from the chloroplast *rpoC1* intron. *Molecular phylogenetics and Evolution* 6: 1-18.

DOWNIE, S. R., D. S. KATZ-DOWNIE, AND M. F. WATSON. 2000. A phylogeny of the flowering plant family Apiaceae based on chloroplast DNA *rpl16* and *rpoC1* intron sequences: towards a suprageneric classification of subfamily Apioideae. *American Journal of Botany* 87: 273-292.

DOWNIE, S. R., AND J. D. PALMER. 1992. Use of chloroplast DNA rearrangements in reconstructing plant phylogeny. *In* Soltis et al. [eds.], Molecular systematics of plants, 14-35. Chapman and Hall, New York, New York, USA.

DOWNIE, S. R., G. M. PLUNKETT, M. F. WATSON, K. SPALIK, D. S. KATZ-DOWNIE, C. M. VALIEJO-ROMAN, E. I. TERENTIEVA, A. V. TROITSKY, B. –Y. LEE, AND A. EL-

OQLAH. 2001. Tribes and clades within Apiaceae subfamily Apioideae: the contribution of molecular data. *Edinburg Journal of Botany* 58: 301-330.

DOYLE, J. J., AND J. L. DOYLE. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* 19: 11-15.

DOYLE, J. J., J. L. DOYLE AND J. D. PALMER. 1995. Multiple independent losses of two genes and one intron from legume chloroplast genomes. *Systematic Botany* 20: 272-294.

DUMOLIN-LAPEGUE, S., M. H. PEMONGE, AND R. J. PETIT. 1997. An enlarged set of consensus primers for the study of organelle DNA in plants. *Molecular Ecology* 6: 393-397.

EDWARDS, K. J., AND P. A. GADEK. 2001. Evolution and biogeography of *Alectryon* (Sapindaceae). *Molecular Phylogenetics and Evolution* 20: 14-26.

EHRENDORFER, F., J. F. MANEN, AND A. NATALI. 1994. cpDNA intergene sequences corroborate restriction site data for reconstructing Rubiaceae phylogeny. *Plant Systematics and Evolution* 190: 245-248.

EL MOUSADIK, A., AND PETIT, R. J. 1996. Chloroplast DNA phylogeography of the Argan tree of Morocco. *Molecular Ecology* 5: 547-555.

FELINER, G. N., J. F. AGUILAR, AND J. A. ROSSELLO. 2002. Reticulation or divergence: the origin of a rare serpentine endemic assessed with chloroplast, nuclear, and RAPD markers. *Plant Systematics and Evolution* 231: 19-38.

FRIESEN, N., R. M. FRITSCH, S. POLLNER, AND F. R. BLATTNER. 2000. Molecular and morphological evidence for an origin of the aberrant genus *Milula* within the Himalayan species of *Allium* (Alliaceae). *Molecular Phylogenetics and Evolution* 17: 209-218.

FUKUDA, T., J. YOKOYAMA, AND H. TSUKAYA. 2003. Phylogenetic relationships among species in the genera *Chisocheton* and *Guarea* that have unique indeterminate leaves as inferred from sequences of chloroplast DNA. *International Journal of Plant Sciences* 164: 13-24.

GASKIN, J. F., AND B. A. SCHAAL. 2003. Molecular phylogenetic investigation of U.S. Invasive *Tamarix*. *Systematic Botany* 28: 86-95.

GE, S., A. LI, B. R. LU, S. Z. ZHANG, AND D. Y. HONG. 2002. A phylogeny of the rice tribe Oryzeae (Poaceae) based on *matK* sequence data. *American Journal of Botany* 89: 1967-1972.

GENE CODES CORP. 1998. Sequencher, version 3.1.1. Gene Codes, Ann Arbor, Michigan, USA.

GIELLY, L. AND P. TABERLET. 1994. The use of chloroplast DNA to resolve plant phylogenies: noncoding versus *rbcL* sequences. *Molecular Biology and Evolution* 11: 769-777.

GOLDBLATT, P., V. SAVOLAINEN, O. PORTEOUS, I. SOSTARIC, M. POWELL, G. REEVES, J. C. MANNING, T. G. BARRACLOUGH, AND M. W. CHASE. 2002. Radiation on the Cape flora and the phylogeny of peacock irises *Moraea* (Iridaceae) based on four plastid DNA regions. *Molecular Phylogenetics and Evolution* 25: 341-360.

GOLENBERG, E. M., M. T. CLEGG, M. L. DURBIN, J. DOEBLEY, AND D. P. MA. 1993. Evolution of a noncoding region of the chloroplast genome. *Molecular Phylogenetics and Evolution* 2: 52-64.

HAHN, W. J. 2002. A phylogenetic analysis of the Arecoid line of palms based on plastid DNA sequence data. *Molecular Phylogenetics and Evolution* 23: 189-204.

HAMILTON, M. B. 1999a. Four primer pairs for the amplification of chloroplast intergenic regions with intraspecific variation. *Molecular Ecology* 8: 521-523.

HAMILTON, M. B. 1999b. Tropical tree gene flow and seed dispersal: deforestation affects the genetic structure of the surviving forest fragments. *Nature* 401: 129-130.

HAMILTON, M. B., J. M. BRAVERMAN, AND D. F. SORIA-HERNANZ. 2003. Patterns and relative rates of nucleotide and insertion/deletion evolution at six chloroplast intergenic regions in New World species of the Lecythidaceae. *Molecular Biology and Evolution* 20: 1710-1721.

HARDIG, T. M., P. S. SOLTIS, AND D. E. SOLTIS. 2000. Diversification of the North American shrub genus *Ceanothus* (Rhamnaceae): conflicting phylogenies from nuclear ribosomal DNA and chloroplast DNA. *American Journal of Botany* 87: 108-123.

HARTMANN, S., J. D. NASON, AND D. BHATTACHARYA. 2002. Phylogenetic origins of *Lophocereus* (Cactaceae) and the senita cactus-senita moth pollination mutualism. *American Journal of Botany* 89: 1085-1092.

HENNEQUIN, S., A. EBIHARA, M. ITO, K. IWATSUKI, AND J. Y. DUBUISSON. 2002. Molecular systematics of the fern genus *Hymenophyllum s.l.* (Hymenophyllaceae) based on chloroplastic coding and noncoding regions. *Molecular Phylogenetics and Evolution* 27: 283-301.

HILU, K. W., AND H. LIANG. 1997. The *matK* gene: sequence variation and application in plant systematics. *American Journal of Botany* 84: 830-839.

HILU, K. W., T. BORSCH, K. MÜLLER, D. E. SOLTIS, P. S. SOLTIS, V. SAVOLAINEN, M. W. CHASE, M. P. POWELL, L. A. ALICE, R. EVANS, H. SAUQUET, C. NEINHUIS, T. A. B. SLOTTA, J. G. ROHWER, C. S. CAMPBELL, AND L. W. CHATROU. 2003. Angiosperm phylogeny based on *matK* sequence information. *American Journal of Botany* 90: 1758-1776.

HIRATSUKA, J., H. SHIMADA, R. WHITTIER, T. ISHIBASHI, M. SAKAMOTO, M. MORI, C. KONDO, Y. HONJI, C. R. SUN, B. Y. MENG, Y. Q. LI, A. KANNO, Y. NISHIZAWA, A. HIRAI, K. SHINOZAKI, AND M. SIGIURA. 1989. The complete sequence of the rice (*Oryza sativa*) chloroplast genome: intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. *Molecular General Genetics* 217: 185-194.

HODGES, S. A., AND M. L. ARNOLD. 1994. Columbines: a geographically widespread species flock. *Proceedings of the National Academy of Sciences, USA* 91: 5129-5132.

HOLDREGGER, R., AND R. J. ABBOTT. 2003. Phylogeography of the Arctic-Alpine *Saxifraga oppositifolia* (Saxifragaceae) and some related taxa based on cpDNA and ITS sequence variation. *American Journal of Botany* 90: 931-936.

HOOT, S. B., A. CULHAM, AND P. R. CRANE. 1995. The utility of *atpB* gene sequences in resolving phylogenetic relationships: comparison with *rbcL* and 18S ribosomal DNA sequences in the Lardizabalaceae. *Annals of the Missouri Botanical Garden* 82: 194-208.

HU, J. –M., M. LAVIN, M. F. WOJCIECHOWSKI, AND M. J. SANDERSON. 2000. Phylogenetic systematics of the tribe Millettieae (Leguminosae) based on chloroplast *trnK/matK* sequences and its implications for evolutionary patterns in Papilionoideae. *American Journal of Botany* 87: 418-430.

HUFFORD, L., M. M. MCMAHON, A. M. SHERWOOD, G. REEVES, AND M. W. CHASE. 2003. The major clades of Loasaceae: phylogenetic analysis using the plastid *matK* and *trnL-F* regions. *American Journal of Botany* 90: 1215-1228.

INGRAM, A. L., AND J. J. DOYLE. 2003. The origin and evolution of *Eragrostis tef* (Poaceae) and related polyploids: evidence from nuclear *waxy* and plastid *rps16*. *American Journal of Botany* 90: 116-122.

JANSEN, R. K., AND J. D. PALMER. 1987. A chloroplast DNA inversion marks an ancient evolutionary split in the sunflower family (Asteraceae). *Proceedings of the National Academy of Sciences, USA* 84: 5818-5822.

JENSEN, U., S. B. HOOT, J. T. JOHANSSON, AND K. KOSUGE. 1995. Systematics and phylogeny of the Ranunculaceae – a revised family concept on the basis of molecular data. *Plant Systematics and Evolution* Suppl. 9: 273-280.

JOBSON, R. W., J. PLAYFORD, K. M. CAMERON, AND V. A. ALBERT. 2003. Molecular phylogenetics of Lentibulariaceae inferred from plastid *rpS16* intron and *trnL-F* DNA sequences: implications for character evolution and biogeography. *Systematic Botany* 28: 157-171.

JOHNSON, L. A. AND D. E. SOLTIS. 1994. *matK* DNA sequence and phylogenetic reconstruction in Saxifragaceae *s.s.* *Systematic Botany* 19: 143-156.

JORDAN, W. C., M. W. COURTNEY, AND J. E. NEIGEL. 1996. Low levels of intraspecific genetic variation at a rapidly evolving chloroplast DNA locus in North American duckweeeds (Lemnaceae). *American Journal of Botany* 83: 430-439.

KAJITA, T., K. KAMIYA, K. NAKAMURA, H. TACHIDA, R. WICKNESWARI, Y. TSUMURA, H. YOSHIMARU, AND T. YAMAZAKI. 1998. Molecular phylogeny of Dipterocarpaceae in southeast Asia based on nucleotide sequences of *matK*, *trnL* intron, and the *trnL-trnF* intergenic spacer region in chloroplast DNA. *Molecular Phylogenetics and Evolution* 10: 202-209.

KAZEMPOUR OSALOO, S., F. H. UTECH, M. OHARA, AND S. KAWANO. 1999. Molecular systematics of Trilliaceae I. Phylogenetic analyses of *Trillium* using *matK* gene sequences. *Journal of Plant Research* 112: 35-49.

KELCHNER, S. A. 2000. The evolution of non-coding chloroplast DNA and its application in plant systematics. *Annals of the Missouri Botanical Garden* 87: 482-498.

KELCHNER, S. A. 2002. Group II introns as phylogenetic tools: structure, function, and evolutionary contraints. *American Journal of Botany* 89: 1651-1669.

KELCHNER, S. A. AND L. G. CLARK. 1997. Molecular evolution and phylogenetic utility of the chloroplast *rpl16* intron in *Chusquea* and the Bambusoideae (Poaceae). *Molecular Phylogenetics and Evolution* 8: 385-397.

KIM, K. J., AND R. K. JANSEN. 1995. *ndhF* sequence evolution and the major clades in the sunflower family. *Proceedings of the National Academy of Sciences, USA* 92: 10379-10383.

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, T., H. IKETANI, K. KOTOBUKI, N. MATSUTA, Y. BAN, T. HAYASHI, AND T. YAMAMOTO. 2003. Genetic characterization of pear varieties revealed by chloroplast DNA sequences. *Journal of Horticultural Science and Biotechnology* 78: 241-247.

KLAK, C., T. A. HEDDERSON, AND P. LINDER. 2003. A molecular systematic study of the *Lampranthus* group (Aizoaceae) based on the chloroplast *trnT-trnF* and nuclear ITS and 5S NTS sequence data. *Systematic Botany* 28: 70-85.

KUSUMI, J., Y. TSUMURA, H. YOSHIMARU, AND H. TACHIDA. 2000. Phylogenetic relationships in Taxodiaceae and Cupressaceae sensu stricto based on *matK* gene, *chlL* gene, *trnL-trnF* IGS region, and *trnL* intron sequences. *American Journal of Botany* 87: 1480-1488.

LEE, C., AND J. WEN. 2004. Phylogeny of *Panax* using chloroplast *trnC-trnD* intergenic region and the utility of *trnC-trnD* in interspecific studies of plants. *Molecular Phylogenetics and Evolution* 31: 894-903.

LEE, J., AND T. HYMOWITZ. 2001. A molecular phylogenetic study of the subtribe Glycininae (Leguminosae) derived from the chloroplast DNA *rps16* intron sequences. *American Journal of Botany* 88: 2064-2073.

LES, D. H., D. J. CRAWFORD, E. LANDOLT, J. D. GABEL, AND R. T. KEMBALL. 2002. Phylogeny and systematics of Lemnaceae, the duckweed family. *Systematic Botany* 27: 221-240.

LEVINSON, G., AND G. A. GUTMAN. 1987. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Molecular Biology and Evolution* 4: 203-221.

LU, S. –Y., C. PENG, Y. –P. CHENG, K. –H. HONG, AND T. –Y. CHIANG. 2001. Chloroplast DNA phylogeography of *Cunninghamia konishii* (Cupressaceae), an endemic conifer of Taiwan. *Genome* 44: 797-807.

MANEN, J. F., A. NATALI, AND F. EHRENDORFER. 1994. Phylogeny of Rubiaceae-Rubieae inferred from the sequence of a cpDNA intergene region. *Plant Systematics and Evolution* 190: 195-211.

MAST, A. R., AND T. J. GIVNISH. 2002. Historical biogeography and the origin of stomatal distributions in *Banksia* and *Dryandra* (Proteaceae) based on their cpDNA phylogeny. *American Journal of Botany* 89: 1311-1323.

MAST, A. R., S. KELSO, A. J. RICHARDS, D. J. LANG, D. M. S. FELLER, AND E. CONTI. 2001. Phylogenetic relationships in *Primula* L. and related genera (Primulaceae) based on noncoding chloroplast DNA. *International Journal of Plant Sciences* 162: 1381-1400.

MAYER, V., M. MOLLER, M. PERRET, AND A. WEBER. 2003. Phylogenetic position and generic differentiation of Epithemateae (Gesneriaceae) inferred from plastid DNA sequence data. *American Journal of Botany* 90: 321-329.

MES, T. H. M., N. FRIESEN, R. M. FRITSCH, M. KLAAS, AND K. BACHMANN. 1997. Criteria for sampling in *Allium* based on chloroplast DNA PCR-RFLP's. *Systematic Botany* 22: 701-712.

MILLER, J. T., AND R. J. BAYER. 2001. Molecular phylogenetics of *Acacia* (Fabaceae: Mimosoideae) based on the chloroplast *matK* coding sequence and flanking *trnK* intron spacer regions. *American Journal of Botany* 88: 697-705.

MILLER, J. T., J. W. GRIMES, D. J. MURPHY, R. J. BAYER, AND P. Y. LADIGES. 2003. A phylogenetic analysis of the Acacieae and Ingeae (Mimosoideae: Fabaceae) based on *trnK, matK, psbA-trnH,* and *trnL/trnF* sequence data. *Systematic Botany* 28: 558-566.

MORT, M. E., D. E. SOLTIS, P. S. SOLTIS, O. J. FRANCISCO, AND A. SANTOS-GUERRA. 2001. Phylogenetics and evolution of the Macaronesian clade of Crassulaceae inferred from nuclear and chloroplast sequence data. *Systematic Botany* 27: 271-288.

MORTON, B. R., AND M. T. CLEGG. 1993. A chloroplast DNA mutational hotspot and gene conversion in a noncoding region near *rbcL* in the grass family (Poaceae). *Current Genetics* 24: 357-365.

MUELLNER, A. N., R. SAMUEL, S. A. JOHNSON, M. CHEEK, T. D. PENNINGTON, AND M. W. CHASE. 2003. Molecular phylogenetics of Meliaceae (Sapindales) based on nuclear and plastid DNA sequences. *American Journal of Botany* 90: 471-480.

MUMMENHOFF, K., H. BRÜGGEMANN, AND J. L. BOWMAN. 2001. Chloroplast DNA phylogeny and biogeography of *Lepidium* (Brassicaceae). *American Journal of Botany* 88: 2051-2063.

NAGANO, Y., R. MATSUNO, AND S. SASAKI. 1991. Sequence and transcriptional analysis of the gene cluster *trnQ-zfpA-psaI-ORF231-petA* in pea chloroplasts. *Current Genetics* 20: 431-436.

NEUHAUS, H., AND G. LINK. 1987. The chloroplast tRNA Lys (UUU) gene from mustard (*Sinapis alba*) contains a class II intron potentially coding for a maturase-related polypeptide. *Current Genetics* 11: 251-257.

NYFFELER, R. 2002. Phylogenetic relationships in the cactus family (Cactaceae) based on evidence from *trnK/matK* and *trnL-trnF* sequences. *American Journal of Botany* 89: 312-326.

O'DONNELL, K. 1992. Ribosomal DNA internal transcribed spacers are highly divergent in the phytopathogenic ascomycete *Fusarium sambucinum* (*Gibberella pulicaris*). *Current Genetics* 22: 213-220.

OHSAKO, T., AND O. OHNISHI. 2000. Intra- and interspecific phylogeny of wild *Fagopyrum* (Polygonaceae) species based on nucleotide sequences of noncoding regions in chloroplast DNA. *American Journal of Botany* 87: 573-582.

OHSAKO, T., AND O. OHNISHI. 2001. Nucleotide sequence variation of the chloroplast *trnK/matK* region in two wild *Fagopyrum* (Polygonaceae) species, *F. leptopodum* and *F. statice*. *Genes and Genetic Systems* 76: 39-46.

OHYAMA, K., H. FUKUZAWA, T. KOHCHI, H. SHIRAI, T. SANO, S. SANO, K. UMESONO, Y. SHIKI, M. TAKEUCHI, Z. CHANG, S. AOTA, H. INOKUCHI, AND H. OZEKI. 1986. Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. *Nature* 322: 522-574.

OLMSTEAD, R. G., AND J. D. PALMER. 1994. Chloroplast DNA systematics: a review of methods and data analysis. *American Journal of Botany* 81: 1205-1224.

OLMSTEAD, R. G., AND P. A. REEVES. 1995. Evidence for the polyphyly of the Scrophulariaceae based on chloroplast *rbcL* and *ndhF* sequences. *Annals of the Missouri Botanical Garden* 82: 176-193.

OLMSTEAD, R. G., AND J. A. SWEERE. 1994. Combining data in phylogenetic systematics: an empirical approach using three molecular data sets in the Solanaceae. *Systematic Biology* 43: 467-481.

OLSON, M. E. 2002a. Intergeneric relationships within the Caricaceae-Moringaceae clade (Brassicales) and potential morphological synapomorphies of the clade and its families. *International Journal of Plant Science* 163: 51-65.

OLSON, M. E. 2002b. Combining data from DNA sequences and morphology for a phylogeny of Moringaceae (Brassicales). *Systematic Botany* 27: 55-73.

OXELMAN, B., M. LIDÉN, AND D. BERGLUND. 1997. Chloroplast *rps16* intron phylogeny of the tribe Sileneae (Caryophyllaceae). *Plant Systematics and Evolution* 206: 393-410.

PACAK, A., AND Z. SZWEYKOWSKA-KULINSKA. 2000. Molecular data concerning alloploid character and the origin of chloroplast and mitochondrial genomes in the liverwort species *Pellia borealis*. *Journal of Plant Biotechnology* 2: 101-108.

PALMER, J. D. 1991. Plastid chromosomes: structure and evolution. *Cell Culture and Somatic Cell Genetics of Plants* 7A: 5-53.

PANERO, J. L., AND B. S. CROZIER. 2003. Primers for PCR amplification of Asteraceae chloroplast DNA. *Lundellia* 6: 1-9.

PEDERSEN, N., AND L. HEDENÄS. 2003. Phylogenetic investigations of a well supported clade within the acrocarpous moss family Bryaceae: evidence from seven chloroplast DNA sequences and morphology. *Plant Systematics and Evolution* 240: 115-132.

PERRET, M., A. CHAUTEMS, R. SPICHIGER, G. KITE, AND V. SAVOLAINEN. 2003. Systematics and evolution of tribe Sinningieae (Gesneriaceae): evidence from phylogenetic analysis of six plastid DNA regions and nuclear *ncpGS*. *American Journal of Botany* 90: 445-460.

PFEIL, B. E., C. L. BRUBAKER, L. A. CRAVEN, AND M. D. CRISP. 2002. Phylogeny of *Hibiscus* and the tribe Hibisceae (Malvaceae) using chloroplast DNA sequences of *ndhF* and the *rpl16* intron. *Systematic Botany* 27: 333-350.

PIRES, J. C., AND K. J. SYTSMA. 2002. A phylogenetic evaluation of a biosystematic framework: *Brodiaea* and related petaloid monocots (Themidaceae). *American Journal of Botany* 89: 1342-1359.

PLUNKETT, G. M., D. E. SOLTIS, AND P. S. SOLTIS. 1996. Evolutionary patterns in Apiaceae: inferences based on *matK* sequence data. *Systematic Botany* 21: 477-495.

POPP, M., AND B. OXELMAN. 2001. Inferring the history of the polyploid *Silene aegaea* (Caryophyllaceae) using plastid and homoeologous nuclear DNA sequences. *Molecular Systematics and Evolution* 20: 474-481.

POSNO, M., A. VAN VLIET, AND G. S. GROOT. 1986. The gene for *Spirodela oligorhiza* chloroplast ribosomal protein homologous to *E. coli* ribosomal protein *L16* is split by a large intron near its 5′ end: structure and expression. *Nucleic Acids Research* 14: 3181-3195.

POTTER, D., F. GAO, S. BAGGETT, J. R. MCKENNA, AND G. H. MCGRANAHAN. 2002. Defining the sources of paradox: DNA sequence markers for North American walnut (*Juglans* L.) species and hybrids. *Scientia Horticulturae* 94: 157-170.

PRIDGEON, A. M., R. SOLANO, AND M. W. CHASE. 2001. Phylogenetic relationships in Pleurothallidinae (Orchidaceae): combined evidence from nuclear and plastid DNA sequences. *American Journal of Botany* 88: 2286-2308.

RENNER, S. S. 1999. Circumscription and phylogeny of the Laurales: evidence from molecular and morphological data. *American Journal of Botany* 86: 1301-1315.

RICHARDSON, J. E., M. F. FAY, Q. C. B. CRONK, D. BOWMAN, AND M. W. CHASE. 2000. A phylogenetic analysis of Rhamnaceae using *rbcL* and *trnL-F* sequences. *American Journal of Botany* 87: 1309-1324.

RITLAND, K., AND M. T. CLEGG. 1987. Evolutionary analysis of plant DNA sequences. *American Naturalist* 130: S74-S100.

SAKAI, M., A. KANAZAWA, A. FUJII, F. S. THSENG, J. ABE, AND Y. SHIMAMOTO. 2003. Phylogenetic relationships of the chloroplast genomes in the genus *Glycine* inferred from four intergenic spacer sequences. *Plant Systematics and Evolution* 239: 29-54.

SALAZAR, G. A., M. W. CHASE, M. A. S. ARENAS, AND M. INGROUILLE. 2003. Phylogenetics of Cranichideae with emphasis on Spiranthinae (Orchidaceae, Orchidoideae): evidence from plastid and nuclear DNA sequences. *American Journal of Botany* 90: 777-795.

SALTONSTALL, K. 2001. A set of primers for amplification of noncoding regions of chloroplast DNA in the grasses. *Molecular Ecology Notes* 1: 76-78.

SAMUEL, R., T. F. STUESSY, K. TREMETSBERGER, C. M. BAEZA, AND S. SILJAK-YAKOVLEV. 2003. Phylogenetic relationships among species of *Hypochaeris* (Asteraceae, Cichorieae) based on ITS, plastid *trnL* intron, *trnL-F* spacer, and *matK* sequences. *American Journal of Botany* 90: 496-507.

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

SCHNABEL, A. AND J. F. WENDEL. 1998. Cladistic biogeography of *Gleditsia* (Leguminosae) based on *ndhF* and *rpl16* chloroplast gene sequences. *American Journal of Botany* 85: 1753-1765.

SCHÖNENBERGER, J., AND E. CONTI. 2003. Molecular phylogeny and floral evolution of the Penaeaceae, Oliniaceae, Rhynchocalycaceae, and Alzateaceae (Myrtales). *American Journal of Botany* 90: 293-309.

SCHULTHEIS, L. M. 2001. Systematics of *Downingia* (Campanulaceae) based on molecular sequence data: implications for floral and chromosome evolution. *Systematic Botany* 26: 603-621.

SCHÜTZE, P., H. FREITAG, AND K. WEISING. 2003. An integrated molecular and morphological study of the subfamily Suaedoideae Ulbr. (Chenopodiaceae). *Plant Systematics and Evolution* 239: 257-286.

SEELANAN T., C. L. BRUBAKER, J. M. STEWART, L. A. CRAVEN, AND J. F. WENDEL. 1999. Molecular systematics of Australian *Gossypium* section *Grandicalyx* (Malvaceae). *Systematic Botany* 24: 183-208.

SHAW, A. J. 2000. Phylogeny of the Sphagnopsida based on chloroplast and nuclear DNA sequences. *Bryologist* 103: 277-306.

SHAW, J., AND R. L. SMALL. 2004. Addressing the "hardest puzzle in American pomology:" Phylogeny of *Prunus* section *Prunocerasus* (Rosaceae) based on seven noncoding chloroplast DNA regions. *American Journal of Botany* 91: 985-996.

SIMPSON, D. A., C. A. FURNESS, T. R. HODKINSON, A. M. MUASYA, AND M. W. CHASE. 2003. Phylogenetic relationships in Cyperaceae subfamily Mapanioideae inferred from pollen and plastid DNA sequence data. *American Journal of Botany* 90: 1071-1086.

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 phylogenetic reconstruction in a recently diverged plant group. *American Journal of Botany* 85: 1301-1315.

SMITH, A. R., AND R. B. CRANFILL. 2002. Intrafamilial relationships of the thelypteroid ferns (Thelypteridaceae). *American Fern Journal* 92: 131-149.

SOLTIS, D. E., R. K. KUZOFF, M. E. MORT, M. ZANIS, M. FISHBEIN, L. HUFFORD, J. KOONTZ, AND M. K. ORROYO. 2001. Elucidating deep-level phylogenetic relationships in Saxifragaceae using sequences for six chloroplastic and nuclear DNA regions. *Annals of the Missouri Botanical Garden* 88: 669-693.

STECH, M., D. QUANDT, AND W. FREY. 2003. Molecular circumscription of the hornworts (Anthocerotophyta) based on the chloroplast DNA *trnL-trnF* region. *Journal of Plant Research* 116: 389-398.

STEELE, K. P., AND R. VILGALYS. 1994. Phylogenetic analyses of Polemoniaceae using nucleotide sequences of the plastid gene *matK*. *Systematic Botany* 19: 126-142.

STEFANOVIC, S., L. KRUEGER, AND R. G. OLMSTEAD. 2002. Monophyly of the Convolvulaceae and circumscription of their major lineages based on DNA sequences of multiple chloroplast loci. *American Journal of Botany* 89: 1510-1522.

STEHLIK, I. 2002. Glacial history of the Alpine herb *Rumex nivalis* (Polygonaceae): a comparison of common phylogeographic methods with nested clade analysis. *American Journal of Botany* 89: 2007-2016.

STEHLIK, I., F. R. BLATTNER, R. HOLDREGGER, AND K. BACHMANN. 2002. Nunatak survival of the high Alpine plant *Eritrichium nanum* (L.) Gaudin in the central Alps during the ice ages. *Molecular Ecology* 11: 2027-2036.

TABERLET, P., L. GIELLY, G. PAUTOU, AND J. BOUVET. 1991. Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology* 17: 1105-1109.

TATE, J. A., AND B. B. SIMPSON. 2003. Paraphyly of *Tarasa* (Malvaceae) and diverse origins of the polyploid species. *Systematic Botany* 28: 723-737.

THOMPSON, J. D., D. G. HIGGINS, AND T. J. GIBSON. 2001. ClustalX. Program available at ftp://ftp-igbmc.u-strasbg.fr/pub/clustalx/.

TSUDZUKI, J., K. NAKASHIMA, T. TSUDZUKI, J. HIRATSUKA, M. SHIBATA, T. WAKASUGI, AND M. SUGIURA. 1992. Chloroplast DNA of Black Pine retains a residual inverted repeat lacking rRNA genes: nucleotide sequences of *trnQ, trnK, psbA, trnI,* and *trnH* and the absence of *rpS16*. *Molecular and General Genetics* 232: 206-214.

VINNERSTEN, A., AND G. REEVES. 2003. Phylogenetic relationships within Colchicaceae. *American Journal of Botany* 90: 1455-1462.

WAKASUGI, T., J. TSUDZUKI, S. ITO, K. NAKASHIMA, T. TSUDZUKI, AND M. SUGIURA. 1994. Loss of all *ndh* genes as determined by sequencing the entire chloroplast genome of the black pine *Pinus thunbergii*. *Proceedings of the National Academy of Sciences, USA* 91: 9794-9798.

WAKASUGI, T., M. SUGITA, T. TSUDZUKI, AND M. SUGIURA. 1998. Updated gene map of Tobacco chloroplast DNA. *Plant Molecular Biology Reporter* 16: 231-241.

WANG, X. R., Y. TSUMURA, H. YOSHIMARU, K. NAGASAKA, AND A. E. SZMIDT. 1999. Phylogenetic relationships of Eurasian Pines (*Pinus*, Pinaceae) based on chloroplast *rbcL*, *matK*, *rpL20-rpS18* spacer, and *trnV* intron sequences. *American Journal of Botany* 86: 1742-1753.

WANNTORP, L., H. E. WANNTORP, B. OXELMAN, AND M. KALLERSJÖ. 2001. Phylogeny of *Gunnera*. *Plant Systematics and Evolution* 226: 85-107.

WESTHOF, E. AND F. MICHEL. 1996. The structural domains of group I introns. *In* R. Green and R. Schroeder [eds.], Ribosomal RNA and group I introns, Chapman and Hall, New York, New York, USA.

WINKWORTH, R. C., J. GRAU, A. W. ROBERTSON, AND P. J. LOCKHART. 2002. The origins and evolution of the genus *Myosotis* L. (Boraginaceae). *Molecular Phylogenetics and Evolution* 24: 180-193.

WOLF, P. G. 1997. Evaluation of *atpB* nucleotide sequences for phylogenetic studies of ferns and other pteridophytes. *American Journal of Botany* 84: 1429-1440.

WOLFE, K. H., W. -H. LI, AND P. M. SHARP. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proceedings of the National Academy of Sciences, USA* 84: 9054-9058.

WOLFE, K. H., C. W. MORDEN, AND J. D. PALMER. 1992. Function and evolution of a minimal plastid genome from a non-photosynthetic parasitic plant. *Proceedings of the National Academy of Science, USA* 89: 10648-10652.

XU, D. H., A. J. SAKAI, M. KANAZAWA, A. SHIMAMOTO, AND Y. SHIMAMOTO. 2000. Sequence variation of non-coding regions of chloroplast DNA of soybean and related wild species and its implications for the evolution of different chloroplast haplotypes. *Theoretical and Applied Genetics* 101: 724-732.

YAMANE, K., Y. YASUI, AND O. OHNISHI. 2003. Intraspecific cpDNA variations of diploid and tetraploid perennial buckwheat, *Fagopyrum cymosum* (Polygonaceae). *American Journal of Botany* 90: 339-346.

YANG, Y. –W., P. –Y. TAI, Y. CHEN, AND W. –H. LI. 2002. A study of the phylogeny of *Brassica rapa, B. nigra, Raphanus sativus*, and their related genera using noncoding regions of chloroplast DNA. *Molecular Phylogenetics and Evolution* 23: 268-275.

ZHANG, W. P. 2000. Phylogeny of the grass family (Poaceae) from *rpL16* intron sequence data. *Molecular Phylogenetics and Evolution* 15: 135-146.

ZUBER, D., AND A. WIDMER. 2000. Genetic evidence for host specificity in the hemi-parasitic *Viscum album* L. (Viscaceae). *Molecular Ecology* 9: 1069-1073.

ZURAWSKI, G., AND M. T. CLEGG. 1987. Evolution of higher plant chloroplast DNA-coded genes: implications for structure-function and phylogenetic studies. *Annual Review of Plant Physiology* 38: 391-418.

APPENDIX; PART 2

Table 1. Sources of cpDNA sequences employed in this study. GenBank accession numbers for each region are shown. O.G. = outgroup taxon.

Species	**Source and voucher**
Gymnosperm (Cupressaceae)	
Taxodium distichum var. *imbricarium* (Nutt.) Croom	S. Bacchus TA-CF-TSRE-S2; USA, FL; TENN
Glyptostrobus pensilis (Staunton) K. Koch	USDA National Arboretum 70.0169; Hong Kong, China
O.G. = *Cryptomeria japonica* (L.f.) Don.	E.B. Lickey T253; cultivated; TENN
Magnoliid (Magnoliaceae)	
Magnolia acuminata L.	J.T. Beck 6000; USA, TN; TENN
Magnolia tripetala L.	J.T. Beck 6001; USA, TN; TENN
O.G. = Liriodendron tulipifera L.	J.T. Beck 6002; USA, TN; TENN
Monocot (Trilliaceae)	
Trillium ovatum Pursh	S. Farmer s.n.; USA, OR; TENN
Trillium texanum Buckl.	S. Farmer and Singhurst s.n.; USA, TX; TENN
O.G. = *Pseudotrillium rivale* (S. Wats.) S.B. Farmer	Graham s.n.: cult. from USA, OR; TENN
Caryophyllid (Caryophyllaceae)	
Minuartia cumberlandensis (B.E. Wofford & Kral) McNeill	C.T. Winder s.n.; USA, TN; TENN
Minuartia glabra(Michx.) Mattf.	C.T. Winder s.n.; USA, TN; TENN
O.G. = *Minuartia uniflora*(Walt.) Mattf.	C.T. Winder s.n.; USA, GA; TENN
Eurosid I (Rosaceae)	
Prunus hortulata Bailey	J. Shaw JSh821-017; USA, TN; TENN
Prunus nigra Ait.	J. Shaw JSh979-125; USA, VT; TENN
O.G. = *Prunus virginiana* L.	J Shaw JSh871-040; USA, NH; TENN
Eurosid II (Malvaceae)	
Hibiscus cannabinus L.	R.L. Small s.n.; USA, FL (cultivar); TENN
Hibiscus mechowii Garcke	R.L. Small s.n.; Zambia; TENN
O.G. = *Hibiscus macrophyllus* Roxb.	L. Craven 10202; Indonesia: CANB
Euasterid I (Scrophulariaceae, Solanaceae)	
Gratiola brevifolia Raf.	D. Estes 02513; USA, TN; EKU
Gratiola virginiana L.	D. Estes 04608; USA, TN; TENN
O.G. = *Gratiola neglecta* Torr.	D. Estes 04609; USA, TN; TENN
Solanum americanum Mill.	E.E. Schilling S-543; USA, FL; TENN
Solanum ptychanthum Dunal	E.E. Schilling S-522; USA, TN; TENN
O.G. = *Solanum physalifolium* Rusby	E.E. Schilling S-548; USA, TN; TENN
Euasterid II (Asteraceae)	
Eupatorium rotundifolium L.	C. Fleming TN002; USA, TN; TENN
Eupatorium hyssopifolium L.	K.C. Siripun 02-Eup-157; USA, NC; TENN
O.G. = *Eupatorium capillifolium* (Lamarck) Small	K. C. Siripun 02-Eup-155: USA, NC; TENN
Carphephorus corymbosus(Nutt.) Torr. & A. Gray	E.E. Schilling 2036; USA, GA; TENN
Trilisa paniculata (Willd.) Cass.	J.B. Nelson 21688; USA, SC; USCH
O.G. = *Eupatorium capillifolium* (Lamarck) Small	K.C. Siripun 02-Eup-155: USA, NC; TENN

Table 1. continued.

trnH-psbA	psbA-3'trnK	3'trnK-matK	matK-5'trnK	rpS16	trnS-trnG	trnG
AY727188	AY727078	-	-	AY727428	AY727521	AY727521
AY727190	AY727076	-	-	AY727429	AY727520	AY727520
AY727189	AY727077	-	-	-	AY727519	AY727519
AY727183	AY727069	AY727069	AY727333	AY727447	AY727516	AY727516
AY727184	AY727068	AY727068	AY727338	AY727436	AY727517	AY727517
AY727182	AY727070	AY727070	AY727340	AY727440	AY727518	AY727518
AY727187	AY727059	AY727059	AY727336	AY727437	-	AY727225
AY727186	AY727060	AY727060	AY727335	AY727448	-	AY727224
AY727185	AY727061	AY727061	AY727334	AY727430	-	AY727223
AY727177	AY727065	AY727065	AY727332	AY727452	AY727514	AY727514
AY727176	AY727066	AY727066	AY727337	AY727451	AY727515	AY727515
AY727178	AY727067	AY727067	AY727339	AY727453	AY727513	AY727513
AY500600	AY727082	AY727050	AY727329	AY500686	AY500705	AY500705
AY500605	AY727083	AY727051	AY727330	AY500691	AY500710	AY500710
AY500634	AY727084	AY727052	AY727331	AY727450	AY500739	AY500739
AY727166	AY727062	AY727062	AY727327	AY727444	AY727504	AY727504
AY727167	AY727064	AY727064	AY727326	AY727435	AY727506	AY727506
AY727165	AY727063	AY727063	AY727328	AY727442	AY727505	AY727505
AY727170	AY727081	AY727054	AY727346	AY727441	AY727510	AY727510
AY727169	AY727080	AY727053	AY727345	AY727445	AY727512	AY727512
AY727168	AY727079	AY727055	AY727344	AY727433	AY727511	AY727511
AY727179	AY727058	AY727058	AY727341	AY727438	AY727508	AY727508
AY727181	AY727056	AY727056	AY727342	AY727431	AY727507	AY727507
AY727180	AY727057	AY727057	AY727343	AY727449	AY727509	AY727509
AY727173	AY727073	AY727073	AY727349	AY727446	-	AY727222
AY727172	AY727071	AY727071	AY727351	AY727434	-	AY727219
AY727175	AY727072	AY727072	AY727350	AY727432	-	AY727218
AY727174	AY727074	AY727074	AY727347	AY727443	-	AY727220
AY727171	AY727075	AY727075	AY727348	AY727439	-	AY727221
AY727175	AY727072	AY727072	AY727350	AY727432	-	AY727218

Table 1. continued.

rpoB-trnC	trnC-ycf6	ycf6-psbM	psbM-trnD	trnD-trnT	trnS-trnfM	trnS-rpS4
-	AY727131	AY727131	AY727095	-	AY727490	AY727243
-	AY727130	AY727130	AY727094	-	AY727492	AY727242
-	AY727132	AY727132	AY727096	-	AY727491	AY727244
AY727422	AY727149	AY727291	AY727107	AY727455	AY727484	AY727249
AY727423	AY727148	AY727290	AY727110	AY727454	AY727486	AY727248
AY727424	AY727147	AY727289	AY727106	AY727456	AY727485	AY727250
AY727421	AY727137	AY727288	AY727118	AY727466	AY727479	AY727253
AY727420	AY727136	AY727286	AY727120	AY727467	AY727480	AY727252
AY727419	AY727138	AY727287	AY727119	AY727468	AY727478	AY727251
AY727418	AY727127	AY727127	AY727091	AY727460	AY727487	AY727266
AY727417	AY727129	AY727129	AY727093	AY727462	AY727489	AY727267
AY727416	AY727128	AY727128	AY727092	AY727461	AY727488	AY727265
AY727413	AY727141	AY727294	AY727109	AY727464	AY727481	AY727269
AY727414	AY727142	AY727293	AY727111	AY727463	AY727482	AY727270
AY727415	AY727143	AY727292	AY727108	AY727465	AY727483	AY727268
AY727410	AY727134	AY727283	AY727114	AY727470	AY727475	AY727245
AY727412	AY727133	AY727284	AY727113	AY727471	AY727477	AY727246
AY727411	AY727135	AY727285	AY727112	AY727469	AY727476	AY727247
AY727426	AY727145	AY727297	AY727116	AY727474	AY727498	AY727255
AY727427	AY727144	AY727298	AY727117	AY727473	AY727496	AY727254
AY727425	AY727146	AY727299	AY727115	AY727472	AY727497	AY727256
AY727408	AY727122	AY727122	AY727086	AY727459	AY727494	AY727262
AY727407	AY727123	AY727123	AY727087	AY727457	AY727495	AY727263
AY727409	AY727121	AY727121	AY727085	AY727458	AY727493	AY727264
-	AY727126	AY727126	AY727090	-	AY727503	AY727260
-	AY727125	AY727125	AY727089	-	AY727501	AY727261
-	AY727124	AY727124	AY727088	-	AY727502	AY727259
-	AY727140	AY727295	-	-	AY727500	AY727257
-	AY727139	AY727296	-	-	AY727499	AY727258
-	AY727124	AY727124	AY727088	-	AY727502	AY727259

Table 1. continued.

rpS4-trnT	trnT-trnL	trnL	trnL-trnF	5'rpS12-rpL20	psbB-psbH	rpL16
AY727151	-	AY727215	AY727215	AY727312	AY727380	AY727401
AY727150	-	AY727217	AY727217	AY727323	AY727378	AY727402
AY727152	-	AY727216	AY727216	AY727325	AY727379	AY727403
AY727154	AY727277	AY727197	AY727230	AY727314	AY727371	AY727381
AY727155	AY727271	AY727195	AY727231	AY727308	AY727372	AY727382
AY727153	AY727272	AY727196	AY727229	AY727306	AY727370	AY727383
AY727160	AY727275	AY727192	AY727233	-	AY727361	AY727405
AY727161	AY727273	AY727191	AY727234	-	AY727362	AY727406
AY727159	AY727274	AY727193	AY727232	-	AY727363	AY727404
AY727097	AY727097	AY727204	AY727240	AY727321	AY727368	AY727399
AY727101	AY727101	AY727206	AY727239	AY727301	AY727369	AY727398
AY727100	AY727100	AY727205	AY727241	AY727313	AY727367	AY727400
AY727103	AY727103	AY500748	AY500767	AY727317	AY727359	AY500643
AY727105	AY727105	AY500753	AY500772	AY727300	AY727360	AY500648
AY727102	AY727102	AY727194	AY727235	AY727303	AY727358	AY500677
AY727162	AY727278	AY727198	AY727227	AY727310	AY727355	AY727397
AY727163	AY727276	AY727200	AY727228	AY727316	AY727357	AY727395
AY727164	AY727279	AY727199	AY727226	AY727319	AY727356	AY727396
AY727156	AY727280	AY727201	AY727237	AY727305	AY727366	AY727384
AY727157	AY727281	AY727202	AY727236	AY727302	AY727364	AY727385
AY727158	AY727282	AY727203	AY727238	AY727320	AY727365	AY727386
AY727097	AY727097	AY727208	AY727208	AY727322	AY727352	AY727387
AY727099	AY727099	AY727209	AY727209	AY727309	AY727353	AY727389
AY727104	AY727104	AY727207	AY727207	AY727324	AY727354	AY727388
AY736007	AY736007	AY727210	AY727210	AY727311	AY727375	AY727393
AY736008	AY736008	AY727211	AY727211	AY727315	AY727374	AY727392
AY736009	AY736009	AY727212	AY727212	AY727304	AY727373	AY727394
AY736010	AY736010	AY727214	AY727214	AY727307	AY727376	AY727390
AY736011	AY736011	AY727213	AY727213	AY727318	AY727377	AY727391
AY736009	AY736009	AY727212	AY727212	AY727304	AY727373	AY727394

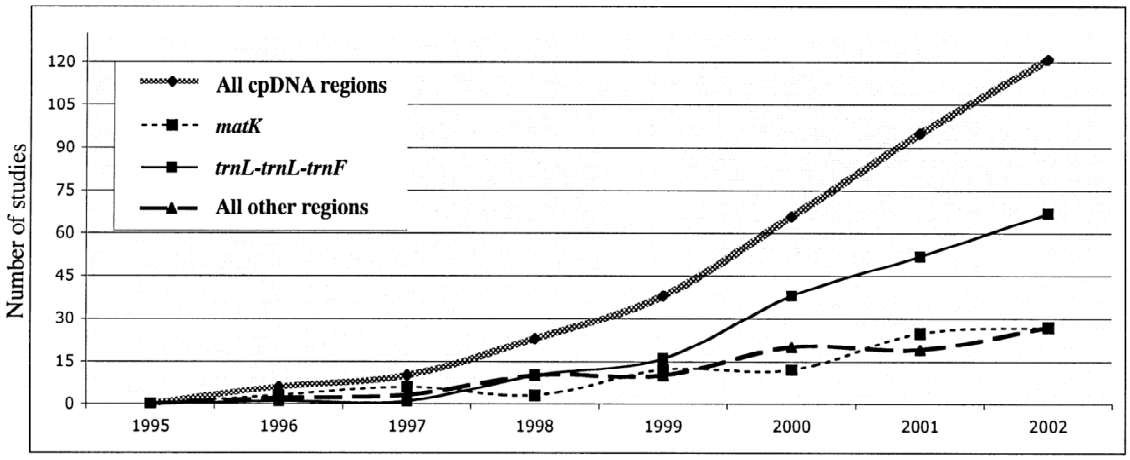


Fig. 1. Phylogenetic studies using cpDNA *matK* and noncoding regions published in *American Journal of Botany*, *Systematic Botany*, *Molecular Phylogenetics and Evolution*, and *Plant Systematics and Evolution* from 1995 through 2002. Lines and symbols represent the number of studies appearing each year (not cumulative). Solid gray line with diamonds = total number of papers published for that year, solid black line with square symbols = studies using some part or all of *trnL-trnL-trnF*, dashed thin line with square symbols = those using some part or all of *trnK-matK-trnK*, and dashed thick line with triangles = studies using regions other than *trnL-trnL-trnF* and *trnK-matK-trnK*.

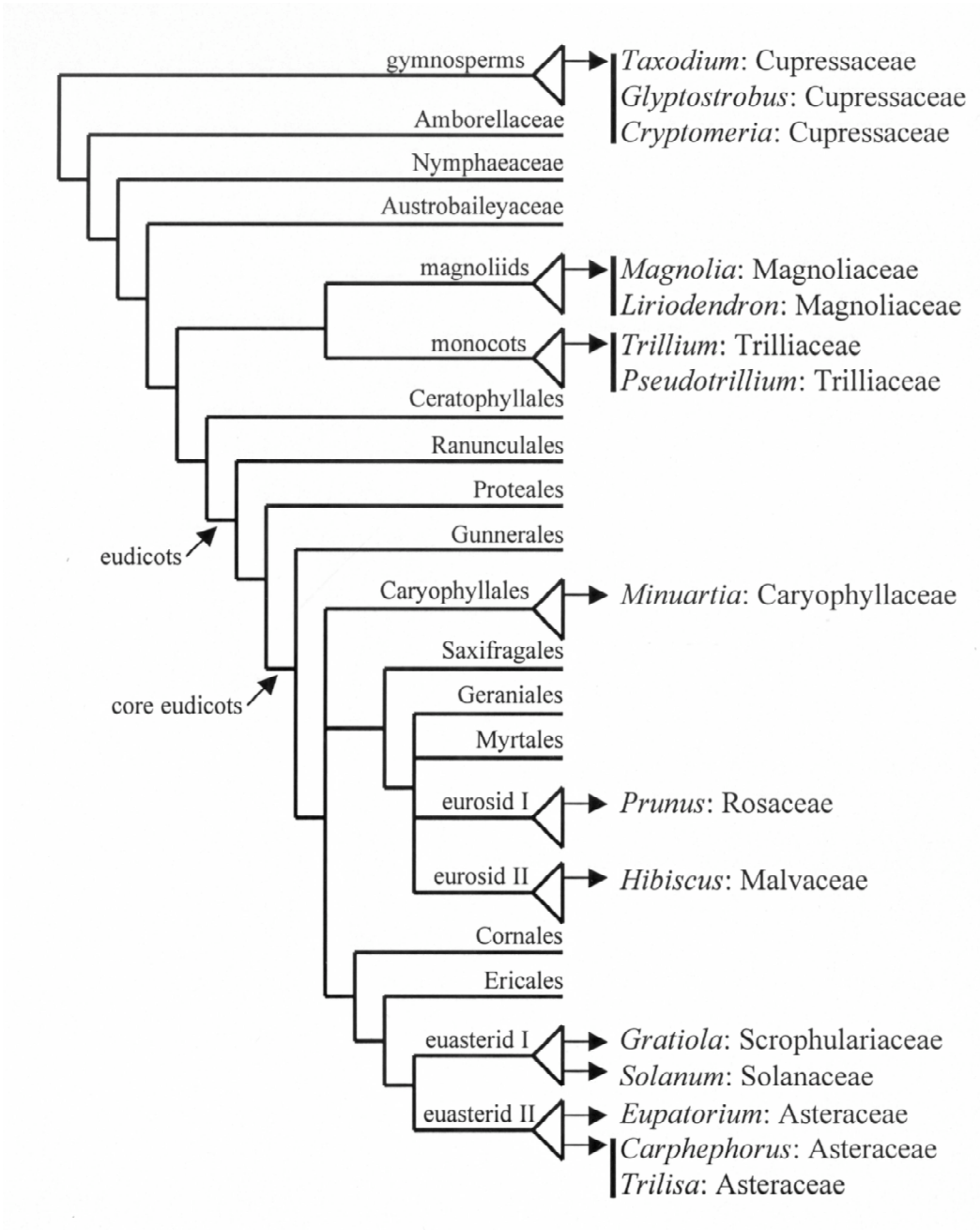
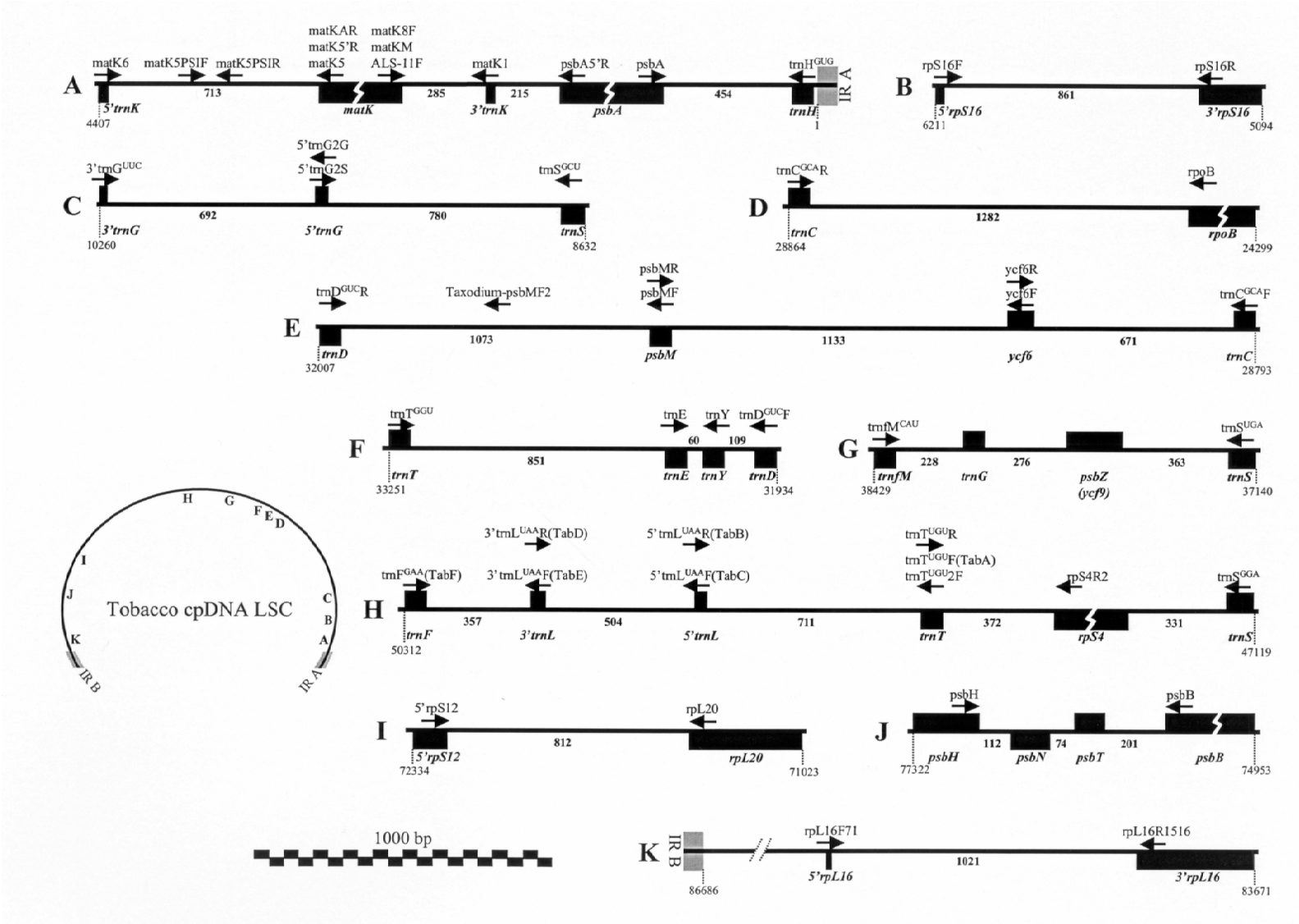


Fig. 2. Simplified phylogenetic representation, modified from APG II (2003), of the 10 lineages used in the survey of 21 noncoding cpDNA regions.

Fig. 3. Scaled map of the 21 noncoding cpDNA regions surveyed in this investigation (based on the *Nicotiana* chloroplast genome [Wakasugi et al., 1998]). The orientation and relative positions of the genes are identified (A - K) along the Large Single Copy (LSC) portion with specific positions denoted by offset numbers at the beginning and end of each region. Gene names are italicized below and amplification and sequencing primer names are in roman typeface above with directional arrows. Lengths of noncoding regions are centered below each intergenic spacer and intron.


A
matK6
matK5PSIF
matK5PSIR
matKAR
matK5'R
matK5
matK8F
matKM
ALS-11F
matK1
psbA5'R
psbA
trnH^GUG
IR A
713
285
215
454
5'trnK
4407
matK
3'trnK
psbA
trnH
1
B
rpS16F
rpS16R
861
5'rpS16
6211
3'rpS16
5094
C
3'trnG^UUC
5'trnG2G
5'trnG2S
trnS^GCU
692
780
3'trnG
10260
5'trnG
trnS
8632
D
trnC^GCAR
rpoB
1282
trnC
28864
rpoB
24299
E
trnD^GUCR
Taxodium-psbMF2
psbMR
psbMF
ycf6R
ycf6F
trnC^GCAF
1073
1133
671
trnD
32007
psbM
ycf6
trnC
28793
F
trnT^GGU
trnE
trnY
trnD^GUCF
60
109
851
trnT
33251
trnE
trnY
trnD
31934
G
trnfM^CAU
trnS^UGA
228
276
363
trnfM
38429
trnG
psbZ
(ycf9)
trnS
37140
H
tmF^GAA(TabF)
3'trnL^UAAR(TabD)
3'trnL^UAAF(TabE)
5'trnL^UAAR(TabB)
5'trnL^UAAF(TabC)
trnT^UGUR
trnT^UGUF(TabA)
trnT^UGU2F
rpS4R2
trnS^GGA
357
504
711
372
331
trnF
50312
3'trnL
5'trnL
trnT
rpS4
trnS
47119
Tobacco cpDNA LSC
H
G
F
E
D
I
J
C
B
A
IR B
IR A
I
5'rpS12
rpL20
812
5'rpS12
72334
rpL20
71023
J
psbH
psbB
112
74
201
psbH
77322
psbN
psbT
psbB
74953
1000 bp
K
IR B
rpL16F71
rpL16R1516
1021
5'rpL16
86686
3'rpL16
83671

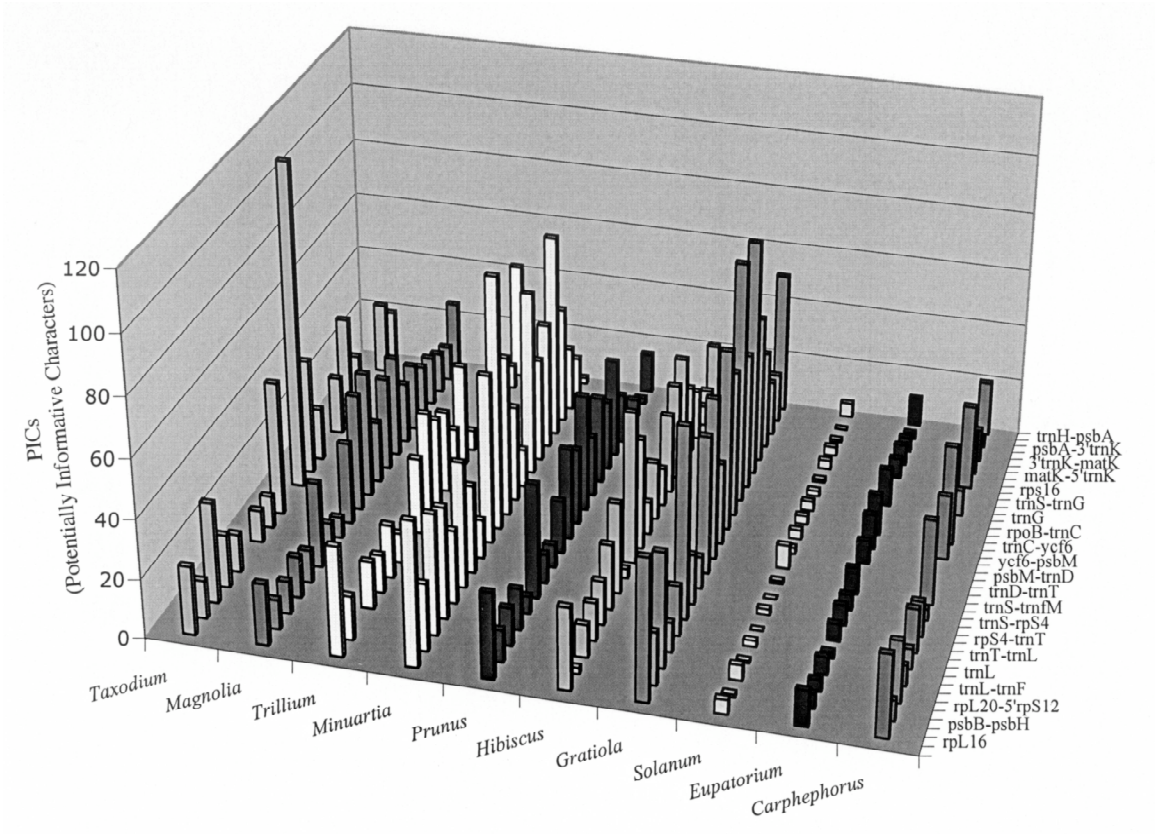


Fig. 4. Representation of the PIC (potentially informative character) values among the noncoding regions (*z*-axis) across taxonomic groups (*x*-axis). These data, summarized from Table 2, indicate that a similar trend exists among the taxonomic groups in the number of PICs provided by each region

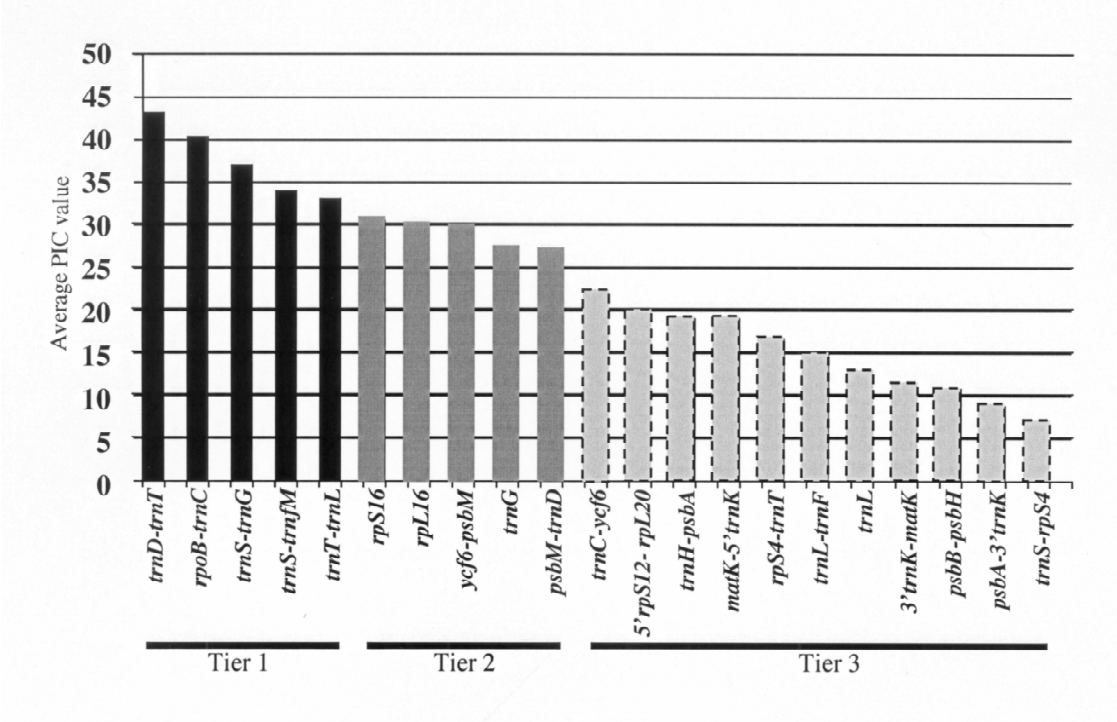


Fig. 5. The average PIC (potentially informative character) value of the six lineages with complete data sets for each region. The 21 regions, oriented in order of most to least number of PICs, are grouped into three tiers based on their qualitative value. The five Tier 1 regions are shown with black bars, the five Tier 2 regions with dark-gray bars, and the Tier 3 regions with light-gray bars with dashed outlines.

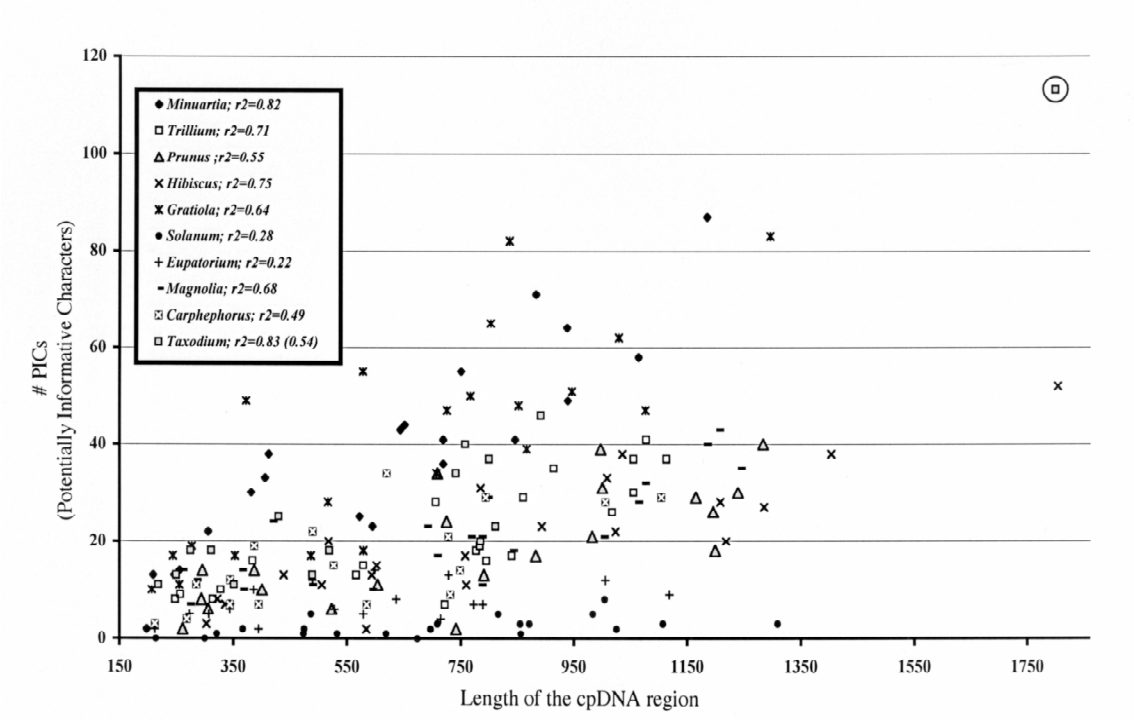


Fig. 6. Scatterplot showing the relationship of region length and its PIC value for each lineage. A key to the symbols representing each lineage is provided, including the r^2 value for each lineage. Because of space constraints, regression lines are not included. An apparent outlier in the *Taxodium* data set is indicated by a circle (upper right). The exclusion of this outlier decreases its r^2 value as shown in parentheses.

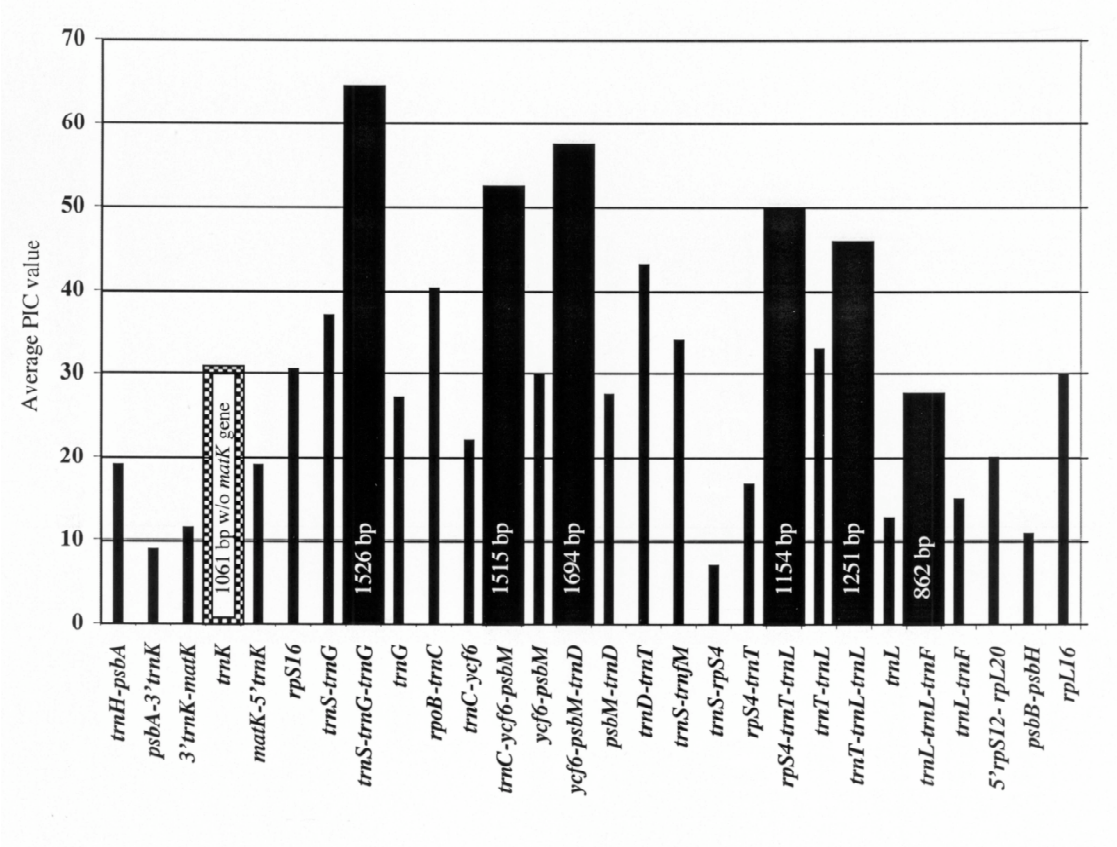


Fig. 7. The average PIC (potentially informative character) value of six lineages with complete data sets for each region compared to easily coamplifiable regions. The 21 single (narrow bars) and seven combined (thick bars) regions are oriented left to right in order as they appear in *Nicotiana* (Wakasugi et al., 1998). The size of the combined regions is included. Additionally, both halves of the *trnK* intron are shown combined (thick checkered bar) because they can be amplified as a single fragment, and each end of the intron can be sequenced completely with one primer each.

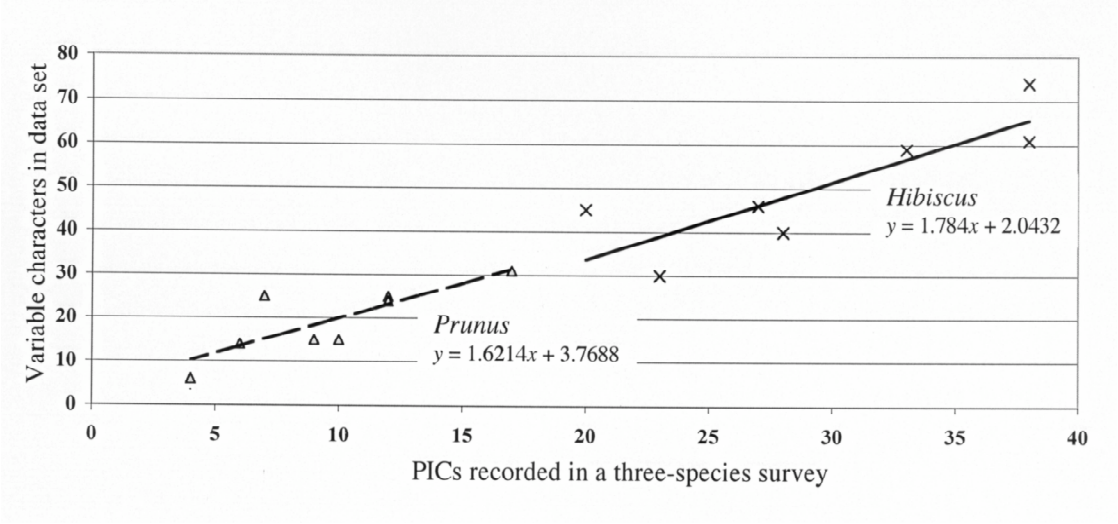


Fig. 8. Comparison of the predictability of a three-species survey with complete data sets used to generate phylogenetic hypotheses for *Prunus* (Shaw and Small, 2004) and *Hibiscus*. Regression lines are shown with their respective equations.

Part 3

ADDRESSING THE "HARDEST PUZZLE IN AMERICAN POMOLOGY:" PHYLOGENY OF *PRUNUS* SECTION *PRUNOCERASUS* (ROSACEAE) BASED ON SEVEN NONCODING CHLOROPLAST DNA REGIONS

This chapter is a lightly revised version of a paper by the same name published in the *American Journal of Botany* in 2004 by Joey Shaw and Randall Small:

Shaw, J., and R. L. Small. 2004. Addressing the "hardest puzzle in American pomology:" Phylogeny of *Prunus* section *Prunocerasus* (Rosaceae) based on seven noncoding chloroplast DNA regions. *American Journal of Botany* 91: 985-996.

The use of "we" in this chapter refers to Randall Small and myself. My primary contributions to this chapter include (1) selection of the topic and development of the problem into a work relevant to study, (2) the lab work to generate data, and (3) most of the writing.

ABSTRACT

Prunus subgenus *Prunus* section *Prunocerasus* (Rosaceae) is a North American taxon with 17 commonly recognized taxa. To test the hypothesis of monophyly for the section we sequenced the *trnG* and *rpL16* introns and the *trnH-psbA* and *trnS-trnG* intergenic spacers for at least two representatives of each of the five subgenera in *Prunus*. Additionally we sampled heavily among *Prunus* subgenus *Prunus* sections *Prunus* and *Armeniaca*, and *Prunus* subgenus *Amygdalus*, because these groups are putatively most closely related to *Prunocerasus*. Once monophyly of section *Prunocerasus* was shown we added the sequences of *trnL* and *rpS16* introns and the *trnL-trnF* spacer in an attempt to increase resolution within the section. The species of section *Prunocerasus* showed an initial split with *P. subcordata*, the only species from western North America, sister to the rest of the group. The remaining species fell into three primary clades. Within each of the three primary clades there was little phylogenetic resolution. Lastly, we present evidence that *P. texana*, previously classified in subgenus *Amygdalus*, may be a plum or at least contain a *Prunocerasus* chloroplast. This is the first phylogenetic hypothesis presented for section *Prunocerasus*, and the clades recovered contrast sharply with previously defined groups based on morphological characters.

INTRODUCTION

Prunus L. (Rosaceae) is a morphologically diverse genus of the Northern Hemisphere with about 200 species of small trees or shrubs with pentamerous flowers and endozoochorous monocarpellate drupes (Rehder, 1940; Robertson, 1974). Many species are economically significant, especially those that are used for ornamentals and food (plums/prunes, peaches, apricots, cherries, and almonds). The Eurasian plums of section *Prunus* (see supplementary data accompanying the online version of this article) are those most heavily used in agriculture, but they have proven difficult to grow in many parts of North America because they are not well adapted to drought, frost, and particularly in the eastern and southern states, fungal, bacterial, and viral diseases (Fogle, 1978). Native North American species of section *Prunocerasus* are currently not cultivated on a large scale, although they have been in the past. Species like *P. americana* Marshall, *P. hortulana* L.H.Bailey, *P. angustifolia* Marshall, and *P. subcordata* Benth. were (Waugh, 1899, 1901), and more recently, *P. maritima* Marshall (R. Uva, Cornell University, personal communication; Dunn, 2002; Karp, 2003) and *P.*

angustifolia (Reid and Gast, 1993) are under study in agricultural trials. Additionally, nearly all *Prunus* section *Prunocerasus* species hybridize, many naturally, not only with themselves but also with section *Prunus* and thus the potential for better-adapted agricultural and horticultural varieties may be unlimited (Wight, 1915). Over 2800 varieties of European plums have been cultivated (Fogle, 1978) and as there are several more North American plum species than there are European plum species, the potential for varieties is much greater (Hedrick et al., 1911).

Several species within section *Prunocerasus* are of conservation concern. Of greatest concern is *P. geniculata* R.M.Harper that has a global distribution of only four counties in central Florida and is Federally Endangered; despite being federally protected it is currently being destroyed by suburban development (personal observation). Seven other species in section *Prunocerasus* are legally protected at the state level, as shown in the supplementary data accompanying the online version of this article.

The most widely accepted classification of *Prunus* is that of Rehder (1940). He divided *Prunus* into five subgenera (*Prunus* [= *Prunophora*], *Amygdalus*, *Cerasus*, *Padus*, and *Laurocerasus*; see supplementary data accompanying the online version of this article) and divided subgenus *Prunus* into three sections: *Prunus* = (*Euprunus*) (Eurasian plums), *Prunocerasus* (North American plums), and *Armeniaca* (apricots). Subgenus *Emplectocladus* was not recognized by Rehder (1940), he placed *P. fasciculata* A.Gray in subgenus *Amygdalus*; however, earlier Mason (1913) believed it to be closely related to section *Armeniaca*, and more recently Bortiri et al. (2001) proposed *P. fasciculata* as sister to an extended subgenus *Prunus*.

Among the first investigators of our native plums was Waugh (1899, 1901), who noted their extensive hybridization and intergradation. He divided the North American plums into four informally named series: Americana, Chickasaw, Hortulana, and Maritima. The Americana Series, centered on *P. americana* var. *americana*, graded into *P. nigra* Aiton to the northeast and *P. americana mollis* Torr. & A.Gray (*P. americana* var. *lanata* Sudw.) to the southwest. The Chickasaw Series only included *P. angustifolia* and a variety. The Hortulana Series was described as "a great congeries of hybrids" (Waugh, 1899, p. 233), whose morphological characters "run into each other in the most puzzling and intricate manner" (Waugh, 1899, p. 233). In this series he noted that *P. hortulana* var. *mineri* L.H.Bailey grades into *P. americana* to the north and into *P. hortulana* var. *waylandii* L.H.Bailey to the south and then also into *P. munsoniana* Wight and Hedrick and eventually *P. angustifolia* "so connectedly as to leave not the slightest break between them" (Waugh, 1899, p. 233). Waugh (1901) also thought that *P. hortulana* and its relatives were of hybrid origin (*P. americana* x *P. angustifolia*). Lastly, he noted that members of this series also grade into *P. maritima* Marshall and *P. injucunda* Small (=*P. umbellata* Elliot var. *injucunda* [Small] Sarg.). His Maritima Series included *P. maritima* and *P. gravesii* Small (=*P. maritima* var. *gravesii* [Small] G.J.Anderson). He also put *P. injucunda* into this section and noted that he could not distinguish the type material from *P. maritima*. He believed *P. gracilis* Engelm. and A.Gray to be a good species and thought it to be closely related to *P. injucunda*. Lastly, Waugh (1899) thought *P. gracilis* to be closely related to *P. glandulosa* (Hook.) Torr. & Gray (not *P. glandulosa* Thunb.), which is a synonym of *P. texana* Dietr. In contrast

Wight (1915) placed this taxon in the genus [subgenus] *Amygdalus* and Correll and Johnston (1970) refer to it as Peach Bush.

Wight (1915) provided the most comprehensive treatment within section *Prunocerasus* (hereafter referred to as *Prunocerasus*) and informally recognized six group*s*: Americana, Angustifolia, Gracilis, Hortulana, Maritima, and Subcordata (see supplementary data accompanying the online version of this article). Similar to Waugh (1899), he placed *P. americana*, *P. nigra*, and *P. mexicana* S.Watson together in the Americana group. He also agreed with Waugh (1899) that *P. angustifolia* and *P. munsoniana* are closely related and placed them together in the Angustifolia group. Again not contradicting Waugh (1899), Wight (1915) saw a relationship between *P. umbellata* var. *umbellata* Elliot, *P. umbellata* var. *injucunda*, and *P. maritima*, but in contrast to Waugh (1899), Wight (1915) placed *P. gracilis* in its own group and never mentioned *P. texana*. He also placed *P. subcordata* in its own group. *Prunus geniculata* was not included in his work because it was described around the same time as Wight's (1915) work.

Morphological taxonomy has long been difficult within *Prunocerasus* because species boundaries are blurred by interspecific similarities and intraspecific variation, and likely by interspecific hybridization (Waugh, 1899, 1901; Hedrick et al., 1911; Wight, 1915; Rehder, 1940). There are 17 commonly recognized taxa in *Prunocerasus* (see supplementary data accompanying the online version of this article); several of which are only differentiated by a few continuous (and often highly variable) characters that different authors may or may not have considered important. This has resulted in a lot of taxonomic confusion. For example, Fernald (1950, p. 877) described *P. americana* var. *americana* as "passing insensibly into *P. americana* var. *lanata*" (which was described as *P. lanata* Sudw.) whereas Gleason and Cronquist (1991) placed *P. americana* var. *lanata* as a synonym of *P. mexicana* (characterized by densely pubescent leaves). Robertson (1974, p. 659) noted that "*P. americana* and *P. mexicana* need to be studied more to ascertain their distinctiveness and distribution." Additionally, *P. umbellata* has been variously subdivided into *P. injucunda*, which has pubescent leaves and twigs, and *P. mitis*, which has glabrous twigs (Radford et al., 1968); both of these putative species differ from *P. umbellata* which is completely glabrous. Sargent (1902) saw *P. tarda* Sarg. as a distinct species, but Wight (1915, p. 54) noted "*P. umbellata* ssp. *injucunda* merges imperceptibly into *P. umbellata* ssp. *tarda*." Duncan and Duncan (1988) believed that *P. umbellata* only differs from *P. alleghaniensis* Porter in distribution. To add to the confusion, Waugh (1899, p. 234), after studying the type material of *P. injucunda*, noted it as "easily being referable to *P. maritima*." Other examples of convoluted taxonomic disagreement abound and excessive "splitting" within this group was noted as early as 1915 (Wight, 1915). Hedrick et al. (1911, p. 5) went so far as to say that *Prunocerasus* is "plastic in all physical characters" and Bailey (1892, p. 90) referred to *Prunocerasus* as "the hardest puzzle in American pomology." Because variable morphological characters have resulted in ambiguous and conflicting species delimitation and classification, many taxa within the section have been variously accepted, ranked, lumped, and split. This has not only created a large amount of taxonomic confusion, but bewildering nomenclatural synonymy. Clearly, morphological data alone are unable to unequivocally define species boundaries and relationships within *Prunocerasus*.

Several authors have employed molecular tools in addressing broader phylogenetic questions in *Prunus* and all of them have included some representatives from *Prunocerasus* (Mowrey and Werner, 1990; Bortiri et al., 2001; Lee and Wen, 2001; Bortiri et al., 2002). Because of what appears to be relatively low divergence within *Prunus*, these studies are only taxonomically suggestive and lack the support or resolution upon which to confidently base a hypothesis of monophyly for *Prunocerasus*. Additionally, some of these earlier studies have suggested, although with weak support, that *Prunocerasus* may not be monophyletic (Mowrey and Werner, 1990; Lee and Wen, 2001; Bortiri et al., 2002).

The earliest work to employ molecular tools (isozymes) suggested that *Prunocerasus* is polyphyletic with *P. americana*, *P. munsoniana*, *P. hortulana*, *P. subcordata*, and *P. angustifolia* belonging to a clade and *P. maritima* and *P. umbellata* belonging to another clade, with the latter clade more closely related to section *Prunus* and subgenus *Cerasus* than to *Prunocerasus* (Mowrey and Werner, 1990). Based on these relationships, Mowrey and Werner (1990) suggested that there were at least two immigrations of plums into North America. More recently, Lee and Wen (2001) included four species of *Prunocerasus* in a nuclear ribosomal internal transcribed spacer (ITS) analysis of *Prunus*. Their data had weak resolution in this part of the tree and were unable to resolve *P. armeniaca* L. of section *Armeniaca* from representatives of *Prunocerasus*. Bortiri et al. (2001) combined data from the cpDNA *trnL-trnF* spacer and ITS sequences to address phylogenetic relationships within *Prunus*. Their data support monophyly of some *eastern* North American *Prunocerasus* representatives, but *P. subcordata*, of northwest North America, was not resolved from species outside *Prunocerasus*. Finally, Bortiri et al. (2002), in a second broad analysis of *Prunus* using the single copy nuclear gene *s6pdh*, showed that *P. subcordata* is sister to other sampled *Prunocerasus* species, and that *P. armeniaca* may be closely related to *Prunocerasus* (in agreement with Lee and Wen, 2001). While these studies conditionally support a close relationship among *Prunocerasus* species, none of them unequivocally support or refute the monophyly of *Prunocerasus* and none of them amply sample within *Prunocerasus* to allow relationships within the section to be inferred.

In summary, *Prunus* subgenus *Prunus* section *Prunocerasus* is a valuable, somewhat vulnerable, and complicated group of organisms that are "misunderstood." Understanding the classification and evolutionary history of the North American plums is of economic importance as well as conservation and taxonomic interest. To date no rigorous phylogenetic analysis addressing the monophyly of, or relationships among, species within *Prunocerasus* has been conducted. Given the lack of resolution in the aforementioned foundational phylogenetic studies, the longstanding taxonomic difficulty of this group, and the potential applicability of known evolutionary relationships to agriculture, horticulture, and conservation, the goals of this research were to (1) test the monophyly of *Prunus* subg *Prunus* section *Prunocerasus* and its relationship to other sections of subgenus *Prunus*, and (2) to produce a phylogenetic hypothesis of the relationships within *Prunocerasus* based on cpDNA sequences.

MATERIALS AND METHODS

***Plant material and outgroup choice*—**We sampled 17 species and subspecific taxa of *Prunocerasus* that represented all of the commonly accepted taxa within the following works: Small (1933), Rehder (1940), Fernald (1950), Bailey and Bailey (1941), Blackburn (1952), Gleason (1952), Steyermark (1963), Radford et al. (1968), Correll and Johnston (1970), Duncan and Duncan (1988), Godfrey (1988), Wunderlin (1988), Gleason and Cronquist (1991), Smith (1994), and Wofford and Chester (2002). Samples of the federally endangered *P. geniculata* were obtained from scientists at the Archbold Biological Station in Lake Placid, Florida and two collections of *P. subcordata* from California were provided by colleagues at Oregon State University. Additionally, we sampled several representative species from each of the other two sections in subgenus *Prunus* along with representatives from each of the other four subgenera: *Amygdalus*, *Cerasus*, *Padus*, and *Laurocerasus* (see supplementary data accompanying the online version of this article). Because none of the taxa within *Prunus* has unequivocally been shown to be sister to the rest of the genus (Mowrey and Werner, 1990; Bortiri et al., 2001, 2002; Lee and Wen, 2001), *Physocarpus opulifolius* (L.) Maxim. was used as an outgroup (see Bortiri et al., 2001). Ingroup sampling of *Prunocerasus* was from wild-collected populations except for one case of a herbarium specimen, and outgroup samples were both from wild-collected populations and from the *Prunus* germplasm facility at the University of California, Davis, USA (DPRU).

***DNA sequences*—**For all 43 taxa in this investigation the *rpL16* and $trnG^{UUC}$ introns as well as the $trnS^{GCU}$-$trnG^{UUC}$ and $trnH^{GUG}$-*psbA* intergenic spacers were polymerase chain reaction (PCR) amplified and sequenced. These regions were chosen because they were shown to be highly informative noncoding cpDNA regions (Shaw et al., 2005). For the ingroup analyses, sequence data were obtained from seven separate chloroplast loci including four introns: $trnL^{UAA}$, $trnG^{UUC}$, *rpS16*, *rpL16* and three intergenic spacers: $trnL^{UAA}$-$trnF^{GAA}$, $trnH^{GUG}$-*psbA*, and $trnS^{GCU}$-$trnG^{UUC}$. The *trnL* and *rpS16* introns along with the *trnL-trnF* intergenic spacer were amplified and sequenced for the seventeen ingroup taxa plus *P. mahaleb* L. before Shaw et al. (2005) showed that they are comparatively less variable than other noncoding cpDNA choices. Primer pairs for the aforementioned regions are listed in Table 1 [all tables and figures are located in the Appendix following the Literature Cited]; primers for the *trnL* intron and the *trnL-trnF* intergenic spacer are from Taberlet et al. (1991), primers for *rpL16* are from Small et al. (1998), and the remaining primers in Table 1 are from Shaw et al. (2005).

***Laboratory procedures*—**DNA was extracted from leaves using the DNeasy Plant Mini Kit (Qiagen, Valencia, California, USA). PCR was performed using Eppendorf Mastercycler gradient or Mastercycler personal thermal cyclers in 50 µL volumes with the following reaction components: 1 µL template DNA (~ 10-100 ng), 1X *ExTaq* buffer (PanVera / TaKaRa, Madison, Wisconsin, USA), 200 µmol/L each dNTP, 3.0 mmol/L $MgCl_2$, 0.1 mmol/L each primer, and 1.25 units *ExTaq* (PanVera/TaKaRa). Reactions included bovine serum albumin at a final concentration of 0.2 µg/µL, which improved amplification from difficult templates. For amplification of the *trnS-trnG-trnG* region, 10 mmol/L of TMACl (tetramethyl ammonium chloride) was added to the reaction

mixture because it is thought to help PCR through polynucleotide runs (see Oxelman et al., 1997) and Shaw et al. (University of Tennessee, unpublished data) have shown that a poly-A/T run exists ~150 base pairs (bp) upstream from the 3′ *trnG* exon across all lineages of phanerogams.

PCR and sequencing primers are given in Table 1 and all PCR protocols described below were preceded by template DNA denaturation at 80°C for 5 min. PCR cycling conditions for *rpL16*: 30 cycles of denaturation at 95°C for 1 min, primer annealing at 50°C for 1 min, followed by a ramp of 0.3°C/s to 65°C, and primer extension at 65°C for 4 min. A final extension step consisted of 5 min at 65°C. PCR cycling conditions for *trnL+trnL-trnF*: 30 cycles of denaturation at 94°C for 1 min, primer annealing at 50°C for 1 min, primer extension at 72°C for 2 min. A final extension step consisted of 5 min at 72°C. PCR cycling conditions for *rpS16*: 30 cycles of denaturation at 94°C for 30 s, primer annealing at 53°C for 30 s, primer extension at 72°C for 2 min. A final extension step consisted of 5 min at 72°C. PCR cycling conditions for *trnH-psbA*: 30 cycles of denaturing at 94°C for 30 s, annealing at 50°C for 30 s, and extension at 72°C for 1 min. PCR cycling conditions for *trnS-trnG-trnG* were by the touchdown method: 15 cycles of denaturation at 96°C for 1 min, primer annealing at 76°C (-0.4°C/cycle) for 45 s, primer extension at 72°C for 2 min and then 30 cycles of denaturation at 96°C for 1 min, primer annealing at 69.5°C for 45 s, primer extension at 72°C for 2 min. A final extension step consisted of 5 min at 72°C.

PCR products were checked on 1% agarose gels before being cleaned with either the QIAquick PCR Purification Kit (Qiagen) or ExoSAP-IT (USB, Cleveland, Ohio, USA). DNA sequencing was performed with the same primers used in amplification (Table 1) with the ABI Prism BigDye Terminator Cycle Sequencing Ready Reaction Kit, v. 2.0 or 3.1 (Perkin-Elmer/Applied Biosystems, Foster City, California, USA), electrophoresed and detected on an ABI Prism 3100 automated sequencer (University of Tennessee Molecular Biology Resource Facility). Sequencher 3.1.1 (Gene Codes, 1998) was used to edit and assemble complementary DNA strands and check for agreement between them. In no cases did the adjoined complementary strands disagree. All sequences have been deposited in GenBank; accession numbers (AY500595-AY500780) are given in the supplementary data accompanying the online version of this article.

***Data analysis and phylogenetic assessment*—**For both data sets, alignment of DNA sequences was initially performed with ClustalX (Thompson et al., 2001), with subsequent manual adjustment by eye in MacClade v. 4.0 (Sinauer). Variable positions in the data matrices were double checked against the original chromatogram files to make sure that all base calls were true at all variable positions. In all cases, alignment of potentially informative positions was unambiguous. Because indels have been shown to provide approximately one-third of the potentially phylogenetically informative information in a cpDNA data set (Shaw et al., University of Tennessee, unpublished data), potentially phylogenetically informative indels were coded as additional binary characters. However, a few indels, such as those within polynucleotide runs, were

omitted from the data set as they may be PCR artifact, and not reflective of the phylogenetic history of the group.

Analysis of phylogenetic relationships was conducted using the optimality criterion of maximum parsimony. Searches for most-parsimonious trees were executed in PAUP* v. 4.0 b10 (Swofford, 2002). Parsimony analyses were carried out by a heuristic search with tree bisection-reconnection (TBR) branch swapping and 1000 random sequence addition replicates. Bootstrap support (Felsenstein, 1985) was estimated with 1000 replications of heuristic search and simple taxon addition with the constraint of 1 000 000 rearrangements per replicate. Both the consistency and retention indices (CI and RI, respectively) were used to assess the amount of homoplasy present in the data set. Additionally, Bayesian analysis was employed as alternative means of phylogenetic assessment. Bayesian analyses of the data were performed using MrBayes 2.01 (Huelsenbeck and Ronquist, 2001) to generate posterior probability distribution using Markov chain Monte Carlo (MCMC) methods. No a priori assumptions about tree topology were made. The statistical model of DNA substitution, Felsenstein 1981 (F81), was that estimated as the best-fitting maximum likelihood model using MrModeltest 1.1b (Nylander, 2002), which is a simplified version of MODELTEST 3.06 (Posada and Crandall, 1998). The Monte Carlo Markov Chain (MCMC) process was set to run one million generations with four chains. Burn-in was estimated visually by plotting log-likelihood values in Microsoft Excel to determine the number of generations that had run before likelihood values reached an asymptote. To calculate the posterior probability of each bipartition a 50% majority-rule consensus tree was constructed from the remaining trees using PAUP*.

RESULTS

***Characterization of the cpDNA data sets*—**For the broad analyses, sequence data were obtained from four different noncoding cpDNA regions: *trnH-psbA*, *rpL16*, *trnG*, and *trnS-trnG* (Table 1). As expected, given the complete linkage among cpDNA regions, data derived from different cpDNA regions were congruent and thus were combined into a single data set.

The combined cpDNA data set for the broad investigation of *Prunus* consisted of 3270 aligned nucleotide positions with 231 variable positions, 134 of which were parsimony informative (including multi-bp indels coded as binary characters). Characteristics of individual cpDNA regions are summarized in Table 1. In the maximum parsimony analysis, the heuristic search found 25 171 most parsimonious trees of 422 steps and high consistency and retention indices, CI = 0.92 and RI = 0.94, which indicates that there was little homoplasy within the data set. Bootstrap values are shown above the branches in Fig. 1. The topology of the tree generated with the Bayesian method (Fig. 2) was consistent with the parsimony tree although in two groups the Bayesian analysis showed more resolution (sections *Prunus* and *Amygdalus*). Support values for equivalent tree branches were consistently higher in the Bayesian analyses than in the parsimony analyses.

Members of section *Microcerasus* contained a relatively high number of homoplasious characters compared to other taxa in this study. Additional Bayesian and bootstrap analyses (not shown) with the same parameters described above were

performed on the broad data set, but excluding members of section *Microcerasus* (*P. pumila* L., *P. glandulosa* Thunb., and *P. tomentosa* Thunb.). Without these species, bootstrap values and posterior probabilities for several of the clades increased. Resolution within the subgenus *Prunus*-subgenus *Amygdalus* clade also increased in the bootstrap analysis as there was 64% bootstrap support for a monophyletic section *Prunus* clade (the same clade shown in the Bayesian analysis in Fig. 2).

Because the results of the broad analysis showed that *Prunocerasus* is a monophyletic assemblage, the *trnL-trnF* intergenic spacer and the *trnL* and *rpS16* introns were added to the aforementioned four cpDNA regions and separate analyses were performed on section *Prunocerasus* + *P. mahaleb* (an outgroup species). The ingroup data set, made up of the seven cpDNA regions consisted of 4375 aligned nucleotide positions, had 88 variable positions, 37 of which were parsimony informative (including multi-bp indels coded as binary characters). Characteristics of individual cpDNA regions are summarized in Table 1. Maximum parsimony analysis found one most parsimonious tree of 128 steps with high consistency and retention indices, CI = 0.98 and RI = 0.98, which indicate a low amount of homoplasy within the data set. The topology of the *Prunocerasus*-only tree is nearly identical to the topology found in the broad analysis and thus is mapped onto the trees generated in the broad analyses. The additional resolution in the American clade provided by this analysis is represented in Figs. 1-3 by dotted lines. Bootstrap values for the *Prunocerasus*-only analysis are marked below the branches in Fig. 1. Similar to the results of the broad analysis, the topology of the tree generated with Bayesian methods was topologically identical to the parsimony tree. Posterior probabilities for the *Prunocerasus*-only analysis are marked below the branches in Fig. 2. Again, support values for all of the tree branches were consistently higher in the Bayesian than in the parsimony analyses.

***Phylogenetic results: broad analyses of Prunus*—**The broad phylogenetic analyses using the four combined cpDNA data sets provided unequivocal support for a monophyletic section *Prunocerasus* with a bootstrap value of 83% and a posterior probability of 1.0. The Bayesian analysis showed section *Prunus* as monophyletic with *P. glandulosa* of subgenus *Cerasus* section *Microcerasus* sister to it with a weak posterior probability of 0.68. The section *Prunus* clade was only supported by two nucleotide substitutions, one of which is likely homoplasious because it disagrees with several other characters in the data set and is possibly the reason why *P. glandulosa* appeared to be sister to the rest of section *Prunus*. Within section *Prunus*, the European-West Asian plums and the Asian plums formed two distinct clades that were supported by the bootstrap analysis; however, in the bootstrap analysis there was no support for these two clades being sister to each other (because of the apparent homoplasy of *P. glandulosa*). Removal of section *Microcerasus* (which includes *P. glandulosa*) from the analysis raised the posterior probability for the section *Prunus* clade from 0.68 to 0.98; the results of the minus-*Microcerasus* bootstrap analysis, which before revealed two separate clades that were unresolved from the backbone of the Pruno-Amygdaloid clade (subgenera *Prunus* and *Amygdalus*), reveals a single section *Prunus* clade supported by a bootstrap value of 64%. The Bayesian analysis showed support, although weak, for a monophyletic subgenus *Amygdalus* as well as for the distinctiveness of the two sections

that comprise it: *Amygdalus* (= *Euamygdalus*) and *Chamaeamygdalus*. As in section *Prunus*, the deeper node was not supported in the bootstrap analysis, leaving only support for a monophyletic section *Amygdalus* (the placement of *P. tenella* Batsch, the section *Chamaeamygdalus* representative was unresolved in the bootstrap analysis).

Both analyses show strong support for a monophyletic Pruno-Amygdaloid clade, with a bootstrap value of 97% and a posterior probability of 1.0. *Prunus fasciculata* of subgenus *Emplectocladus* and subgenus *Cerasus* section *Mahaleb* were consecutively sister to the Pruno-Amygdaloid clade with high support. Subgenus *Cerasus* section *Microcerasus*, *P. pumila*, *P. tomentosa*, and *P. glandulosa*, appeared not be a monophyletic group (see Discussion), but its constituent species were strongly supported within the Pruno-Amygdaloid clade. Lastly, subgenera *Padus* and *Laurocerasus* appeared to be closely related, paraphyletic, and sister to the remaining members of *Prunus*.

***Phylogenetic results: section Prunocerasus*—**Within the *Prunocerasus* clade, the backbone of the tree indicated an initial split in the section between the northwestern species, *P. subcordata*, and the remaining species. *Prunus texana*, classified in subgenus *Amygdalus* (Wight, 1913), was positioned with strong support within section *Prunocerasus* as sister to the remaining species (in a grade between *P. subcordata* and the rest of the species in the section). Within these remaining species the data distinguished three major clades: an "American clade," a "Beach clade," and a "Chickasaw clade" (see Fig. 3). The American clade was supported with bootstrap value of 100% and a posterior probability of 1.0. Within this clade relationships were weakly resolved because of the lack of sequence divergence but showed support for *P. hortulana* and *P. mexicana* as being sister taxa. The Chickasaw clade was highly supported with a bootstrap value of 98% and a posterior probability of 1.0. Within this clade, there was little resolution because the sequences are nearly identical. Of the 4375 total aligned nucleotides used in this study, the sequences of *P. angustifolia*, *P. munsoniana*, *P. umbellata*, and *P. alleghaniensis* var. *davisii* (Wight) Sarg. were identical, while *P. gracilis* and *P. alleghaniensis*, and *P. nigra* each contained a single nucleotide difference. Within this clade however, there was weak support for *P. alleghaniensis* being sister to the rest of the clade. The Beach clade was also highly supported with a bootstrap value of 100% and a posterior probability of 1.0 and there was support for *P. geniculata* being sister to *P. maritima* var. *maritima* plus *P. maritima* var. *gravesii*.

DISCUSSION

***Molecular evolution of the cpDNA regions*—**As phylogenetic studies move to focus on lower-level taxonomic groups, it has become apparent that a multi-locus approach is necessary to obtain a sufficient number of phylogenetically informative characters, especially when using the relatively slowly evolving chloroplast genome. Many recent investigations have used several noncoding cpDNA regions to obtain sufficient characters for phylogenetic resolution (e.g., Kusumi et al., 2000; Bayer et al., 2002; Cronn et al., 2002; Goldblatt et al., 2002; Hartmann et al., 2002; Schönenberger and Conti, 2003; Yamane et al., 2003). At low taxonomic levels, some noncoding cpDNA regions may show sufficient variation for phylogenetic resolution while others do

not (e.g., McCauley, 1995; Dumolin-Lapegue et al., 1997; Demesure et al., 1996; Ohsako and Ohnishi, 2000). The seven chloroplast regions chosen for this phylogenetic investigation, the *rpS16*, *rpL16*, *trnL*, and *trnG* introns as well as the *trnL-trnF*, *trnS-trnG*, and *trnH-psbA* intergenic spacers, showed dissimilar characteristics with respect to the amount of variability present within a particular region (Table 1). Comparison of the *trnH-psbA* and *trnS-trnG* spacers and *trnG* and *rpL16* introns in the broad analysis showed that the *trnS-trnG* intergenic spacer provided the greatest number (43) of informative characters in the data set, followed by the *trnG* (36), and *rpL16* (31) introns. However, from a cost-benefit point of view, because the *trnS-trnG* spacer and the *trnG* intron were amplified and sequenced together with the same two primers, this *trnS-trnG-trnG* region was by far the most valuable as it provided 79 out of 134 informative characters in the broad analysis (outlined in Table 1).

In the narrow phylogenetic investigation of only *Prunocerasus*, the *rpL16* region provided more informative characters than did *trnS-trnG-trnG* (13 and 10, respectively). These two regions were the most informative ones used in this investigation and the addition of the *rpS16* and *trnL* introns and the *trnL-trnF* and *trnH-psbA* spacers provided little more resolution but did help to raise support for some nodes.

***Relationships within Prunus*—**The combined cpDNA data sets provide insight into the phylogenetic relationships within *Prunus*. It is clear that *Prunocerasus* is a monophyletic assemblage of taxa and this is discussed below. We, in agreement with Bortiri et al. (2001), show support for a monophyletic section *Prunus* inside of which there is support for both Asian and European-West Asian lineages. Interestingly, although the data set contains several characters uniting section *Prunus*, *P. domestica* L. and *P. insititia* L. have identical sequences and share three autapomorphic positions with respect to *P. spinosa* L. and *P. cerasifera* Erhr. (*P. insititia* has been described as a subspecies, variety, and nothomorph of *P. domestica*). Subgenus *Amygdalus* is a monophyletic assemblage and we further show support for members of section *Amygdalus* grouping together with the single representative of section *Chamaeamygdalus* sister to the rest of the species in that subgenus, again in agreement with Bortiri et al. (2001). Among the monophyletic groups, *Prunocerasus*, section *Prunus*, and subgenus *Amygdalus*, there is no support for relationships among them, except that they all belong to a monophyletic polytomy. Other investigators have shown this close relationship between subgenus *Prunus* and subgenus *Amygdalus* (Watkins, 1976; Mowrey and Werner, 1990; Badenes and Parfitt, 1995; Bortiri et al., 2001; Lee and Wen, 2001; Jung et al., 2002), which we refer to as the Pruno-Amygdaloid clade. Interestingly, species of subgenus *Cerasus* section *Microcerasus* are well supported within the Pruno-Amygdaloid clade, a position that has been shown previously (Mowrey and Werner, 1990; Bortiri et al., 2001; Lee and Wen, 2001). The Pruno-Amygdaloid clade, including several unresolved *Microcerasus* species, is strongly supported here by a 1.0 posterior probability and 100% bootstrap support. Sister to the Pruno-Amygdaloid clade is *P. fasciculata*, which like many taxa within the genus has been subjected to fluctuating nomenclature. Rehder (1940) placed it in subgenus *Amygdalus* while Mason (1913) had it in subgenus *Emplectocladus*; in agreement with Bortiri et al. (2001), our data place *P. fasciculata* as sister to the Pruno-Amygdaloid clade. Outside of *P. fasciculata*,

representative species of subgenus *Cerasus* section *Mahaleb* form a clade sister to the Pruno-Amygdaloid clade + *P. fasciculata*. Outside of section *Mahaleb* we show paraphyly of subgenera *Padus* and *Laurocerasus*; this is in agreement with both Bortiri et al. (2001) and Lee and Wen (2001). However, Lee and Wen (2001) showed *P. virginiana* L. and *P. serotina* Ehrh. sister, with *P. laurocerasus* L. outside of several other *Padus* and *Maddenia hypoleuca* Koehne. They also showed *P. caroliniana* Aiton (and *P. grayana* Maxim., which is not included here) as sister to all of *Prunus*. This is in contrast to Bortiri et al. (2001) who showed *P. serotina* and *P. mahaleb* as sister taxa and *P. virginiana* and *P. laurocerasus* as more closely related to each other than either is to *P. serotina* + *P. mahaleb*. Our data do not support the relationship between *P. serotina* and *P. mahaleb*.

Taxa of subgenus *Cerasus* section *Microcerasus* (*P. pumila*, *P. tomentosa*, and *P. glandulosa*) have been difficult to place phylogenetically (Bortiri et al., 2001; Lee and Wen, 2001) although they appear to be associated with the Pruno-Amygdaloids. Mowrey and Werner (1990) showed section *Microcerasus* as a polyphyletic group with *P. pumila* and others nested between sections *Prunocerasus* and *Prunus*, while *P. glandulosa* was sister to the cherries and *P. tomentosa* was sister to the remaining Pruno-Amygdaloid group. Lee and Wen (2001) also showed section *Microcerasus* as a polyphyletic group with *P. besseyii* L.H.Bailey, a variety of *P. pumila* (Rohrer, 2000), in three very different parts of their trees in three different analyses. Additionally, they suggested that *P. tomentosa* is either related to section *Prunocerasus*, subgenus *Amygdalus* or sister to some Pruno-Amygdaloids in their three different analyses. Lastly, Bortiri et al. (2001) have shown section *Microcerasus* as a paraphyletic group with *P. besseyii* sister to members of section *Penarmeniaca* (not sampled in this study); they also showed *P. glandulosa* and *P. tomentosa* as unresolved within the Pruno-Amygdaloids. Possibly the most striking result of our broad analysis is the amount of apparent homoplasy among species of section *Microcerasus*. *Prunus tomentosa*, *P. pumila*, and especially *P. glandulosa* share molecular characters with several of the other groups that are resolved in our analysis. For example, *P. glandulosa* shares a 5-bp insertion with *P. mume* Siebold & Zucc. + *P. armeniaca*, a nucleotide substitution with *P. mume* + *P. mahaleb* + *P. pensylvanica* L., a substitution with section *Prunus*, another with *P. subcordata* and the American clade of section *Prunocerasus*, and at two other separate positions it shares substitutions with either *P. geniculata* or *Physocarpus opulifolius*. The homoplasy within, or at least paraphyly of, section *Microcerasus* has been noted previously (Mowrey and Werner, 1990; Bortiri et al., 2001; Lee and Wen, 2001).

Bortiri et al. (2001) noted that *Microcerasus* is not related to the rest of subgenus *Cerasus*, a finding that we report here as well; they also speculated that this section appears not to constitute a natural group. Waugh (1901) was ambiguous with respect to the placement of *P. pumila* var. *pumila* and *P. pumila* var. *besseyii* with the cherries or with the plums, perhaps indicating morphological intermediacy. Based on hybridization studies, Watkins (1976) believed in a closely related Pruno-Amygdaloid group and noted that members of subgenus *Cerasus* would not hybridize directly with them but genetic transfer was possible via the "*Microcerasus* bridge." Members of *Microcerasus* hybridize with both the Pruno-Amygdaloids and subgenus *Cerasus*. Based on this, and the apparent extensive homoplasy within section *Microcerasus*, we hesitantly speculate

that it is possible that this group represents an assemblage of taxa that carries the results of ancient hybridization like pleisiomorphic, homoplasious scars— as might be possible if section *Microcerasus* represents a polyphyletic assemblage of hybrids with much of the maternal parentage being via several different ancient Pruno-Amygdaloids.

***Relationships within Prunus section Prunocerasus*—**This study provides the first phylogenetic hypothesis for the species of *Prunus* section *Prunocerasus*. We report that the North American plums, section *Prunocerasus*, are monophyletic and include *P. texana*, which has previously been classified as a peach (subgenus *Amygdalus*). *Prunus subcordata*, the only species within section *Prunocerasus* of western North America, is sister to the rest of the section. We show strong support for the inclusion of *P. texana* within the North American plums—it may be a North American plum or possibly may have captured a *Prunocerasus* chloroplast. Sister to *P. texana* we report the existence of three strongly supported clades here designated the "American Clade," the "Chickasaw Clade," and the "Beach Clade" (Fig. 3). The American clade includes *P. americana* var. *americana, P. americana* var. *lanata, P. mexicana, P. rivularis* Scheele*, P. hortulana, P. umbellata* var. *injucunda*; the Chickasaw clade includes *P. angustifolia, P. munsoniana, P. gracilis, P. nigra, P. umbellata* var. *umbellata, P. alleghaniensis* var. *alleghaniensis,* and *P. alleghaniensis* var. *davisii*; and the Beach clade includes *P. geniculata, P. maritima* var. *maritima,* and *P. maritima* var. *gravesii*.

Classically, species within *Prunocerasus* have been defined and grouped based on relatively few qualitative characters; the lack of invariant quantitative characters in the section has led to much taxonomic confusion. Characters such as the presence or absence of glands terminating the leaf teeth or marginal on the calyx lobes, leaf or calyx pubescence, whether or not the flowers appear before or with the leaves, habit (tree or shrub), ability to root sprout, petal length (greater or less than 7.5 mm long), and geography have been the main characters used for species delimitation and grouping.

Taxa like *P. americana* var. *americana*, *P. americana* var. *lanata*, and *P. mexicana* have been largely separated using the characters of leaf and calyx pubescence; these three taxa in addition to *P. nigra* are often grouped because they are large trees with large flowers, comparatively. Additionally, pubescence is the main character separating some other closely related taxa like *P. umbellata* var. *umbellata* and *P. umbellata* var. *injucunda*. Small (1898 p.150) described *P. injucunda* as a distinct species and noted that is easily confused with *P. umbellata* from which it differs by having "a more rigid habit" and tomentose leaves and twigs. Additionally, it is difficult to separate *P. umbellata* and *P. umbellata* var. *injucunda* from *P. alleghaniensis* and *P. alleghaniensis* var. *davisii* and to a lesser extent *P. maritima* without some knowledge of their geography as these species are all morphologically overlapping. *Prunus alleghaniensis* and *P. umbellata* differ only in distribution according to Duncan and Duncan (1988) and Waugh (1899) reported that he could not morphologically distinguish the type material of *P. injucunda* from *P. maritima*. Lastly, *P. alleghaniensis* and *P. maritima* are also morphologically very close; to paraphrase T. C. Porter's (1877) description of *P. alleghaniensis*: it is nearly allied to *P. maritima*, but its remoteness from the seaboard, habitat on bluffs and mountains, narrower and longer more acuminate leaves, smaller fruit, and character of the stone, entitle it to the rank as a distinct species. Lastly, *P. angustifolia*, *P.*

munsoniana, *P. hortulana*, and *P. rivularis* can easily be confused. Their leaf margins are all distinctly serrate glandular and with the exception of *P. angustifolia*, their calyx lobes are glandular also. Furthermore, there seem to be many intermediates among them (Waugh, 1899; personal observation).

Within *Prunocerasus* there are a few morphologically coherent "groups" that can relatively easily be delimited; it is within these "groups" that species determination is difficult. For example, it may be difficult to distinguish *P. umbellata* from *P. alleghaniensis*, which are members of Wight's (1915) Maritima group; however, it is much easier to separate *P. alleghaniensis* from *P. americana*, which are members of Wight's (1915) Maritima and Americana groups, respectively. The gaps between these morphologically coherent groups are what previous authors like Waugh (1899) and Wight (1915) used to divide the North American plums into four series/six groups, respectively. Neither of these two authors clearly defined the delimiting characteristics of their series/groups although it is clear to us that their groups are based on overall similarity using the suite of aforementioned morphological characters (unless cited otherwise, in the following discussion Waugh = Waugh, 1899 and Wight = Wight, 1915). Waugh believed each series to be a group of hybrids sharing overall similarity. Into the Americana series, he placed *P. americana*, *P. americana mollis* (*P. americana* var. *lanata* and/or *P. mexicana*), and *P. nigra*. He thought *P. americana* to be a species of the central eastern USA that becomes increasingly glandular on the leaves and calyx lobes to the northeast, grading to *P. nigra*, and less glandular and more pubescent on the leaves and calyx lobes to the southwest, grading to *P. americana* var. *lanata* or *P. mexicana*. Waugh's Maritima series includes several forms of *P. maritima* in addition to *P. maritima* var. *gravesii*, *P. umbellata* var. *injucunda*, *P. gracilis*, and interestingly *P. texana*, all of which are smaller in stature and have smaller flowers than members of the Americana series. His Chickasaw series only includes *P. angustifolia* and his Hortulana series is centered on *P. hortulana*, its varieties, and questionably *P. munsoniana*. Waugh described the Hortulana series as a hybrid swarm created through various hybridizations between the other series (see Waugh, 1901), which intergraded with them where their ranges overlap. Although he did not include as many species as Wight, the series he described were very similar to the groups proposed by Wight whose delimitation of groups is shown in Fig. 3.

Our cpDNA-based phylogenetic hypothesis for section *Prunocerasus* is strongly supported and largely in disagreement with the previous morphological classifications described above (Waugh, 1899; Wight, 1915). While the groups that Waugh and Wight proposed make sense based on morphological similarity, and morphologically we do not disagree with them, they are not supported phylogenetically (see Fig. 3 and the supplementary data accompanying the online version of this article). Figure 3 shows both a comparison of previous classifications and the homoplasious distribution of the common morphological characters compared to the cpDNA phylogeny. The commonly used morphological characters such as habit, leaf teeth glands, calyx lobe glands, flower size, and time of flowering are all homoplasious with respect to the cpDNA phylogeny.

We show that, although many of the species within Waugh's series or Wight's groups are hardly morphologically distinguishable, they are split among three of the major phylogenetic clades of section *Prunocerasus*. Because the groups that Waugh and

Wight proposed are similar, and because Wight recognized more taxa, the following discussion is based on Wight's classification (see supplementary data accompanying the online version of this article or Fig. 3).

Our data place *P. americana*, *P. americana* var. *lanata*, and *P. mexicana* together in the American clade, in agreement with Wight; however, *P. nigra*, which both Waugh and Wight believed to be related to *P. americana*, is here placed in the Chickasaw clade. Additionally, although *P. americana* and *P. mexicana* are both in the American clade, they are separated by species that both Waugh and Wight placed in other groups and are therefore apparently not as closely related as Waugh and Wight had thought.

The morphologically coherent Maritima group of Wight is not supported here. Members of this group are nearly evenly split among the American, Chickasaw, and Beach clades. The close relationship between *P. maritima* and *P. geniculata* is unexpected with respect to morphology. *Prunus geniculata* is the most easily distinguishable species in the section and impossible to confuse with any of the other taxa. Morphologically, it resembles either a dwarfed *P. angustifolia* or *P. texana*. The only apparent commonality between *P. geniculata* and *P. maritima* is that both species are found in sandy soils, although they are separated by hundreds of miles in central Florida and the east coast north of Maryland through to Canada, respectively.

Within the Chickasaw clade we report a noticeable lack of genetic divergence between taxa; the sequences of members of this clade are nearly identical. While morphologically, some of the taxa are nearly identical (*P. angustifolia* and *P. munsoniana* or *P. umbellata* and *P. alleghaniensis*), *P. gracilis* and *P. nigra* are quite easily distinguished from the rest of the taxa in the clade. Additionally, it is quite easy to differentiate between *P. angustifolia* or *P. munsoniana* and *P. umbellata* or *P. alleghaniensis*. The lack of genetic divergence within the Chickasaw clade suggests that either (1) the species in this clade are the result of recent, rapid radiation, (2) frequent hybridization between these species has resulted in chloroplast sharing among species, or (3) several of the members of this clade are of hybrid origin and share a common chloroplast lineage. Based on the morphological incongruence between taxa within this clade, we suspect that reason number three may be the most likely (see discussion below).

Wight noted that *P. hortulana* and *P. rivularis* are closely related and although the close relationship is also not supported here since these two species are separated by several other species, both species belong to the American clade. Morphologically, these two species are most similar to *P. munsoniana* and *P. angustifolia* of the Chickasaw clade. This is yet another example of morphologically very similar species being separated by less similar ones.

Members of each of the American and Chickasaw clades are unequivocally closely related with respect to evidence provided by the maternally-inherited cpDNA molecule. Yet, morphologically, each of the two clades contains taxa that appear to be more closely related to taxa of the other clade, based on morphological evidence.

Although strongly supported, our cpDNA-based phylogenetic hypothesis for *Prunocerasus* is largely in disagreement with the previous morphological classifications described above (Waugh, 1899; Wight, 1915). This appears to be due to homoplastic morphological characters classically used to delineate the taxa within *Prunocerasus* (see

Fig. 3). We suspect that the apparent homoplasy, with respect to morphological characters, may be indicative of hybrid origin for several of the species of the section as (1) origin via hybridization has been shown to occur within the genus (Watkins, 1976; Brettin et al., 2000; Mohanty et al., 2000), (2) much of *Prunocerasus* is capable of natural hybridization (Hedrick et al., 1911; Wight, 1915; Flory, 1938; Rehder, 1940), (3) *Prunocerasus* is a group of obligate outcrossers, and (4) other researchers have proposed species within the section to be of hybrid origin (Waugh, 1899, 1901; Steyermark, 1963). For example, the close relationship between *P. hortulana* and *P. mexicana* reported here is unusual because: (1) *P. mexicana* and *P. hortulana* do not share many of the characters classically used to distinguish species within the section and are relatively easily discernible, and (2) *P. hortulana* and *P. rivularis* are very morphologically similar as are *P. mexicana* and *P. americana* var. *lanata*, yet relationships shown in our phylogenetic hypothesis do not agree with morphological similarity. Then again, Steyermark (1963) noted that some individuals of *P. hortulana* appear to be hybrids between it and *P. mexicana*. Based on our cpDNA-based phylogenetic evidence being incongruous with morphology and the idea of speciation via hybridization being common within the genus we reason that some members of the American and Chickasaw clades are of hybrid origin. Furthermore, an analysis of nuclear encoded ITS in addition to preliminary data from a granule bound starch synthase gene (GBSSI-2) reveals that several of the taxa in each of these two clades section contain polymorphic nucleotide positions and indels (unpublished data). Further evidence from the nuclear genome as well as analysis of cpDNA from additional accessions from each of the putative taxa is being pursued to resolve this uncertainty.

A biogeographical note—*Prunus* is thought to have originated in Central Asia (Watkins, 1976). The work of Bortiri et al. (2001) does not refute an Asian origin for the genus but it does not unambiguously support it either. Because there is no clear sister taxon to *Prunus* and both Old and New world taxa are possible candidates (Bortiri et al. 2001), the geographic origin of the genus is equivocal. If *Prunus* originated in Asia, then based on *P. subcordata* being the only western species and phylogenetically sister to the rest of *Prunocerasus*, we would suggest that the common ancestor of *Prunocerasus* arrived in North America via Beringia. But, more work is needed within the genus for us to root a phylogeographic hypothesis. In any case, our data refute the hypothesis of Mowrey and Werner (1990) that suggested *Prunocerasus* is paraphyletic and arrived in North America via two immigration events.

Conclusions and future directions—Four main conclusions can be drawn from the data and analyses presented in this paper. First, phylogenetic analysis of cpDNA sequences demonstrates that *Prunus* subgenus *Prunus* section *Prunocerasus* is a monophyletic assemblage. Second, phylogenetic analysis of this noncoding cpDNA data set provides a maternally-inherited framework on which we can further develop a hypothesis of relationships for the species within *Prunocerasus.* Third, our data counter the hypothesis of Mowrey and Werner (1990) that *Prunocerasus* is polyphyletic and experienced more than one immigration event to North America. Finally, the results presented in this paper highlight the continuing need, especially at low taxonomic levels,

to assay the relative utility of multiple cpDNA regions before investigation instead of just using commonly employed tools.

Future investigations of this section will include: (1) sampling cpDNA sequences from additional populations of each putative species to determine how much variation there is within species and more importantly if chloroplasts are monophyletic within each putative species and (2) the addition of a nuclear marker to this analysis to further illuminate relationships, especially those that may be reticulating, within section *Prunocerasus*.

LITERATURE CITED

AZUMA, H., J. G. GARCÍA-FRANCO, V. RICO-GRAY, AND L. B. THIEN. 2001. Molecular phylogeny of the Magnoliaceae: the biogeography of tropical and temperate disjunctions. *American Journal of Botany* 88: 2275-2285.

BADENES, M. L., AND D. E. PARFITT. 1995. Phylogenetic relationships of cultivated *Prunus* species from an analysis of chloroplast DNA variation. *Theoretical and applied genetics* 90: 1035-1041.

BAILEY, L. H. 1892. *Prunus hortulana*. *Garden and Forest* 5: 90.

BAILEY, L. H., AND E. Z. BAILEY. 1941. Hortus Second: a concise dictionary of gardening, general horticulture, and cultivated plants in North America. The Macmillan Company, New York, New York, USA.

BAYER, R. J., D. G. GREBER, AND N. H. BAGNALL. 2002. Phylogeny of Australian Gnaphalieae (Asteraceae) based on chloroplast and nuclear sequences, the *trnL* intron, *trnL-trnF* intergenic spacer, *matK*, and ETS. *Systematic Botany* 27: 801-814.

BLACKBURN, B. 1952. Trees and Shrubs in eastern North America: keys to the wild and cultivated woody plants growing in the temperate regions exclusive of conifers. Oxford University Press, New York, New York, USA.

BORTIRI, E., S. H. OH, F. Y. GAO, AND D. POTTER. 2002. The phylogenetic utility of nucleotide sequences of sorbitol 6-phosphate dehydrogenase in *Prunus* (Rosaceae). *American Journal of Botany* 89: 1697-1708.

BORTIRI, E., S. H. OH, J. G. JIANG, S. BAGGETT, A. GRANGER, C. WEEKS, M. BUCKINGHAM, D. POTTER, AND D. E. PARFITT. 2001. Phylogeny and systematics of *Prunus* (Rosaceae) as determined by sequence analysis of ITS and the chloroplast *trnL-trnF* spacer DNA. *Systematic Botany* 26: 797-807.

BRETTIN, T. S., R. KARLE, E. L. CROWE, AND A. F. IEZZONI. 2000. Chloroplast inheritance and DNA variation in sweet, sour, and ground cherry. *Journal of Heredity* 91: 75-79.

CORRELL, D. S., AND M. C. JOHNSTON. 1970. Manual of the vascular plants of Texas. Texas Research Foundation, Renner, Texas, USA.

CRONN, R. C., R. L. SMALL, T. HASELKORN, AND J. F. WENDEL. 2002. Rapid diversification of the cotton genus (*Gossypium*: Malvaceae) revealed by analysis of sixteen nuclear and chloroplast genes. *American Journal of Botany* 89: 707-725.

DEMESURE, B., B. COMPS, AND R. PETIT. 1996. Chloroplast DNA phylogeography of the common beech (*Fagus sylvatica* L.) in Europe. *Evolution* 50: 2515-2520.

DUMOLIN-LAPEGUE, S., M. H. PEMONGE, AND R. J. PETIT. 1997. An enlarged set of consensus primers for the study of organelle DNA in plants. *Molecular Ecology* 6: 393-397.

DUNCAN, W. H., AND M. B. DUNCAN. 1988. Trees of the southeastern United States. University of Georgia Press, Athens, Georgia, USA.

DUNN, K. L. 2002. Growing ripe for development: beach plums as a regional crop. *Farming: The Journal of Northeast Agriculture.* September, pp. 19, 30-32. (www.farmingmagazine.com)

FELSENSTEIN, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of Molecular Evolution* 17: 368-376.

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

FERNALD, M. L. 1950. Gray's manual of botany, 8th ed. American Book Company, New York, New York, USA.

FLORY, W. S. 1938. Cross sterility in hybrid plum varieties. *Proceedings of the American Society for Horticultural Science*. Contribution No. 409: 43-46.

FOGLE, H. W. 1978. Plum improvement in the United States. *Acta Horticulturae* 74: 35-40.

GENE CODES. 1998. Sequencher, version 3.1.1. Gene Codes, Ann Arbor, Michigan, USA.

GLEASON, H. A. 1952. The new Britton and Brown illustrated flora of the United States and adjacent Canada, vol. 2. Lancaster Press, Lancaster, Pennsylvania, USA.

GLEASON, H. A., AND A. CRONQUIST. 1991. Manual of vascular plants of northeastern United States and adjacent Canada, 2nd ed. New York Botanical Garden, Bronx, New York, USA.

GODFREY, R. K. 1988. Trees, shrubs, and woody vines of northeastern Florida and adjacent Georgia and Alabama. The University of Georgia Press, Athens, Georgia, USA.

GOLDBLATT, P., V. SAVOLAINEN, O. PORTEOUS, I. SOSTARIC, M. POWELL, G. REEVES, J. C. MANNING, T. C. BARRACLOUGH, AND M. W. CHASE. 2002. Radiation in the Cape flora and the phylogeny of peacock irises *Moraea* (Iridaceae) based on four plastid DNA regions. *Molecular Phylogenetics and Evolution* 25: 341-360.

HARTMANN, S., J. D. NASON, AND D. BHATTACHARYA. 2002. Phylogenetic origins of *Lophocereus* (Cactaceae) and the Senita Cactus-Senita Moth pollination mutualism. *American Journal of Botany* 89: 1085-1092.

HEDRICK, U. P., W. H. ALDERMAN, M. J. DORSEY, O. M. TAYLOR, AND R. WELLINGTON. 1911. The plums of New York. J. B. Lyon Company State Printers, Albany, New York, USA.

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

JUNG, Y. H., S. H. HAN, AND M. Y. OH. 2002. Phylogenetic analysis of Korean *Prunus* (Rosaceae) based on ITS sequences of nuclear ribosomal DNA. *Korean Journal of Genetics* 24: 247-258.

KARP, D. 2003. The beach plums: hard to tame, but growers don't give up. The New York Times website: http://www.nytimes.com/2003/09/03/dining/03PLUM.html.

KUSUMI, J., Y. TSUMURA, H. YOSHIMARU, AND H. TACHIDA. 2000. Phylogenetic relationships in Taxodiaceae and Cupressaceae sensu stricto based on *matK* gene, *chlL* gene, *trnL-trnF* IGS region, and *trnL* intron sequences. *American Journal of Botany* 87: 1480-1488.

LEE, S., AND J. WEN. 2001. A phylogenetic analysis of *Prunus* and the Amygdaloideae (Rosaceae) using ITS sequences of nuclear ribosomal DNA. *American Journal of Botany* 88: 150-60.

MASON, S. C. 1913. The pubescent-fruited species of *Prunus* of the Southwestern States. *Journal of Agricultural Research* 1: 147-179.

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

MOHANTY, A., J. P. MARTÍN, AND I AGUINAGALDE. 2000. Chloroplast DNA diversity within and among populations of the allotetraploid *Prunus spinosa* L. *Theoretical and Applied Genetics* 100: 1304-1310.

MOWREY, B. D., AND D. J. WERNER. 1990. Phylogenetic relationships among species of *Prunus* as inferred by isozyme markers. *Theoretical and Applied Genetics* 80: 129-133.

NYLANDER, J. A. A. 2002. Testing models of evolution - MrModeltest version 1.1b. Computer program and documentation distributed by author, website: http://www.ebc.uu.se/systzoo/staff/nylander.html.

OHSAKO, T., AND O. OHNISHI. 2000. Intra- and interspecific phylogeny of wild *Fagopyrum* (Polygonaceae) species based on nucleotide sequences of noncoding regions in chloroplast DNA. *American Journal of Botany* 87: 573-582.

OXELMAN, B., M. LIDÉN, AND D. BERGLUND. 1997. Chloroplast *rps16* intron phylogeny of the tribe Sileneae (Caryophyllaceae). *Plant Systematics and Evolution* 206: 393-410.

PORTER. T. C. 1877. *Prunus alleghaniensis* n. sp. *Coulter's Botanical Gazette* 2: 85.

POSADA, D., AND K. A. KRANDALL. 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14: 817-818.

RADFORD, A. E., H. E. AHLES, AND C. R. BELL. 1968. Manual of the vascular flora of the Carolinas. The University of North Carolina Press, Chapel Hill, North Carolina, USA.

REHDER, A. 1940. Manual of cultivated trees and shrubs hardy in North America, exclusive of the subtropical and warmer temperate regions, 2nd revised and enlarged edition. The Macmillan Company, New York, New York, USA.

REID, W., AND K. L. B. GAST. 1993. The potential for domestication and utilization of native plums in Kansas. *In* J. Janick, and J. E. Simon [eds.], New crops, 520-523. Wiley, New York, New York, USA.

ROBERTSON, K. R. 1974. The genera of Rosaceae in the southeastern United States. *Journal of the Arnold Arboretum* 55: 654-662.

ROHRER, J. R. 2000. The sand cherry in Wisconsin and neighboring states. *Michigan Botanist* 39: 59-69.

SARGENT, C. S. 1902. Texas, Louisiana, and Arkansas [*P. tarda*]. *Botanical Gazette* 33: 108.

SCHÖNENBERGER, J., AND E. CONTI. 2003. Molecular phylogeny and floral evolution of the Peneaceae, Olinaceae, Rhynchocalycaceae, and Alzateaceae (Myrtales). *American Journal of Botany* 90: 293-309.

SHAW, J., E. LICKEY, J. BECK, S. FARMER, W. LIU, J. MILLER, K. C. SIRIPUN, C. WINDER, E. E. SCHILLING, AND R. SMALL. 2005. The tortoise and the hare II: relative utility of 21 non-coding chloroplast DNA sequences for phylogenetic analysis. *American Journal of Botany* 92: 142-166.

SMALL, J. K. 1898. Studies in the botany of the southern United States-XIII. *Bulletin of the Torrey Botanical Club* 25: 134-151.

SMALL, J. K. 1933. Manual of the southeastern flora. Published by the author, The Science Press Printing Company, Lancaster, Pennsylvania, USA.

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 phylogenetic reconstruction in a recently diverged plant group. *American Journal of Botany* 85: 1301-1315.

SMITH, E. B. 1994. Keys to the flora of Arkansas. University of Arkansas Press, Fayetteville, Arkansas, USA.

STEYERMARK, J. A. 1963. Flora of Missouri. Iowa State University Press, Ames, Iowa, USA.

SWOFFORD, D. L. 2002. PAUP*: phylogenetic analysis using parsimony (*and other methods) version 4.0b10. Sinauer Associates, Sunderland, Massachusetts, USA.

TABERLET, P., L. GIELLY, G. PAUTOU, AND J. BOUVET. 1991. Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology* 17: 1105-1109.

THOMPSON, J. D., T. J. GIBSON, F. PLEWNIAK, F. JEANMOUGIN, AND D. G. HIGGINS. 2001. The Clustal_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* 25: 4876-4882.

USDA, NRCS. 2002. The PLANTS database, version 3.5. National Plant Data Center, website: http://plants.usda.gov.

WATKINS, R. 1976. Cherry, plum, peach, apricot, and almond. *In* N. W. Simmons [ed.], Evolution of crop plants, 242-247. Longman, London, UK.

WAUGH, F. A. 1899. The pollination of plums. *Vermont Agricultural Experimentation Station, Annual Report* 12: 189-209.

WAUGH, F. A. 1901. Plums and plum culture. Orange Judd, New York, New York, USA.

WIGHT, W. F. 1913. North American species of the genus *Amygdalus*. Dudley Memorial Volume. Leland Stanford Junior University Publications, University Series, California, USA.

WIGHT, W. F. 1915. Native American species of *Prunus*. *Bulletin of the U.S. Department of Agriculture* 179: 1-75.

WOFFORD, B. E., AND E. W. CHESTER. 2002. Guide to the trees, shrubs, and woody vines of Tennessee. The University of Tennessee Press, Knoxville, Tennessee, USA.

WUNDERLIN, R. P. 1988. Guide to the vascular plants of Florida. University of Florida Press, Gainesville, Florida, USA.

YAMANE, K., Y. YASUI, AND O. OHNISHI. 2003. Intraspecific cpDNA variations of diploid and tetraploid perennial buckwheat, *Fagopyrum cymosum* (Polygonaceae). *American Journal of Botany* 90: 339-346.

APPENDIX; PART 3

Table 1. List of taxa used in this investigation along with current classification scheme, source and voucher numbers, status of species of conservation concern, and GenBank accession numbers. Classification follows Rehder (1940) and Wight (1915).

Subgenus/Section		Species	Cons. concern
Rehder (1940)			
Prunus L.			
Prunocerasus Koehne		*P. geniculata* R.M.Harper	FE
	Americana	*P. americana* Marshall	T-NH, VT
		P. a. Marshall var. *lanata* Sudw.	
		P. mexicana S.Watson	PE-OH
		P. nigra Aiton	E-IA, PE-OH
	Angustifolia	*P. angustifolia* Marshall	E-NJ
		P. munsoniana Wight & Hedrick	
	Gracilis	*P. gracilis* Engelm. & A.Gray	
	Hortulana	*P. hortulana* L.H.Bailey	
		P. rivularis Scheele	
	Maritima	*P. alleghaniensis* Porter	E-NJ, T-MD, PA
		P.a. var. *davisii* (Wight.) Sarg.	
		P. maritima Marshall	E-ME, MD, PA
		P. m. var. *gravesii* (Small) G.I.Anderson	PE-CT
		P. umbellata Elliot	
		P. u. Elliott var. *injucunda* (Small) Sarg.	
	Subcordata	*P. subcordata* Benth.	
Prunus L.		*P. cerasifera* Erhr.	
		P. domestica L.	
		P. insititia L.	
		P. salicina Lindl.	
		P. simonii Carr.	
		P. spinosa L.	
Armeniaca (Lam.) Koch		*P. armeniaca* L.	
		P. mandshurica (Maxim.) Koehne	
		P. mume Siebold & Zucc	
Amygdalus (L.) Focke			
		* *P. texana* Dietr.	
Amygdalus		*P. amygdalus* Batsch	
		P. davidiana (Carriere) N.E.Brown	
		P. persica (L.) Batsch	
		P. mira Koehne	
Chamaeamygdalus Focke		*P. tenella* Batsch	
**Emplectocladus* (Torr.) Sargent			
		P. fasciculata A.Gray	
Cerasus Pers.			
Microcerasus Webb		*P. pumila* L.	
		P. glandulosa Thunb.	
		P. tomentosa Thunb.	
Mahaleb Focke		*P. mahaleb* L.	
		P. pensylvanica L.	
Padus (Moench) Koehne			
		P. serotina Ehrh.	
		P. virginiana L.	
Laurocerasus Koehne			
		P. caroliniana Aiton	
		P. laurocerasus L.	
Outgroup		*Physocarpus opulifolius* (L.) Maxim.	

Table 1. continued.

Source and Voucher	Genbank Accession Numbers					
	trnL	*trnL-trnF*	*rbS16*	*rbL16*	*trnH-psbA*	*trnS-trn-trnG*
ISh898; FL: TENN	AY500756	AY500775	AY500694	AY500651	AY500608	AY500713
ISh038; TN: TENN	AY500743	AY500762	AY500681	AY500638	AY500595	AY500700
I.Beck 4955; TN: TENN	AY500744	AY500763	AY500682	AY500639	AY500596	AY500701
ISh919; TX: TENN	AY500747	AY500766	AY500685	AY500642	AY500599	AY500704
ISh979; VT: TENN	AY500753	AY500772	AY500691	AY500648	AY500605	AY500710
ISh785; GA: TENN	AY500749	AY500768	AY500687	AY500644	AY500601	AY500706
ISh810; TN: TENN	AY500750	AY500769	AY500688	AY500645	AY500602	AY500707
ISh936; TX: TENN	AY500751	AY500770	AY500689	AY500646	AY500603	AY500708
ISh821; TN: TENN	AY500748	AY500767	AY500686	AY500643	AY500600	AY500705
Endquist 3372; TX: BRIT	AY500745	AY500764	AY500683	AY500640	AY500597	AY500702
ISh837; PA: TENN	AY500755	AY500774	AY500693	AY500650	AY500607	AY500712
G. Schmidt; MI: TENN	AY500754	AY500773	AY500692	AY500649	AY500606	AY500711
ISh877-045; MA: TENN	AY500757	AY500776	AY500695	AY500652	AY500609	AY500714
Conn Greenhouse: TENN	AY500758	AY500777	AY500696	AY500653	AY500610	AY500715
ISh774-003; FL: TENN	AY500752	AY500771	AY500690	AY500647	AY500604	AY500709
ISh958-108; GA: TENN	AY500746	AY500765	AY500684	AY500641	AY500598	AY500703
I.Syring; CA: TENN	AY500760	AY500779	AY500698	AY500655	AY500612	AY500717
DPRU 563: TENN				AY500659	AY500616	AY500721
DPRU 350: TENN				AY500657	AY500614	AY500719
DPRU 2054: TENN				AY500656	AY500613	AY500718
DPRU 791: TENN				AY500660	AY500617	AY500722
DPRU 545: TENN				AY500661	AY500618	AY500723
DPRU 2289.22: TENN				AY500658	AY500615	AY500720
DPRU 1372.2: TENN				AY500663	AY500620	AY500725
DPRU 2311.1: TENN				AY500662	AY500619	AY500724
DPRU 1588: TENN				AY500664	AY500621	AY500726
ISh924-077; TX: TENN	AY500759	AY500778	AY500697	AY500654	AY500611	AY500716
DPRU 1463.5: TENN				AY500668	AY500625	AY500730
DPRU 581: TENN				AY500669	AY500626	AY500731
ISh992; TN: TENN				AY500671	AY500628	AY500733
DPRU 2228.3: TENN				AY500670	AY500627	AY500732
DPRU 2225.6: TENN				AY500672	AY500629	AY500734
DPRU 2033: TENN				AY500673	AY500630	AY500735
Horn 2001-02; TN: TENN				AY500666	AY500623	AY500728
DPRU 403.1: TENN				AY500665	AY500622	AY500727
DPRU 2316.4: TENN				AY500667	AY500624	AY500729
ISh966-116; TN: TENN	AY500761	AY500780	AY500699	AY500674	AY500631	AY500736
ISh865; VT: TENN				AY500675	AY500632	AY500737
ISh1013; TN: TENN				AY500676	AY500633	AY500738
ISh871-040; NH: TENN				AY500677	AY500634	AY500739
E. Lickey; FL: TENN				AY500679	AY500636	AY500741
ISh1014; TN: TENN				AY500678	AY500635	AY500740
ISh1015, TN: TENN				AY500680	AY500637	AY500742

Fig. 1. Strict consensus of 25 171 equally parsimonious trees resulting from maximum parsimony analysis using the combined data from the *rpL16* and *trnG* introns as well as the *trnS-trnG* and *trnH-psbA* intergenic spacers (3270 aligned nucleotide positions, 134 are parsimony informative; tree length = 422 steps; consistency index [CI] = 0.92; retention index [RI] = 0.94). Bootstrap values greater than 50% are above the branches. Branches shown with dotted lines are not resolved in the broad analysis but show up in the *Prunocerasus*-only analysis with the addition of the *trnL-trnF* intergenic spacer and the *trnL* and *rpS16* introns to the aforementioned four cpDNA regions. A separate, *Prunocerasus*-only maximum parsimony analysis of the seven cpDNA regions found one most parsimonious tree of 128 steps (4375 aligned nucleotide positions, 37 are parsimony informative; tree length = 128 steps; CI = 0.98; RI = 0.98). Bootstrap values for the *Prunocerasus*-only analysis greater than 50% are shown below the branches. Classification into subgenera and sections follows Rehder (1940). *Cs = Cerasus, Amyg = Amygdalus, E = Emplectocladus, , Pd = Padus*, *Lc = Laurocerasus, Armen = Armeniaca, Mc = Microcerasus*, *Ca = Chamaeamygdalus, Mah = Mahaleb.*

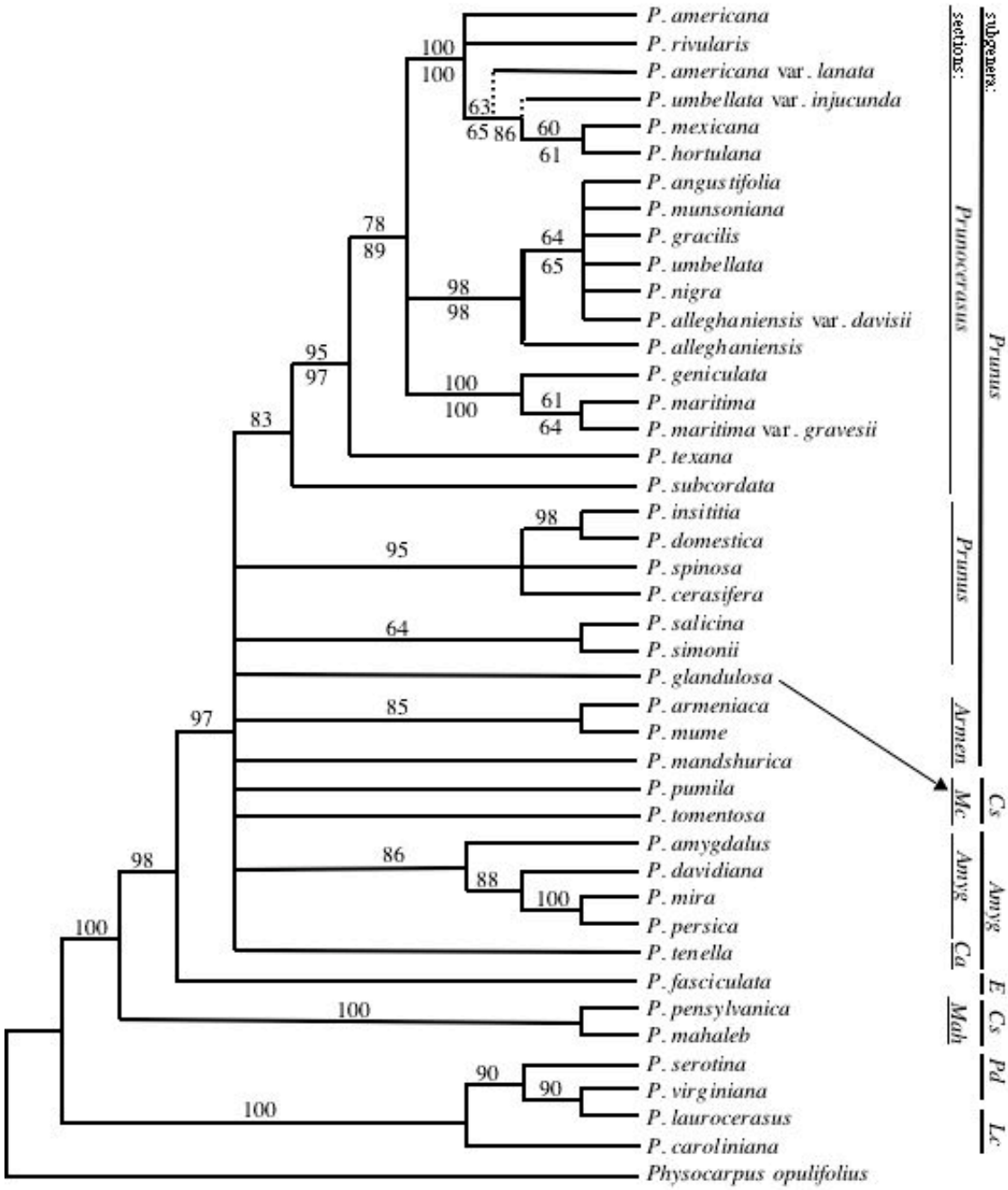
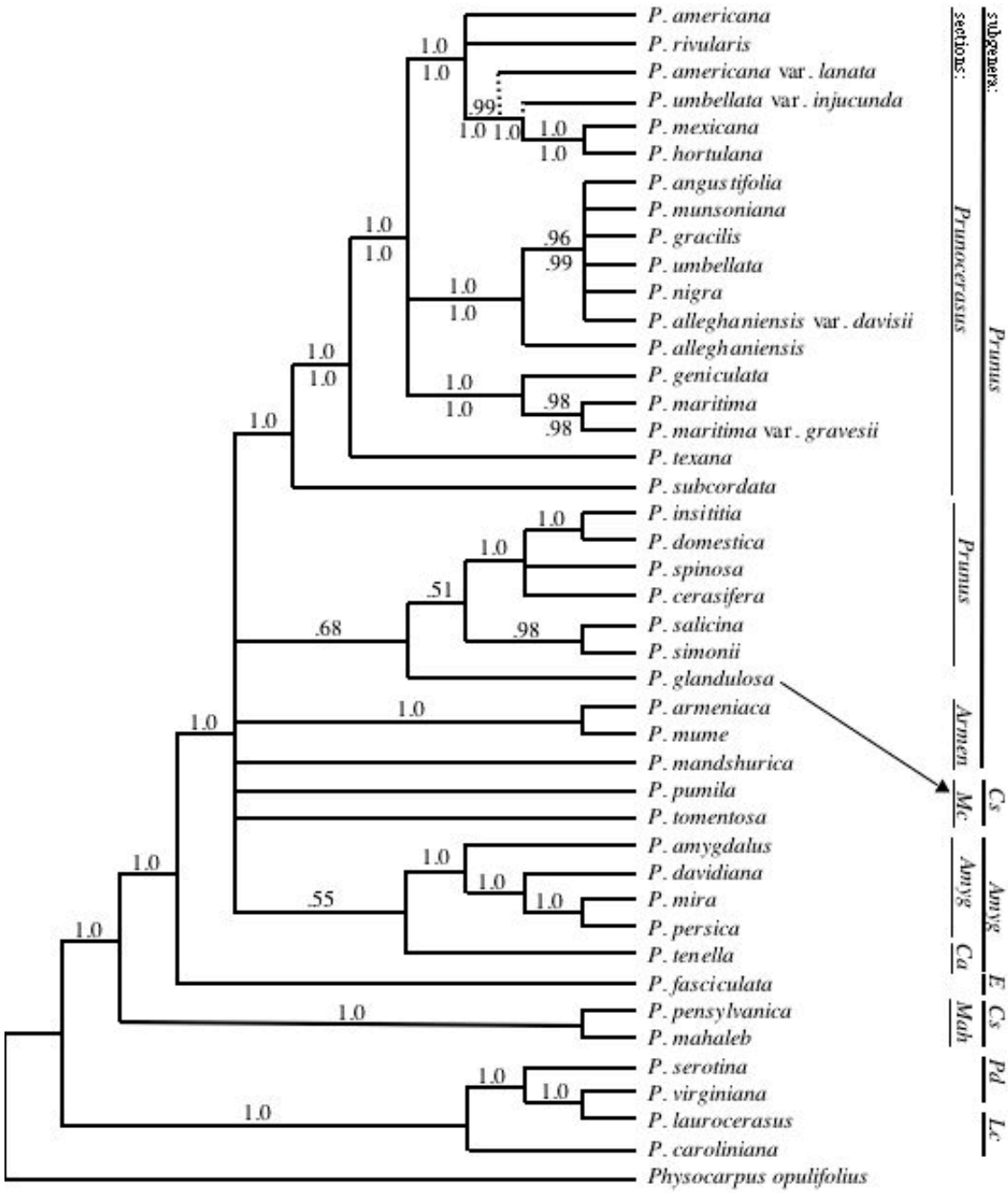
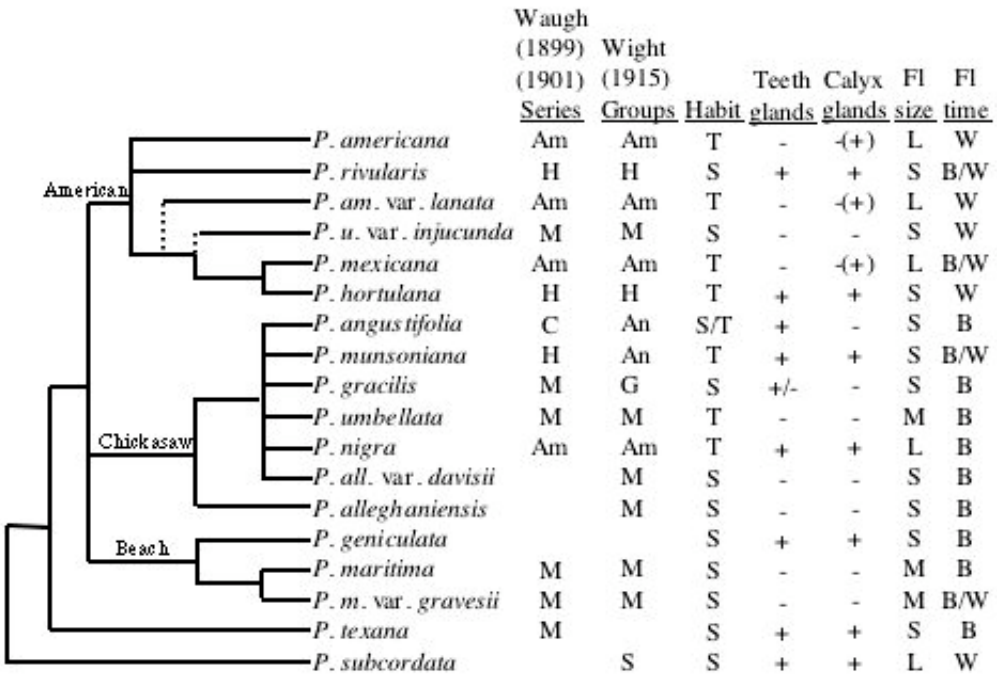

P. americana
P. rivularis
P. americana var. lanata
P. umbellata var. injucunda
P. mexicana
P. hortulana
P. angustifolia
P. munsoniana
P. gracilis
P. umbellata
P. nigra
P. alleghaniensis var. davisii
P. alleghaniensis
P. geniculata
P. maritima
P. maritima var. gravesii
P. texana
P. subcordata
P. insititia
P. domestica
P. spinosa
P. cerasifera
P. salicina
P. simonii
P. glandulosa
P. armeniaca
P. mume
P. mandshurica
P. pumila
P. tomentosa
P. amygdalus
P. davidiana
P. mira
P. persica
P. tenella
P. fasciculata
P. pensylvanica
P. mahaleb
P. serotina
P. virginiana
P. laurocerasus
P. caroliniana
Physocarpus opulifolius
100
100
63
65
86
60
61
78
89
64
65
98
98
95
97
100
100
61
64
83
98
95
64
97
85
98
86
88
100
100
100
90
90
100
subgenera:
Prunus
sections:
Prunocerasus
Prunus
Armen
Mc
Amyg
Ca
E
Mah
Cs
Amyg
Cs
Pd
Lc

Fig. 2. Bayesian analysis tree for *Prunus* using the combined data from the *rpL16* and *trnG* introns as well as the *trnS-trnG* and *trnH-psbA* intergenic spacers (3270 aligned nucleotide positions, 134 are parsimony informative). Numbers above the branches are posterior probabilities derived in the broad analysis. Branches shown with dotted lines are not resolved in the broad analysis but show up in the *Prunocerasus*-only analysis with the addition of the *trnL-trnF* intergenic spacer and the *trnL* and *rpS16* introns to the aforementioned four cpDNA regions (4375 aligned nucleotide positions, 37 are parsimony informative). Posterior probabilities for the *Prunocerasus*-only analysis are below the branches. Classification into subgenera and sections follows Rehder (1940).


subgenera:
sections:
Prunus
Prunocerasus
Prunus
Armen
Cs
Mc
Amyg
Amyg
Ca
E
Cs
Mah
Pd
Lc
P. americana
P. rivularis
P. americana var. lanata
P. umbellata var. injucunda
P. mexicana
P. hortulana
P. angustifolia
P. munsoniana
P. gracilis
P. umbellata
P. nigra
P. alleghaniensis var. davisii
P. alleghaniensis
P. geniculata
P. maritima
P. maritima var. gravesii
P. texana
P. subcordata
P. insititia
P. domestica
P. spinosa
P. cerasifera
P. salicina
P. simonii
P. glandulosa
P. armeniaca
P. mume
P. mandshurica
P. pumila
P. tomentosa
P. amygdalus
P. davidiana
P. mira
P. persica
P. tenella
P. fasciculata
P. pensylvanica
P. mahaleb
P. serotina
P. virginiana
P. laurocerasus
P. caroliniana
Physocarpus opulifolius



	Waugh (1899) (1901) Series	Wight (1915) Groups	Habit	Teeth glands	Calyx glands	Fl size	Fl time
P. americana	Am	Am	T	-	-(+)	L	W
P. rivularis	H	H	S	+	+	S	B/W
P. am. var. *lanata*	Am	Am	T	-	-(+)	L	W
P. u. var. *injucunda*	M	M	S	-	-	S	W
P. mexicana	Am	Am	T	-	-(+)	L	B/W
P. hortulana	H	H	T	+	+	S	W
P. angustifolia	C	An	S/T	+	-	S	B
P. munsoniana	H	An	T	+	+	S	B/W
P. gracilis	M	G	S	+/-	-	S	B
P. umbellata	M	M	T	-	-	M	B
P. nigra	Am	Am	T	+	+	L	B
P. all. var. *davisii*		M	S	-	-	S	B
P. alleghaniensis		M	S	-	-	S	B
P. geniculata			S	+	+	S	B
P. maritima	M	M	S	-	-	M	B
P. m. var. *gravesii*	M	M	S	-	-	M	B/W
P. texana	M		S	+	+	S	B
P. subcordata		S	S	+	+	L	W

Fig. 3. Phylogenetic relationships of *Prunocerasus* shown with the classifications of Waugh (1899, 1901) and Wight (1915) in addition to five morphological characters classically used to delimit taxonomic boundaries within the section. The three major clades, the American, Chickasaw, and Beach, are shown on the respective branches. Abbreviations for classification by Waugh (1899, 1901) and Wight (1915): Am = Americana, An = Angustifolia, C = Chickasaw (same as An), H = Hortulana, M = Maritima, G = Gracilis, and S = Subcordata. Abbreviations for morphological characters: Habit: T = tree, S = shrub; Teeth glands: + = glands present on each tooth of the leaf blade, - = glands absent on each tooth of the leaf blade; Calyx glands: + = glands present on margin of the calyx lobes, - = glands absent on margin of the calyx lobes; Flower size (Fl size): flowers large (L) = flowers greater than 15 mm wide, flowers small (S) = flowers less than 15 mm wide, flowers midsized (M) = flowers ~ 15 mm wide; Time of flowering (Fl time): W = flowers emerging with the leaves, B = flowers present before leaves emerge.

Part 4

CHLOROPLAST DNA PHYLOGENY AND PHYLOGEOGRAPHY OF THE NORTH AMERICAN PLUMS (*PRUNUS* SUBGENUS *PRUNUS* SECTION *PRUNOCERASUS*; ROSACEAE)

This chapter is a version of a paper by the same name submitted to the *American Journal of Botany* in March of 2005 by Joey Shaw and Randall Small:

Chloroplast DNA phylogeny and phylogeography of the North American plums (*Prunus* subgenus *Prunus* section *Prunocerasus*; Rosaceae).

The use of "we" in this chapter refers to Randall Small and myself. My contributions to this chapter include (1) selection of the topic and development of the problem into a work relevant to study, (2) the lab work to generate data, and (3) most of the writing.

ABSTRACT

Prunus L. subgenus *Prunus* section *Prunocerasus* Koehne (Rosaceae) is a North American taxon with 17 commonly recognized taxa that are not easily circumscribed, have overlapping morphologies, and are known to hybridize. In a previous study, we showed that the North American plums are a monophyletic group comprised of three primary clades with little to no sequence divergence within each of those clades. An unanticipated result of that study was that the inferred relationships contrast sharply with previously defined morphologically-based groups. Here we determined if the three primary cpDNA haplotypes are confined to the taxa in which they were initially observed. Approximately 800bp of the cpDNA *rpL16* intron were sequenced for 207 accessions of 18 taxa. The results showed that 14 of 18 section *Prunocerasus* taxa contain more than one of the three primary cpDNA haplotypes. Aside from the results found in section *Prunocerasus*, this study has broader implications for phylogenetics in general. The common practice of choosing a single exemplar to represent a taxon in closely related groups can be profoundly misleading. In hindsight, the possibility existed in our earlier study that we could have chosen a different combination of exemplars, which could have resulted in a different inferred phylogeny.

INTRODUCTION

Morphological taxonomy has been notoriously difficult within the North American plums, *Prunus* L. subgenus *Prunus* section *Prunocerasus* Koehne (Rosaceae), because species boundaries are blurred by interspecific similarities and intraspecific variation, and likely by interspecific introgressive hybridization (Waugh, 1899, 1901; Hedrick et al., 1911; Wight, 1915; Rehder, 1940; Shaw and Small, 2004). This has historically led to species oversplit within section *Prunocerasus* where species were named based on characters as taxonomically questionable as pubescence and geography. However, relatively little taxonomic work has been done on the North American plums in the last ca. 60 years and throughout North America taxonomists have settled on names in a fashion that brings to mind lineage sorting. Because there are few discrete morphological characters available for constructing a cladistic hypothesis for section *Prunocerasus*, we are relegated to DNA sequence data to provide insight into the relationships among the taxa within the section.

In a previous study (Shaw and Small, 2004) that employed seven noncoding cpDNA regions (totaling 4375 bp), we showed that the North American plums are a monophyletic assemblage of taxa, but that there is little genetic divergence between species. In that

study we selected single representatives of each of the 17 commonly accepted taxa and showed that the majority segregate into three primary clades sister to *P. subcordata*, the only species from western North America (Fig. 1; all tables and figures are located in the Appendix following the Literature Cited). Although these three primary clades were each supported by high bootstrap values, there was little to no phylogenetic resolution within clades. An unexpected result of that study was that composition of the three clades contrasted sharply with previously defined groups based on morphological characters (Waugh, 1899, 1901; Wight, 1915; Rehder, 1940; see Fig. 3 in Shaw and Small, 2004) — each of the three clades contained taxa that are morphologically more similar to taxa of other clades. For example, *P. umbellata* var. *umbellata* and its putative variety *P. umbellata* var. *injucunda*, which are only distinguished from each other by pubescence differences, appeared in different clades. Comparison of the species within the Chickasaw clade (Fig. 1, see also Fig. 3 of Shaw and Small, 2004) reveals another example of the discontinuity between the cpDNA phylogeny and classical morphological groups. Of six species within the Chickasaw clade, the 4375 bp of noncoding cpDNA sequences were identical with the exception of two autapomorphies. While some of the taxa within the Chickasaw clade are nearly identical morphologically (e.g., *P. angustifolia* and *P. munsoniana* or *P. umbellata* var. *umbellata* and *P. alleghaniensis* var. *alleghaniensis*), *P. gracilis* and especially *P. nigra* are easily distinguished from the rest of the taxa in the clade. It is also easy to differentiate between *P. angustifolia* or *P. munsoniana* and *P. umbellata* var. *umbellata* or *P. alleghaniensis* var. *alleghaniensis*. The lack of genetic divergence within these three primary clades coupled with the obvious discontinuity with respect to longstanding morphological groups leads to the question of whether or not chloroplast lineages are monophyletic within the taxa of section *Prunocerasus*. This question can be addressed though the additional sampling of multiple accessions of each species within the section in a phylogeographic-level sampling scheme.

While it has been recognized that intraspecific cpDNA variation may be common in higher plants (Harris and Ingram, 1991; Soltis et al., 1992), relatively few studies have addressed the frequency of chloroplast sharing among species in closely related groups. To date most phylogenetic investigations have used the strategy of sampling one or few representatives per species (as opposed to "population" studies that involve sampling multiple accessions within a single species or closely related species pair). In sampling a single representative per species, one makes the following assumptions: (1) the study species are monophyletic, (2) the study species represent genetically and reproductively isolated lineages that are not reticulating through introgressive hybridization (which is important because the lack of recombination of the chloroplast genome means that it introgresses as a linked block and can therefore be a source of error in molecular systematics [Doyle, 1992; Rieseberg and Wendel, 1993; Riesberg et al., 1996]), (3) the boundaries between the study species are well characterized, and (4) the exemplar specimens are accurately identified. Therefore, in closely related groups of species, where one or more of these assumptions may be violated, sampling multiple individuals from each of the representative species is the only means of accurately assessing phylogenetic relationships. For example, Whittemore and Schaal (1991) cautioned that a single accession of each species could have been misleading in studying hybridizing oak

species. Funk and Omland (2003, p. 414) wrote that ideally one should "include all species believed a priori to be closely related (e.g., congeners), maximize the geographic diversity of samples and the number of samples collected from areas of sympatry between study species [this same suggestion was made by Matos and Schaal (2000)], and sample broadly from known sources of biological variation (subspecies, ecotypes, morphological variants, etc.)." Furthermore, Funk and Omland (2003) call for a new tradition of "congeneric phylogeography" where phylogenetics and phylogeography are united yielding a more population-level sampling scheme that will improve the resolution of evolutionary relationships among closely related entities.

While the concept of phylogeography (Avise, 1987) has had a large impact on research in animal systems over the last 18 years it has been relatively slowly applied to botanical systems because of the relatively slow nucleotide substitution rate of organellar DNA in plants (Schaal et al., 1998). This may be remedied to some extent once more informative regions of the cpDNA molecule have been better characterized (e.g., Shaw et al., 2005). Botanical phylogeography has successfully been applied to plant systems in Eurasia/North Africa (e.g., Comes and Abott, 2001; Besnard et al., 2002; Palme and Vendramin, 2002; Burban and Petit, 2003; Dane et al., 2003; Grivet and Petit, 2003; Palme et al., 2003) and Japan and East Asia (Chiang et al., 2001; Fujii et al., 2002; Honjo et al., 2004; Kanno et al., 2004). Comparatively, the botanical phylogeography of North America has been much less studied, and of those studies most have been restricted either in the number of species or in geography (e.g., Parks et al., 1994; Sewell et al., 1996; Manos et al., 1999; Tremblay and Schoen, 1999; Abbott et al., 2000; Sanjur et al., 2002; Dobes et al., 2004; JØrgensen and Mauricio, 2004).

Most studies to investigate interspecific versus intraspecific cpDNA variation are "population" in nature and confined to relatively few species (e.g., Mason-Gamer et al., 1995; Comes and Abbott, 1998, 1999; Maskas and Cruzan, 2000; Matos and Schaal, 2000), but a few studies have compared interspecific versus intraspecific cpDNA variation among a number of hybridizing species (e.g., Whittemore and Schaal, 1991; Bain and Jansen, 1997; Dumolin-Lapegue et al., 1997; Manos et al., 1999; Comes and Abbott, 2001; Gardner et al., 2004; Kanno et al., 2004). A common thread throughout these studies is the occurrence of chloroplast sharing among closely related species. Furthermore, it has been shown that if hybridization is frequent, chloroplasts may be distributed geographically instead of taxonomically (*Quercus*: Whittemore and Schaal, 1991; *Quercus*: Dumolin-Lapegue et al., 1997; *Eucalyptus*: Jackson et al., 1999; the *Pinus montezumae* complex: Matos and Schaal, 2000).

The present study addresses two issues in *Prunus* subgenus *Prunus* section *Prunocerasus*. Hybrid speciation has been hypothesized to occur within the genus *Prunus* (Watkins, 1976; Brettin et al., 2000; Mohanty et al., 2000) in general, and specifically in section *Prunocerasus* (Waugh, 1899, 1901; Steyermark, 1963). Because the species of section *Prunocerasus* are capable of natural hybridization (Hedrick et al., 1911; Wight, 1915; Flory, 1938; Rehder, 1940), we sampled extensively within each of the putative taxa to determine if the 3 primary cpDNA haplotypes observed in Shaw and Small (2004) (Fig. 1) are confined to the taxa in which they were initially observed. The second issue is phylogeographic in nature. Because we are sampling heavily within each

of the putative taxa of section *Prunocerasus*, we can analyze the distribution of cpDNA haplotypes to assess geographic structure in their distribution.

MATERIALS AND METHODS

***Taxon sampling*—**Using both wild collected and herbarium specimens (TENN, APSU, BRIT, MICH, BH), leaf material was selected from multiple accessions of each taxon in an attempt to sample each taxon throughout its respective range. A total of 207 accessions from 18 taxa were included (see Table 1). All of the commonly accepted members of section *Prunocerasus* were sampled, according to: Small (1933); Rehder (1940); Fernald (1950); Bailey and Bailey (1941); Blackburn (1952); Gleason (1952); Steyermark (1963); Radford et al. (1968); Correll and Johnston (1970); Duncan and Duncan (1988); Godfrey (1988); Wunderlin (1988); Gleason and Cronquist (1991); Smith (1994); Wofford and Chester (2002). A single sample of the Federally Endangered *P. geniculata* was obtained from Carl Weakley at the Archbold Biological Station in Lake Placid, Florida.

Much of the material for this study was taken from herbarium specimens. At date of DNA isolation, herbarium material ranged in age from 6 to 82 years with an average age of 34 years. Herbarium vouchers were labeled with a material notation for future reference. Voucher information for each of the accessions is in Table 1.

***DNA sequences*—**The *rpL16* intron region was chosen because Shaw and Small (2004) showed it to contain at least one synapomorphic character for each of the three primary clades revealed in that study. Also, the characters identifying the three primary clades are all within approximately 650 bp from the 3′ end of the intron and could therefore be obtained by one sequencing reaction using the rpL1516R primer (see Laboratory procedures, below).

***Laboratory procedures*—**DNA was extracted from leaves using the DNeasy Plant Mini Kit (Qiagen, Valencia, California, USA). The polymerase chain reaction (PCR) was performed using Eppendorf Mastercycler gradient or Mastercycler personal thermal cyclers in 50 µl volumes with the following reaction components: 1 µL template DNA (~ 10-100 ng), 1X *ExTaq* buffer (PanVera / TaKaRa, Madison, Wisconsin, USA), 200 mmol/L each dNTP, 3.0 mmol/L $MgCl_2$, 0.1 µmol/L each primer, and 1.25 units *ExTaq* (PanVera / TaKaRa). Reactions included bovine serum albumin at a final concentration of 0.2 mg/mL, which is known to improve amplification from difficult templates.

PCR and sequencing primers for the *rpL16* intron **rpL16F71** (GCT ATG CTT AGT GTG TGA CTC GTT G) and **rpL16R1516** (CCC TTC ATT CTT CCT CTA TGT TG) are from Small et al. (1998) and Shaw et al. (2005). The PCR protocol described below was preceded by template DNA denaturation at 80°C for 5 minutes and followed by a final extension step of 5 min at 65°C. The PCR cycling conditions were 30 cycles of: denaturation at 95°C for 1 min, primer annealing at 50°C for 1 min, followed by a ramp of 0.3°C/sec to 65°C, and primer extension at 65°C for 4 min.

PCR products were checked on 1% agarose gels before being cleaned with ExoSAP-IT (USB, Cleveland, Ohio, USA). In most cases DNA sequencing was performed using only the reverse primer rpL16R1516, however in a few cases the forward primer was

needed. All DNA sequencing was performed with the ABI Prism BigDye Terminator Cycle Sequencing Ready Reaction Kit, v. 2.0 or 3.1 (Perkin-Elmer / Applied Biosystems, Foster City, California, USA) and electrophoresed and detected on an ABI Prism 3100 automated sequencer (University of Tennessee Molecular Biology Resource Facility). Sequencher 3.1.1 (Gene Codes) was used to edit the DNA strands. All sequences have been deposited in GenBank; accession numbers AY773482-AY773670 and AY500638-AY500655 and are listed in Table 1.

***Data analysis*—**Alignment of DNA sequences was initially performed with ClustalX (Thompson et al., 2001), with subsequent manual adjustment by eye in MacClade v. 4.0 (Sinauer, Sunderland, Massachusetts, USA). Variable positions in the data matrix were double checked against the original chromatogram files to make sure that all base calls were true at all variable positions. In all cases, alignment of potentially informative positions was unambiguous. Indels were coded as binary characters except in the case of a single poly-A/T run, which was omitted from the data set as it may be PCR artifact and not reflective of the phylogenetic history of the group.

Analysis of phylogenetic relationships was conducted using the optimality criterion of maximum parsimony. Searches for most-parsimonious trees were executed in PAUP* v. 4.0 b10 (Swofford, 2002) by a heuristic search with tree bisection-reconnection (TBR) branch swapping and 1000 random sequence addition replicates with the "collapse zero-length branches" option in effect. Bootstrap support (Felsenstein, 1985) was estimated with 1000 replications of heuristic search and simple taxon addition with the constraint of 10 000 000 rearrangements per replicate. Both the consistency and retention indices (CI and RI, respectively) were used to assess the amount of homoplasy present in the data set.

Relationships among the chloroplast haplotypes were also inferred using the software TCS v. 1.13 (Clement et al., 2000), which implements a statistical parsimony approach to estimating gene genealogies. TCS was performed as an alternative means of analysis because it can infer ancestral or intermediate haplotypes (as opposed to assuming that these haplotypes are extinct). For the TCS analysis we used the same sequence alignment that was used in the PAUP* analysis described above.

***Geographic distribution of haplotypes*—**To assess the distribution of haplotypes within each of the putative taxa of section *Prunocerasus*, distribution maps were created using range information from USDA national PLANTS database (USDA, 2002), Little (1971, 1976, 1977), Gleason and Cronquist (1991), and Correll and Johnston (1970). Haplotypes found within each of the taxa were then mapped on to the distribution maps. Additional maps were created irrespective of taxon to show overall haplotype distribution.

RESULTS

RpL16 *intron analysis and inference of chloro-haplotypes*—The aligned data set consisted of 797 bp from the 3′ end of the ~ 1 kb *rpL16* intron. Within the aligned matrix there were 23 parsimony informative characters (including multi-bp indels coded as binary characters) and 10 variable but parsimony uninformative characters. In the maximum parsimony analysis, a heuristic search found three equally parsimonious trees

of 34 steps with high consistency and retention indices, 0.97 and 0.99, respectively. Bootstrap values are shown above the branches in Figs. 2a-2c. The topology of the genealogy generated with TCS (Fig. 3) is consistent with the phylogeny generated with PAUP* (Fig. 2).

Twenty-three unique haplotypes were identified from the 207 section *Prunocerasus* accessions (Fig. 3). Ten different haplotypes were associated with the American clade, two haplotypes with the Beach clade, and seven haplotypes with the Chickasaw clade. Additionally, one Texana haplotype, one Subcordata haplotype, and one peculiar Umbellata haplotype were identified (Fig. 3). The primary haplotypes: Americana, Beach, Chickasaw, Subcordata, Texana, and the peculiar Umbellata are hereafter referred to as A, B, C, S, T, and pU, respectively. The A and C haplotypes were found in the highest frequencies. The observed haplotype for each accession is shown in Table 1.

***Phylogenetic results and haplotypic distribution*—**Because 207 accessions could not be placed on the same figure, the phylogeny of section *Prunocerasus* is shown in the strict consensus tree in Figs. 2a-2c. Each of the three figures details one of the three primary clades. Within section *Prunocerasus*, the backbone of the tree shows an initial split in the section between the northwestern species, *P. subcordata*, and the remaining species. A single accession of *P. umbellata* var. *umbellata* was placed in a grade between *P. subcordata* and *P. texana* (see also the gene genealogy in Fig. 3). This haplotype is unique and therefore interesting; despite other sampled accessions being from the same general geographic location, this haplotype was observed only once. Four accessions of *P. texana*, currently classified in subgenus *Amygdalus* (Wight, 1913), were positioned with strong support as sister to the remaining taxa. The 199 remaining accessions segregated into three clades (A=American, B=Beach and C=Chickasaw) that are each strongly supported. These three primary clades directly correspond to those found in Shaw and Small (2004) and the exemplars used in that study are denoted in Figs. 2a-2c with an asterisk. Within each of the three primary clades, relationships are weakly resolved because of the lack of sequence divergence. However, the three figures show that many of the taxa of section *Prunocerasus* are para- or polyphyletic. Representatives of most taxa were found in two or three of the primary clades (compare Figs. 2a-2c or see Table 1). In fact, of those taxa of the three primary clades, *P. angustifolia* and *P. alleghaniensis* var. *davisii* are the only ones where multiple accessions all resolve in the same clade.

The geographical position of accessions and their observed haplotypes were mapped and are shown in Figs. 4a-4k. The individual taxon maps show that most taxa in the section contain more than one cpDNA haplotype. Only two taxa, *P. subcordata* and *P. texana*, were found to possess unique haplotypes that were not found in any other taxon (Fig. 4a), excluding the one peculiar Umbellata (pU) haplotype found in one accession of *P. umbellata*.

DISCUSSION

***Phylogenetic relationships within section* Prunocerasus—**In a previous study (Shaw and Small, 2004) we reported that the majority of section *Prunocerasus* species

resolve in three primary clades sister to *P. subcordata* (see Fig. 1). Because the relationships revealed in that study were incongruous with classical morphological assemblages (Waugh, 1899, 1901; Wight, 1915; Rehder, 1940) we were motivated to survey several accessions of each of the taxa to test taxon monophyly. Here we report the organellar para- and polyphyly of several of the taxa of section *Prunocerasus*, the North American plums.

Prunus subcordata is unique among the species of section *Prunocerasus* because it is the only one with a western North American distribution. All of the other taxa in the section have ranges east of the Rocky Mountains. The sampled accessions of *P. subcordata* have identical *rpL16* sequences, the S haplotype (Fig. 4a). This haplotype was supported by three characters and was not observed in any other taxon. Four accessions of *Prunus texana*, a species endemic to Texas with pubescent fruits that is currently classified as a peach in subgenus *Amygdalus* (Wight, 1913), had identical *rpL16* sequences, haplotype T (Fig. 4a). Two synapomorphic characters support the monophyly of *P. texana* and this clade is sister to the remaining species, phylogenetically inside of *P. subcordata* and a single *P. umbellata* var. *umbellata* accession (haplotype pU). More work needs to be done to determine whether *P. texana* is a pubescent-fruited plum or contains a plum chloroplast in peach clothing.

Prunus geniculata is a federally endangered species restricted to four counties in the sand-pine scrub habitat of the Lake Wales Ridge region of central Florida — an ancient dune system habitat known to harbor many endemic species in addition to *P. geniculata* (Christman and Judd, 1990; Dobson et al., 1997). This species was only sampled once because of its protected status and the rarity of herbarium material. The single accession contained the B haplotype (Fig. 4a), but there is an autapomorphy (possible synapomorphy?) in *P. geniculata* just beyond where the data set had to be trimmed. This is worth mentioning because only two different B haplotypes (B and B1) were identified, compared to seven different C and ten different A haplotypes. This possible additional B haplotype is also significant because it may prove useful in our future work concerning gene flow and hybridization of this federally protected species. Another taxon that was only sampled once is *P. maritima* var. *gravesii* (Fig. 4a). This is a particularly interesting taxon because since its discovery in ~1894 it has only been known as a single clonal individual near Groton, Connecticut, USA (Anderson, 1980). *Prunus maritima* var. *gravesii* is currently presumed extirpated from the wild and the single accession of this taxon contained the B haplotype, a haplotype shared by several (but not all) accessions of *P. maritima* var. *maritima*. Aside from *P. geniculata* and *P. maritima* var. *gravesii*, only *P. angustifolia* and *P. alleghaniensis* var. *davisii*, were invariant and both contained only the C haplotype (Figs. 4b and 4d).

The remaining 172 accessions of 12 taxa are scattered among the three primary clades (Fig. 2a-2c). Within each of these clades relationships are weakly resolved because of the lack of sequence divergence. Even still, this study clearly shows that chloroplast sharing is occurring in most of the North American plum taxa of section *Prunocerasus.* Examination of the phylogeny (Figs. 2a-c) and the haplotype maps (Figs. 4a-4k) for many of the taxa reveals that most taxa contain more than one of the three primary haplotypes. All three primary haplotypes were found in *P. gracilis* (Fig. 4e). The three primary haplotypes plus one basal haplotype (pU) were found in *P. umbellata* (including

P. u. var. *injucunda*) (Fig. 4k). Haplotypes A and C were found in *P. alleghaniensis* var. *alleghaniensis* (Fig. 4b), *P. americana* var. *americana* and *P. americana* var. *lanata* (Fig. 4c), *P. hortulana* (Fig. 4f), *P. mexicana* (Fig. 4g), *P. munsoniana* (Fig. 4h), *P. nigra* (Fig. 4i), and *P. rivularis* (Fig. 4j). Haplotypes B and C were found in *P. maritima* var. *maritima* (Fig. 4a).

The findings of this study prompt speculation regarding the distinctiveness of several of the taxa in the section. Many workers have commented on the inability to reliably morphologically distinguish many of the species. For example, Duncan and Duncan (1988, p. 303) wrote that *P. umbellata* "varies considerably and has been divided into varieties by others, based mainly on pubescence. Under such a division *P. mitis* Beadle has twigs glabrous and leaves hairy on both surfaces, *P. injucunda* Small has twigs and both surfaces of the leaves hairy, and *P. umbellata* has twigs and leaves glabrous or leaves hairy only on the lower surface. Much like *P. alleghaniensis* and apparently only separated by their different distributions." In another example, Robertson (1974, p. 659) wrote of *P. americana* that "plants from the southern part of the range westward to Texas and northern Mexico have the leaves pubescent below to varying degrees, the plants evidently do not often produce root suckers, and the fruit is bluish to purple-red; these plants perhaps are best placed in *P. mexicana* [incl. *P. americana* var. *lanata*]. These two species need to be studied in detail to ascertain their distinctiveness and distribution." Because of the level of para- and polyphyly observed in many of the taxa in section *Prunocerasus*, and the lack of distinctive morphological characters in the section, we feel that taxonomic "lumping" may be in order. We are currently accumulating a nuclear DNA data set to gather more evidence to further address this question.

Within several of the 12 taxa that have more than one haplotype, there is a haplotype that is found with a higher frequency (Table 1). For example, 35 accessions of *P. americana* var. *americana* contain the A haplotype (resolve in the American clade) while six accessions have the C haplotype and resolve in the Chickasaw clade. If we generate a phylogeny based on the most commonly found haplotype in each taxon, the phylogeny more closely matches classical morphological assemblages than the one illustrated in Fig. 3 of Shaw and Small (2004) (compare Fig. 5 to Fig. 1, which is modified from Shaw and Small [2004]). Of the 12 taxa with more than one haplotype, two (*P. nigra* and *P. hortulana*) had haplotypic frequencies close to 50% and are therefore shown in Fig. 5 with dashed lines as they could have been positioned in either the American or Chickasaw clades. This is noteworthy in the context of their morphology. *Prunus nigra* shares morphological characters with some members of the American clade, *P. americana* var. *americana* and *P. mexicana*—all three are trees, have relatively large flowers, and similar leaf shape. These three species were allied by earlier workers (Waugh 1899, 1901; Wight 1915). But, *P. nigra* also has glandular leaf teeth and calyx lobes, characteristics shared by *P. angustifolia* and *P. munsoniana* of the Chickasaw clade. Because *P. nigra* shares morphological characters with the dominant members of both the American and the Chickasaw clades, it is not surprising that *P. nigra* contains the American and Chickasaw chloroplast types in equal frequency.

In *P. hortulana* we also found an approximately equal frequency of A and C haplotypes and *P. hortulana* is morphologically intermediate between several section *Prunocerasus* species — in fact, Waugh (1899, 1901) thought that this species was the

center taxon of a hybrid swarm and Steyermark (1963) noted that some individuals appear to be hybrids between it and *P. mexicana*.

In our earlier study, *P. maritima* var. *maritima* and *P. maritima* var. *gravesii* were shown to be sister to *P. geniculata* in the Beach clade (Fig. 1). These results were unexpected because *P. maritima* and *P. geniculata* are morphologically very different. *Prunus geniculata* is a diminutive species unlike any other in the section and is morphologically closer to *P. angustifolia* than to any other taxon in the section. Furthermore, *P. maritima* var. *maritima* is morphologically closer to *P. umbellata* var. *umbellata*, *P. alleghaniensis* var. *alleghaniensis*, and *P. gracilis*, and was allied to these species by earlier workers (Waugh 1899, 1901; Wight 1915). In this study we sampled 14 accessions of *P. maritima* var. *maritima*, nine of which contained the C haplotype and five had the B haplotype (Fig. 4a). Interestingly, while most *P. maritima* var. *maritima* accessions contain the C haplotype and morphologically this species more closely resembles the species of the Chickasaw clade, the variety *P. maritima* var. *gravesii* (known as a single clonal individual since ~1894) contains the B haplotype and is geographically embedded in the *P. maritima* var. *maritima* B haplotype region.

***Phylogeographic inference*—**The strength of the phylogeographic approach stems from an increased sampling strategy, as compared to most phylogenetic studies. This allows evolutionary questions to be addressed in more detail and to be placed in a geographical context. The phylogeographic approach will doubtless further our knowledge of closely related species, like the North American plums, where factors like hybridization, introgression, or lineage sorting can mask "true" evolutionary relationships in studies where sampling is not thorough. This study highlights the need for a new tradition of "congeneric phylogeography" (Funk and Omland, 2003) where phylogenetics and phylogeography are joined yielding a more population-level sampling scheme that will improve the resolution of evolutionary relationships among closely related biological entities.

Prunus is a morphologically diverse genus of the northern hemisphere with about 200 species of small trees or shrubs (Rehder, 1940; Robertson, 1974). Although three broad phylogenetic studies have focused on *Prunus* (Bortiri et al., 2001; Lee and Wen, 2001; Shaw and Small, 2004), none have adequately sampled outside *Prunus* to establish a root and pinpoint the center of origin for the genus. Furthermore, in studies spanning Rosaceae, *Prunus* has been allied with a clade that includes *Exochorda* Lindl., *Oemleria* Reichb., and *Prinsepia* Royle (other genera of Takhtajan's [1997] Amygdaloideae) (Morgan et al., 1994), within an array of "spiraeoid" lineages, or as sister to Maloideae s.l. (Potter, 2003). Without knowledge of the sister taxon to *Prunus*, the phylogeography of the North American plum section of the genus cannot be unequivocally established. Nevertheless, based on hybridization studies and a broad knowledge of the genus, Watkins (1976) hypothesized that *Prunus* originated in central Asia. This hypothesis is consistent with inferences from phylogenetic analyses (Bortiri et al., 2001; Lee and Wen, 2001; Shaw and Small, 2004) although they do not support it to the exclusion of other hypotheses. In the context of Watkins' (1976) hypothesis, immigration of the North American plums to North America may have occurred via the Bering land bridge (Fig. 7). This inference is supported by our earlier study (Shaw and Small, 2004), which showed

that *P. subcordata*, the only western species, is sister to all of the other species in section *Prunocerasus*. Furthermore, *P. subcordata* is the only section *Prunocerasus* taxon that contains a haplotype other than A, B, or C that was found more than once; suggesting it has been reproductively isolated (geographically) for longer periods of time than the other taxa. The pU haplotype, which was positioned in a grade between *P. subcordata* and *P. texana*, was only found once in a single *P. umbellata* var. *umbellata* accession in northwest Loiusiana. This haplotype is noteworthy because it is a possible link to the past—before the *P. texana* speciation event or before *P. texana* acquired a "plum" chloroplast. Currently, we have insufficient data to further address questions surrounding this peculiar haplotype. Phylogenetically inside of *P. subcordata* and the pU accession of *P. umbellata* and sister to the remaining species is *P. texana*, which is endemic to south central Texas. In the context of geography, this phylogenetic grade may suggest migration of section *Prunocerasus* to North America via Beringia and from western North America to eastern North America through a south central corridor (Fig. 7), although these data do not support this hypothesis to the exclusion of an eastern migration to and across North America.

Multiple accessions of the remaining taxa are scattered among three different clades whose relationships to one another are unresolved despite sequencing seven noncoding cpDNA regions (4375 aligned nucleotide positions) (Shaw and Small, 2004). This polytomy may be "real" (a hard polytomy) in the sense that all three primary haplotypes appear to be very close in age. Based on the philosophy that refugial plant populations have higher levels of genetic diversity (Demesure et al., 1996; Dumolin-Lapegue et al., 1997; King and Ferris, 1998), and because the three primary haplotypes were only found intermingled around eastern Texas (along with *P. texana*), this area appears to have been a refugium, or point of origin, for many section *Prunocerasus* members. Based on the evidence presented in the previous paragraphs, we speculate that the North American plum lineage entered North America via the Bering land bridge, migrated southeastward toward Texas, and subsequently radiated through eastern North America in a manner illustrated in Fig. 7.

Over the last 2 million years there have been ~16-20 cycles of glaciation during which plants species underwent southern migrations and northern reintroductions (Hays et al., 1969; Davis, 1983). Cyclical glaciation could account for multiple bouts of range expansion and divergence followed by range contraction and hybridization. This may explain both the morphological complexity and the randomness of the taxonomic distribution of the three primary haplotypes.

Even though individual taxon haplotype maps reveal no obvious pattern, a clear pattern emerges if all of the haplotypes are mapped together regardless of the taxa in which they were observed (Figs. 6a,b). The A (American) haplotype was found in 11 taxa (*P. alleghaniensis* var. *alleghaniensis, P. americana* var. *americana, P. americana* var. *lanata, P. gracilis, P. hortulana, P. mexicana, P. munsoniana, P. nigra, P. rivularis, P. umbellata* var. *umbellata, P. umbellata* var. *injucunda*) (Figs. 4b, 4c, 4e-4k). This haplotype is mostly confined to the interior of the USA. The B (Beach) haplotype was only found in eastern central Texas, the panhandle of Florida, and the easternmost portions of Cape Cod, Massachusetts, and Long Island, New York (Fig. 6a). This haplotype was found in five different taxa: *P. geniculata, P. gracilis, P. maritima var.*

maritima, P. maritima var. gravesii, and *P. umbellata* var. *umbellata* (Figs. 4a, 4e, 4k). The distribution of the B haplotype is engaging because it was only found in restricted, isolated areas of the Atlantic and Gulf Coastal Plain floristic province (Fig. 6a).

The C (Chickasaw) haplotype, the most abundant haplotype observed in this study and was observed in 14 taxa (*P. alleghaniensis* var. *alleghaniensis, P. alleghaniensis* var. *davisii, P. americana* var. *americana, P. americana* var. *lanata, P. angustifolia, P. gracilis, P. hortulana, P. maritima* var. *maritima, P. mexicana, P. munsoniana, P. nigra, P. rivularis, P. umbellata* var. *umbellata, P. umbellata* var. *injucunda*) (Figs. 4a-4k). The distribution of the C haplotype (Fig. 6b) follows a distribution pattern that has long intrigued botanists — the Atlantic and Gulf Coastal Plain disjunctions to the Great Lakes region. Peattie (1922) wrote that there were approximately 60 species that fit this distribution pattern and that most are of conservative habits and habitats (not weedy and therefore not likely to have spread there because of human-mediated introductions). While explanation for this distribution pattern is still unclear, Reznicek (1994) proposed that many species migrated into the Great Lakes Region through dispersal jumps between areas of suitable habitat created along major postglacial drainage channels. It is interesting to point out that the distribution of the C haplotype, although found in 14 taxa in section *Prunocerasus*, is nearly identical to the distribution of *Rhexia virginica* (see Fig. 4 in Reznicek, 1994). This distribution was also recently observed in *Trillium grandiflorum* cpDNA (Griffin and Barrett, 2004). Furthermore, according to Jackson and Singer (1997), species whose ranges are disjunct between the Coastal Plain and the Great Lakes region are typically rare or absent in the North American interior, a pattern that is also congruent with the distribution of the C haplotype in section *Prunocerasus*.

Plum genetic structure is not only related to the occurrence of refugial zones and particular biogeographic conditions of North America, but also to human influence in terms of dispersal or disturbance. In the early stages of this study, we were concerned that the inconsistency of haplotypes observed (while we were accumulating this data set) may have been due to human mediated dispersal and hybridization. However, it is our impression that sampling was thorough enough to place confidence in the larger patterns observed in the data. Because plums have undoubtedly been spread by humans we are less inclined to comment on the less frequent haplotypes and their distributions.

Broader ramifications for phylogenetics—Earlier workers cautioned that, in phylogenetic studies of closely related species, using only a single accession of each species could be misleading (Whittemore and Schaal, 1991; Matos and Schaal, 2000; Funk and Olmland, 2003), and they recommended more thorough sampling to circumvent potential problems. This study highlights that reality. In closely related groups of species, sampling multiple individuals from each of the representative taxa is the only means of accurately assessing phylogenetic relationships.

In *Prunus* section *Prunocerasus*, 14 of 18 taxa contained more than one chloro-haplotype. In hindsight, the possibility existed that in our earlier study we could have chosen different combinations of exemplars, each of which could have resulted in a different phylogeny. In fact, because nine taxa contain both the A and C haplotypes, one taxon contains the A and B haplotypes, one taxon contains the A, B, and C haplotypes, and one taxon contains the A, B, C, and pU haplotypes, the possibility exists that we

could have generated 12 288 different "species" trees depending on the accessions chosen.

The high degree of chloroplast sharing that was observed in this study may be more common than currently recognized. Chat et al. (2004) found that five of eight species of Kiwi (*Actinidia*, Actinidiaceae) were polyphyletic in their cpDNA. In another example, McKinnon et al. (1999) showed that five of seven closely related *Eucalyptus* (Myrtaceae) species were polyphyletic in their cpDNA and Jackson et al. (1999) showed that cpDNA variation did not conform with subspecies boundaries but was instead geographically distributed in *Eucalyptus globulus* Labill. The results of these studies highlight the need for increased sampling in phylogenetic studies of closely related taxa.

Conclusions—Four main conclusions can be drawn from this study. The first concerns the para- and polyphyly of chloro-haplotypes observed in 12 of 17 taxa in *Prunus* subgenus *Prunus* section *Prunocerasus*. Although lineage sorting from a polymorphic ancestor cannot be ruled out as the cause for the observed pattern, we feel that a hypothesis of chloroplast sharing through past and current hybridization is a more likely explanation because most of the species in the section are known to be able to hybridize. Secondly, four accessions of the Texas Peachbush, *P. texana*, all contain the same haplotype (T), which is strongly supported as being embedded within section *Prunocerasus* species. Our future nDNA work may help to further illuminate the taxonomic position of this species. Third, the data presented in this study are consistent with a biogeographic hypothesis that the ancestors of section *Prunocerasus* entered North America via the Bering land bridge, migrated south and east, and subsequently radiated from Texas throughout eastern North America. Lastly, and perhaps the most important general conclusion, this study highlights the need for more thorough sampling in phylogenetic investigations. Most phylogenetic studies utilizing cpDNA data commonly include one or a few individual(s) per taxon; therefore, in closely related groups, the true phylogeny might not be reflected but rather the obscuring effects of hybridization and introgression or lineage sorting from ancestral polymorphisms.

LITERATURE CITED

ABBOTT, R. J., L. C. SMITH, R. I. MILNE, R. M. M. CRAWFORD, K. WOLFF, J. BALFOUR. 2000. Molecular analysis of plant migration and refugia in the arctic. *Science* 289: 1343-1346.

ANDERSON, G. J. 1980. The status of the very rare *Prunus gravesii* Small. *Rhodora* 82: 113-129.

AVISE, J. C., J. ARNOLD, R. MARTIN 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.

BAILEY, L. H., AND E. Z. BAILEY. 1941. Hortus Second: a concise dictionary of gardening, general horticulture, and cultivated plants in North America. The Macmillan Company, New York, New York, USA.

BESNARD, G., B. KHADIRI, P. BARADAT, AND A. BERVILLE. 2002. *Olea europaea* (Oleaceae) phylogeography based on chloroplast DNA polymorphism. *Theoretical and Applied Genetics* 104: 1353-1361.

BIRKY, C. W. JR. 2001. The inheritance of genes in mitochondria and chloroplasts: Laws, Mechanisms, and Models. *Annual Review of Genetics* 35: 125-148.

BLACKBURN, B. 1952. Trees and Shrubs in eastern North America: keys to the wild and cultivated woody plants growing in the temperate regions exclusive of conifers. Oxford University Press, New York, USA

BORTIRI, E., S. H. OH, J. G. JIANG, S. BAGGETT, A. GRANGER, C. WEEKS, M. BUCKINGHAM, D. POTTER, AND D. E. PARFITT. 2001. Phylogeny and systematics of *Prunus* (Rosaceae) as determined by sequence analysis of ITS and the chloroplast *trnL-trnF* spacer DNA. *Systematic Botany* 26: 797-807.

BRETTIN, T. S., R. KARLE, E. L. CROWE, AND A. F. IEZZONI. 2000. Chloroplast inheritance and DNA variation in sweet, sour, and ground cherry. *The Journal of Heredity* 91: 75-79.

BURBAN, C., AND R. J. PETIT. 2003. Phylogeography of maritime pine inferred with organelle markers having contrasted inheritance. *Molecular Ecology* 12: 1487-1495.

CHAT, J., B. JAUREGUI, R. J. PETIT, AND S. NADOT. 2004. Reticulate evolution in Kiwifruit (*Actinidia*, Actinidiaceae) identified by comparing their maternal and paternal phylogenies. *American Journal of Botany* 91: 736-747.

CHIANG, T. Y., Y. C. CHIANG, Y. J. CHEN, C. H. CHOU, S. HAVANOND, T. N. HONG, AND S. HUANG. 2001. Phylogeography of *Kandelia candel* in East Asiatic mangroves based on nucleotide variation of chloroplast and mitochondrial DNAs. *Molecular Ecology* 10: 2697-2710.

CHIANG, T. Y., AND B. A. SCHAAL. 1998. Phylogeography of North American populations of the moss species *Hylocomium splendens* based on the nucleotide sequence of internal transcribed spacer 2 of nuclear ribosomal DNA. *Molecular Ecology* 8: 1037-1042.

CHRISTMAN, S. P., AND W. S. JUDD. 1990. Notes on plants endemic to Florida scrub. *Florida Scientist* 53: 52-73.

CLEMENT, M., D. POSADA, AND K. A. CRANDALL. 2000. TCS: a computer program to estimate gene genealogies. *Molecular Ecology* 9: 1657-1659.

COMES, H. P., AND R. J. ABBOTT. 1998. The relative importance of historical events and gene flow on the population structure of a Mediterranean Ragwort, *Senecio gallicus* (Asteraceae). *Evolution* 52: 355-367.

COMES, H. P., AND R. J. ABBOTT. 1999. Population genetic structure and gene flow across arid versus mesic environments: a comparative study of two parapatric *Senecio* species from the near east. *Evolution* 53: 36-54.

COMES, H. P., AND R. J. ABBOTT. 2001. Molecular phylogeography, reticulation, and lineage sorting in Mediterranean *Senecio* section *Senecio* (Asteraceae). *Evolution* 55: 1943-1962.

CORRELL, D. S., AND M. C. JOHNSTON. 1970. Manual of the vascular plants of Texas. Texas Research Foundation, Renner, Texas, USA.

DANE, F., P. LANG, AND R. BAKHTIYAROVA. 2003. Comparative analysis of chloroplast DNA variability in wild and cultivated *Citrullus* species. *Theoretical and Applied Genetics* 108: 958-966.

DAVIS, M. B. 1983. Quaternary history of deciduous forests of eastern North America and Europe. *Annals of the Missouri Botanical Garden* 70: 550-563.

DEMESURE, B., COMPS, B., AND R. J. PETIT. 1996. Chloroplast DNA phylogeography of the Common Beech (*Fagus sylvatica* L.) in Europe. *Evolution* 50: 2515-2520.

DOBES, C. H., T. MITCHELL-OLDS, AND M. A. KOCH. 2004. Extensive chloroplast haplotype variation indicates Pleistocene hybridization and radiation of North American *Arabis drummondii, A. x divaricarpa,* and *A. holboelii* (Brassicaceae). *Molecular Ecology* 13: 349-370.

DOBSON, A. P., J. P. RODRIGUEZ, W. M. ROBERTS, AND D. S. WILCOX. 1997. Geographic distribution of the endangered species in the United States. *Science* 275: 550-553.

DOYLE, J. J. 1992. Gene trees and species trees: Molecular systematics as one-character taxonomy. *Systematic Botany* 17: 144-163.

DUMOLIN-LAPEGUE, S., B. DEMESURE, S. FINESCHI, V. LE CORRE, AND R. J. PETIT. 1997. Phylogeographic structure of White Oaks throughout the European continent. *Genetics* 146: 1475-1487.

DUNCAN, W. H., AND M. B. DUNCAN. 1988. Trees of the southeastern United States. University of Georgia Press, Athens, Georgia, USA.

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

FERNALD, M. L. 1950. Gray's Manual of Botany, 8th ed. American Book Company, New York, New York, USA.

FLORY, W. S. 1938. Cross sterility in hybrid plum varieties. *Proceedings of the American Society for Horticultural Science*. Contribution No. 409: 43-46.

FUJII, N., N. TOMARU, K. OKUYAMA, T. KOIKE, T. MIKAMI, AND K UEDA. 2002. Chloroplast DNA phylogeography of *Fagus crenata* (Fagaceae) in Japan. *Plant Systematics and Evolution* 232: 21-33.

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 and Evolutionary Systematics* 34: 397-423.

GARDNER, R. C., P. J. DELANG, D. J. KEELING, T. BOWALA, H. A. BROWN, AND S. D. WRIGHT. 2004. A late quaternary phylogeography for *Metrosideros* (Myrtaceae) in New Zealand inferred from chloroplast DNA haplotypes. *Biological Journal of the Linnaean Society* 83: 399-412.

GLEASON, H. A. 1952. The New Britton and Brown Illustrated Flora of the United States and Adjacent Canada, volume 2. Lancaster Press, Inc., Lancaster, Pennsylvania, USA.

GLEASON, H. A., AND A. CRONQUIST. 1991. Manual of vascular plants of northeastern United States and adjacent Canada, 2nd ed. New York Botanical Garden, Bronx, New York, USA.

GODFREY, R. K. 1988. Trees, shrubs, and woody vines of northeastern Florida and adjacent Georgia and Alabama. The University of Georgia Press, Athens, Georgia, USA.

GRIFFIN, S. R., AND S. C. H. BARRETT. 2004. Post-glacial history of *Trillium grandiflorum* (Melanthaceae) in eastern North America: inferences from phylogeography. *American Journal of Botany* 91: 465-473.

GRIVET, D., AND R. J. PETIT. Chloroplast DNA phylogeny of the hornbeam in Europe: evidence for a bottleneck at the outset of postglacial colonization. *Conservation Genetics* 4: 47-56.

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

HAYS, J. D., T. SAINTO, N. D. OPDYKE, AND L. H. BURCKLE. 1969. Pliocene-Pleistocene sediments of the equatorial Pacific: Their paleomagnetic, biostratigraphic, and climatic record. *Geological Society of America Bulletin* 80: 1481-1514.

HEDRICK, U. P., W. H. ALDERMAN, M. J. DORSEY, O. M. TAYLOR, AND R. WELLINGTON. 1911. The plums of New York. J. B. Lyon Company state printers, Albany, New York, USA.

JACKSON, H. D., D. A. STEANE, B. M. POTTS, AND R. E. VAILLANCOURT. 1999. Chloroplast DNA evidence for reticulate evolution in *Eucalyptus* (Myrtaceae). *Molecular Ecology* 8: 739-751.

JACKSON, S. T., AND D. K. SINGER. 1997. Climate change and the development of Coastal Plain disjunctions in the central Great Lakes region. *Rhodora* 99: 101-117.

JØRGENSEN, S., AND R. MAURICIO. 2004. Neutral genetic variation among North American populations of the weedy plant *Arabidopsis thaliana* is not geographically structured. *Molecular Ecology* 13: 3403-3413.

KING, R. A., AND C. FERRIS. 1998. Chloroplast DNA phylogeography of *Alnus glutinosa* (L.) Gaertn. *Molecular Ecology* 7: 1151-1161.

LEE, S., AND J. WEN. 2001. A phylogenetic analysis of *Prunus* and the Amygdaloideae (Rosaceae) using ITS sequences of nuclear ribosomal DNA. *American Journal of Botany* 88: 150-60.

LITTLE, E.L. JR. 1971. Atlas of United States Trees. Vol. 1. Conifers and important hardwoods. US Dept. of Agric. Misc. Publ. 1146.

LITTLE, E.L. JR. 1976. Atlas of the United States Trees, Vol. 3. Minor western hardwoods. US Dept. Agric. Misc. Publ. 1314.

LITTLE, E.L. JR. 1977. Atlas of the United States Trees, Vol. 4, Minor eastern hardwoods. US Dept. Agric. Misc. Publ. 1342.

MANOS, P. S., J. J. DOYLE, AND K. C. NIXON. 1999. Phylogeny, biogeography, and processes of molecular differentiation in *Quercus* subgenus *Quercus* (Fagaceae). *Molecular Phylogenetics and Evolution* 12: 333-349.

MASKAS, S. D., AND MITCHELL B. CRUZAN. 2000. Patterns of intraspecific diversification in the *Piriqueta caroliniana* complex in southeastern North America and the Bahamas. *Evolution* 54: 815-827.

MASON-GAMER, R. J., K. E. HOLSINGER, AND R. K. JANSEN. 1995. Chloroplast DNA haplotype variation within and among populations of *Coreopsis grandiflora* (Asteraceae). *Molecular Biology and Evolution* 12:371-381.

MATOS, J. A., AND B. A. SCHAAL. 2000. Chloroplast evolution in the *Pinus montezumae* complex: a coalescent approach to hybridization. *Evolution* 54: 1218-1233.

McKINNON, G. E., D. A. STEANE, B. M. POTTS, AND R. E. VAILLANCOURT. 1999. Incongruence between chloroplast and species phylogenies in *Eucalyptus* subgenus *Monocalyptus* (Myrtaceae). *American Journal of Botany* 86: 1038-1046.

MOHANTY, A., J. P. MARTÍN, AND I AGUINAGALDE. 2000. Chloroplast DNA diversity within and among populations of the allotetraploid *Prunus spinosa* L. *Theoretical and Applied Genetics* 100: 1304-1310.

MORGAN, D. R., D. E. SOLTIS, AND K. R. ROBERTSON. 1994. Systematic and evolutionary implications of *rbcL* sequence variation in Rosaceae. *American Journal of Botany* 81: 890-903.

PALME, A. E., AND G. G. VENDRAMIN. 2002. Chloroplast DNA variation, postglacial recolonization, and hybridization in hazel, *Corylus avellana*. *Molecular Ecology* 11: 1769-1779.

PALME, A. E., Q. SU, A. RAUTENBERG, F. MANNI, AND M. LASCOUX. 2003. Postglacial recolonization and cpDNA variation of silver birch, *Betula pendula*. *Molecular Ecology* 12: 201-212.

PARKS, C. R., J. F. WENDEL, M. M. SEWELL, AND Y.-L. QIU. 1994. The significance of allozyme variation and introgression in the *Liriodendron tulipifera* complex (Magnoliaceae). *American Journal of Botany* 81: 878-889.

PEATTIE, D. C. 1922. The Atlantic Coastal Plain element in the flora of the Great Lakes. *Rhodora* 24: 57-70.

POTTER, D. 2003. Molecular phylogenetic studies in Rosaceae. Pp. 319-351 in: Sharma, A. K., and Sharma, A., eds., *Plant Genome: Biodiversity and Evolution, Vol. I, Pt. A: Phanerogams*. Science Publishers, Inc., Enfield, New Hampshire, USA.

RADFORD, A. E., H. E. AHLES, AND C. R. BELL. 1968. Manual of the vascular flora of the Carolinas. The University of North Carolina Press, Chapel Hill, North Carolina, USA.

REHDER, A. 1940. Manual of cultivated trees and shrubs hardy in North America, exclusive of the subtropical and warmer temperate regions, 2d, rev. and enl. edition. The Macmillan Company, New York, New York, USA.

REZNICEK, A. A. 1994. The disjunct coastal plain flora in the Great Lakes region. *Biological Conservation* 68: 203-215.

RIESEBERG, L. H., S. M. BECKSTROM-STERNBERG, AND K. DOAN. 1996. Molecular markers incongruence in plant hybrid zones and phylogenetic trees. *Acta Botanica Neerlandica* 45: 243-262.

RIESEBERG, L. H., AND J. F. WENDEL. 1993. Introgression and its consequences in plants. Pp. 70-109 *In* R. G. Harrison, [ed.], Hybrid zones and the evolutionary process. Oxford University Press, Oxford, UK.

ROBERTSON, K. R. 1974. The genera of Rosaceae in the southeastern United States. *Journal of the Arnold Arboretum* 55: 654-662.

SANJUR, O. I., D. R. PIPERNO, T. C. ANDRES, AND L. WESSEL-BEAVER. 2002. Phylogenetic relationships among the domesticated and wild species of *Cucurbita* (Cucurbitaceae) inferred from a mitochondrial gene: Implications for crop plant evolution and areas of origin. *Proceeding of the National Academy of Sciences, USA* 99: 535-540.

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.

SEWELL, M. M., C. R. PARKS, AND M. CHASE. 1996. Intraspecific chloroplast DNA variation and biogeography of North American *Liriodendron* L. (Magnoliaceae). *Evolution* 50: 1147-1154.

SHAW, J., AND R. L. SMALL. 2004. Addressing the "hardest puzzle in American pomology:" Phylogeny of *Prunus* section *Prunocerasus* (Rosaceae) based on seven noncoding chloroplast DNA regions. *American Journal of Botany* 91: 985-996.

SHAW, J., E. B. LICKEY, J. T. BECK, S. B. FARMER, W. LIU, J. MILLER, K. C. SIRIPUN, C. T. WINDER, E. E. SCHILLING, 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.

SMALL, J. K. 1933. Manual of the southeastern flora. Published by the author, The Science Press Printing Company, Lancaster, Pennsylvania, USA.

SMITH, E. B. 1994. Keys to the flora of Arkansas. University of Arkansas Press, Fayetteville, Arkansas, USA.

SOLTIS, D. E., P. S. SOLTIS, AND B. M. MILLIGAN. 1992. Intraspecific chloroplast DNA variation: systematic and phylogenetic implications. Pp. 117-150 *In* P. S. Soltis, D. E. Soltis, and J. J. Doyle, [eds.], Molecular systematics of plants. Chapman and Hall, New York, New York, USA.

STEYERMARK, J. A. 1963. Flora of Missouri. Iowa State University Press, Ames, Iowa, USA.

SWOFFORD, D. L. 2002. PAUP*: phylogenetic analysis using parsimony (*and other methods) version 4.0b10. Sinauer Associates, Sunderland, Massachusetts, USA.

TAKHTAJAN, A. L. 1997. Diversity and Classification of Flowering Plants. Columbia University Press, New York, New York, USA.

THOMPSON, J. D., T. J. GIBSON, F. PLEWNIAK, F. JEANMOUGIN, AND D. G. HIGGINS. 2001. The Clustal_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* 25: 4876-4882.

TREMBLAY, N. O., AND D. J. SCHOEN. 1999. Molecular phylogeography of *Dryas integrifolia*: glacial refugia and postglacial recolonization. *Molecular Ecology* 8: 1187-1198.

USDA, NRCS. 2002. The PLANTS Database, Version 3.5 (http://plants.usda.gov). National Plant Data Center, Baton Rouge, Louisiana, USA.

WATKINS, R. 1976. Cherry, plum, peach, apricot, and almond. *In* N. W. Simmons [ed.], Evolution of crop plants, 242-247. Longman, London, England.

WAUGH, F. A. 1899. The pollination of plums. *Vermont Agricultural Experimentation Station, Annual Report* 12: 189-209.

WAUGH, F. A. 1901. Plums and plum culture. Orange Judd, New York, New York, USA.

WIGHT, W. F. 1913. North American species of the genus *Amygdalus*. Dudley Memorial Volume. Leland Stanford Junior University Publications, University Series, California, USA.

WHITTEMORE, A. T., AND B. A. SCHAAL. 1991. Interspecific gene flow in sympatric oaks. *Proceedings of the National Academy of Sciences, USA* 88: 2540-2544.

WIGHT, W. F. 1915. Native American species of *Prunus*. *Bulletin of the U.S. Department of Agriculture* 179: 1-75.

WOFFORD, B. E. AND E. W. CHESTER. 2002. Guide to the Trees, Shrubs, and Woody Vines of Tennessee. The University of Tennessee Press, Knoxville, Tennessee, USA.

WUNDERLIN, R. P. 1988. Guide to the vascular plants of Florida. University of Florida Press, Gainesville, Florida, USA.

APPENDIX; PART 4

Table. 1. List of taxa used in this investigation: Chloroplast DNA phylogeny and phylogeography of the North American plums (*Prunus* subgenus *Prunus* section *Prunocerasus*; Rosaceae). Ht = cpDNA haplotype; Herb. = Herbarium.

Ht	Species	Collector; Number; Year	Herb.	State and County	GenBank
A	*P. alleghaniensis* Porter	J. Shaw; JSh835; 2001	TENN	USA; WV; Greenbriar	AY773484
C	*P. alleghaniensis* Porter (umbellata?)	J. Shaw; JSh1008; 2004	TENN	USA; NC; Rutherford	AY773482
C3	*P. alleghaniensis* Porter	J. Shaw; JSh836; 2001	TENN	USA; MD; Allegany	AY773485
C *	*P. alleghaniensis* Porter	J. Shaw; JSh837; 2001	TENN	USA; PA; Fulton	AY500650
C	*P. alleghaniensis* Porter	J. Shaw; JSh846; 2001	TENN	USA; PA; Centre	AY773486
C	*P. alleghaniensis* Porter	S.R. Hill; 10353; 1981	BH	USA; MD; Washington	AY773487
Ca	*P. alleghaniensis* Porter	J. Shaw; JSh834; 2001	TENN	USA; WV; Greenbriar	AY773483
C	*P.all.*var. *davisii* (Wight.) Sarg.	B.A. Daubendiek; s.n.; 1980	MICH	USA; MI; Oceana	AY773488
C	*P.all.*var. *davisii* (Wight.) Sarg.	E.G. Voss; 7139; 1958	MICH	USA; MI; Mason	AY773489
C	*P.all.*var. *davisii* (Wight.) Sarg.	F.C. Gates; 16642; 1931	BRIT	USA; MI; Cheboygan	AY773490
C *	*P.all.* var. *davisii* (Wight.) Sarg.	G. Schmidt; s.n.; 2001	TENN	USA; MI; Oscoda	AY500649
C	*P.all.*var. *davisii* (Wight.) Sarg.	R. McVaugh; 11067; 1949	MICH	USA; MI; Montmorency	AY773491
C	*P.all.*var. *davisii* (Wight.) Sarg.	S. Taylor; s.n.; 1992	MICH	USA; MI; Crawford	AY773492
C	*P.all.*var. *davisii (Wight.) Sarg.*	S. Taylor; s.n.; 1992	MICH	USA; MI; Ogemaw	AY773493
A	*P. americana* Marshall	A.W. Cusick; 30155; 1992	MICH	USA; OH; Lawrence	AY773496
A	*P. americana* Marshall	A.W. Cusick; 35818; 2001	MICH	USA; MI; Wayne	AY773497
A	*P. americana* Marshall	C.A. Weatherby; 961; 1937	BH	USA; CT; Tolland	AY773502
A	*P. americana* Marshall	C.A. Weatherby; s.n.; 1937	BRIT	USA; CT; Tolland	AY773499
A	*P. americana* Marshall	C.A. Weatherby; s.n.; 1937	MICH	USA; CT; Tolland	AY773500
A	*P. americana* Marshall	G.E.D.; 76; 1926	BH	USA; NY; Albany	AY773505
A	*P. americana* Marshall	J. Shaw; JSh1039; 2004	TENN	USA; KS; Saline	AY773515
A	*P. americana* Marshall	J. Shaw; JSh1040; 2004	TENN	USA; KS; McPherson	AY773516
A	*P. americana* Marshall (x umbellata?)	J. Shaw; JSh911; 2002	TENN	USA; AL; Sumter	AY773533
A	*P. americana* Marshall	J.K. Bissell; 1993:006; 1993	MICH	USA; OH; Ashtabula	AY773508
A	*P. americana* Marshall	J.K. Bissell; 1993:018; 1993	MICH	USA; PA; Erie	AY773509
A	*P. americana* Marshall	J.R. Bozeman; 10711; 1967	MICH	USA; NC; Durham	AY773512
A	*P. americana* Marshall	J.W. Thieret; 59607; 1995	MICH	USA; KY; Bracken	AY773521
A	*P. americana* Marshall	M.Oldham; 13676; 1992	MICH	Canada; Ontario; Middlesex Co.	AY773503
A	*P. americana* Marshall	R. Kral; 43361; 1971	BRIT	USA, AL; Madison	AY773528
A	*P. americana* Marshall	R.B. Clarkson; 2418; 1958	BRIT	USA; WV; Pocahontas	AY773525
A	*P. americana* Marshall	R.G. Koch; 4495; 1968	BRIT	USA; NE; Seward	AY773526
A	*P. americana* Marshall	R.W. Smith; 1497; 1986	MICH	USA; MI; Lenawee	AY773529
A1	*P. americana* Marshall	C.E. Garton; 14885; 1972	MICH	Canada; Ontario; Thunder Bay	AY773501
A1	*P. americana* Marshall	J. Shaw; JSh1038; 2004	TENN	USA; CO; Boulder	AY773514
A1	*P. americana* Marshall	J.H. Ehlers; 8061; 1941	VDB	USA; CO; Jefferson	AY773507
A1	*P. americana* Marshall	L.S. Ehlers; 321; 1922	MICH	USA; CO; Jefferson	AY773524
A2	*P. americana* Marshall	J. Shaw; JSh1042; 2004	TENN	USA; KS; Elk	AY773517
A3	*P. americana* Marshall	J. Shaw; JSh1032C; 2004	TENN	USA; AR; Perry	AY773513
A4	*P. americana* Marshall	R. Kral; 42121; 1971	BRIT	USA; AL; Jackson	AY773527
A5	*P. americana* Marshall	A.W. Cusick; 28053; 1989	MICH	USA; OH; Lucas	AY773495
A6	*P. americana* Marshall	L.D. Estes; 5753; 2004	TENN	USA; TN; Hickman	AY773523

Table 1. continued

Ht		Species	Collector; Number; Year	Herb.	State and County	GenBank
Aa		*P. americana* Marshall	A.W. Cusick; 26313; 1987	MICH	USA; OH; Adams	AY773494
Aa		*P. americana* Marshall	F.B. Cotner; s.n.; 1949	BH	USA; MT; Carter	AY773504
Aa	*	*P. americana* Marshall	J. Shaw; JSh959; 2001	TENN	USA; GA; Catoosa	AY500638
Aa		*P. americana* Marshall	J. Shaw; JSh967; 2002	TENN	USA; TN; Washington	AY773519
Aa		*P. americana* Marshall	J. Shaw; JSh969; 2002	TENN	USA; TN; Hamblen	AY773520
Aa		*P. americana* Marshall	L.R. Phillipe; 27114; 1996	MICH	USA; IL; Vermilion	AY773531
Aa		*P. americana* Marshall	S. Hill; 26475; 1995	BRIT	USA; IL; Kane	AY773530
Aa		*P. americana* Marshall	T.G. Lammers; 10603; 1998	BRIT	USA; IL; Cook	AY773532
C		*P. americana* Marshall (alleghaniensis?)	H.A. Wahl; 68; 1938	BH	USA; PA; Centre	AY773506
C		*P. americana* Marshall	J. Shaw; JSh1043; 2004	TENN	USA; KS; Elk	AY773518
C1		*P. americana* Marshall	B.U. Gates; 24915; 1948	BRIT	USA; MA; Worcester	AY773498
C2		*P. americana* Marshall	L. Bougere; 1372; 1949	BRIT	USA; CO; Boulder	AY773522
C4		*P. americana* Marshall	J.K. Morton; 13130; 1980	MICH	Canada; Ontario; Manitoulin Isl.	AY773510
Cb		*P. americana* Marshall	J.K. Morton; NA5747; 1974	BRIT	Canada; Ontario; Brant	AY773511
A		*P. a.* Marshall var. *lanata* Sudw.	C.T. Bryson; 5710; 1987	MICH	USA; MS; Washington	AY773534
A		*P. a.* Marshall var. *lanata* Sudw.	H.H. Iltis; 25300; 1966	BRIT	USA; LA; Morehouse	AY773536
A		*P. a.* Marshall var. *lanata* Sudw.	J. Shaw; JSh569; 2000	TENN	USA; TN; Scott	AY773539
A		*P. a.* Marshall var. *lanata* Sudw.	L.H. Shinners; 13115; 1951	BRIT	USA; TX; Dallas	AY773541
A2		*P. a.* Marshall var. *lanata* Sudw.	R.D. Thomas; 143876; 1995	BRIT	USA; AR; Logan	AY773535
A7		*P. a.* Marshall var. *lanata* Sudw.	R.W. Smith; 2424; 1988	MICH	USA; MI; Lenawee	AY773543
Aa	*	*P. a.* Marshall var. *lanata* Sudw.	J. Beck; 4955; 2002	TENN	USA; TN; Marion	AY500639
Aa		*P. a.* Marshall var. *lanata* Sudw.	J. Shaw; JSh1003; 2002	TENN	USA; WI; Sauk	AY773537
C		*P. a.* Marshall var. *lanata* Sudw.	K.M. Wiegand; 971; 1935	BH	USA; SD; Custer	AY773540
C		*P. a.* Marshall var. *lanata* Sudw.	R. Kral; 53264; 1974	VDB	USA; AL; Coosa	AY773542
C		*P. a.* Marshall var. *lanata* Sudw.	S.C. Gunn; 490; 1982	BRIT	USA; AL; Wilcox	AY773544
A		*P. a.* Marshall var. *lanata* Sudw.	J. Shaw; JSh1044; 2004	TENN	USA; KS; Montgomery	AY773538
C		*P. angustifolia* Marshall	J. Beck; 5283; 2004	TENN	USA; GA; Chattahoochee	AY773546
C		*P. angustifolia* Marshall	J. Shaw; JSh1015; 2004	TENN	USA; SC; Laurens	AY773547
C		*P. angustifolia* Marshall	J. Shaw; JSh1018; 2004	TENN	USA; SC; Orangeburg	AY773548
C		*P. angustifolia* Marshall	J. Shaw; JSh1020; 2004	TENN	USA; SC; Sumter	AY773549
C		*P. angustifolia* Marshall	J. Shaw; JSh1024; 2004	TENN	USA; GA; DeKalb	AY773550
C		*P. angustifolia* Marshall	J. Shaw; JSh1030; 2004	TENN	USA; GA; Walton	AY773544
C		*P. angustifolia* Marshall	J. Shaw; JSh1032A; 2004	TENN	USA; AR; Prairie	AY773551
C		*P. angustifolia* Marshall	J. Shaw; JSh1033; 2004	TENN	USA; OK; McIntosh	AY773552
C		*P. angustifolia* Marshall	J. Shaw; JSh1037; 2004	TENN	USA; OK; Beckham	AY773553
C		*P. angustifolia* Marshall	J. Shaw; JSh1041; 2004	TENN	USA; KS; Cowley	AY773554
C		*P. angustifolia* Marshall	J. Shaw; JSh1045; 2004	TENN	USA; KS; Montgomery	AY773555
C		*P. angustifolia* Marshall	J. Shaw; JSh1049; 2004	TENN	USA; MO; Shannon	AY773556
C		*P. angustifolia* Marshall	J. Shaw; JSh776; 2001	TENN	USA; GA; Tift	AY773557
C		*P. angustifolia* Marshall	J. Shaw; JSh777-778; 2001	TENN	USA; GA; Irwin	AY773558
C	*	*P. angustifolia* Marshall	J. Shaw; JSh785; 2001	TENN	USA; GA; Floyd	AY500644

Table 1. continued.

Ht		Species	Collector; Number; Year	Herb.	State and County	GenBank
C		*P. angustifolia* Marshall	J. Shaw; JSh910; 2002	TENN	USA; AL; Greene	AY773559
C		*P. angustifolia* Marshall	J. Shaw; JSh913; 2002	TENN	USA; MS; Clarke	AY773560
C		*P. angustifolia* Marshall	J. Shaw; JSh914; 2002	TENN	USA; MS; Pearl River	AY773561
C		*P. angustifolia* Marshall	J. Shaw; JSh916; 2002	TENN	USA; TX; Waller	AY773562
C		*P. angustifolia* Marshall	J. Shaw; JSh947; 2002	TENN	USA; AR; Nevada	AY773563
C		*P. angustifolia* Marshall	J. Shaw; JSh952; 2002	TENN	USA; TN; Henderson	AY773564
B	*	*P. geniculata* R.M. Harper	J. Shaw; JSh898; 2002	TENN	USA; FL; Highlands	AY500651
A		*P. gracilis* Engelm. & A. Gray	E.L. Richards; 2585; 1961	BRIT	USA; KS; Morton	AY773568
A		*P. gracilis* Engelm. & A. Gray	L.H. Shinners; 12194; 1950	BRIT	USA; TX; Parker	AY773570
A		*P. gracilis* Engelm. & A. Gray	P. Nighswonger; 1368; 1976	BRIT	USA; OK; Caddo	AY773574
B1		*P. gracilis* Engelm. & A. Gray	B.L. Lipscomb; 3140; 1981	BRIT	USA; TX; Henderson	AY773565
C		*P. gracilis* Engelm. & A. Gray	D. Thomas; 80853; 1982	BRIT	USA; LA; Caddo	AY773567
C		*P. gracilis* Engelm. & A. Gray	D.S. Correll; 27194; 1963	BRIT	USA; TX; Smith	AY773566
C	*	*P. gracilis* Engelm. & A. Gray	J. Shaw; JSh936; 2002	TENN	USA; TX; Milam	AY500646
C		*P. gracilis* Engelm. & A. Gray	J. Shaw; JSH946; 2002	TENN	USA; LA; Bossier	AY773569
C		*P. gracilis* Engelm. & A. Gray	L.H. Shinners; 32122; 1968	BRIT	USA; TX; Collingsworth	AY773571
C		*P. gracilis* Engelm. & A. Gray	M. Hopkins; 186; 1944	BH	USA; OK; Cleveland	AY773572
C		*P. gracilis* Engelm. & A. Gray	P. Nighswonger; 1360; 1976	BRIT	USA; OK; Major	AY773573
C		*P. gracilis* Engelm. & A. Gray	R. Kral; 71427; 1984	BRIT	USA; TX; Upshur	AY773575
C		*P. gracilis* Engelm. & A. Gray	S.L. Orzell; 10515; 1989	BRIT	USA; TX: Limestone	AY773576
C		*P. gracilis* Engelm. & A. Gray	S.L. Orzell; 1919; 1985	BRIT	USA; AR; Miller	AY773577
A		*P. hortulana* L.H. Bailey	C.T. Bryson; 10040; 1990	BRIT	USA; MS; Bolivar	AY773580
A		*P. hortulana* L.H. Bailey	R.R. Haynes; 8757; 1985	BRIT	USA; TN; Rutherford	AY773587
A3		*P. hortulana* L.H. Bailey	R. Kral; 66924; 1981	BRIT	USA; TN; Hickman	AY773586
Aa	*	*P. hortulana* L.H. Bailey	J. Shaw; JSh821; 2002	TENN	USA; TN; Clay	AY500643
Aa		*P. hortulana* L.H. Bailey	J. Shaw; JSh971; 2002	TENN	USA; TN; Clay	AY773584
C		*P. hortulana* L.H. Bailey (rivularis?)	C.M. Sladowski; W0438; 1989	BRIT	USA; MO; Pulaski	AY773578
C		*P. hortulana* L.H. Bailey	C.S. Wallis; 6346; 1958	BRIT	USA; OK; Ottawa	AY773579
C		*P. hortulana* L.H. Bailey	J. Shaw; JSh1046; 2004	TENN	USA; MO; Oregon	AY773581
C		*P. hortulana* L.H. Bailey	J. Shaw; JSh1047; 2004	TENN	USA; MO; Oregon	AY773582
C		*P. hortulana* L.H. Bailey	J. Shaw; JSh1050; 2004	TENN	USA; MO; Shannon	AY773583
C		*P. hortulana* L.H. Bailey	J.W. Thieret; 59602; 1995	MICH	USA; KY; Campbell	AY773585
B		*P. maritima* Marshall	J. Shaw; JSh876; 2001	TENN	USA; MA; Plymouth	AY773589
B	*	*P. maritima* Marshall	J. Shaw; JSh877; 2001	TENN	USA; MA; Plymouth	AY500652
B		*P. maritima* Marshall	J. Shaw; JSh882; 2001	TENN	USA; NY; Suffolk	AY773590
B		*P. maritima* Marshall	R. Uva; 90; 2001	TENN	USA; NY; Suffolk	AY773599
B		*P. maritima* Marshall	R. Uva; 98; 2001	TENN	USA; NY; Suffolk	AY773600
C		*P. maritima* Marshall	J. Shaw; JSh874; 2001	TENN	USA; MA; Essex	AY773588
C		*P. maritima* Marshall	R. Uva; 100; 2001	TENN	USA; MA; Plymouth	AY773591

Table 1. continued.

Ht		Species	Collector; Number; Year	Herb.	State and County	GenBank
C		*P. maritima* Marshall	R. Uva; 107; 2001	TENN	USA; ME; York	AY773592
C		*P. maritima* Marshall	R. Uva; 111; 2001	TENN	USA; MA; Essex	AY773593
C		*P. maritima* Marshall	R. Uva; 33; 2001	TENN	USA; NJ; Ocean	AY773594
C		*P. maritima* Marshall	R. Uva; 39; 2001	TENN	USA; DE; Sussex	AY773595
C		*P. maritima* Marshall	R. Uva; 55; 2001	TENN	USA; NJ; Cape May	AY773596
C		*P. maritima* Marshall	R. Uva; 60; 2001	TENN	USA; NJ; Burlington	AY773597
C		*P. maritima* Marshall	R. Uva; 69; 2001	TENN	USA; NY; Suffolk	AY773598
B	*	*P. m.* var. *gravesii* (Small) G.J. Anderson	Conn Greenhouse; 2001	TENN	USA; CT; Tolland	AY500653
A		*P. mexicana* S. Watson	C.T. Bryson; 15396; 1996	TENN	USA; MS; Leflore	AY773601
A		*P. mexicana* S. Watson	J. Marroquin 4137; 1988	BRIT	Mexico	AY773602
A		*P. mexicana* S. Watson	J. Marroquin 4153; 1988?	BRIT	Mexico	AY773603
A		*P. mexicana* S. Watson	J. Shaw; JSh1032B; 2004	TENN	USA; AR; Perry	AY773604
A		*P. mexicana* S. Watson	J. Shaw; JSh1036; 2004	TENN	USA; OK; Seminole	AY773605
A		*P. mexicana* S. Watson	J. Shaw; JSh1048; 2004	TENN	USA; MO; Oregon	AY773606
A		*P. mexicana* S. Watson	R. Kral; 17274; 1963	BRIT	USA; LA; Union	AY773612
Aa	*	*P. mexicana* S. Watson	J. Shaw; JSh919; 2002	TENN	USA; TX; Waller	AY500642
Aa		*P. mexicana* S. Watson	J. Shaw; JSh928; 2002	TENN	USA; TX; Comal	AY773607
Aa		*P. mexicana* S. Watson	J. Shaw; JSh933; 2002	TENN	USA; TX; Llano	AY773608
Aa		*P. mexicana* S. Watson	L.H. Shinners; 13122; 1951	BRIT	USA; TX; Grayson	AY773610
C		*P. mexicana* S. Watson (americana?)	K.E. Rogers 9154; 1973	BRIT	USA; MS; Forrest	AY773609
C		*P. mexicana* S. Watson	P.H. Raven; 27845; 1989	MICH	USA; MO; Jefferson	AY773611
C		*P. mexicana* S. Watson	S. B. Jones; 19669; 1970	TENN	USA; MS; Pike	AY773613
A2		*P. munsoniana* Wight & Hedrick	G. Tucker; 26022; 1984	BRIT	USA; AR; Benton	AY773615
A8		*P. munsoniana* Wight & Hedrick	R.D. Thomas; 106953; 1988	TENN	USA; LA; Bossier	AY773618
C		*P. munsoniana* Wight & Hedrick	A.W. Cusick; 31462; 1994	MICH	USA; KY; Ohio	AY773614
C	*	*P. munsoniana* Wight & Hedrick	J. Shaw; JSh810; 2001	TENN	USA; TN; Anderson	AY500645
C		*P. munsoniana* Wight & Hedrick	J. Shaw; JSh960; 2002	TENN	USA; TN; Rhea	AY773616
C		*P. munsoniana* Wight & Hedrick	J. Shaw; JSh961; 2002	TENN	USA; TN; Carter	AY773617
A		*P. nigra* Ait.	A.W. Cusick; 33787; 1997	MICH	Canada; Ontario; Plympton	AY773620
A		*P. nigra* Ait.	P.M. Catling et al.; 4624; 1982	MICH	Canada; Ontario; Frontenac Co.	AY773629
A		*P. nigra* Ait.	W.J. Crins; 8783; 1992	MICH	Canada; Ontario; Timiskaming	AY773631
Aa		*P. nigra* Ait.	E.G. Voss; 2620; 1955	BRIT	USA; MI; Cheboygan	AY773623
Aa		*P. nigra* Ait.	E.G. Voss; 6254; 1958	MICH	USA; MI; Gogebic	AY773625
Aa		*P. nigra* Ait.	J. Looman; 14581; 1970	TENN	Canada; Mannitoba -St. Malo	AY773627
Aa		*P. nigra* Ait.	J. Shaw; JSh1004; 2002	TENN	USA; WI; Sauk	AY773628
C		*P. nigra* Ait.	F. Marie-Victorin; 34180; 1930	BH	Canada; Quebec; Mont-Royal	AY773626
C1		*P. nigra* Ait.	A.J. Eames; 21018; 1944	BH	USA; NY; Tompkins	AY773619
C1		*P. nigra* Ait.	C.E. Garton; 16394; 1975	MICH	Canada; Ontario; Thunder Bay	AY773621
C1		*P. nigra* Ait.	C.E. Garton; 19384; 1980	MICH	Canada; Ontario; Thunder Bay	AY773622
C1		*P. nigra* Ait.	E.G. Voss; 6094; 1958	MICH	USA; MI; Baraga	AY773624

Table 1. continued.

Ht		Species	Collector; Number; Year	Herb.	State and County	GenBank
C1	*	*P. nigra* Ait.	J. Shaw; JSh979; 2002	TENN	USA; VT; Chittendon	AY500648
C1		*P. nigra* Ait.	R.B. MacFarlane; 5051; 1993	MICH	USA; MI; Ontonagon	AY773630
A		*P. rivularis*	L.H. Shinners; 12102; 1950	BRIT	USA; TX; Hill	AY773633
A		*P. rivularis* var. *rivularis*	M. Endquist; 2205; 1992	BRIT	USA; TX; Hays	AY773635
A	*	*P. rivularis* Scheele var. *pubescens*	M. Endquist; 3372; 1997	BRIT	USA; TX; Sutton	AY500640
A		*P. rivularis* var. *pubescens*	M. Endquist; 3408; 1997	BRIT	USA; TX; Coke	AY773636
A		*P. rivularis* var. *pubescens (P. murrayana)*	M. Endquist; 3500; 1997	BRIT	USA; TX; Jeff Davis	AY773637
A		*P. rivularis*	R. McVaugh; 8305; 1947	BRIT	USA; TX; Kimble	AY773638
C		*P. rivularis*	J. Shaw; JSh931; 2002	TENN	USA; TX; Mason	AY773632
C		*P. rivularis*	L. Pace; 39; 1975	BRIT	USA; TX; Palo Pinto	AY773634
Cb		*P. rivularis*	R.W. Sanders; 3495; 1995	BRIT	USA; TX; Hunt	AY773639
S	*	*P. subcordata* Benth.	J.Syring; s.n.; 2001	TENN	USA; CA; Shasta	AY500655
S		*P. subcordata* Benth.	J.Syring; s.n.; 2001	TENN	USA; CA; Shasta	AY773641
S		*P. subcordata* Benth.	W.B. Cooke; 16263; 1941	BH	USA; CA; Siskiyou	AY773640
T		*P. texana* Dietr.	C.L. Lundell; 14971; 1949	MICH	USA; TX; Brooks	AY773642
T	*	*P. texana* Dietr.	J. Shaw; JSh924; 2002	TENN	USA; TX; Goliad	AY500654
T		*P. texana* Dietr.	M.H. Mayfield; 2143; 1995	BRIT	USA; TX; Goliad	AY773643
T		*P. texana* Dietr.	R. McVaugh; 7034; 1945	BRIT	USA; TX; Wilson	AY773644
Aa		*P. umbellata* Elliott	W. Muenscher; 14290; 1939	BH	USA, FL; Alachua	AY773657
B		*P. umbellata* Elliott	J. Shaw; JSh908; 2002	TENN	USA; FL; Marion	AY773653
B1		*P. umbellata* Elliott (*P. mitis*)	C.L. Lundell; 10949; 1942	BRIT	USA; TX; Tyler	AY773665
B1		*P. umbellata* Elliott	J. Beck; 5291; 2004	TENN	USA; FL: Alachua	AY773649
C		*P. umbellata* Elliott (*P. mitis*)	A.E. Radford; 44801; 1966	BRIT	USA; SC; McCormick	AY773664
C		*P. umbellata* Elliott (x angustifolia?)	B. Hansen; 10347; 1985	TENN	USA; FL; Sumter	AY773645
C		*P. umbellata* Elliott	C.L. Lundell; 10533; 1941	MICH	USA; TX; Jasper	AY773646
C		*P. umbellata* Elliott	D. Demaree; 6377?.3; 1971	BRIT	USA; AR; Ouachita	AY773647
C		*P. umbellata* Elliott (*P. mitis*)	G.E. Tucker; 9218; 1972	BRIT	USA; AR; Union	AY773666
C	*	*P. umbellata* Elliott (x angustifolia?)	J. Shaw; JSh774-775; 2001	TENN	USA; FL; Leon	AY500647
C		*P. umbellata* Elliott (*P. mitis*)	J. Shaw; JSh787; 2001	TENN	USA; MS; Perry	AY773667
C		*P. umbellata* Elliott	J. Shaw; JSh896; 2001	TENN	USA; FL; Highlands	AY773651
C		*P. umbellata* Elliott	J. Shaw; JSh903; 2002	TENN	USA; FL; Lake	AY773652
C		*P. umbellata* Elliott	J.B. Nelson; 15459; 1994	BH	USA; SC; Barnwell	AY773650
C		*P. umbellata* Elliott (*P. mitis*)	M. Pine; 92-079; 1992	TENN	USA; TN; White	AY773669
C		*P. umbellata* Elliott (*P. mitis*)	R. Kral; 16712; 1963	BRIT	USA; LA; Lincoln	AY773668
C		*P. umbellata* Elliott	R. Kral; 35156; 1969	BRIT	USA; AL; Escambia	AY773654
C		*P. umbellata* Elliott (*P. mitis*)	R. McVaugh; 6807; 1945	BRIT	USA; TX; Panola	AY773670
C		*P. umbellata* Elliott	R. McVaugh; 8419; 1947	BRIT	USA; TX; Sabine	AY773656
C		*P. umbellata* Elliott	R.K. Godfrey; 80803; 1994	BRIT	USA; FL; Leon	AY773655
pU		*P. umbellata* Elliott	R.D. Thomas; 21172; 1970	BRIT	USA; LA; Jackson	AY773648

Table 1. continued.

Ht		Species	Collector; Number; Year	Herb.	State and County	GenBank
Aa	*	*P. u.* Elliott var. *injucunda* (Small) Sarg.	J. Shaw; JSh958; 2002	TENN	USA; GA; Cobb	AY500641
C		*P. u.* Elliott var. *injucunda* (Small) Sarg.	R. Kral; 31807; 1968	BRIT	USA; AL; Russell	AY773658
C		*P. u.* Elliott var. *injucunda* (Small) Sarg.	R. Kral; 59510; 1977	BRIT	USA; GA; Cobb	AY773659
C		*P. u.* Elliott var. *injucunda* (Small) Sarg.	R. Kral; 71376; 1984	BRIT	USA; AL; Escambia	AY773660
C		*P. u.* Elliott var. *injucunda* (Small) Sarg.	S.B. Jones; 22094; 1972	BRIT	USA; GA; Harris	AY773661
C		*P. u.* Elliott var. *injucunda* (Small) Sarg.	S.B. Jones; 22177 1972	BRIT	USA; GA; Meriwether	AY773662
C		*P. u.* Elliott var. *injucunda* (Small) Sarg.	W.Duncan; 3797; 1941	TENN	USA; GA; Oglethorpe	AY773663

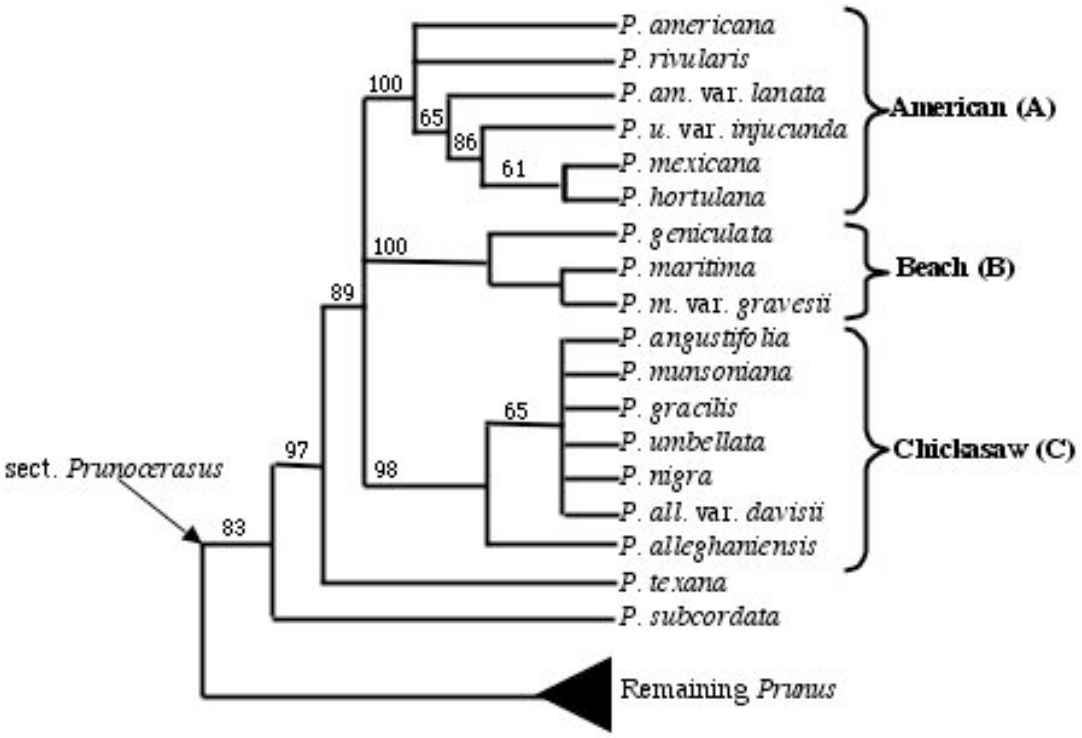


Fig. 1. A maximum parsimony cladogram from seven cpDNA regions, 4375 aligned nucleotide positions (modified from Shaw and Small, 2004). This single most parsimonious tree was 128 steps (CI = 0.98; RI = 0.98). Bootstrap values greater than 50% are shown above the branches. The North American plums of section *Prunocerasus* are monophyletic. Within *P. subcordata* and *P. texana*, the rest of the species resolve in three clades (labeled American, Beach, Chickasaw).

Fig. 2. (a) to (c). Strict consensus of 3 equally parsimonious trees resulting from maximum parsimony analysis using the *rpL16* intron (797 aligned bp from the 3′ end of the intron). The data set contained 23 parsimony informative characters (including multi-bp indels coded as binary characters) and 10 variable but parsimony uninformative characters; tree length = 34 steps; consistency index [CI] = 0.97; retention index [RI] = 0.99). Bootstrap values greater than 50 % are above the branches. This phylogeny was split into three trees (a, b, c) because of the number of accessions. Fig. 3a details the relationships of the Americana clade, Fig. 3b details the relationships of the Beach clade), and Fig. 3c details the relationships of the Chickasaw clade (All 21 identical accessions of *P. angustifolia* represented by: **P. ANGUSTIFOLIA**).

Fig. 2a

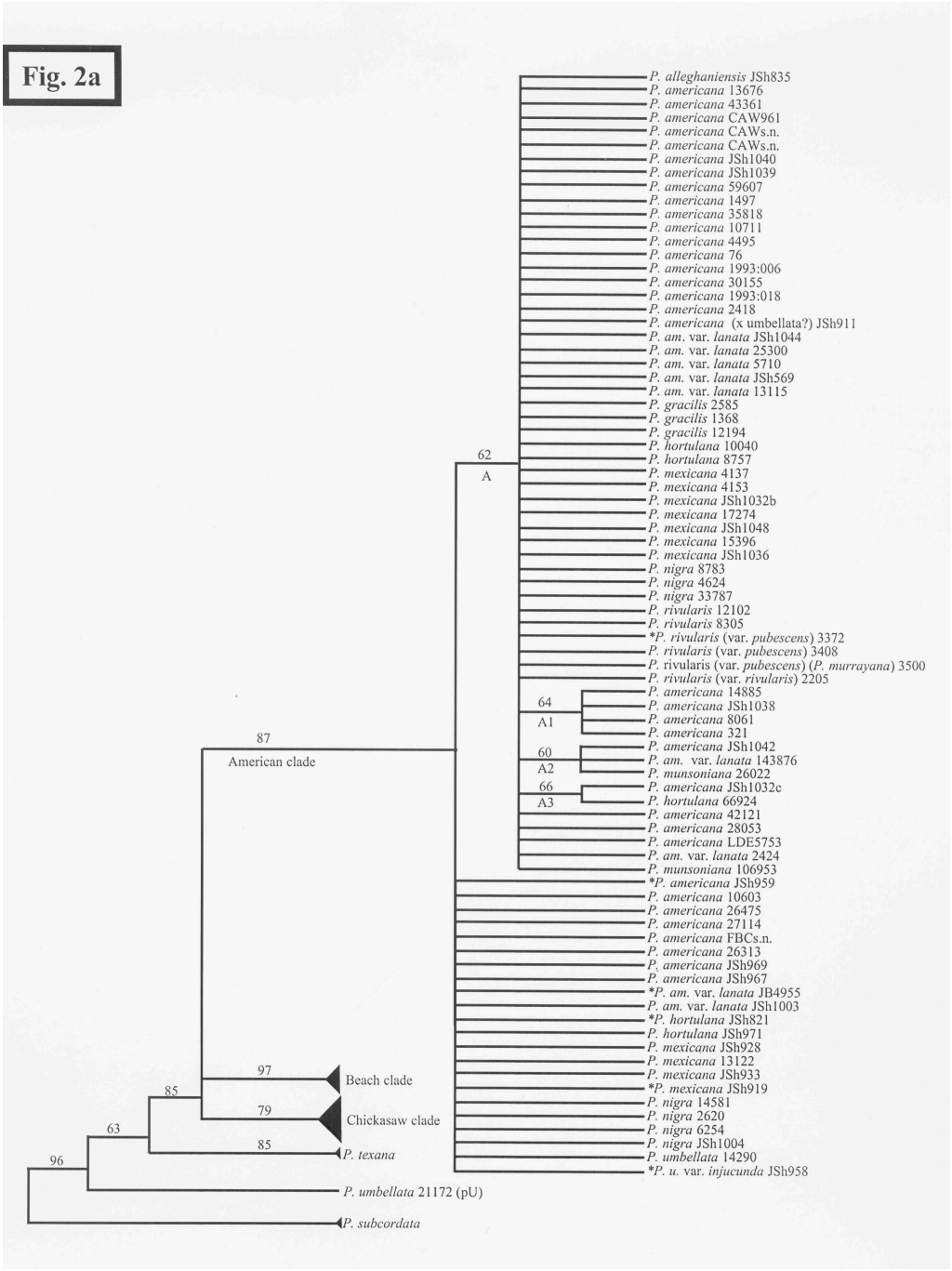

P. alleghaniensis JSh835
P. americana 13676
P. americana 43361
P. americana CAW961
P. americana CAWs.n.
P. americana CAWs.n.
P. americana JSh1040
P. americana JSh1039
P. americana 59607
P. americana 1497
P. americana 35818
P. americana 10711
P. americana 4495
P. americana 76
P. americana 1993:006
P. americana 30155
P. americana 1993:018
P. americana 2418
P. americana (x umbellata?) JSh911
P. am. var. lanata JSh1044
P. am. var. lanata 25300
P. am. var. lanata 5710
P. am. var. lanata JSh569
P. am. var. lanata 13115
P. gracilis 2585
P. gracilis 1368
P. gracilis 12194
P. hortulana 10040
P. hortulana 8757
P. mexicana 4137
P. mexicana 4153
P. mexicana JSh1032b
P. mexicana 17274
P. mexicana JSh1048
P. mexicana 15396
P. mexicana JSh1036
P. nigra 8783
P. nigra 4624
P. nigra 33787
P. rivularis 12102
P. rivularis 8305
*P. rivularis (var. pubescens) 3372
P. rivularis (var. pubescens) 3408
P. rivularis (var. pubescens) (P. murrayana) 3500
P. rivularis (var. rivularis) 2205
P. americana 14885
P. americana JSh1038
P. americana 8061
P. americana 321
P. americana JSh1042
P. am. var. lanata 143876
P. munsoniana 26022
P. americana JSh1032c
P. hortulana 66924
P. americana 42121
P. americana 28053
P. americana LDE5753
P. am. var. lanata 2424
P. munsoniana 106953
*P. americana JSh959
P. americana 10603
P. americana 26475
P. americana 27114
P. americana FBCs.n.
P. americana 26313
P. americana JSh969
P. americana JSh967
*P. am. var. lanata JB4955
P. am. var. lanata JSh1003
*P. hortulana JSh821
P. hortulana JSh971
P. mexicana JSh928
P. mexicana 13122
P. mexicana JSh933
*P. mexicana JSh919
P. nigra 14581
P. nigra 2620
P. nigra 6254
P. nigra JSh1004
P. umbellata 14290
*P. u. var. injucunda JSh958
62
A
64
A1
60
A2
66
A3
87
American clade
97
Beach clade
85
79
Chickasaw clade
63
85
P. texana
96
P. umbellata 21172 (pU)
P. subcordata

Fig. 2b

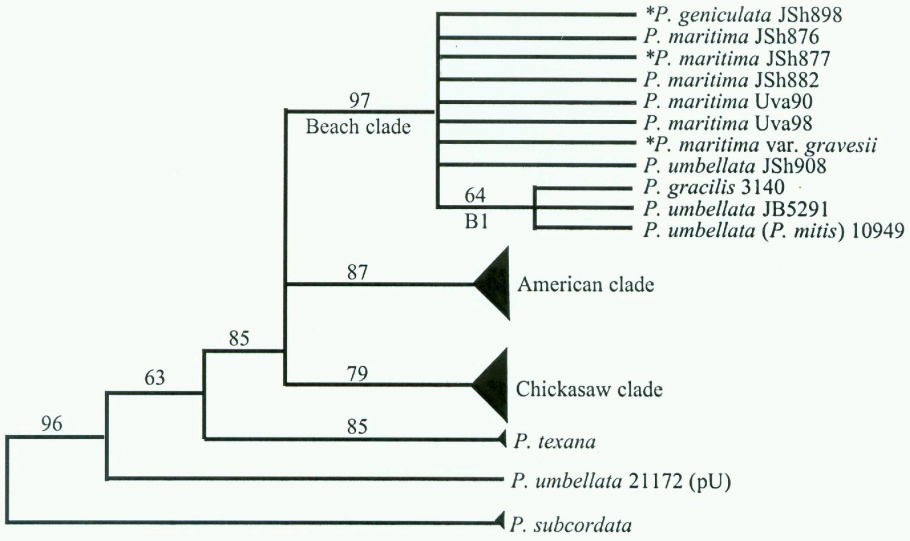
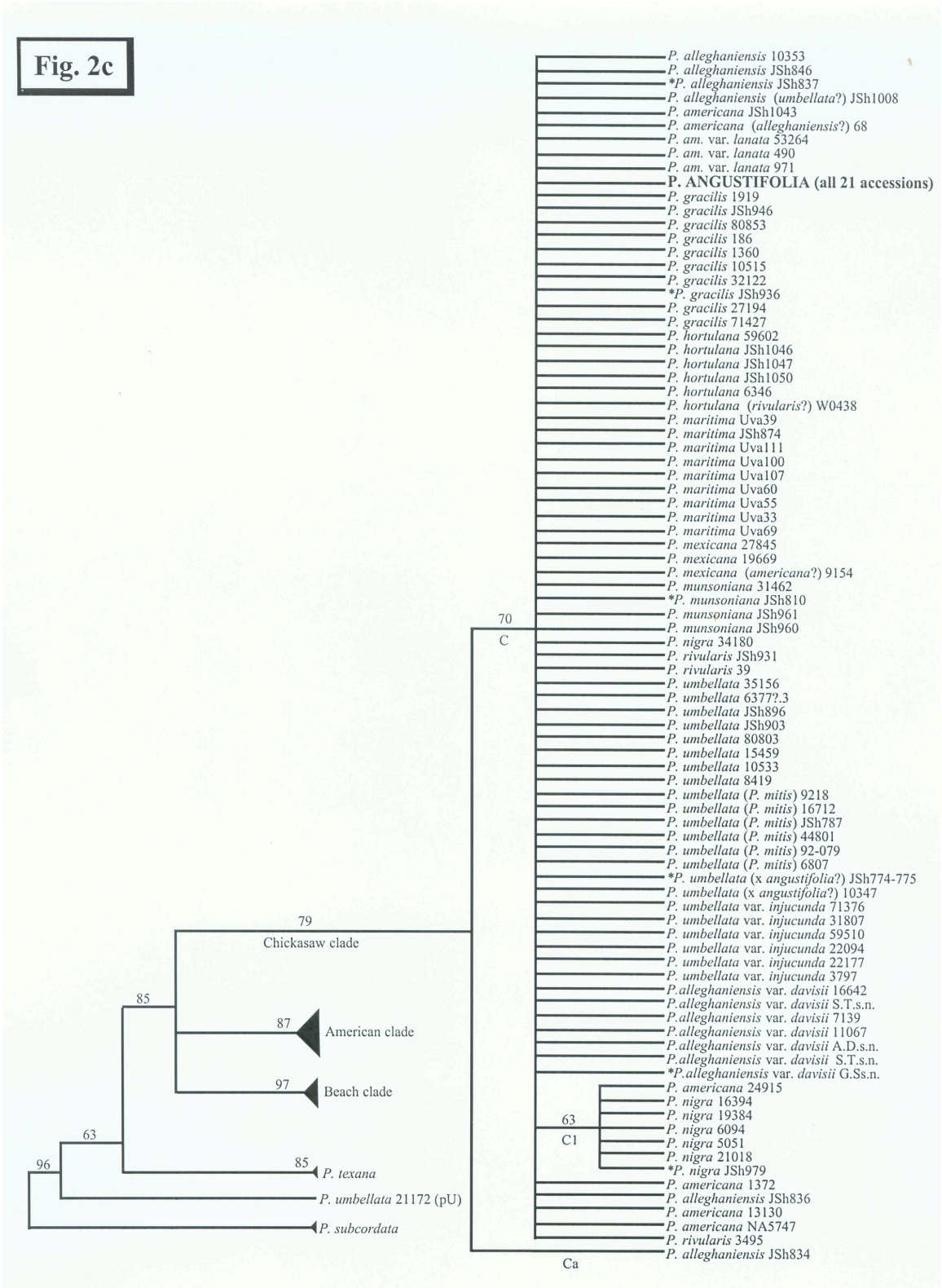

*P. geniculata JSh898
P. maritima JSh876
*P. maritima JSh877
P. maritima JSh882
P. maritima Uva90
P. maritima Uva98
*P. maritima var. gravesii
P. umbellata JSh908
P. gracilis 3140
P. umbellata JB5291
P. umbellata (P. mitis) 10949
97
Beach clade
64
B1
87
American clade
85
63
79
Chickasaw clade
96
85
P. texana
P. umbellata 21172 (pU)
P. subcordata

Fig. 2c


P. alleghaniensis 10353
P. alleghaniensis JSh846
*P. alleghaniensis JSh837
P. alleghaniensis (umbellata?) JSh1008
P. americana JSh1043
P. americana (alleghaniensis?) 68
P. am. var. lanata 53264
P. am. var. lanata 490
P. am. var. lanata 971
P. ANGUSTIFOLIA (all 21 accessions)
P. gracilis 1919
P. gracilis JSh946
P. gracilis 80853
P. gracilis 186
P. gracilis 1360
P. gracilis 10515
P. gracilis 32122
*P. gracilis JSh936
P. gracilis 27194
P. gracilis 71427
P. hortulana 59602
P. hortulana JSh1046
P. hortulana JSh1047
P. hortulana JSh1050
P. hortulana 6346
P. hortulana (rivularis?) W0438
P. maritima Uva39
P. maritima JSh874
P. maritima Uva111
P. maritima Uva100
P. maritima Uva107
P. maritima Uva60
P. maritima Uva55
P. maritima Uva33
P. maritima Uva69
P. mexicana 27845
P. mexicana 19669
P. mexicana (americana?) 9154
P. munsoniana 31462
*P. munsoniana JSh810
P. munsoniana JSh961
P. munsoniana JSh960
P. nigra 34180
P. rivularis JSh931
P. rivularis 39
P. umbellata 35156
P. umbellata 6377?.3
P. umbellata JSh896
P. umbellata JSh903
P. umbellata 80803
P. umbellata 15459
P. umbellata 10533
P. umbellata 8419
P. umbellata (P. mitis) 9218
P. umbellata (P. mitis) 16712
P. umbellata (P. mitis) JSh787
P. umbellata (P. mitis) 44801
P. umbellata (P. mitis) 92-079
P. umbellata (P. mitis) 6807
*P. umbellata (x angustifolia?) JSh774-775
P. umbellata (x angustifolia?) 10347
P. umbellata var. injucunda 71376
P. umbellata var. injucunda 31807
P. umbellata var. injucunda 59510
P. umbellata var. injucunda 22094
P. umbellata var. injucunda 22177
P. umbellata var. injucunda 3797
P.alleghaniensis var. davisii 16642
P.alleghaniensis var. davisii S.T.s.n.
P.alleghaniensis var. davisii 7139
P.alleghaniensis var. davisii 11067
P.alleghaniensis var. davisii A.D.s.n.
P.alleghaniensis var. davisii S.T.s.n.
*P.alleghaniensis var. davisii G.Ss.n.
P. americana 24915
P. nigra 16394
P. nigra 19384
P. nigra 6094
P. nigra 5051
P. nigra 21018
*P. nigra JSh979
P. americana 1372
P. alleghaniensis JSh836
P. americana 13130
P. americana NA5747
P. rivularis 3495
P. alleghaniensis JSh834
70
C
63
C1
Ca
79
Chickasaw clade
85
87
American clade
97
Beach clade
63
96
85
P. texana
P. umbellata 21172 (pU)
P. subcordata

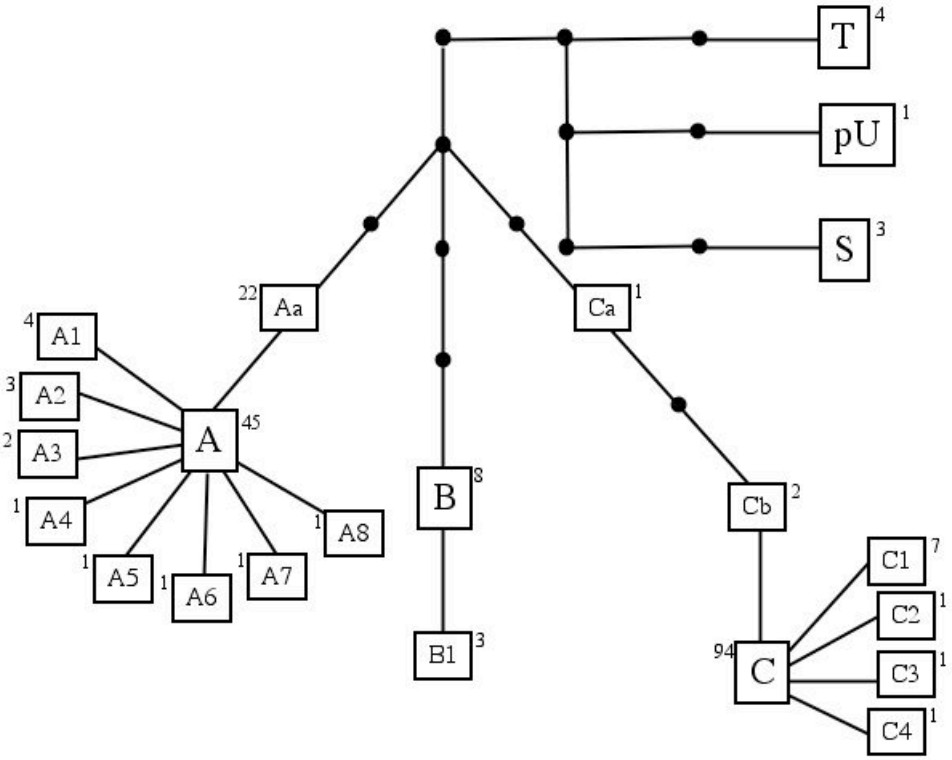


Fig. 3. TCS gene genealogy of *Prunus* subgenus *Prunus* section *Prunocerasus* from the *rpL16* intron. Haplotypes observed in this study are represented by squares. The number of times that the haplotype was observed is indicated by a number beside each of the squares. Lines connecting the haplotypes represent a single mutation (nucleotide substitution or indel) with solid circles representing inferred mutational steps not observed in this study. A = American haplotypes, B = Beach haplotypes, C = Chickasaw haplotypes, T = Texana haplotype, S = Subcordata haplotype, pU = a peculiar Umbellata haplotype.

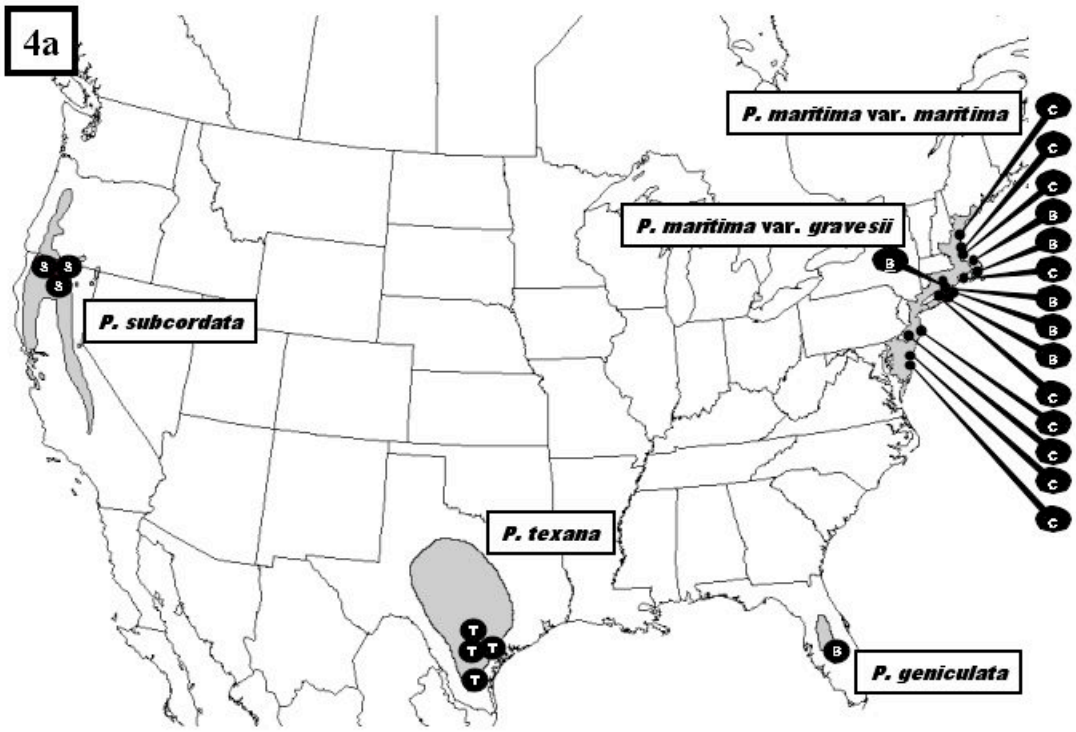


Fig. 4. (a) to (k). Geographical locations of haplotypes observed throughout the ranges of the 18 taxa of this study. Individual taxon ranges are shaded. Map 4a illustrates four species with relatively narrow and separated ranges and maps 4b-4k contain one species each. Where appropriate, varieties are denoted with an underline. A = American haplotypes, B = Beach haplotypes, C = Chickasaw haplotypes, T = Texana haplotype, S = Subcordata haplotype, pU = a peculiar Umbellata haplotype.

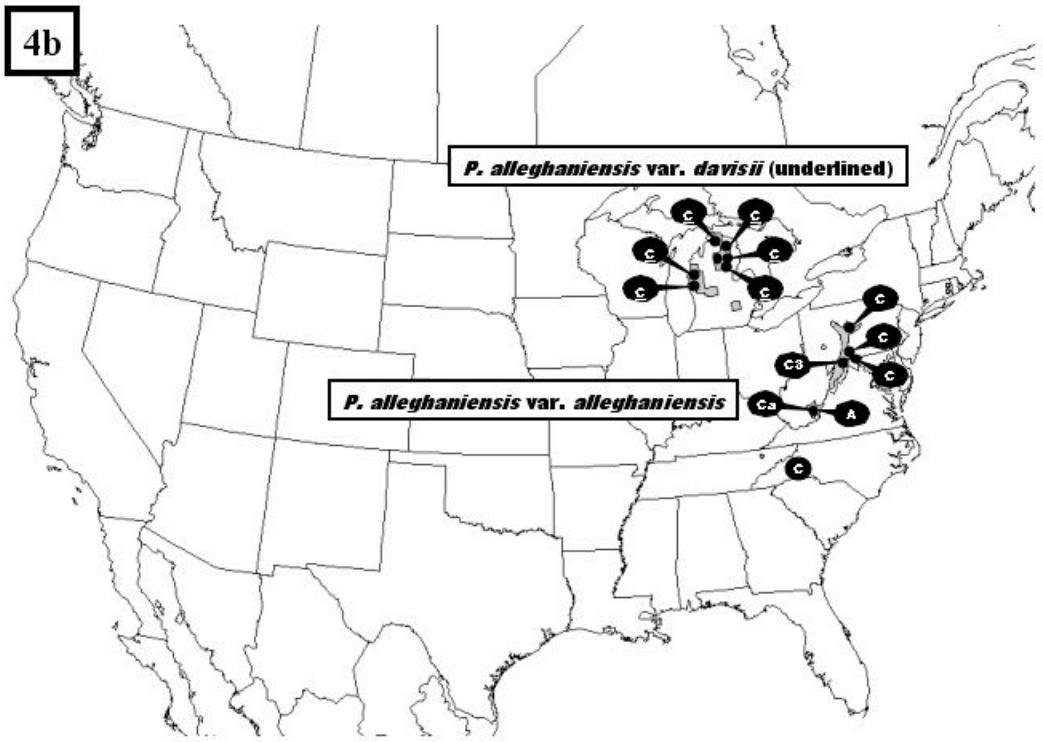


Fig. 4. continued.

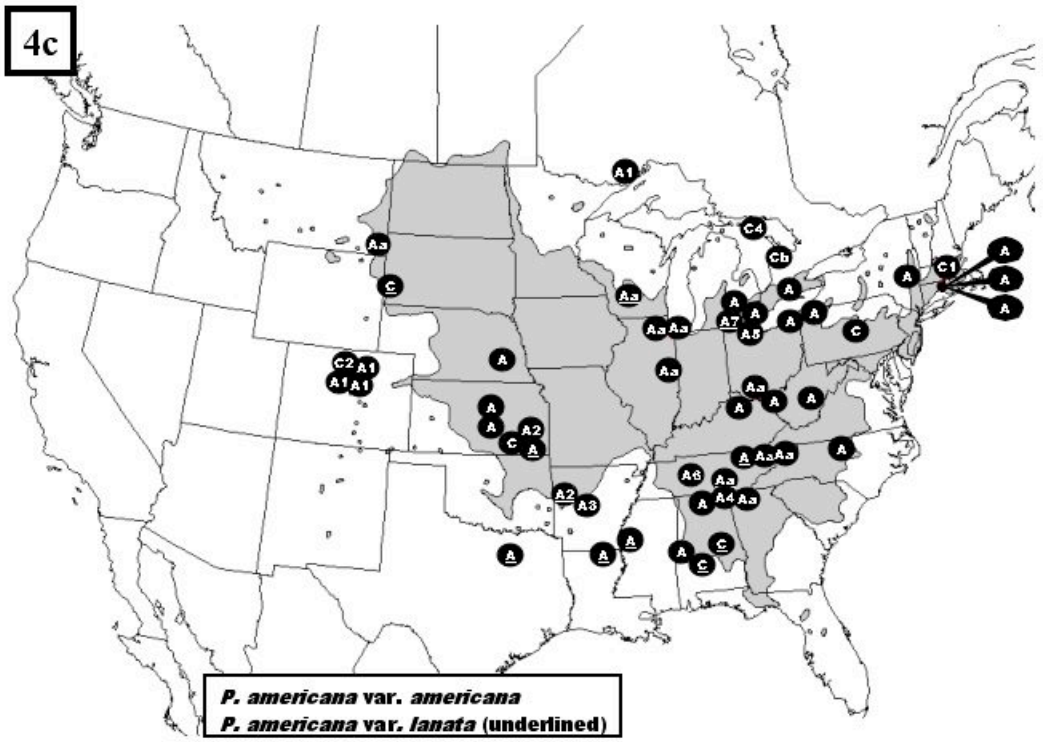


Fig. 4. continued.

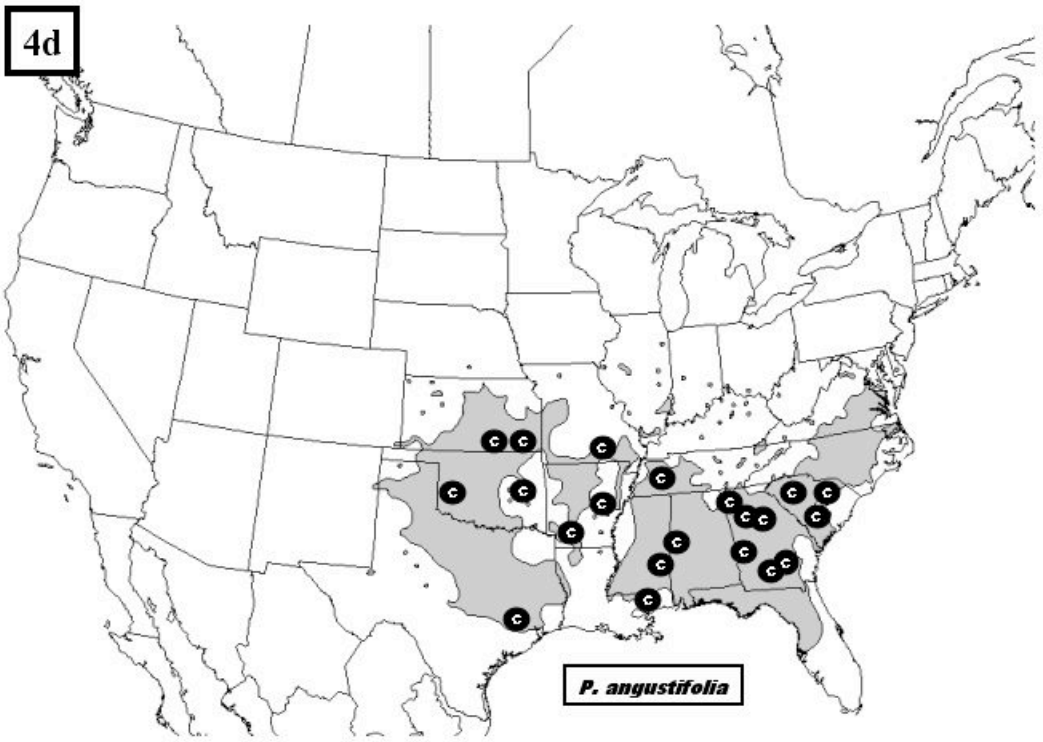


Fig. 4. continued.

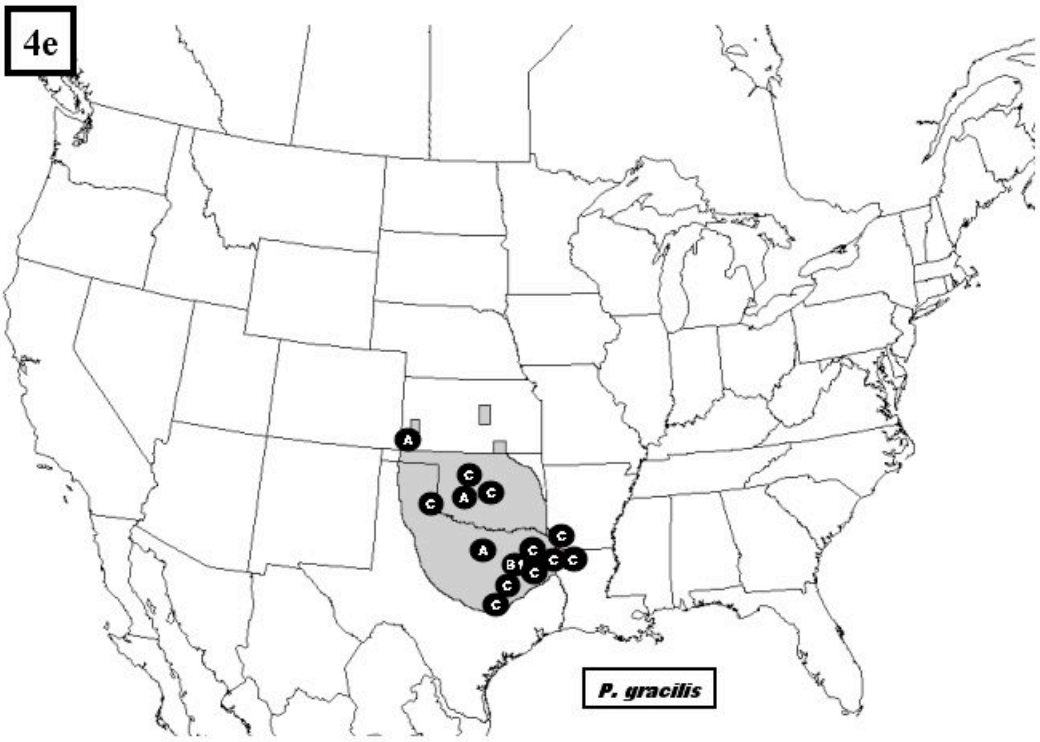


Fig. 4. continued.

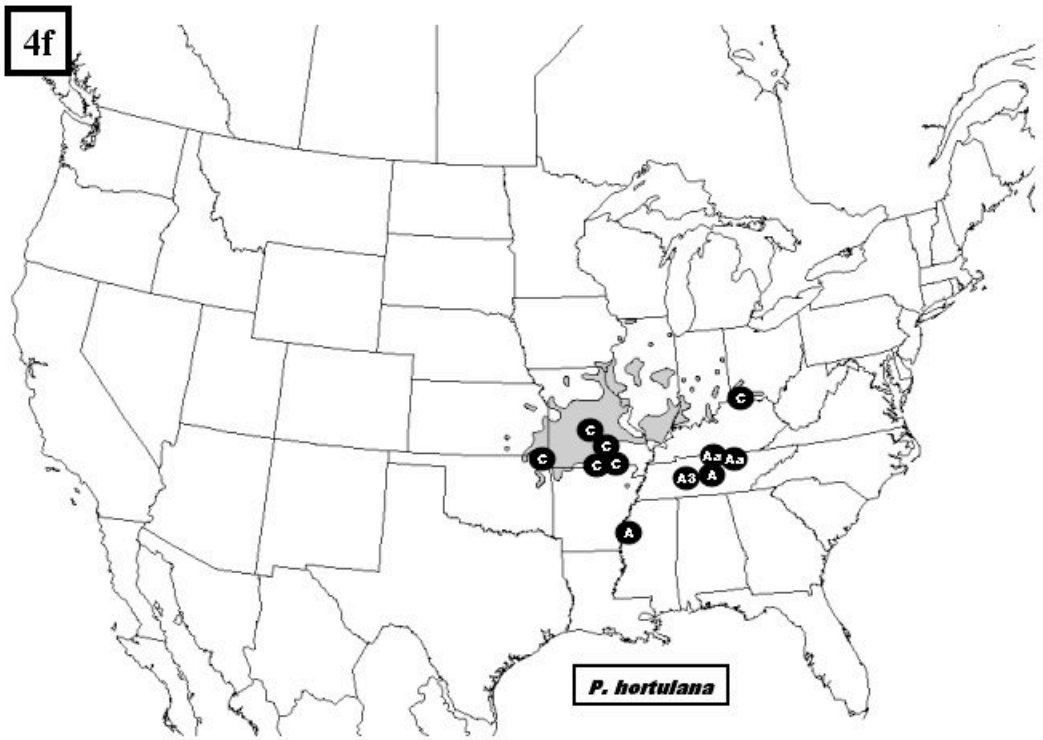


Fig. 4. continued.

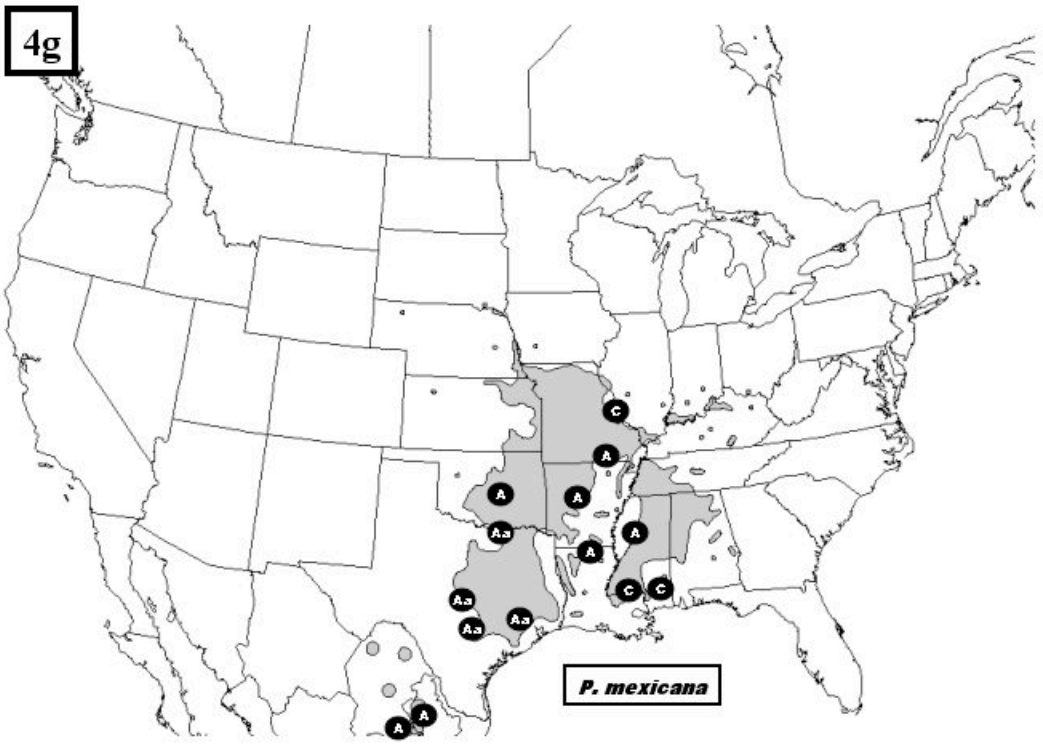


Fig. 4. continued.

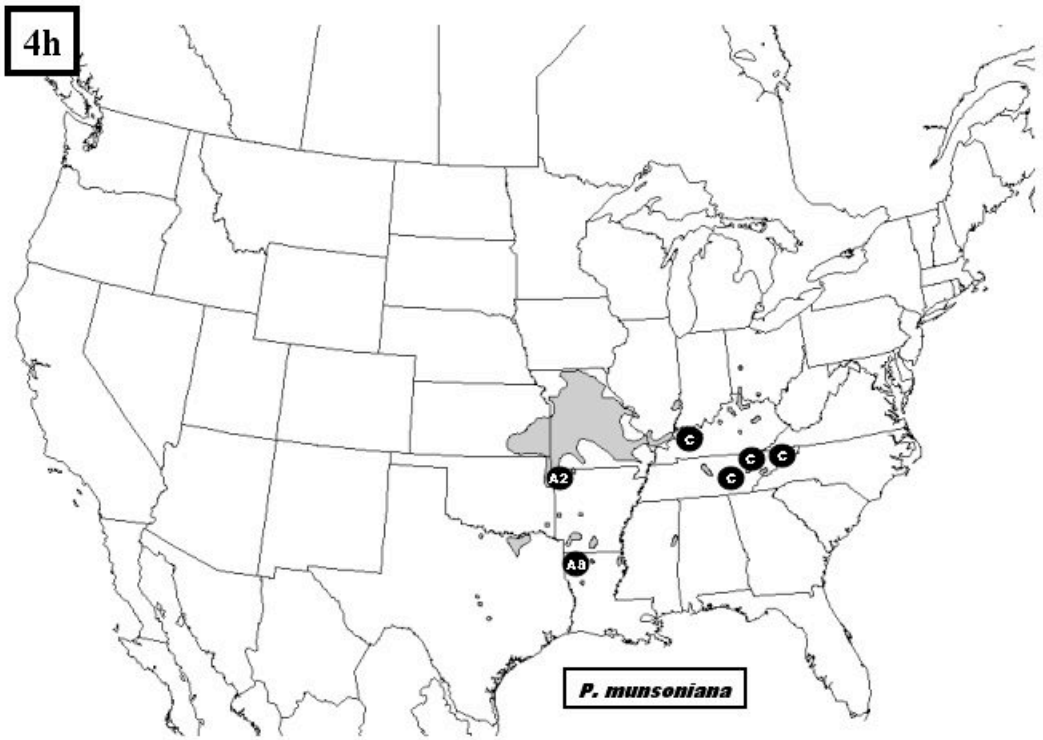


Fig. 4. continued.

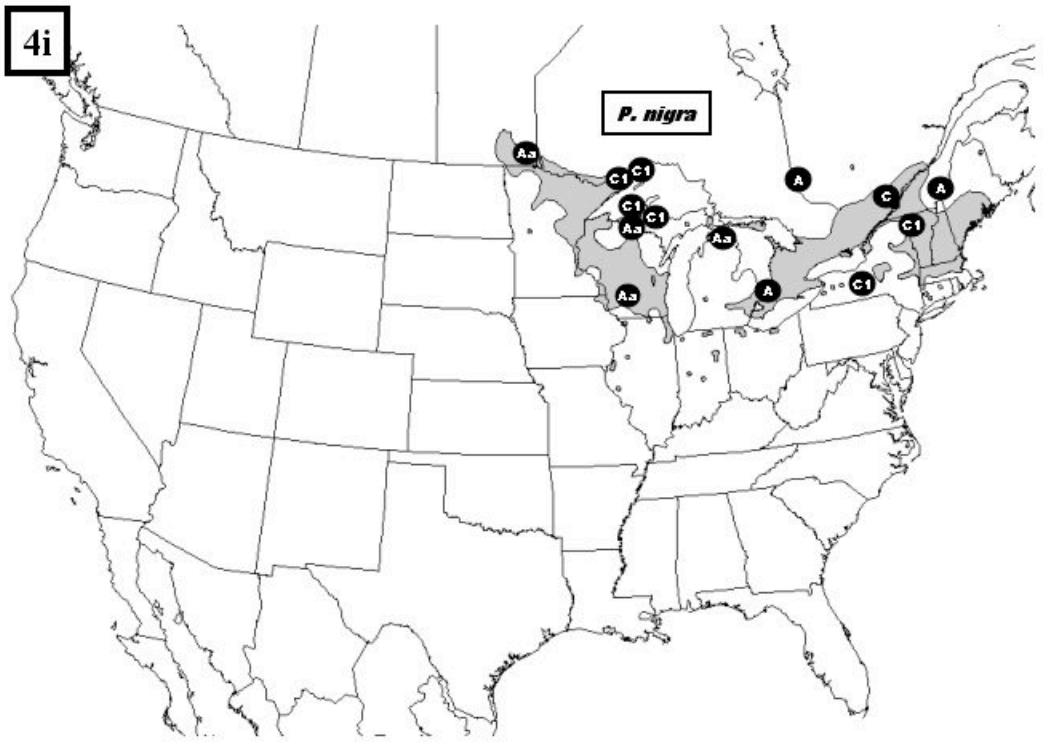


Fig. 4 continued.

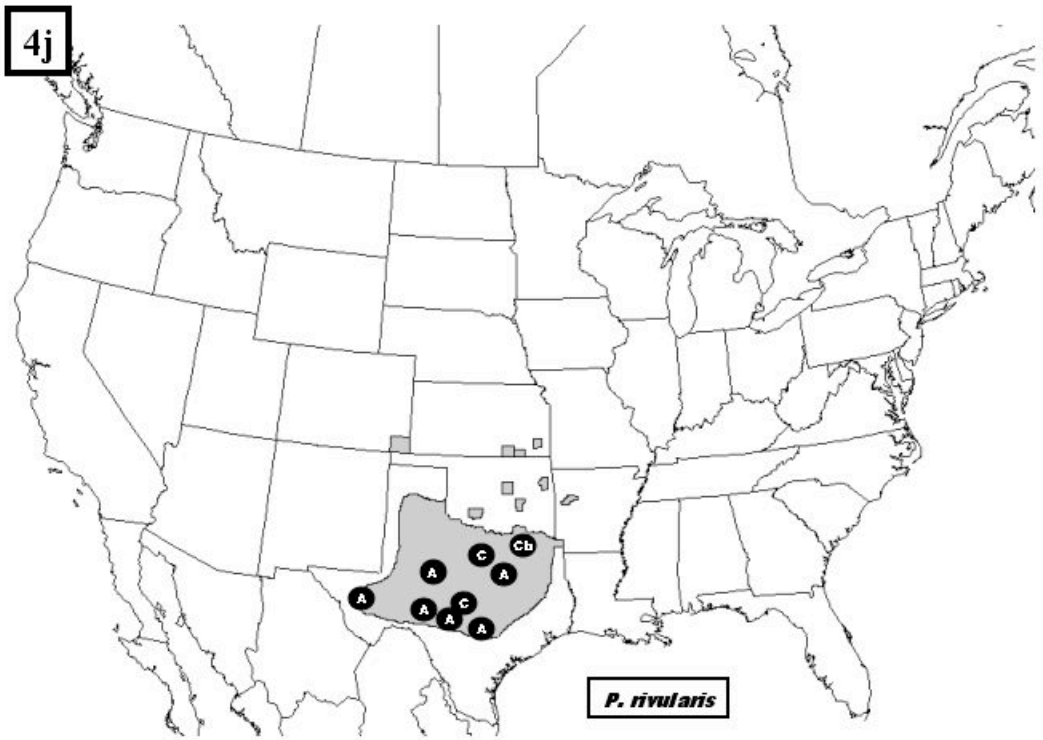


Fig. 4. continued.

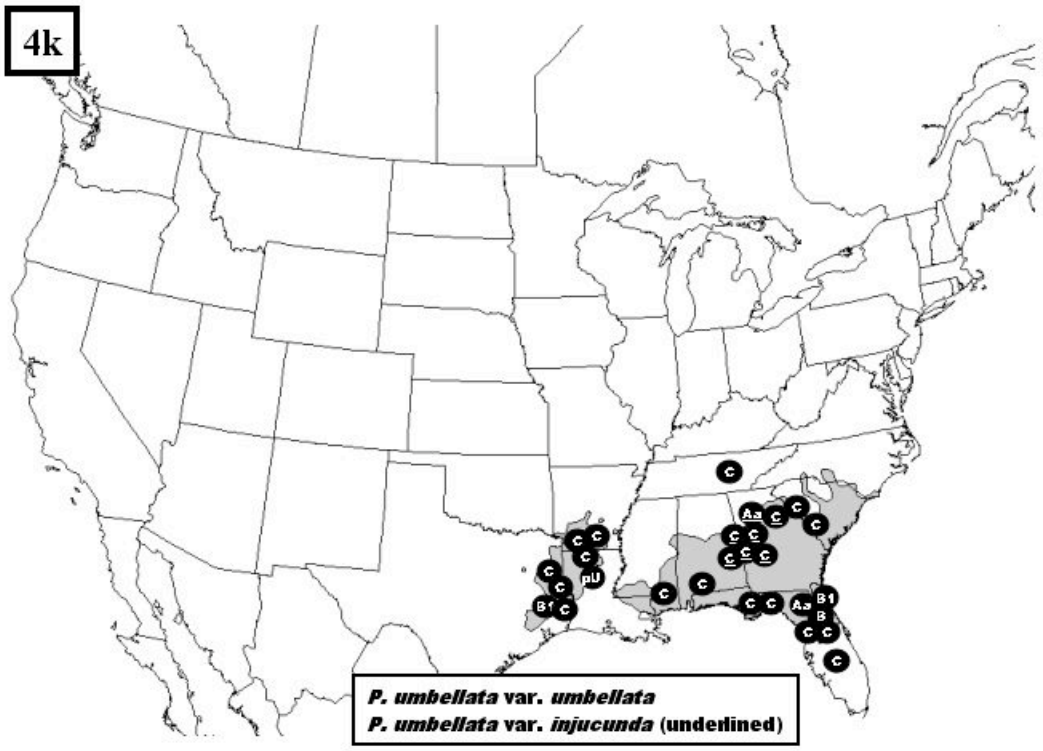
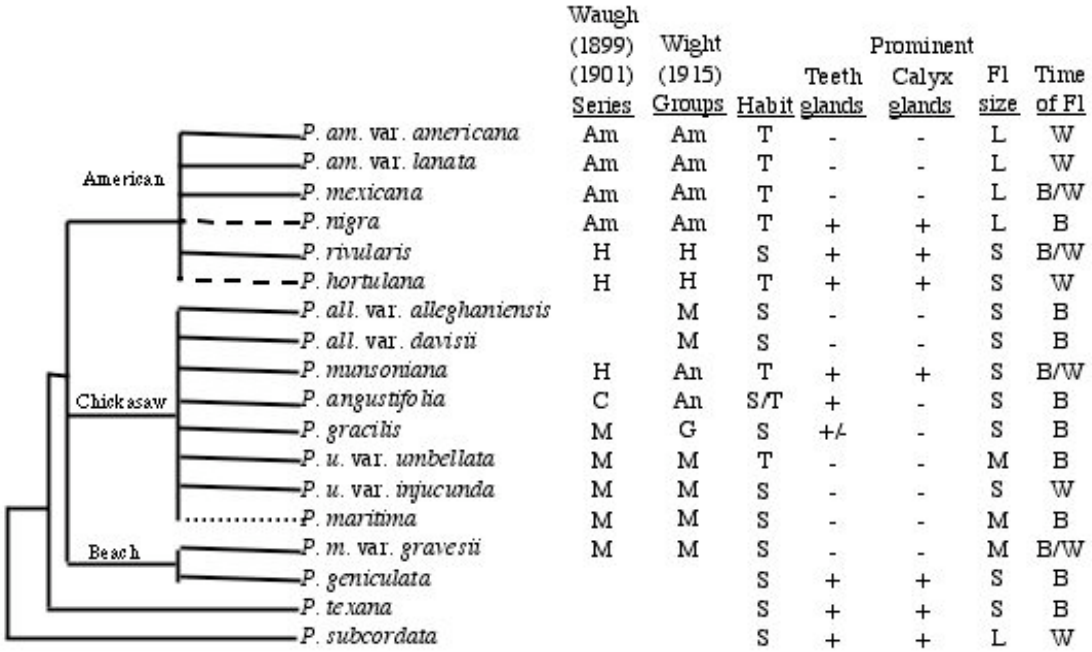


Fig. 4. continued.



	Waugh (1899) (1901)	Wight (1915)		Prominent			
	Series	Groups	Habit	Teeth glands	Calyx glands	Fl size	Time of Fl
P. am. var. *americana*	Am	Am	T	-	-	L	W
P. am. var. *lanata*	Am	Am	T	-	-	L	W
P. mexicana	Am	Am	T	-	-	L	B/W
P. nigra	Am	Am	T	+	+	L	B
P. rivularis	H	H	S	+	+	S	B/W
P. hortulana	H	H	T	+	+	S	W
P. all. var. *alleghaniensis*		M	S	-	-	S	B
P. all. var. *davisii*		M	S	-	-	S	B
P. munsoniana	H	An	T	+	+	S	B/W
P. angustifolia	C	An	S/T	+	-	S	B
P. gracilis	M	G	S	+/-	-	S	B
P. u. var. *umbellata*	M	M	T	-	-	M	B
P. u. var. *injucunda*	M	M	S	-	-	S	W
P. maritima	M	M	S	-	-	M	B
P. m. var. *gravesii*	M	M	S	-	-	M	B/W
P. geniculata			S	+	+	S	B
P. texana			S	+	+	S	B
P. subcordata			S	+	+	L	W

Fig. 5. Cladogram of *Prunus* subgenus *Prunus* section *Prunocerasus* based on the most frequent haplotype observed in each taxon. Two taxa, *P. nigra*, *P. hortulana*, had a haplotypic frequency near 50% and therefore could have been positioned in either the American or Chickasaw clades. The dotted line (*P. maritima* var. *maritima*) represents a taxon that may be positioned in either the Beach or Chickasaw clades even though the majority of accessions contained the C haplotype (see text). Also shown here are the classification schemes of Waugh (1899, 1901) and Wight (1915) in addition to five morphological characters often used to delimit taxonomic boundaries in the section. The classification schemes and morphological characters and are shown here because if the most frequent haplotype observed in a particular taxon is selected to represent that taxon, morphological characters and classical species assemblages more closely fit the phylogeny (compare to Fig. 3 in Shaw and Small, 2004).

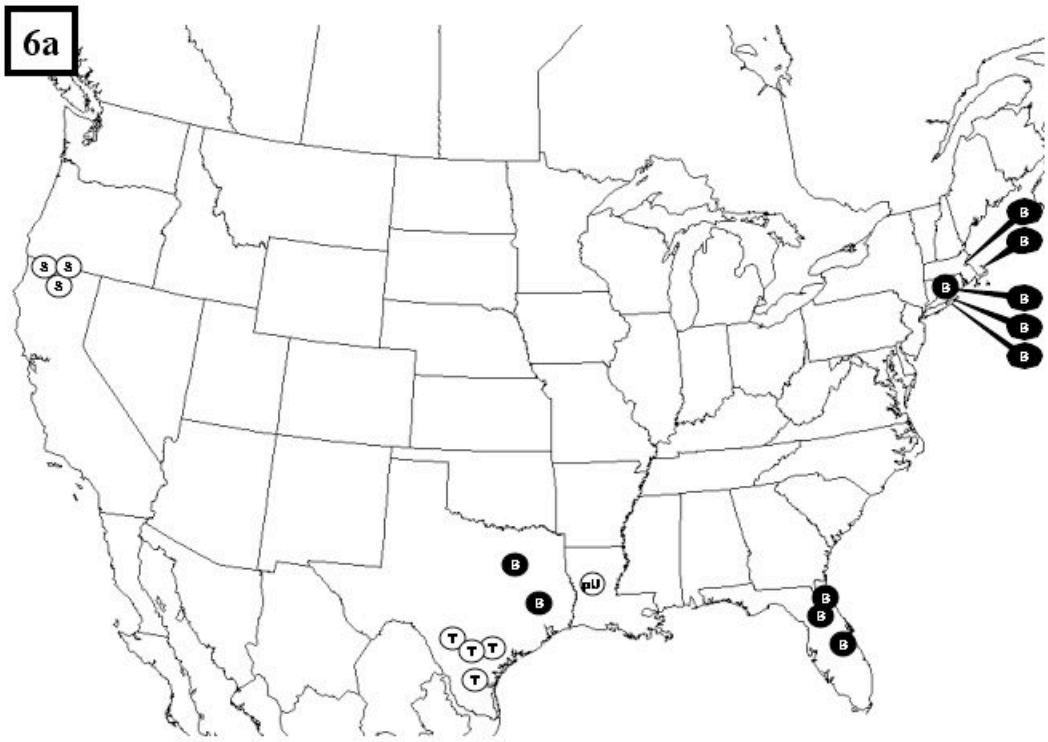


Fig. 6. (a) and (b). The distribution of the six major haplotypes (A, B, C, S, T, and pU) observed in this study. Fig. 6a shows the distribution of the B (Beach), S (Subcordata), T (Texana), and pU (peculiar Umbellata) haplotypes. Fig. 6b shows the distribution of the A (American) and C (Chickasaw) haplotypes.

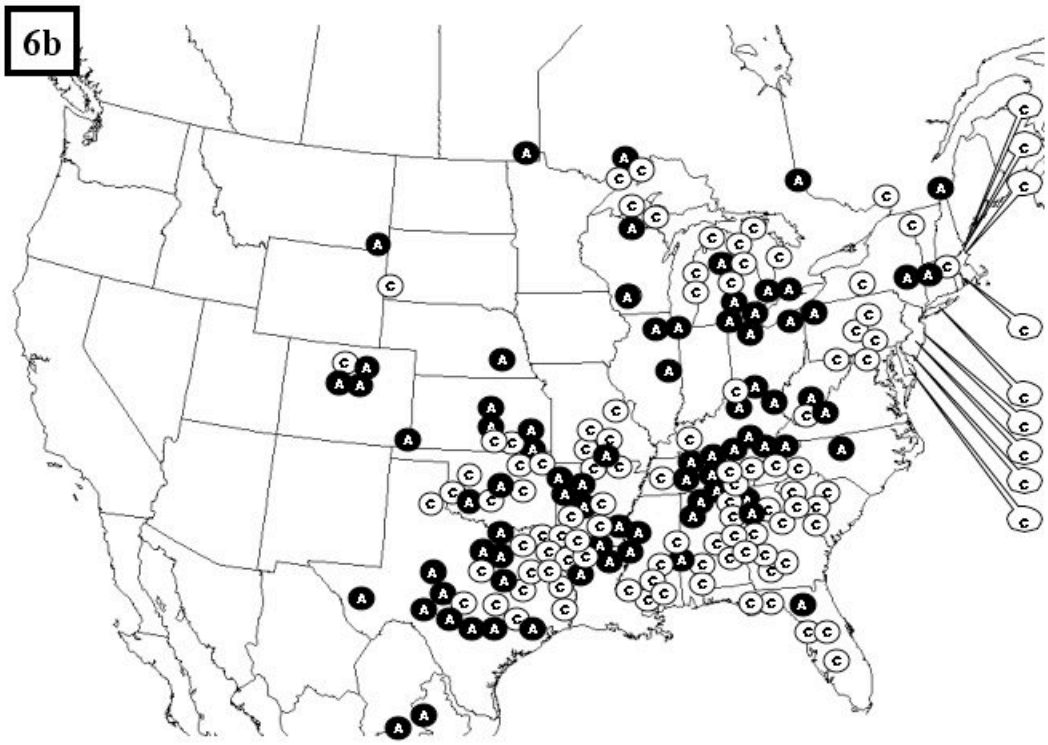


Fig. 6. continued.

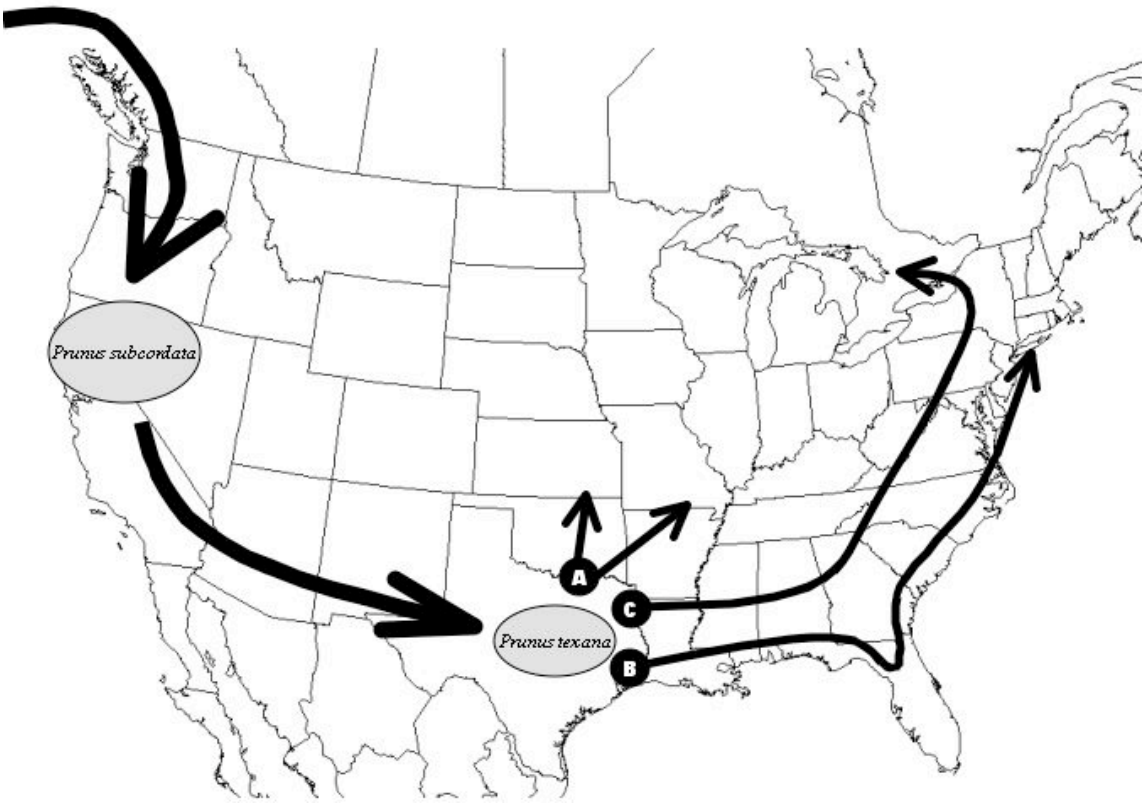


Fig. 7. Proposed migration route of the North American plums through North America.

PART 5

CONCORDANT PHYLOGEOGRAPHIC PATTERNS OF THE CHLOROPLAST AND NUCLEAR GENOMES IN THE NORTH AMERICAN PLUM SYNGAMEON (*PRUNUS* SUBGENUS *PRUNUS* SECTION *PRUNOCERASUS*; ROSACEAE)

This chapter is a manuscript, likely by the same name, that is being prepared for the *American Journal of Botany* by Joey Shaw and Randall Small. We anticipate its submission during the summer of 2005:

Concordant Phylogeographic patterns of the chloroplast and nuclear genomes in the North American plum syngameon (*Prunus* subgenus *Prunus* Section *Prunocerasus*; Rosaceae)

The use of "we" in this chapter refers to Randall Small and myself. My contributions to this chapter include (1) selection of the topic and development of the problem into a work relevant to study, (2) the lab work to generate data, and (3) most of the writing.

ABSTRACT

The North American plums (*Prunus* subgenus *Prunus* section *Prunocerasus*; Rosaceae) are infamous for their very poor development of reproductive isolating barriers. Many North American plum species are interfertile in many combinations and natural hybrids blur nearly all morphologically-based taxonomic boundaries. In previous studies, we showed that the North American plums are monophyletic and that chloroplasts were geographically, rather than taxonomically distributed. Conclusions from our earlier work showed that frequent hybridization has allowed regional sharing of the three primary chloroplast haplotypes among the 17 putative taxa in the section. Here we examined the relationships of the North American plums using the nuclear encoded NADP+ dependent sorbitol-6-phosphate dehydrogenase gene (*s6pdh*) to test the phylogenetic / phylogeographic conclusions of our earlier studies. The results showed that most section *Prunocerasus* taxa are clearly not monophyletic. Although some nuclear haplotypes are more common in some taxa, few are exclusive. Coupled with evidence from our previous studies, these results support the hypothesis that North American plums are an evolutionarily complicated species complex, or syngameon. Another important contribution of this study is to highlight the problem of undersampling in closely related groups of species. However, our results also show that problems can be circumvented by increased sampling.

INTRODUCTION

The foundation underlying the field of phylogenetics is the assumption that species evolve by the *complete* process of branching followed by reproductive, and therefore genetic, isolation — in theory, genetic similarity is due to recent common ancestry. Because of this assumption, polymorphisms that are shared across species boundaries, or trans-specific shared polymorphisms, pose significant problems to phylogenetic reconstruction. Trans-specific shared polymorphisms may occur for any of three reasons: (1) homoplasy, or the stochastic possibility that the same mutation has occurred in independent lineages, (2) hybridization, or (3) incomplete lineage sorting from ancestral polymorphism. The latter two phenomena can be difficult to distinguish from one another (Vollmer and Palumbi, 2004). In any case, the result is either decreased phylogenetic resolution or alteration in tree topology (McDade, 1992). To date there are no good statistical means of dealing with introgression or incomplete lineage sorting in

phylogenetic investigations (Linder and Rieseberg, 2004); therefore, phylogenetic inference among closely related species is often misled by these two phenomena (Harrison, 1991; Redenbach and Taylor, 2002; Funk and Olmland, 2003; Vollmer and Palumbi, 2004; Shaw and Small, submitted [Part 4]), especially if only a few samples are used to represent each of the taxa under study. In spite of these negative effects on phylogenetic inference, a phylogeographic sampling scheme can yield insights into the evolutionary processes that have occurred within a group of closely related species (Olsen, 2002; Funk and Olmland, 2003; Shaw and Small, submitted [Part 4]).

Hybridization is known to be an important process in plant evolutionary history (Anderson, 1949; Stebbins, 1959; Arnold, 1997; Reiseberg et al., 2003). Most evolutionary studies that have focused on hybridization are restricted in terms of focusing on relatively narrow zones of contact between closely related pairs of taxa (Barton and Hewitt, 1985; Roelofs and Bachmann, 1997; Martinsen et al., 2001). It has also been shown that alleles can introgress far outside the zone of morphologically recognizable intermediates and genetic material can travel through species across broad geographic regions and yet morphology might *not* reveal this transference (Dodd and Afzal-Rafii, 2004). With few exceptions (Whittemore and Schaal, 1991; McKinnon et al., 1999; Shaw and Small, submitted [Part 4]), little work has been done with widely hybridizing species complexes or syngameons in the sense of Grant (1981). A common finding of broadly sampled studies is that genetic material is often laterally transferred (Whittemore and Schaal, 1991; Kornkven et al., 1999; Shaw and Small, submitted [Part 4]). If lateral gene transfer is frequent enough, this can result in a phylogenetic pattern that may not reflect species boundaries but rather is concordant with geographical location instead (e.g., Whittemore and Schaal, 1991; Jackson et al., 1999; Shaw and Small, submitted [Part 4]). As more thoroughly sampled studies among species complexes are conducted, trans-specific shared polymorphisms appear to be more common than previously thought (Jackson et al., 1999; McKinnon et al., 1999; Shaw and Small, submitted [Part 4]; Chat et al., 2004).

The North American plums of *Prunus* subgenus *Prunus* section *Prunocerasus* (Rosaceae) are a closely related group with approximately 17 commonly recognized species and lesser taxa. They are infamous for their very poor development of reproductive isolating barriers and are comparable with other infamously taxonomically difficult genera like *Amelanchier* Medikus, *Crataegus* L., *Quercus* L., *Salix* L., and *Viola* L. Many North American plum species are interfertile in many combinations and natural hybrids blur nearly all morphologically-based taxonomic boundaries. Even still, morphological variation is such that when intermediates are ignored several taxa may be seen as being reasonably different from one another with respect to both morphology and ecology. Additionally, the ranges of most of the North American plum taxa overlap with the ranges of several others. The only exception is *P. subcordata*, which is the only species in the section whose range is west of the Rocky Mountains—it is allopatric from the rest.

In a previous study of the North American plum section of *Prunus* (Shaw and Small, 2004) that employed seven noncoding cpDNA regions (totaling 4375 bp), we showed that the North American plums are monophyletic; however, there was little genetic differentiation among taxa. Within that study, we selected single exemplars of

each of the taxa and showed that the majority segregate into three primary clades. Although these three primary clades were each supported with very high bootstrap values, within clades there was little to no phylogenetic resolution. An unexpected result of that study was that composition of the three clades contrasts sharply with previously defined groups based on morphological characters (Waugh, 1899, 1901; Wight, 1915; Rehder, 1940; see Fig. 3 in Shaw and Small, 2004). In other words, each of the three clades contains taxa that are morphologically more similar to taxa of other clades.

The discontinuity between morphology and the cpDNA-based phylogeny led to another cpDNA-based study (Shaw and Small, submitted [Part 4]) in which we employed the *rpL16* intron and a phylogeographic-level sampling scheme to analyze over 200 accessions of the 17 putative taxa. Shaw and Small (submitted [Part 4]), showed that three primary chloroplast types are scattered among most of the North American plum taxa leading to the conclusion that either the three primary chloro-haplotypes are shared among all North American plum taxa (through frequent hybridization) or incomplete lineage sorting from ancestral polymorphism has resulted in multiple chloro-haplotypes being present in most taxa. Because (1) frequent hybridization is known to occur in the group, (2) the two lineages were invariant in their respectively unique chloro-haplotypes, and (3) the distribution of the chloro-haplotypes shows strong geographic structure, the former explanation of chloroplast sharing through hybridization is more likely. This investigation not only highlights the strong effect of hybridization within the North American plums and cautions against narrow sampling in closely related species groups, but it has dramatically improved our understanding of section *Prunocerasus* evolution.

To date, plant molecular phylogenetic studies have mostly relied on either nuclear ribosomal DNA ITS (Baldwin et al., 1995) or a few chloroplast DNA regions (e.g., Taberlet, 1991; Chase et al., 1993; Johnson and Soltis, 1994). Increasingly, low-copy nuclear regions (e.g., Mason-Gamer et al., 1998; Small et al., 1998; Sang, 2002; Small et al., 2004) are being used in phylogenetic investigations because they can evolve more quickly than organellar DNA in plants (Gaut, 1998; Wolfe et al., 1987) and therefore should provide greater numbers of characters to phylogenetic investigations. Since independent phylogenies inferred from different genomes are necessary to both corroborate hypotheses of relationships and identify speciation via hybridization, a low-copy nuclear DNA data set is integral to understanding the evolution of the North American plums. In a previous study, Bortiri et al. (2002) characterized the nuclear DNA NADP+ -dependent sorbitol-6-phosphate dehydrogenase gene (*s6pdh*) and developed *Prunus*-specific primers. In that study, they pointed out that intron one (~ 0.9 kb) was highly variable among distantly related *Prunus* species; therefore, this intron is ideal for the study of closely related *Prunus* species.

In the present study, we examine the phylogenetic and phylogeographic relationships within the North American plums of *Prunus* section *Prunocerasus* with *s6pdh* gene sequences. Using a sampling scheme parallel to Shaw and Small (submitted [Part 4]), the aim of this study was to test the phylogenetic / phylogeographic observations of our earlier study and possibly gain new insights into the evolutionary history of the North American plums.

MATERIALS AND METHODS

***Taxon sampling*—**Our sampling strategy aimed to cover both the range of each of the putative species and also the multiple chloro-haplotypes observed in each of the species in Shaw and Small (submitted [Part 4]). Therefore, our taxonomic sampling in this study is a subset of that of Shaw and Small (submitted [Part 4]) and includes multiple representatives of each of the commonly accepted members of section *Prunocerasus*, according to: Small (1933); Rehder (1940); Fernald (1950); Bailey and Bailey (1941); Blackburn (1952); Gleason (1952); Steyermark (1963); Radford et al. (1968); Correll and Johnston (1970); Duncan and Duncan (1988); Godfrey (1988); Wunderlin (1988); Gleason and Cronquist (1991); Smith (1994); Wofford and Chester (2002). A single sample of the Federally Endangered *P. geniculata* was obtained from Carl Weakley at the Archbold Biological Station in Lake Placid, Florida. Voucher information for each of the accessions is in Table 1.

***The nuclear-encoded* s6pdh *gene*—**In Rosaceae, excluding Rosoideae sensu stricto, sucrose and the sugar alcohol sorbitol are the most important products of photosynthesis (Webb and Burley, 1962). Sorbitol-6-phosphate is produced by the reduction of glucose-6-phosphate by the nuclear encoded NADP+ -dependant sorbitol-6-phosphate dehydrogenase gene, or *s6pdh*. *S6pdh* consists of six exons and 5 introns in both *Malus domestica* (Bains et al., 1998) and many *Prunus* species (Bortiri et al., 2002). In *Malus domestica* the entire region is approximately 3400 bp; intron one is by far the largest at 1819 bp, while the other five introns range between 114-266 bp. This size difference between intron 1 and the others is also the case in *Prunus*, but intron one is approximately 800 - 900 bp. Intron one of *s6pdh* was chosen for phylogenetic analysis of the North American plums because Bortiri et al. (2002) suggested that this region was highly divergent and therefore difficult to align among a subset of divergent *Prunus* species. Because of the putative quick mutation rate in intron one, this region may provide an adequate number of characters to generate a well-resolved phylogenetic hypothesis among the closely related North American plum taxa.

***Laboratory procedures*—**DNA was extracted from leaves using the DNeasy Plant Mini Kit (Qiagen, Valencia, California, USA). PCR was performed using Eppendorf Mastercycler gradient or Mastercycler personal thermal cyclers in 50 µl volumes with the following reaction components: 1 µL template DNA (~ 10-100 ng), 1X *ExTaq* buffer (PanVera / TaKaRa, Madison, Wisconsin, USA), 200 mmol/L each dNTP, 3.0 mmol/L $MgCl_2$, 0.2 µmol/L each primer, and 1.25 units *ExTaq* (PanVera / TaKaRa). Reactions included bovine serum albumin at a final concentration of 0.2 mg/mL, which is known to improve amplification from difficult templates.

PCR was performed with the s6pdh-a (ATG CCG GTC ATC GGT CTC GG) and 77L25 (TGG CCT TGG AGG TGA TGA AAA GTT C) primers that were originally published in Bortiri et al. (2002), see that study for a diagram of *s6pdh* and primer placement. The PCR amplification protocol was preceded by template DNA denaturation at 80°C for 5 minutes and followed by a final extension step of 10 min at 65°C. The PCR cycling conditions were 25-30 cycles of denaturation at 95°C for 1 min, primer annealing

at 56°C for 2 min, followed by a ramp of 0.3°C/sec to 65°C, and primer extension at 65°C for 4 min.

Amplification products were visualized by agarose gel electrophoresis before being cleaned with the Wizard Purification System (Promega, Madison, Wisconsin, USA). Cleaned amplification products were ligated into the pGEM-T vector (Promega, Madison, Wisconsin, USA) according to the manufacturers instructions. Competent cells (Promega, Madison, Wisconsin, USA) were transformed via heat shock and 8-15 of the resulting colonies were screened for plasmids via PCR using the vector specific T7 and SP6 primers. This PCR amplification protocol was preceded by template DNA denaturation at 80°C for 5 minutes and followed by a final extension step of 10 min at 65°C. The PCR cycling conditions for amplifying the cloned fragments were 25-30 cycles of: denaturation at 94°C for 1 min, primer annealing at 50°C for 1 min, and primer extension at 72°C for 2 min. For each accession, five colonies yielding bands of the appropriate size (~850 bp) were chosen for sequencing in an attempt to obtain both alleles in the heterozygotes. The PCR amplicons were then cleaned using ExoSAP-IT (USB, Cleveland, Ohio, USA) and sequenced with a single primer (77L25). All DNA sequencing was performed with the ABI Prism BigDye Terminator Cycle Sequencing Ready Reaction Kit, v. 3.1 (Perkin-Elmer / Applied Biosystems, Foster City, California, USA) and electrophoresed and detected on an ABI Prism 3100 automated sequencer (University of Tennessee Molecular Biology Resource Facility). Sequencher 3.1.1 (Gene Codes) was used to edit the DNA strands. All sequences will be deposited in GenBank.

***Data analysis*—**Alignment of DNA sequences was initially performed with ClustalX (Thompson et al., 2001), with subsequent manual adjustment by eye in MacClade v. 4.0 (Sinauer, Sunderland, Massachusetts, USA). Variable positions in the data matrix were double checked against the original chromatogram files to make sure that all base calls were true at all variable positions. Approximately 100 bp were omitted from the middle portion of the sequenced region because this region is A/T rich with various repeats was unalignable. In all other cases, alignment of potentially informative positions was unambiguous. Indels were coded as binary characters except in the case of a single poly-A/T run, which was omitted from the data set because variation in length of this run may be PCR-artifact and not reflective of the phylogenetic history of the group.

Phylogenetic analyses were performed with both PAUP* v. 4.0 b10 (Swofford, 2002), based on maximum parsimony and with MrBayes 2.01 (Huelsenbeck and Ronquist, 2001), based on Bayesian inference. Parsimony analysis was carried out by a heuristic search with tree bisection-reconnection (TBR) branch swapping. Bootstrap support (Felsenstein, 1985) was estimated with 1000 replications of heuristic search and simple taxon addition with the constraint of 1 000 000 rearrangements per replicate. Both the consistency and retention indices (CI and RI, respectively) were used to assess the amount of homoplasy present in the data set.

Bayesian analysis was performed using MrBayes 2.01 to generate posterior probability distribution using Markov chain Monte Carlo (MCMC) methods. No a priori assumptions about tree topology were made. The statistical model of DNA substitution, Kimura 1980 (K80), was estimated as the best-fitting maximum likelihood model using MrModeltest 1.1b (Nylander, 2002), which is a simplified version of MODELTEST 3.06

(Posada and Crandall, 1998). The Monte Carlo Markov Chain (MCMC) process was set to run one million generations with four chains. Burn-in was estimated visually by plotting log-likelihood values in Microsoft Excel to determine the number of generations that had run before likelihood values reached an asymptote. To calculate the posterior probability of each bipartition a 50% majority-rule consensus tree was constructed from the remaining trees using PAUP*.

Relationships among the nDNA haplotypes were also inferred using the software TCS v. 1.13 (Clement et al., 2000), which implements a statistical parsimony approach to estimating gene genealogies. TCS was performed as an alternative means of analysis because it can infer ancestral or intermediate haplotypes (as opposed to assuming that these haplotypes are extinct) and loops (possible recombinants). For the TCS analysis we used the same sequence alignment that was used in the PAUP* analysis described above.

***Geographic distribution of haplotypes*—**To assess the distribution of haplotypes within each of the putative taxa of section *Prunocerasus*, distribution maps were created using range information from the USDA national PLANTS database (USDA, 2002), Little (1971, 1976, 1977), Gleason and Cronquist (1991), and Correll and Johnston (1970). Haplotypes found within each of the taxa were then mapped on to the distribution maps.

RESULTS

***Characterization of the* s6pdh *data set*—**The aligned data set consisted of 955 bp. Within the aligned matrix there were 85 parsimony informative characters (including multi-bp indels coded as binary characters) and 11 variable but parsimony uninformative characters. In the maximum parsimony analysis, a heuristic search found 653 equally parsimonious trees of 163 steps with average consistency and retention indices of 0.72 and 0.93, respectively (indicative of some homoplasy in the data set). Bootstrap values are shown above the branches in Fig. 1. The topology of the tree generated with Bayesian analysis is largely congruent with the maximum parsimony tree, but support values were higher (Fig. 2). The TCS analysis revealed 58 unique haplotypes among the ingroup taxa excluding *P. subcordata* (Fig. 3). This unrooted network is largely congruent with the topologies of the maximum parsimony and Bayesian phylogenies (compare Fig. 3 to Figs. 1 and 2). In contrast to the phylogenetic analyses several haplotypes were identified as internal or ancestral (e.g., haplotypes 1, 30, and 38) and the analysis revealed eight reticulations. Because many of the 58 unique haplotypes were only observed once and are closely associated with more frequently observed haplotypes, we cannot rule out that they are the result of cloning-amplified *Taq* error. Because of this uncertainty and to facilitate interpretation and discussion, the 58 haplotypes were grouped into three primary haplotype groups, designated □, □, and □, respectively (Figure 3). These three primary haplotypes correspond to the three primary clades of Fig. 1 and are indicated in Figs. 1, 2, and Table 1.

An additional haplotype, 29, found in the allelic accessions *P. hortulana* 821 B and *P. umbellata* 787 B was not allied to any of these primary haplotype groups because (1) an easy decision could not be made as to whether it should be grouped with □ or □

and it was positioned in two different places in the parsimony and Bayesian analyses (shown in bold in Figs. 1 and 2).

Before the abovementioned analyses, other parsimony, Bayesian, and TCS analyses were performed that included five additional allelic accessions: *P. munsoniana* 810 A, *P. munsoniana* 960 B, *P. munsoniana* 961 A, *P. americana* 959 A, and *P. americana* 1038 A (designated with an "-" in Table 1 and Fig. 5d and drawn on to Figs. 1, 2, and 3 with dashed lines where they were positioned in those previous analyses). These five accessions comprised two closely related alleles that appear to be recombinants. Because we cannot distinguish if these recombinant alleles are true recombinants or the result of PCR-mediated recombination and because they reduced the resolution of our phylogenetic inference, we ultimately excluded them from our analyses.

We sampled 57 accessions of the 18 taxa of this investigation (17 North American plum taxa plus *P. texana*). Forty-six of the 57 accessions were heterozygous (this is not truly reflected in Table one because only primary haplotype groups are shown in Table 1).. In a few cases (e.g., *P. maritima* 877 A-D), more than two haplotypes were identified (and therefore subsequently included in the abovementioned analysis). In all cases, these "extra" haplotypes came out sister to one of the commonly found haplotypes (Figs. 1 and 2) and therefore are likely artifacts from clone-amplified *Taq* error.

***Phylogenetic results and allelic distribution*—**The most significant results of this study show that three primary haplotypes are shared among the North American plums taxa. Of the 18 taxa (17 North American plums + *P. texana*) all three primary haplotype groups (□, □, and □) were found in *P. maritima* var. *maritima*, two were found in *P. alleghaniensis* var. *alleghaniensis*, *P. alleghaniensis* var. *davisii*, *P. americana* var. *americana*, *P. munsoniana*, *P. nigra*, and *P. umbellata* var. *umbellata*, and *P. nigra*. Restating these results in a different light, the □ haplotype was found in 6/17 taxa, the □ haplotype in 11/17 taxa, and the □ haplotype in 8/17 taxa.

The inferred phylogeny (Figs. 1 and 2) shows that *P. subcordata* is sister to the remaining taxa in the group. Multiple accessions of most of the other taxa under study are scattered across the topology of the inferred trees. For example, alleles that were found in *P. maritima* var. *maritima* are located on all three branches of the phylogeny and alleles from the same individual are not necessary in the same clade. The only major difference in topology between the parsimony and Bayesian trees concerns the phylogenetic placement of *P. hortulana* 821 B and *P. umbellata* 787 B (in bold in Figs. 1 and 2). These two alleles are central in the TCS analysis (haplotype 29, Fig. 3). Lastly, *P. texana*, currently classified as a peach in subgenus *Amygdalus* is strongly embedded within the North American plum section of subgenus *Prunus*.

Because this data set does not support the long-standing morphologically-based intrasectional concepts of section *Prunocerasus*, haplotype maps were generated to look for geographic structure in haplotypic distribution. When individual species haplotype maps were created, no obvious pattern emerged (Figs. 4a-k). However, when all haplotypes were mapped irrespective of the taxa in which they were found a pattern emerged that corroborates the pattern observed in the distribution of the three primary chloroplast haplotypes (compare Fig. 5e to Figs. 5a-d). Haplotype □ was found distributed throughout the interior of the USA and was rare in the Coastal Plain (Fig. 5a) .

Haplotype was locally abundant in central Texas, the panhandle of Florida, the northeastern coast, and the interior highlands (Fig. 5b). Haplotype was distributed through the Atlantic and Gulf Coastal Plain, rare in the interior, and in the Great Lakes region (Fig 5d). Lastly, the two putative recombinant haplotypes groups, "-" and 29, respectively, were mostly clustered together around Tennessee.

DISCUSSION

Phylogenetic relationships within section **Prunocerasus—**Hybridization is an important mechanism in the evolution of plants and its key effect is an increase in genetic variation. This can lead to the formation of ecological races and in the end new species (Welch and Rieseberg, 2002). Here we report that the North American plums lack the genetic differentiation necessary to conclude that most taxa are reproductively isolated or even only occasionally hybridizing. The primary result of this study shows that most North American plum taxa share more than one of the three primary nDNA haplotypes observed in this study. This result in combination with the facts that the North American plums also contain multiple chloroplast haplotypes (Shaw and Small, submitted [Part 4]) that do not necessary correlate to nDNA haplotypes (Table 1) and that hybridization is well-documented in the group (Waugh, 1899, 1901; Wight, 1915; Rehder, 1940; Shaw and Small, submitted [Part 4]) support the hypothesis that the North American plums form a syngameon in the sense of Grant (1981).

Earlier studies have shown that *Prunus* consists of two primary clades, the *Amygdalus-Prunus* (plum-peach-apricot) clade and the *Cerasus-Laurocerasus-Padus* (the cherry clade) clade (Bortiri et al. 2001, Lee and Wen, 2001; Bortiri et al., 2002; Shaw and Small 2004). In the aforementioned studies, accumulated evidence has adequately shown the monophyly of North American plums; therefore, we did not heavily sample outgroup taxa. However, the outgroup taxa that we did choose are members of the *Amygdalus-Prunus* (plum-peach-apricot) clade and are therefore relatively closely related to the ingroup. Not surprisingly, the two Old World plum species (four alleles) group together and the two apricot alleles group together. The phylogeny was rooted with *P. davidiana*, a peach.

Prunus subcordata is unique among the North American plum species because it is the only one with a western North American distribution. All of the other taxa have ranges east of the Rocky Mountains. The accession of *P. subcordata* was heterozygous and contained alleles that were both sister to the remaining North American plum species and sister to each other (Figs. 1 and 2). The *P. subcordata* alleles were too different for TCS to incorporate into the network (Fig. 3) even with the connection limit reduced to 90%. This phylogenetic position for *P. subcordata* is now strongly supported by several lines of evidence including cpDNA (Bortiri et al., 2001; Shaw and Small, 2004) and nDNA (Bortiri et al., 2002; this study).

Prunus texana, a species endemic to sandy soils in central Texas with pubescent fruits that is currently classified as a peach in subgenus *Amygdalus* (Wight, 1913), was positioned with strong support in a derived clade embedded in the North American plums (clade in Figs. 1 and 2). This species was originally chosen for study because of its striking resemblance to *P. geniculata*, a morphologically unique species that is endemic to sandy soils in Florida. In earlier studies, we showed that *P. texana* contains a

chloroplast haplotype that is unique among the plums and phylogenetically positioned in a grade between *P. subcordata* and the A, B, and C tritomy (Shaw and Small, 2004; Shaw and Small, submitted [Part 4]). Interestingly, in this study the single accession of *P. texana* is embedded in a derived clade that also contains *P. geniculata* (Figs. 1 and 2). More work needs to be done with this taxon, specifically the addition of several more acceccions from throughout its range, to more accurately determine its phylogenetic position. Based on three studies and both cpDNA and nDNA data (Shaw and Small, 2004; Shaw and Small, submitted [Part 4]; this study) it appears that *P. texana* is possibly a plum in peach clothing (pubescent-fruits).

In a previous study (Shaw and Small, 2004), we reported that the majority of the North American plums form a tritomy sister to *P. subcordata* plus *P. texana*. Because the sectional relationships revealed in that study were incongruous with classical morphological assemblages (Waugh, 1899, 1901; Wight, 1915; Rehder, 1940) we surveyed 207 accessions of each of the 17 taxa to test taxon monophyly (Shaw and Small, submitted [Part 4]). Shaw and Small (submitted [Part 4]) showed that three primary chloroplast haplotypes (A, B, and C) were scattered among the majority of the North American plum taxa. These results suggested that these three chloroplast haplotypes are shared among the majority of the North American plum species because of extensive hybridization.

Here we show that the nDNA *s6pdh* haplotypes are shared among many North American plum taxa. Of the 18 North American plum taxa (17 + *P. texana*) all three primary haplotype groups (□, □, and □) were found in *P. maritima* var. *maritima* and two were found in *P. alleghaniensis* var. *alleghaniensis*, *P. alleghaniensis* var. *davisii*, *P. americana* var. *americana*, *P. munsoniana*, *P. nigra*, and *P. umbellata* var. *umbellata*, and *P. nigra*. Because most of the taxa contain multiple nDNA haplotypes, phylogenetic analysis reveals an evolutionary history that is not reflective of morphologically-defined taxonomic groups (the *s6pdh* gene tree is not reflective of a morphologically-based species tree). While this data set is not as broadly sampled as the cpDNA data set of Shaw and Small (submitted [Part 4]), there is strong congruence between the two studies. For example, in many accessions where the C chloroplast haplotype was found we also found the nDNA *s6pdh* □ haplotype and the same is true for the A chloroplast and □ *s6pdh* haplotypes (Table 1). To compare and contrast the relationships revealed in both studies the primary cpDNA and nDNA haplotypes are shown next to the vouchers from which they were obtained in Table 1. This is not to suggest perfect correlation between the *s6pdh* and cpDNA data sets. Several cases can be found where a C chloroplast haplotype was found in an individual with an □ *s6pdh* haplotype. The data presented here, combined with the results of other studies (Rohrer et al., 2004; Shaw and Small, 2004; Shaw and Small, submitted [Part 4]), show that section *Prunocerasus* has a complicated and reticulating evolutionary history. In light of these data, we submit that there is no phylogeny of the North American plum taxa in the sense that none of the putative species have been non-recombinant for a period of time long enough to allow each to accumulate neutral mutations independently and therefore reveal a bifurcating phylogenetic pattern that is reflective of morphology.

According to Grant (1981, p. 37), "geographical races can live in different habitats in the same area and interbreed and intergrade in numerous zones of contact, but

preserve their distinctive racial characteristics in their respective habitats." Grant (1981, p. 74) went on to say that a "syngameon is the most inclusive interbreeding population system in a hybridizing species group. The syngameon behaves like a well-isolated biological species on its outer boundary, but differs in its more complex internal structure" (Grant, 1981, p 74).

Evidence from the data presented here and in Shaw and Small (submitted [Part 4]) shows that the genetic diversity of the North American plums lies in a region from east Texas stretching northeastward to west Tennessee (Fig. 4a-4k and Figs. 4a-4k in Shaw and Small, submitted [Part 4]). This is where the majority of the cpDNA and nDNA *s6pdh* haplotypes were found together. It is not surprising that in this same region species boundaries are also the most inconspicuous. *Prunus hortulana*, whose general range is in this region, was described as "a great congeries of hybrids…that are notably variable…and run into each other in the most puzzling and intricate manner" (Waugh, 1899, p. 233). Waugh (1899) further wrote that *P. hortulana* grades into *P. americana* to the northwest and to the south and ultimately grades into *P. angustifolia* "so connectedly as to leave not the slightest break between them" (Waugh, 1899, p. 233). Other taxa not recognized by Waugh (1899) also have ranges in this general area and are likely part of this same hybridizing assemblage. For example, *P. rivularis* is very similar in morphology to *P. hortulana* except for it shrubby habit. In another example, Waugh (1899, p. 232) wrote that *P. americana* grades into *P. nigra* to the north and into *P. mexicana* to the south in a continuous cline "with no break either in the geographical distribution or gradual morphological modification" from New Brunswick Canada to the border of Mexico. Not only is the interior of the eastern USA the center of genetic diversity for the North American plums, it is also where the majority of named taxa overlap (Fig. 6). The evidence put forward in this paragraph mirrors the "complex internal structure" of a syngameon as put forth by Grant (1981).

According to Grant's definition of a syngameon, nearer the periphery of the syngameon the taxa should behave like "well-isolated biological species." This is also true for the North American plums. There is little taxonomic difficulty in species determination nearer the periphery of the combined ranges of the North American plums. For example, in the southeastern extent of their range, in Florida, there are four species (*P. americana*, *P. angustifolia*, *P. umbellata*, and *P. geniculata*) and each are relatively easily separable. Also, throughout the northern midwest in Minnesota and Wisconsin there are only two species (*P. americana* and *P. nigra*) and these are also relatively distinct. In one more example, *P. maritima* var. *maritima*, *P. americana*, and *P. nigra* are the only species from the northeast and they are also relatively easily distinguished.

Considering their regional morphological differences and the lack of genetic differentiation within the group, a species concept that relies on total genetic isolation between species to delimit their distinctness is not appropriate in section *Prunocerasus*. This is not to say that all eastern species of North American plums should be lumped under one name. According to Grant (1981, p. 235), "the components of a syngameon ("species") are generally treated, and properly so, as species in formal systematics; the components are usually good taxonomic species." Based on total evidence, we believe that the majority of North American plum taxa are behaving as a syngameon comprised of several morphogeographic taxa that intergrade from one to another and radiate from a

center of diversity from east Texas through Oklahoma, Arkansas, northern Louisiana, southern Missouri to west Tennessee. However, two other genetic "hotspots" should be noted, namely the planhandle of central Florida and the northeast coast of the USA.

***Phylogeographic inference*—**The strength of phylogeography stems from an increased sampling strategy, as compared to traditional phylogenetic studies. This allows questions concerning evolutionary history to be addressed with increased accuracy and placed in a geographic context. As more studies between closely related taxa employ the method of "congeneric phylogeography" of Funk and Olmland (2003), we may begin to gain insight into the complex evolutionary histories of taxonomically problematic groups of plants, which are in many cases hybridizing swarms, or syngameons.

It has been suggested that the point of origin for *Prunus* is in Asia (Watkins, 1976; Bortiri et al., 2001). Furthermore, *P. subcordata*, the only western species, may represent the oldest lineage within section *Prunocerasus* because this species is sister to a hard polytomy containing the rest; this phylogenetic position for *P. subcordata* is now strongly supported by several lines of evidence including cpDNA (Bortiri et al., 2001; Shaw and Small, 2004) and nDNA (Bortiri et al., 2002). Based on the abovementioned evidence, we hypothesized that the North American plums immigrated to North America via the Bering land bridge (see Fig. 7 of Shaw and Small, submitted [Part 4]). Corroborating that hypothesis the nDNA *s6pdh*-based phylogeny presented here (Figs. 1 and 2) shows *P. subcordata* to represent the sister lineage to the rest of the North American plum species.

Excluding *P. subcordata*, we identified 58 unique haplotypes that we segregated into three primary haplotype groups (□, □, and □) centered on ancestral alleles revealed in the TCS analysis and the clades revealed in the phylolgenetic analyses. Both the cpDNA phylogeny (Shaw and Small, 2004; Shaw and Small, submitted [Part 4]) and the *s6pdh* phylogeny (Figs. 1 and 2) show little to no resolution among the three primary clades of North American plums. This tritomy, now corroborated by independent data sets (cpDNA and *s6pdh*), may be real if the three haplotypes are close in age, perhaps as the result of a radiation.

The majority of the taxa of the tritomy are poly- or paraphyletic, as evidenced by their being spread across the *s6pdh* phylogeny (Figs. 1 and 2). To gain insight into which taxa might be hybridizing and where areas of hybridization may occur we mapped the haplotypes found within each of the respective taxa (Figs. 4a-4k). Observation of the individual maps reveals no obvious pattern; however, a clearer pattern emerges if each haplotype is mapped regardless of the taxa in which it was observed (Figs. 5a-5d). The □ haplotype was found in *P. alleghaniensis* var. *alleghaniensis*, *P. americana* var. *americana*, *P. americana* var. *lanata*, *P. maritima* var. *maritima*, *P. mexicana*, and *P. nigra*. This haplotype was generally found throughout the interior USA and was rare or absent from the southeastern Coastal Plain and Great Lakes region (Fig. 5a). The □ haplotype was the most abundant haplotype found in this study. It was found in *P. alleghaniensis* var. *alleghaniensis*, *P. alleghaniensis* var. *davisii*, *P. americana* var. *americana*, *P. gracilis*, *P. hortulana*, *P. maritima* var. *maritima*, *P. maritima* var. *gravesii*, *P. munsoniana*, *P. rivularis*, *P. umbellata* var. *umbellata*, and *P. umbellata* var. *injucunda*. This haplotype was observed in central Texas, the panhandle of Florida, the

northeast coast, and the interior highlands (Fig. 5b). One allele was identified in a heterozygous individual from the Great Lakes region. The haplotype was found in *P. alleghaniensis* var. *davisii*, *P. angustifolia*, *P. geniculata*, *P. maritima* var. *maritima*, *P. munsoniana*, *P. nigra*, *P. texana*, and *P. umbellata* var. *umbellata*. It was found throughout the Atlantic and Gulf Coastal Plain, rare in the interior, and around the Great Lakes region. Other than the three primary haplotypes, seven putative recombinant haplotypes were observed (Fig. 5d). Although we cannot be certain that recombination was not PCR-mediated, the observation that they too show geographic (as well as taxonomic) patterning points to their being real. Of the seven recombinant alleles, two were found in *P. americana* var. *americana* (-), three in *P. munsoniana* (-), one in *P. hortulana* (29), and one in *P. umbellata* var. *umbellata* (29). Most of the putative recombinants were found near Tennessee.

Haplotypic distribution observed in this study corroborates the more detailed patterns observed in chloroplast distribution (compare Fig. 7 to Figs 5a-5c). While these data do not ultimately solve what has been called the "hardest puzzle in American pomology" (Bailey and Bailey, 1941), they certainly provide insight into the reticulating and complex evolutionary history of the North American plum syngameon, a complex of approximately 18 morphogeographically variable entities.

Conclusions—Three main conclusions can be drawn from this study. First, the results presented here coupled with the results of a previous study (Shaw and Small, submitted [Part 4]) show that most section *Prunocerasus* taxa are clearly not monophyletic with respect to either their chloroplasts or the nuclear encoded *s6pdh* gene. Although some nuclear and chloroplast haplotypes are more common in some taxa, few are exclusive. In fact, evidence from this and our previous studies (Shaw and Small, 2004; Shaw and Small, submitted [Part 4]) strongly supports the hypothesis of the North American plums behaving as a syngameon, or hybridizing species group. Although lineage sorting from a polymorphic ancestor cannot be ruled out as the cause for the observed patterns, we feel that a hypothesis of gene sharing through past and current hybridization is a more likely explanation because: (1) most of the species in the section are known to hybridize, (2) the haplotypes show geographic patterning, and (3) *P. subcordata* the oldest geographically separated lineage of the section does not possess these polymorphisms.

Second, *P. texana*, currently classified as a peach in subgenus *Amygdalus* (Wight, 1913), was positioned with strong support within the North American plums. This placement coupled with the results of our earlier studies (Shaw and Small, 2004; Shaw and Small, submitted [Part 4]) show that this taxon is likely a North American plum and not a peach. Because limited sampling here, *P. texana* warrants further study to unequivocally determine its correct taxonomic placement.

Lastly, had only single representatives of taxa been chosen as exemplars in this study, the complex genetic history of the North American plums would not have been illuminated. Earlier workers cautioned that in low-level plant investigations using only a single accession of each species could be misleading (Whittemore and Schaal, 1991; Matos and Schaal, 2000; Funk and Olmland, 2003), so they recommended more thorough sampling to circumvent potential problems. This study highlights that reality.

Phylogenetic estimation at this low level (population-like) presents a number of difficulties to traditional methods of phylogeny reconstruction because of invalid assumptions; for example, that ancestral haplotypes are extinct or genetic recombination has not occurred. In closely related groups of species, sampling multiple individuals from each of the representative taxa is the only means of completely assessing the phylogenetic or evolutionary history of a group.

LITERATURE CITED

ANDERSON, E. 1949. Introgressive hybridization. Wiley, New York, New York, USA.

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

BAILEY, L. H., AND E. Z. BAILEY. 1941. Hortus Second: a concise dictionary of gardening, general horticulture, and cultivated plants in North America. The Macmillan Company, New York, New York, USA.

Bains, H. S., R. Tao, S. L. Uratsu, and M. Dandekar. 1998. Genomic nucleotide sequence of NADP sorbitol-6-phosphate dehydrogenase (NADP-s6pdh) gene from apple (accession No. AF057134) (PGR98-193). *Plant Physiology* 118: 1533.

BALDWIN, B. G., M. J. SANDERSON, J. M. PORTER, M. F. WOJCIECHOWSKI, C. S. CAMPBELL, AND M. J. DONOGHUE. 1995. The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. *Annals of the Missouri Botanical Garden* 82: 247-277.

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

BLACKBURN, B. 1952. Trees and Shrubs in eastern North America: keys to the wild and cultivated woody plants growing in the temperate regions exclusive of conifers. Oxford University Press, New York, USA

BORTIRI, E., S. H. OH, F. Y. GAO, AND D. POTTER. 2002. The phylogenetic utility of nucleotide sequences of sorbitol 6-phosphate dehydrogenase in *Prunus* (Rosaceae). *American Journal of Botany* 89: 1697-1708.

BORTIRI, E., S. H. OH, J. G. JIANG, S. BAGGETT, A. GRANGER, C. WEEKS, M. BUCKINGHAM, D. POTTER, AND D. E. PARFITT. 2001. Phylogeny and systematics of *Prunus* (Rosaceae) as determined by sequence analysis of ITS and the chloroplast *trnL-trnF* spacer DNA. *Systematic Botany* 26: 797-807.

CHASE, M. W., ET AL. 1993. Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbcL*. *Annals of the Missouri Botanical Garden* 80: 528-580.

CHAT, J., B. JAUREGUI, R. J. PETIT, AND S. NADOT. 2004. Reticulate evolution in Kiwifruit (*Actinidia*, Actinidiaceae) identified by comparing their maternal and paternal phylogenies. *American Journal of Botany* 91: 736-747.

CLEMENT, M., D. POSADA, AND K. A. CRANDALL. 2000. TCS: a computer program to estimate gene genealogies. *Molecular Ecology* 9: 1657-1659.

CORRELL, D. S., AND M. C. JOHNSTON. 1970. Manual of the vascular plants of Texas. Texas Research Foundation, Renner, Texas, USA.

DODD, R. S. AND Z. AFZAL-RAFII. 2004. Selection and dispersal in a multispecies oak hybrid zone. *Evolution* 58: 261-269.

DUNCAN, W. H., AND M. B. DUNCAN. 1988. Trees of the southeastern United States. University of Georgia Press, Athens, Georgia, USA.

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

FERNALD, M. L. 1950. Gray's Manual of Botany, 8th ed. American Book Company, New York, New York, USA.

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 and Evolutionary Systematics* 34: 397-423.

GAUT, B.S. 1998. Molecular clocks and nucleotide substitution rates in higher plants. Evolutionary Biology 30:93-120

GLEASON, H. A. 1952. The New Britton and Brown Illustrated Flora of the United States and Adjacent Canada, volume 2. Lancaster Press, Inc., Lancaster, Pennsylvania, USA.

GLEASON, H. A., AND A. CRONQUIST. 1991. Manual of vascular plants of northeastern United States and adjacent Canada, 2nd ed. New York Botanical Garden, Bronx, New York, USA.

GODFREY, R. K. 1988. Trees, shrubs, and woody vines of northeastern Florida and adjacent Georgia and Alabama. The University of Georgia Press, Athens, Georgia, USA.

GRANT, V. 1981. *Plant speciation*, second ed. Columbia University Press, New York, New York, USA.

HARRISON, R. 1991. Molecular changes at speciation. *Annual Review of Ecology and Systematics* 22: 281-308.

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

JACKSON, H. D., D. A. STEANE, B. M. POTTS, AND R. E. VAILLANCOURT. 1999. Chloroplast DNA evidence for reticulate evolution in *Eucalyptus* (Myrtaceae). *Molecular Ecology* 8: 739-751.

JACKSON, S. T., AND D. K. SINGER. 1997. Climate change and the development of Coastal Plain disjunctions in the central Great Lakes region. *Rhodora* 99: 101-117.

JOHNSON, L. A. AND D. E. SOLTIS. 1994. *matK* DNA sequence and phylogenetic reconstruction in Saxifragaceae *s.s.* *Systematic Botany* 19: 143-156.

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.

KORNKVEN, A. B., L. E. WATSON, AND J. R. ESTES. 1999. Molecular phylogeny of *Artemesia* section *Tridentatae* (Asteraceae) based on chloroplast DNA restriction site variation. *Systematic Botany* 24: 69-84.

LEE, S., AND J. WEN. 2001. A phylogenetic analysis of *Prunus* and the Amygdaloideae (Rosaceae) using ITS sequences of nuclear ribosomal DNA. *American Journal of Botany* 88: 150-60.

LINDER, C. R., AND L. H. RIESEBERG. 2004. Reconstructing patterns of reticulate evolution in plants. *American Journal of Botany* 91: 1700-1708.

LITTLE, E.L. JR. 1971. Atlas of United States Trees. Vol. 1. Conifers and important hardwoods. US Dept. of Agric. Misc. Publ. 1146.
LITTLE, E.L. JR. 1976. Atlas of the United States Trees, Vol. 3. Minor western hardwoods. US Dept. Agric. Misc. Publ. 1314.
LITTLE, E.L. JR. 1977. Atlas of the United States Trees, Vol. 4, Minor eastern hardwoods. US Dept. Agric. Misc. Publ. 1342.
MARTINSEN, G. D., T. G. WHITHAM, R. J. TUREK, AND P. KIEM. 2001. Hybrid populations selectively filter gene introgression between species. *Evolution* 55: 1325-1335.
MASON-GAMER, R. J., C. F. WEIL, AND A. KELLOGG. 1998. Granule-bound starch synthase: structure, function, and phylogenetic utility. Molecular Biology and Evolution 15: 1658-1673.
MATOS, J. A., AND B. A. SCHAAL. 2000. Chloroplast evolution in the *Pinus montezumae* complex: a coalescent approach to hybridization. *Evolution* 54: 1218-1233.
MCDADE, L. A., 1992. Hybrids and phylogenetic systematics II. The impact of hybrids on cladistic analysis. *Evolution* 46: 1329-1346.
MCKINNON, G. E., D. A. STEANE, B. M. POTTS, AND R. E. VAILLANCOURT. 1999. Incongruence between chloroplast and species phylogenies in *Eucalyptus* subgenus *Monocalyptus* (Myrtaceae). *American Journal of Botany* 86: 1038-1046.
OLSEN, K. M. 2002. Population history of *Manihot esculenta* (Euphorbiaceae) inferred from nuclear DNA sequences. *Molecular Ecology* 11: 901-911.
POSADA, D., AND K. A. KRANDALL. 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14: 817-818.
RADFORD, A. E., H. E. AHLES, AND C. R. BELL. 1968. Manual of the vascular flora of the Carolinas. The University of North Carolina Press, Chapel Hill, North Carolina, USA.
REHDER, A. 1940. Manual of cultivated trees and shrubs hardy in North America, exclusive of the subtropical and warmer temperate regions, 2d, rev. and enl. edition. The Macmillan Company, New York, New York, USA.
RIESEBERG, L. H., O. RAYMOND, D. M. ROSENTHAL, Z. LAI, K. LIVINGSTONE, T. NAKAZATO, J. L. DURPHY, A. E. SCHWARZBACH, L. A. DONOVAN, AND C. LEXER. 2003. Major ecological transitions in annual sunflowers facilitated by hybridization. *Science* 301: 1211-1216.
ROELOFS, D., AND K. BACHMAN. 1997. Comparison of chloroplast and nuclear phylogeny in the autogamous annual *Microseris douglasii* (Asteraceae: Lactuceae). *Plant Systematics and Evolution* 204: 49-63.
ROHRER, J. R., R. AHMAD, S. M. SOUTHWICK, AND D. POTTER. 2004. Microsatellite analysis of relationships among North American plums (*Prunus* section *Prunocerasus*, Rosaceae). *Plant Systematics and Evolution* 244: 69-75.
SANG, T. 2002. Utility of low-copy nuclear gene sequences in plant phylogenetics. Critical Reviews in Biochemistry and Molecular Biology 37: 121-147.
SHAW, J. AND R. L. SMALL. submitted [Part 4]. Chloroplast DNA phylogeny and phylogeography of the North American plum (*Prunus* subgenus *Prunus* section *Prunocerasus*; Rosaceae)

SHAW, J., AND R. L. SMALL. 2004. Addressing the "hardest puzzle in American pomology:" Phylogeny of *Prunus* section *Prunocerasus* (Rosaceae) based on seven noncoding chloroplast DNA regions. *American Journal of Botany* 91: 985-996.

SMALL, J. K. 1933. Manual of the southeastern flora. Published by the author, The Science Press Printing Company, Lancaster, Pennsylvania, USA.

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 phylogenetic reconstruction in a recently diverged plant group. *American Journal of Botany* 85: 1301-1315.

SMALL, R. L., R. C. CRONN, AND J. F. WENDEL. 2004. Use of nuclear genes for phylogeny reconstruction in plants. *Australian Journal of Botany* 17: 145-170.

SMITH, E. B. 1994. Keys to the flora of Arkansas. University of Arkansas Press, Fayetteville, Arkansas, USA.

STEBBINS, G. L. 1959. The role of hybridization in evolution. *Proceedings of the American Philosophical Society* 103: 231-251.

STEYERMARK, J. A. 1963. Flora of Missouri. Iowa State University Press, Ames, Iowa, USA.

SWOFFORD, D. L. 2002. PAUP*: phylogenetic analysis using parsimony (*and other methods) version 4.0b10. Sinauer Associates, Sunderland, Massachusetts, USA.

TABERLET, P., L. GIELLY, G. PAUTOU, AND J. BOUVET. 1991. Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology* 17: 1105-1109.

THOMPSON, J. D., T. J. GIBSON, F. PLEWNIAK, F. JEANMOUGIN, AND D. G. HIGGINS. 2001. The Clustal_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* 25: 4876-4882.

USDA, NRCS. 2002. The PLANTS Database, Version 3.5 (http://plants.usda.gov). National Plant Data Center, Baton Rouge, Louisiana, USA.

VOLLMER, S. V., AND S. R. PALUMBI. 2004. Testing the utility of internally transcribed spacer sequences in coral phyhlogenetics. *Molecular Ecology* 13: 2763-2772.

WATKINS, R. 1976. Cherry, plum, peach, apricot, and almond. *In* N. W. Simmons [ed.], Evolution of crop plants, 242-247. Longman, London, England.

WAUGH, F. A. 1899. The pollination of plums. *Vermont Agricultural Experimentation Station, Annual Report* 12: 189-209.

WAUGH, F. A. 1901. Plums and plum culture. Orange Judd, New York, New York, USA.

Webb, K. L., and J. W. A. Burley. 1962. Sorbitol translocation in apple. *Science* 13: 766.

WELCH, M. E., AND L. H. RIESEBERG. 2002. Patterns of genetic variation suggest a single, ancient orogin for the diploid hybrid species *Helianthus paradoxus*. *Evolution* 56: 2126-2137.

WHITTEMORE, A. T., AND B. A. SCHAAL. 1991. Interspecific gene flow in sympatric oaks. *Proceedings of the National Academy of Sciences, USA* 88: 2540-2544.

WIGHT, W. F. 1913. North American species of the genus *Amygdalus*. Dudley Memorial Volume. Leland Stanford Junior University Publications, University Series, California, USA.

WIGHT, W. F. 1915. Native American species of *Prunus*. *Bulletin of the U.S. Department of Agriculture* 179: 1-75.

WOFFORD, B. E. AND E. W. CHESTER. 2002. Guide to the Trees, Shrubs, and Woody Vines of Tennessee. The University of Tennessee Press, Knoxville, Tennessee, USA.

WOLFE, K. H., W. -H. LI, AND P. M. SHARP. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proceedings of the National Academy of Sciences, USA* 84: 9054-9058.

WUNDERLIN, R. P. 1988. Guide to the vascular plants of Florida. University of Florida Press, Gainesville, Florida, USA.

APPENDIX; PART 5

Table 1. List of taxa used in this investigation and nDNA *s6pdh* and cpDNA haplotypes.

s6pdh	cpDNA		Species	Collector; Number; Year	Herb.	State and County
subg. *Prunus* sect. *Prunocerasus*						
/	A		*P. alleghaniensis* Porter (x americana?)	J. Shaw; JSh835; 2001	TENN	USA; WV; Greenbriar
/	C		*P. alleghaniensis* Porter (umbellata?)	J. Shaw; JSh1008; 2004	TENN	USA; NC; Rutherford
/	C3		*P. alleghaniensis* Porter	J. Shaw; JSh836; 2001	TENN	USA; MD; Allegany
/	C		*P. alleghaniensis* Porter	J. Shaw; JSh846; 2001	TENN	USA; PA; Centre
/	C	*	*P.all.* var. *davisii* (Wight.) Sarg.	G. Schmidt; s.n.; 2001	TENN	USA; MI; Oscoda
/	A		*P. americana* Marshall	J. Shaw; JSh1039; 2004	TENN	USA; KS; Saline
/	A		*P. americana* Marshall (x umbellata?)	J. Shaw; JSh911; 2002	TENN	USA; AL; Sumter
/	A1		*P. americana* Marshall	J. Shaw; JSh1038; 2004	TENN	USA; CO; Boulder
/	A2		*P. americana* Marshall	J. Shaw; JSh1042; 2004	TENN	USA; KS; Elk
/	Aa	*	*P. americana* Marshall	J. Shaw; JSh959; 2001	TENN	USA; GA; Catoosa
/	Aa		*P. americana* Marshall	J. Shaw; JSh967; 2002	TENN	USA; TN; Washington
/	C		*P. americana* Marshall	J. Shaw; JSh1043; 2004	TENN	USA; KS; Elk
/	Aa	*	*P. am.* Marshall var. *lanata* Sudw.	J. Beck; 4955; 2002	TENN	USA; TN; Marion
/	C		*P. angustifolia* Marshall	J. Shaw; JSh1018; 2004	TENN	USA; SC; Orangeburg
/	C		*P. angustifolia* Marshall	J. Shaw; JSh1030; 2004	TENN	USA; GA; Walton
/	C		*P. angustifolia* Marshall	J. Shaw; JSh1045; 2004	TENN	USA; KS; Montgomery
/	C		*P. angustifolia* Marshall	J. Shaw; JSh913; 2002	TENN	USA; MS; Clarke
/	C		*P. angustifolia* Marshall	J. Shaw; JSh947; 2002	TENN	USA; AR; Nevada
/	B	*	*P. geniculata* R.M. Harper	J. Shaw; JSh898; 2002	TENN	USA; FL; Highlands
/	C	*	*P. gracilis* Engelm. & A. Gray	J. Shaw; JSh936; 2002	TENN	USA; TX; Milam
/	C		*P. gracilis* Engelm. & A. Gray	J. Shaw; JSH946; 2002	TENN	USA; LA; Bossier
/ 29	Aa	*	*P. hortulana* L.H. Bailey	J. Shaw; JSh821; 2002	TENN	USA; TN; Clay
/	C		*P. hortulana* L.H. Bailey	J. Shaw; JSh1046; 2004	TENN	USA; MO; Oregon
/	C		*P. hortulana* L.H. Bailey	J. Shaw; JSh1050; 2004	TENN	USA; MO; Shannon
/	B	*	*P. maritima* Marshall	J. Shaw; JSh877; 2001	TENN	USA; MA; Plymouth
/	B		*P. maritima* Marshall	J. Shaw; JSh882; 2001	TENN	USA; NY; Suffolk
/	B		*P. maritima* Marshall	R. Uva; 98; 2001	TENN	USA; NY; Suffolk
/	C		*P. maritima* Marshall	R. Uva; 107; 2001	TENN	USA; ME; York
/	C		*P. maritima* Marshall	R. Uva; 111; 2001	TENN	USA; MA; Essex
/	C		*P. maritima* Marshall	R. Uva; 39; 2001	TENN	USA; DE; Sussex
/	C		*P. maritima* Marshall	R. Uva; 69; 2001	TENN	USA; NY; Suffolk
	B	*	*P. m.* var. *gravesii* (Small) Anderson	Conn Greenhouse; 2001	TENN	USA; CT; Tolland
/	A		*P. mexicana* S. Watson	J. Shaw; JSh1032B; 2004	TENN	USA; AR; Perry
/	A		*P. mexicana* S. Watson	J. Shaw; JSh1048; 2004	TENN	USA; MO; Oregon
/	Aa	*	*P. mexicana* S. Watson	J. Shaw; JSh919; 2002	TENN	USA; TX; Waller

Table 1. continued.

s6pdh	cpDNA		Species	Collector; Number; Year	Herb.	State and County
subg. *Prunus* sect. *Prunocerasus*						
/	Aa		*P. mexicana* S. Watson	J. Shaw; JSh928; 2002	TENN	USA; TX; Comal
/	Aa		*P. mexicana* S. Watson	J. Shaw; JSh933; 2002	TENN	USA; TX; Llano
/			*P. mexicana* S. Watson	J. Shaw; JSh942; 2002	TENN	USA; TX; Smith
/			*P. mexicana* S. Watson	J. Shaw; JSh943; 2002	TENN	USA; LA; Bossier
/	C	*	*P. munsoniana* Wight & Hedrick	J. Shaw; JSh810; 2001	TENN	USA; TN; Anderson
/	C		*P. munsoniana* Wight & Hedrick	J. Shaw; JSh960; 2002	TENN	USA; TN; Rhea
/	C		*P. munsoniana* Wight & Hedrick	J. Shaw; JSh961; 2002	TENN	USA; TN; Carter
/	C		*P. munsoniana* Wight & Hedrick	J. Shaw; JSh1049; 2004	TENN	USA; MO; Shannon
/	Aa		*P. nigra* Ait.	J. Shaw; JSh1004; 2002	TENN	USA; WI; Sauk
/	C1	*	*P. nigra* Ait.	J. Shaw; JSh979; 2002	TENN	USA; VT; Chittendon
/		*	*P. rivularis* Scheele	J. Shaw; JSh935; 2002	TENN	USA; TX; Williamson
/	C		*P. rivularis* Scheele	J. Shaw; JSh931; 2002	TENN	USA; TX; Mason
/			*P. rivularis* Scheele	J. Shaw; JSh929; 2002	TENN	USA; TX; Kerr
	S	*	*P. subcordata* Benth.	J.Syring; s.n.; 2001	TENN	USA; CA; Shasta
/	T	*	*P. texana* Dietr.	J. Shaw; JSh924; 2002	TENN	USA; TX; Goliad
/	B		*P. umbellata* Elliott	J. Shaw; JSh908; 2002	TENN	USA; FL; Marion
/	B1		*P. umbellata* Elliott	J. Beck; 5291; 2004	TENN	USA; FL: Alachua
/	C	*	*P. umbellata* Elliott (x angustifolia?)	J. Shaw; JSh774; 2001	TENN	USA; FL; Leon
/	C		*P. umbellata* Elliott	J. Shaw; JSh896; 2001	TENN	USA; FL; Highlands
/	C		*P. umbellata* Elliott	J. Shaw; JSh903; 2002	TENN	USA; FL; Lake
/ 29	C		*P. umbellata* Elliott	J. Shaw; JSh787; 2001	TENN	USA; MS; Perry
/	Aa	*	*P. u.* var. *injucunda* (Small) Sarg.	J. Shaw; JSh958; 2002	TENN	USA; GA; Cobb
OUTGROUP TAXA						
subg. *Prunus* sect. *Prunus*						
			P. salicina Lindl.	DPRU 791: TENN	TENN	
			P. spinosa L.	DPRU 2289.22: TENN	TENN	
subg. *Prunus* sect. *Armeniaca*						
			P. armeniaca L.	DPRU 1372.2: TENN	TENN	
subg. *Amygdalus* sect. *Amygdalus*						
			P. davidiana (Carriere) N.E.Brown	DPRU 581: TENN	TENN	

Fig. 1. Strict consensus of 653 equally parsimonious trees resulting from maximum parsimony analysis using the nDNA *s6pdh* intron one (955 aligned nucleotide positions, 85 are parsimony informative; tree length = 165 steps; consistency index [CI] = 0.72; retention index [RI] = 0.93). Bootstrap values greater than 50% are above the branches and primary haplotype groups (□, □, and □) are designated on the appropriate branches. Accessions included in the first analysis (not shown) but excluded from the second because they may be recombinant are underlined and drawn in with dashed lines. Haplotype 29 is illustrated in two accessions by bold face text. The letters A-D following the taxon name (e.g., *P. angustifolia* 913 B) designate the alleles found in that particular individual. A, B, and C marked along the branches just before they terminate refers to the primary chloroplast type found within that individual in our previous study (Shaw and Small, submitted). Classification into subgenera and sections follows Rehder (1940).

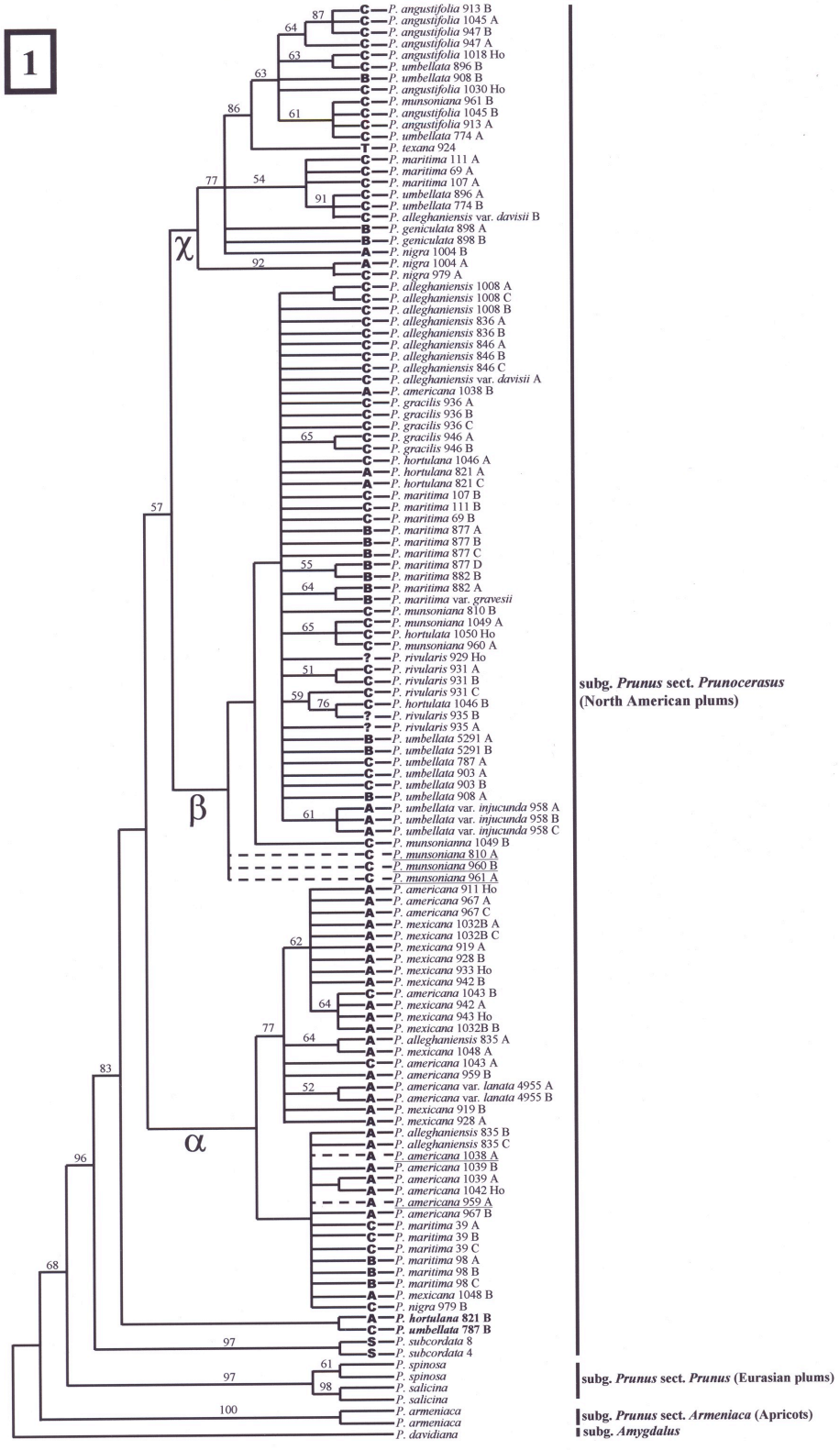

1
C — P. angustifolia 913 B
C — P. angustifolia 1045 A
C — P. angustifolia 947 B
C — P. angustifolia 947 A
C — P. angustifolia 1018 Ho
C — P. umbellata 896 B
B — P. umbellata 908 B
C — P. angustifolia 1030 Ho
C — P. munsoniana 961 B
C — P. angustifolia 1045 B
C — P. angustifolia 913 A
C — P. umbellata 774 A
T — P. texana 924
C — P. maritima 111 A
C — P. maritima 69 A
C — P. maritima 107 A
C — P. umbellata 896 A
C — P. umbellata 774 B
C — P. alleghaniensis var. davisii B
B — P. geniculata 898 A
B — P. geniculata 898 B
A — P. nigra 1004 B
A — P. nigra 1004 A
C — P. nigra 979 A
C — P. alleghaniensis 1008 A
C — P. alleghaniensis 1008 C
C — P. alleghaniensis 1008 B
C — P. alleghaniensis 836 A
C — P. alleghaniensis 836 B
C — P. alleghaniensis 846 A
C — P. alleghaniensis 846 B
C — P. alleghaniensis 846 C
C — P. alleghaniensis var. davisii A
A — P. americana 1038 B
C — P. gracilis 936 A
C — P. gracilis 936 B
C — P. gracilis 936 C
C — P. gracilis 946 A
C — P. gracilis 946 B
C — P. hortulana 1046 A
A — P. hortulana 821 A
A — P. hortulana 821 C
C — P. maritima 107 B
C — P. maritima 111 B
C — P. maritima 69 B
B — P. maritima 877 A
B — P. maritima 877 B
B — P. maritima 877 C
B — P. maritima 877 D
B — P. maritima 882 B
B — P. maritima 882 A
B — P. maritima var. gravesii
C — P. munsoniana 810 B
C — P. munsoniana 1049 A
C — P. hortulata 1050 Ho
C — P. munsoniana 960 A
? — P. rivularis 929 Ho
C — P. rivularis 931 A
C — P. rivularis 931 B
C — P. rivularis 931 C
C — P. hortulata 1046 B
? — P. rivularis 935 B
? — P. rivularis 935 A
B — P. umbellata 5291 A
B — P. umbellata 5291 B
C — P. umbellata 787 A
C — P. umbellata 903 A
C — P. umbellata 903 B
B — P. umbellata 908 A
A — P. umbellata var. injucunda 958 A
A — P. umbellata var. injucunda 958 B
A — P. umbellata var. injucunda 958 C
C — P. munsonianna 1049 B
C — P. munsoniana 810 A
C — P. munsoniana 960 B
C — P. munsoniana 961 A
A — P. americana 911 Ho
A — P. americana 967 A
A — P. americana 967 C
A — P. mexicana 1032B A
A — P. mexicana 1032B C
A — P. mexicana 919 A
A — P. mexicana 928 B
A — P. mexicana 933 Ho
A — P. mexicana 942 B
C — P. americana 1043 B
A — P. mexicana 942 A
A — P. mexicana 943 Ho
A — P. mexicana 1032B B
A — P. alleghaniensis 835 A
A — P. mexicana 1048 A
C — P. americana 1043 A
A — P. americana 959 B
A — P. americana var. lanata 4955 A
A — P. americana var. lanata 4955 B
A — P. mexicana 919 B
A — P. mexicana 928 A
A — P. alleghaniensis 835 B
A — P. alleghaniensis 835 C
A — P. americana 1038 A
A — P. americana 1039 B
A — P. americana 1039 A
A — P. americana 1042 Ho
A — P. americana 959 A
A — P. americana 967 B
C — P. maritima 39 A
C — P. maritima 39 B
C — P. maritima 39 C
B — P. maritima 98 A
B — P. maritima 98 B
B — P. maritima 98 C
A — P. mexicana 1048 B
C — P. nigra 979 B
A — P. hortulana 821 B
C — P. umbellata 787 B
S — P. subcordata 8
S — P. subcordata 4
P. spinosa
P. spinosa
P. salicina
P. salicina
P. armeniaca
P. armeniaca
P. davidiana
χ
β
α
87
64
63
63
86
61
77
54
91
92
57
65
55
64
65
51
59
76
61
62
64
77
64
52
83
96
68
97
61
97
98
100
subg. Prunus sect. Prunocerasus (North American plums)
subg. Prunus sect. Prunus (Eurasian plums)
subg. Prunus sect. Armeniaca (Apricots)
subg. Amygdalus

Fig. 2. Bayesian analysis tree using the nDNA *s6pdh* intron one (955 aligned nucleotide positions, 85 are parsimony informative). Numbers above the branches are posterior probabilities. Primary haplotype groups (□, □, and □) are designated on the appropriate branches. Accessions included in the first analysis (not shown) but excluded from the second because they may be recombinant are drawn in with dashed lines. Haplotype 29 is illustrated in two accessions by bold face text. The letters A-D following the taxon name (e.g., *P. angustifolia* 913 B) designate the alleles found in that particular individual. A, B, and C marked along the branches just before they terminate refers to the primary chloroplast type found within that individual in our previous study (Shaw and Small, submitted). Classification into subgenera and sections follows Rehder (1940).

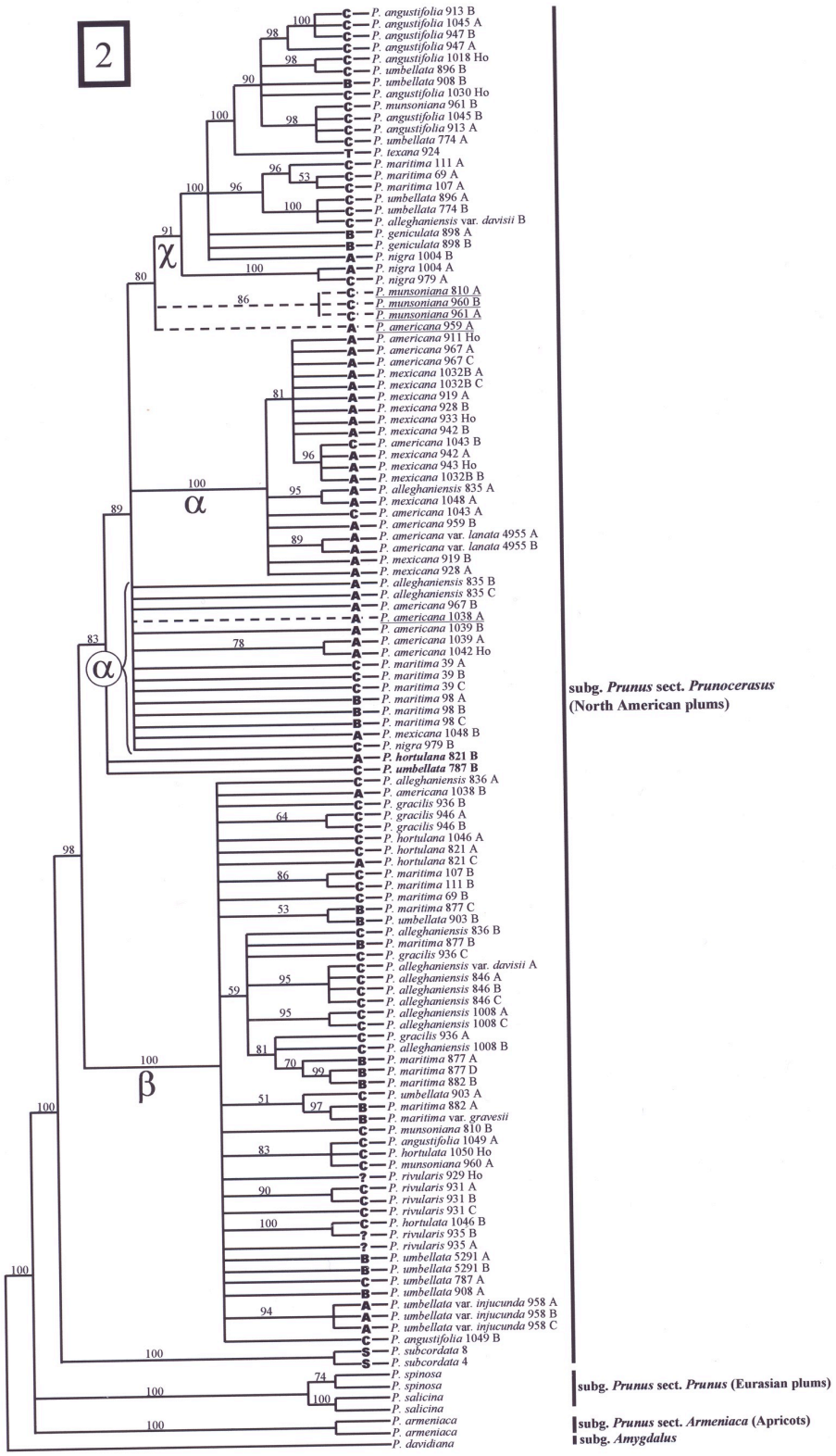
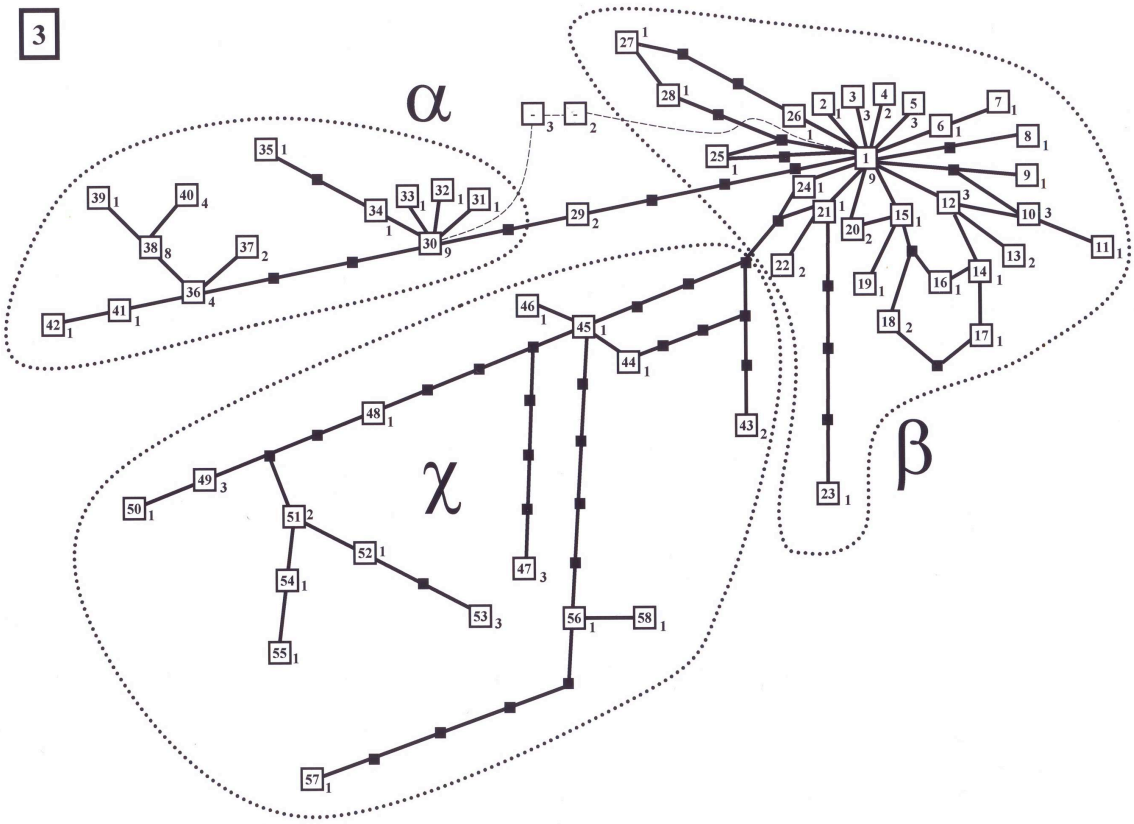

2
P. angustifolia 913 B
P. angustifolia 1045 A
P. angustifolia 947 B
P. angustifolia 947 A
P. angustifolia 1018 Ho
P. umbellata 896 B
P. umbellata 908 B
P. angustifolia 1030 Ho
P. munsoniana 961 B
P. angustifolia 1045 B
P. angustifolia 913 A
P. umbellata 774 A
P. texana 924
P. maritima 111 A
P. maritima 69 A
P. maritima 107 A
P. umbellata 896 A
P. umbellata 774 B
P. alleghaniensis var. davisii B
P. geniculata 898 A
P. geniculata 898 B
P. nigra 1004 B
P. nigra 1004 A
P. nigra 979 A
P. munsoniana 810 A
P. munsoniana 960 B
P. munsoniana 961 A
P. americana 959 A
P. americana 911 Ho
P. americana 967 A
P. americana 967 C
P. mexicana 1032B A
P. mexicana 1032B C
P. mexicana 919 A
P. mexicana 928 B
P. mexicana 933 Ho
P. mexicana 942 B
P. americana 1043 B
P. mexicana 942 A
P. mexicana 943 Ho
P. mexicana 1032B B
P. alleghaniensis 835 A
P. mexicana 1048 A
P. americana 1043 A
P. americana 959 B
P. americana var. lanata 4955 A
P. americana var. lanata 4955 B
P. mexicana 919 B
P. mexicana 928 A
P. alleghaniensis 835 B
P. alleghaniensis 835 C
P. americana 967 B
P. americana 1038 A
P. americana 1039 B
P. americana 1039 A
P. americana 1042 Ho
P. maritima 39 A
P. maritima 39 B
P. maritima 39 C
P. maritima 98 A
P. maritima 98 B
P. maritima 98 C
P. mexicana 1048 B
P. nigra 979 B
P. hortulana 821 B
P. umbellata 787 B
P. alleghaniensis 836 A
P. americana 1038 B
P. gracilis 936 B
P. gracilis 946 A
P. gracilis 946 B
P. hortulana 1046 A
P. hortulana 821 A
P. hortulana 821 C
P. maritima 107 B
P. maritima 111 B
P. maritima 69 B
P. maritima 877 C
P. umbellata 903 B
P. alleghaniensis 836 B
P. maritima 877 B
P. gracilis 936 C
P. alleghaniensis var. davisii A
P. alleghaniensis 846 A
P. alleghaniensis 846 B
P. alleghaniensis 846 C
P. alleghaniensis 1008 A
P. alleghaniensis 1008 C
P. gracilis 936 A
P. alleghaniensis 1008 B
P. maritima 877 A
P. maritima 877 D
P. maritima 882 B
P. umbellata 903 A
P. maritima 882 A
P. maritima var. gravesii
P. munsoniana 810 B
P. angustifolia 1049 A
P. hortulata 1050 Ho
P. munsoniana 960 A
P. rivularis 929 Ho
P. rivularis 931 A
P. rivularis 931 B
P. rivularis 931 C
P. hortulata 1046 B
P. rivularis 935 B
P. rivularis 935 A
P. umbellata 5291 A
P. umbellata 5291 B
P. umbellata 787 A
P. umbellata 908 A
P. umbellata var. injucunda 958 A
P. umbellata var. injucunda 958 B
P. umbellata var. injucunda 958 C
P. angustifolia 1049 B
P. subcordata 8
P. subcordata 4
P. spinosa
P. spinosa
P. salicina
P. salicina
P. armeniaca
P. armeniaca
P. davidiana
χ
α
α
β
subg. Prunus sect. Prunocerasus (North American plums)
subg. Prunus sect. Prunus (Eurasian plums)
subg. Prunus sect. Armeniaca (Apricots)
subg. Amygdalus

Fig. 3. TCS gene genealogy of *Prunus* subgenus *Prunus* section *Prunocerasus* from the *s6pdh* intron one. Haplotypes observed in this study are represented by numbered squares. The number of times that the respective haplotype was observed is indicated by a number to the bottom right of each of the squares. Lines connecting the haplotypes represent a single mutation (nucleotide substitution or indel) with solid squares representing inferred mutational steps not observed in this study. The haplotypes were grouped into three primary haplotype groups encircled by dashed lines and labeled: □, □, and □. Accessions included in the first analysis (not shown) but excluded from the second because they may be recombinant are drawn in with dashed lines. *Prunus texana* is allele number 48.


3
α
β
χ

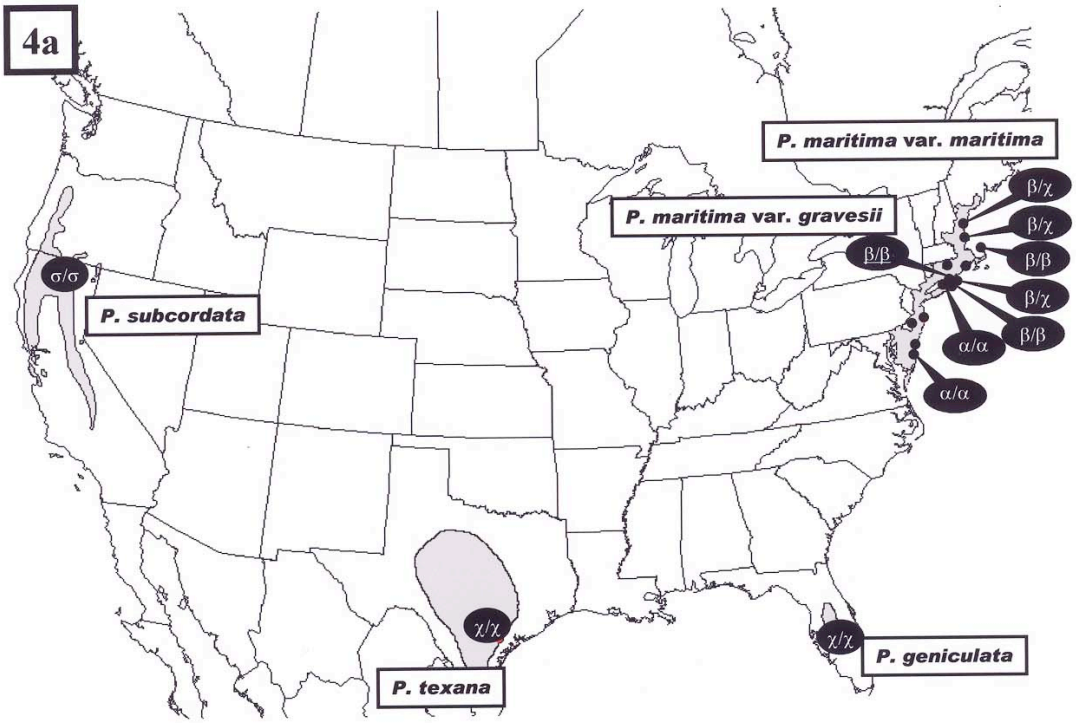


Fig. 4. (a) to (k). Geographical locations of haplotypes observed throughout the ranges of the 18 taxa of this study. Individual taxon ranges are shaded. Map 4a illustrates four species with relatively narrow and separated ranges and maps 4b-4k contain one species each. Where appropriate, varieties are denoted with an underline.

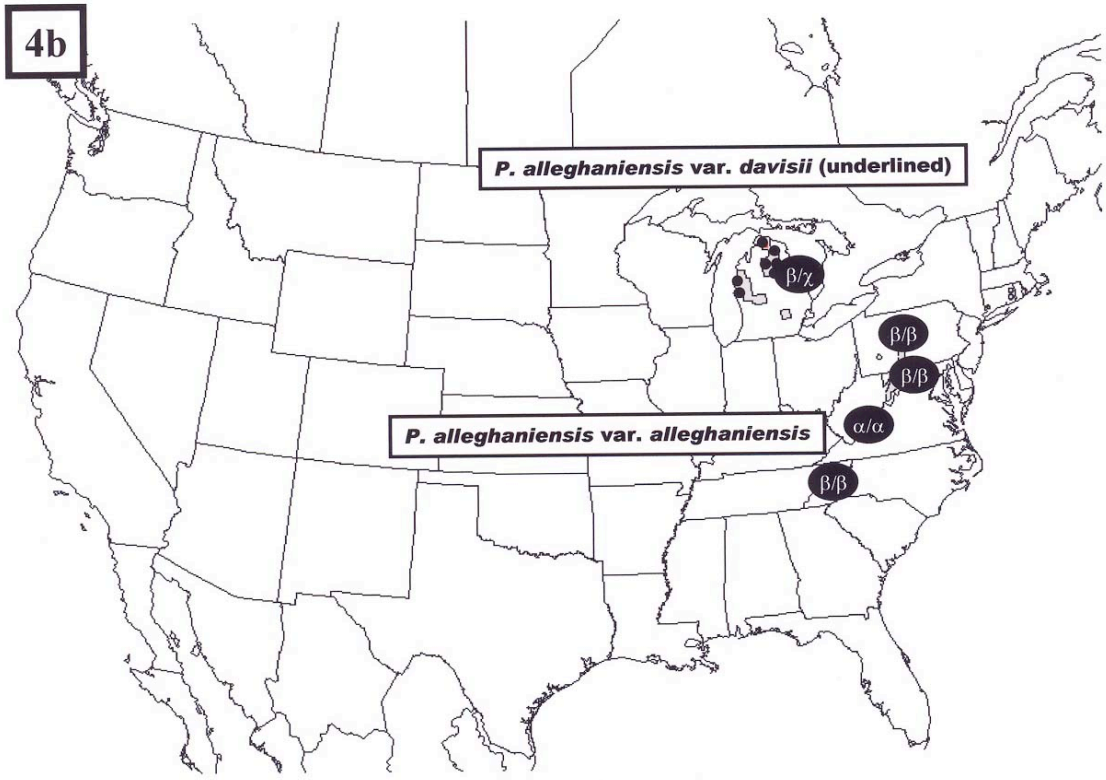


Fig. 4. continued.

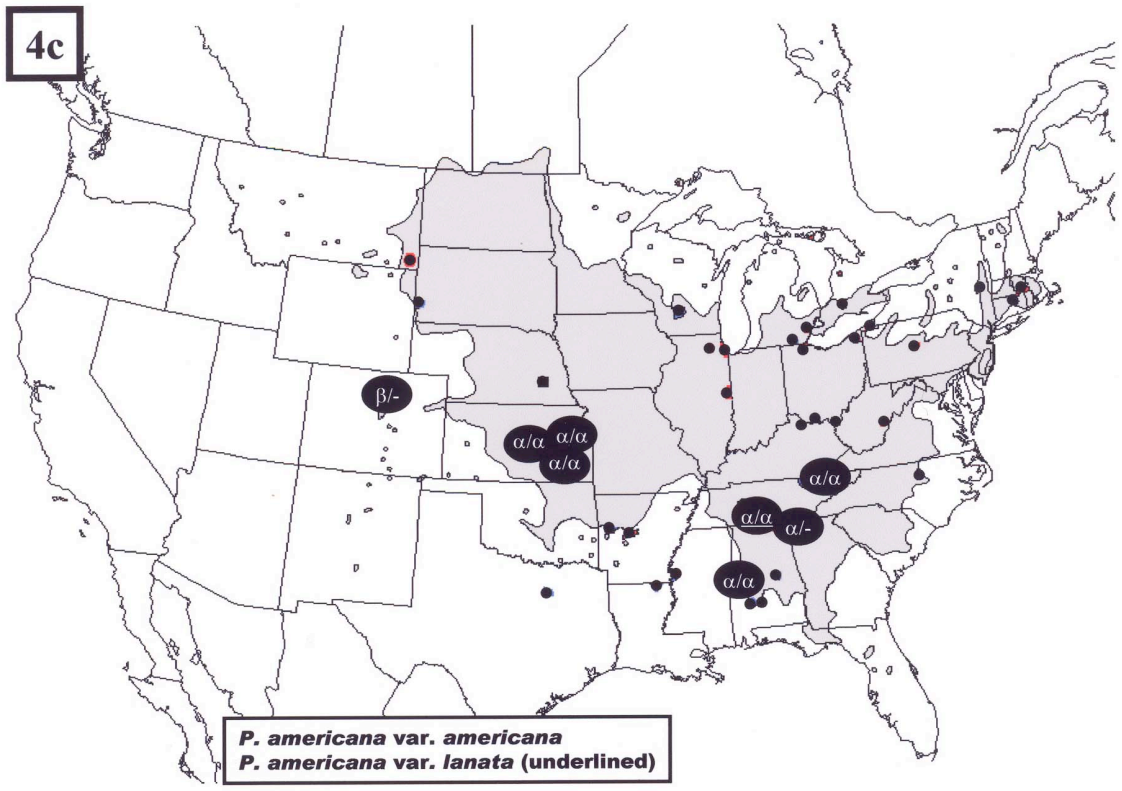


Fig. 4. continued.

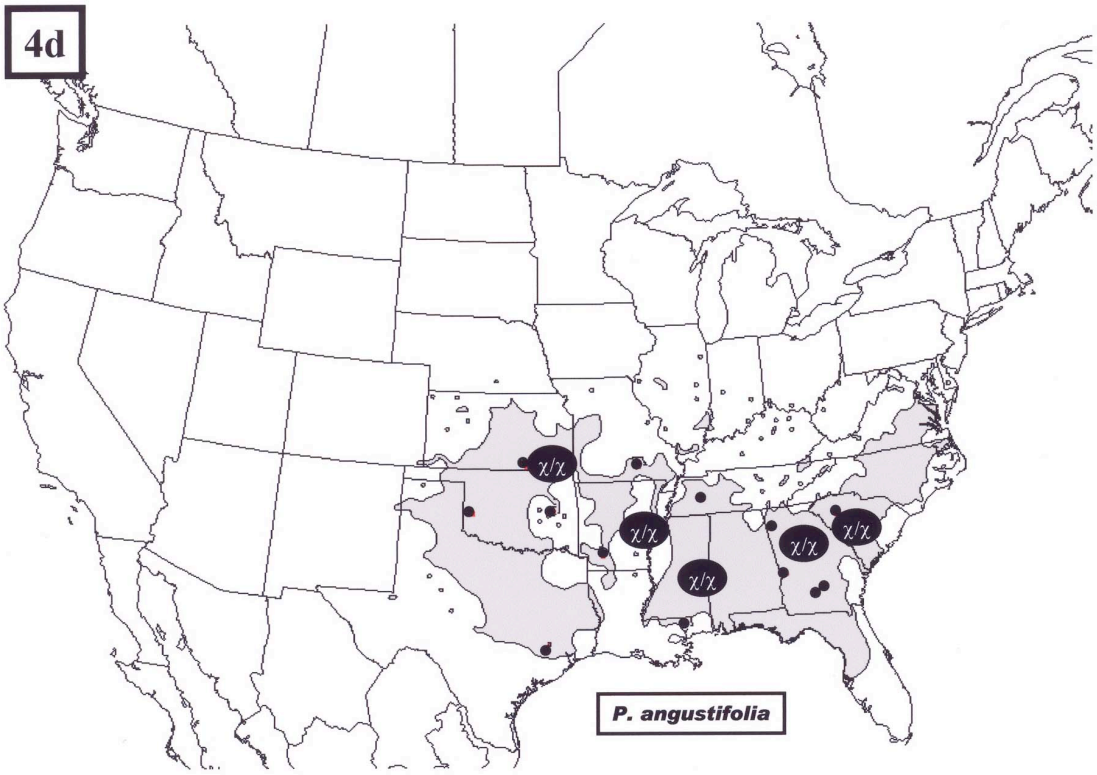


Fig. 4. continued.

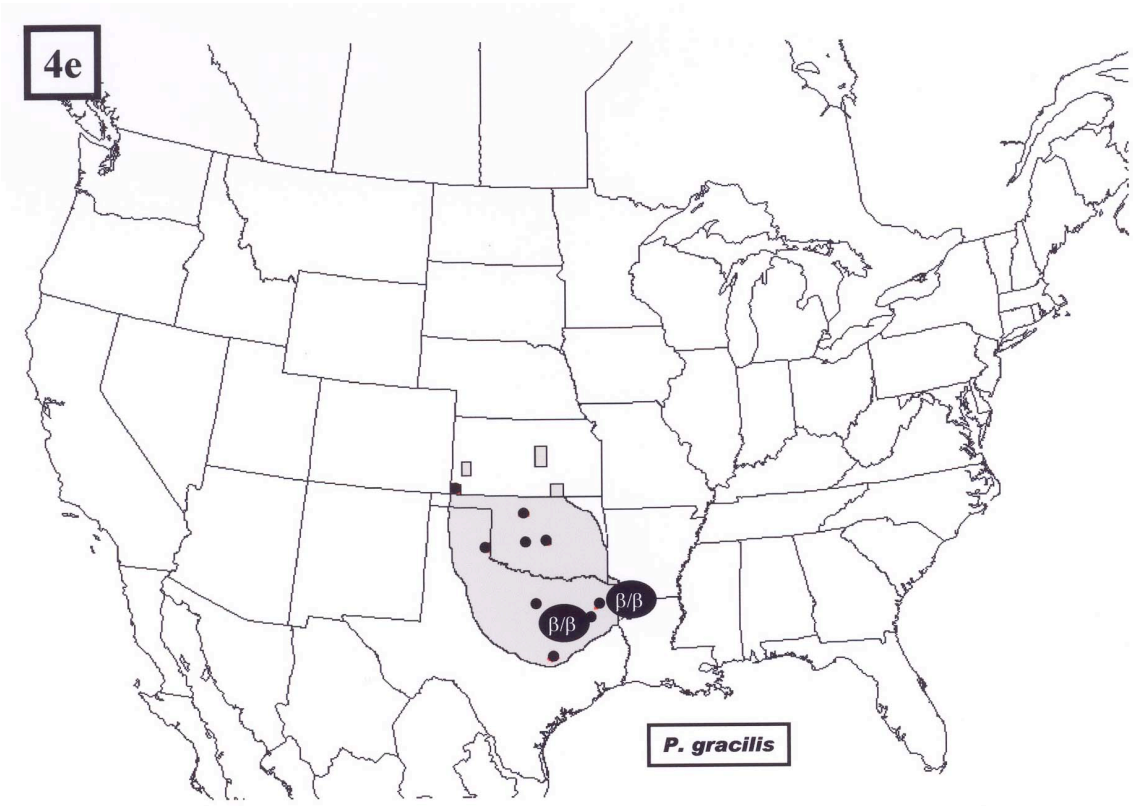


Fig. 4. continued.

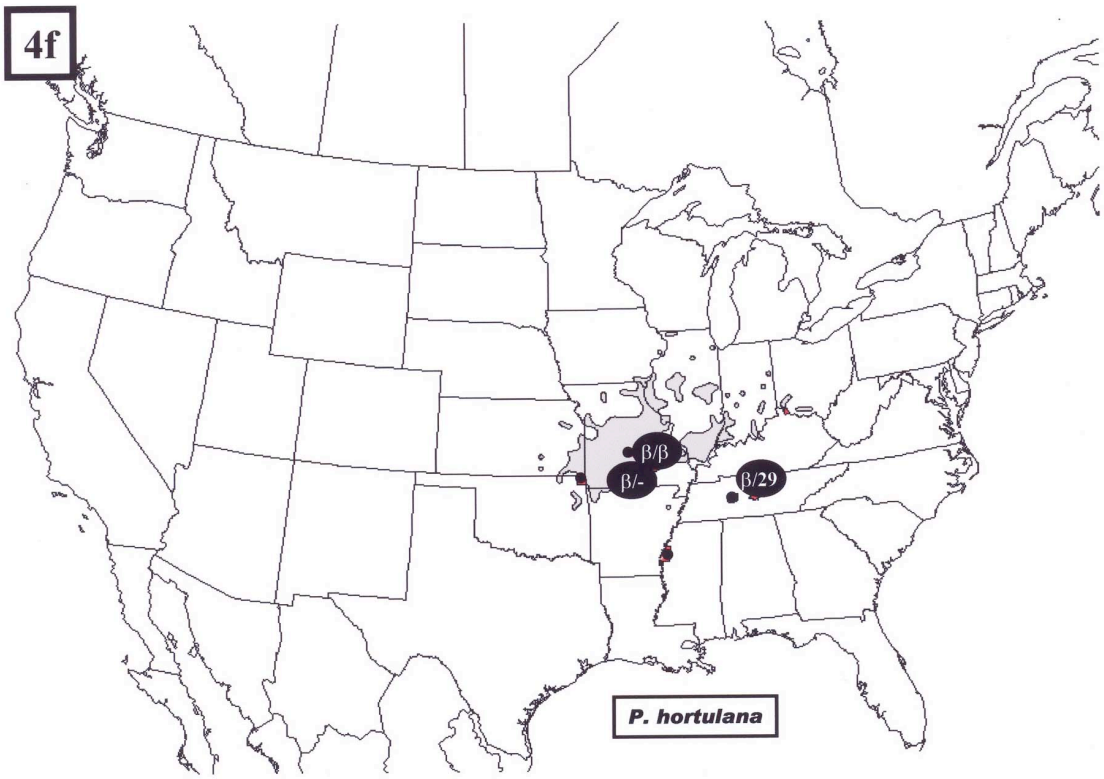


Fig. 4. continued.

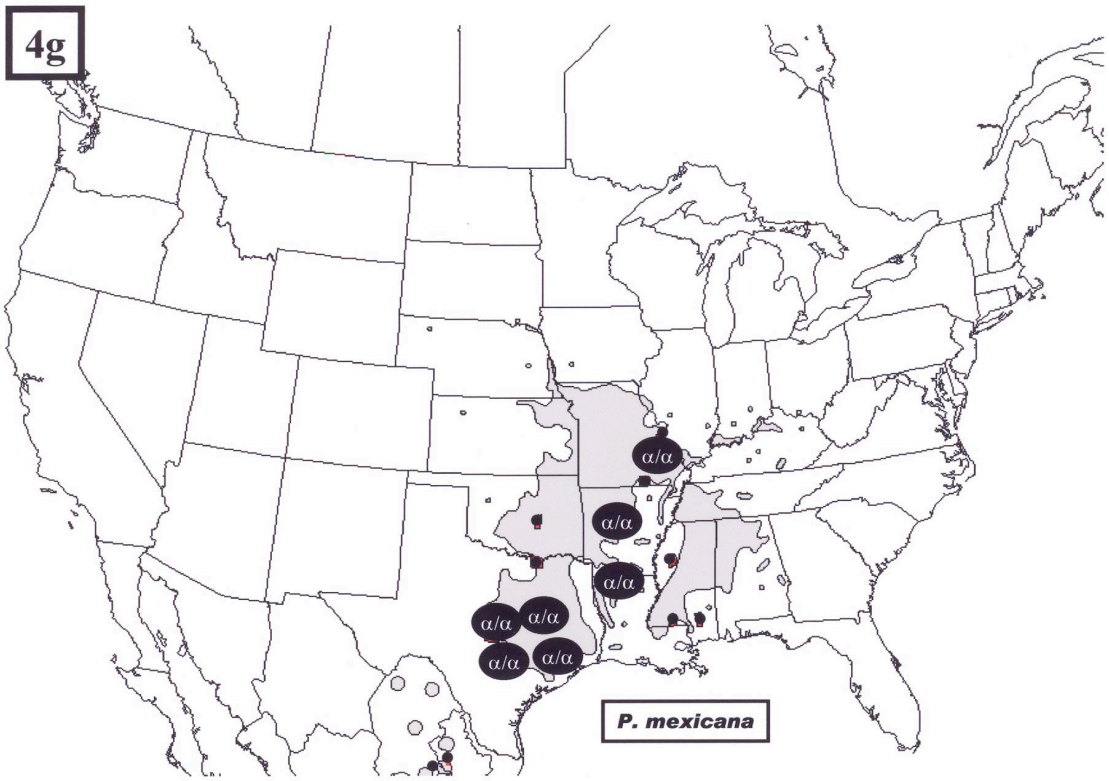


Fig. 4. continued.

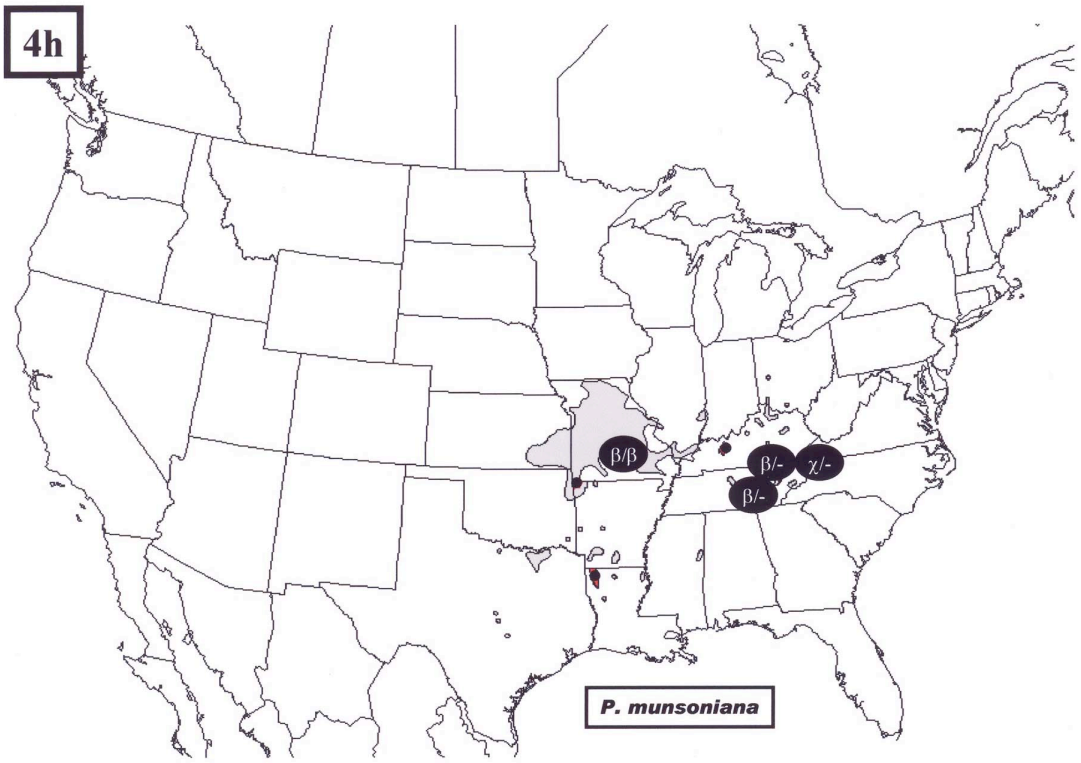


Fig. 4. continued.

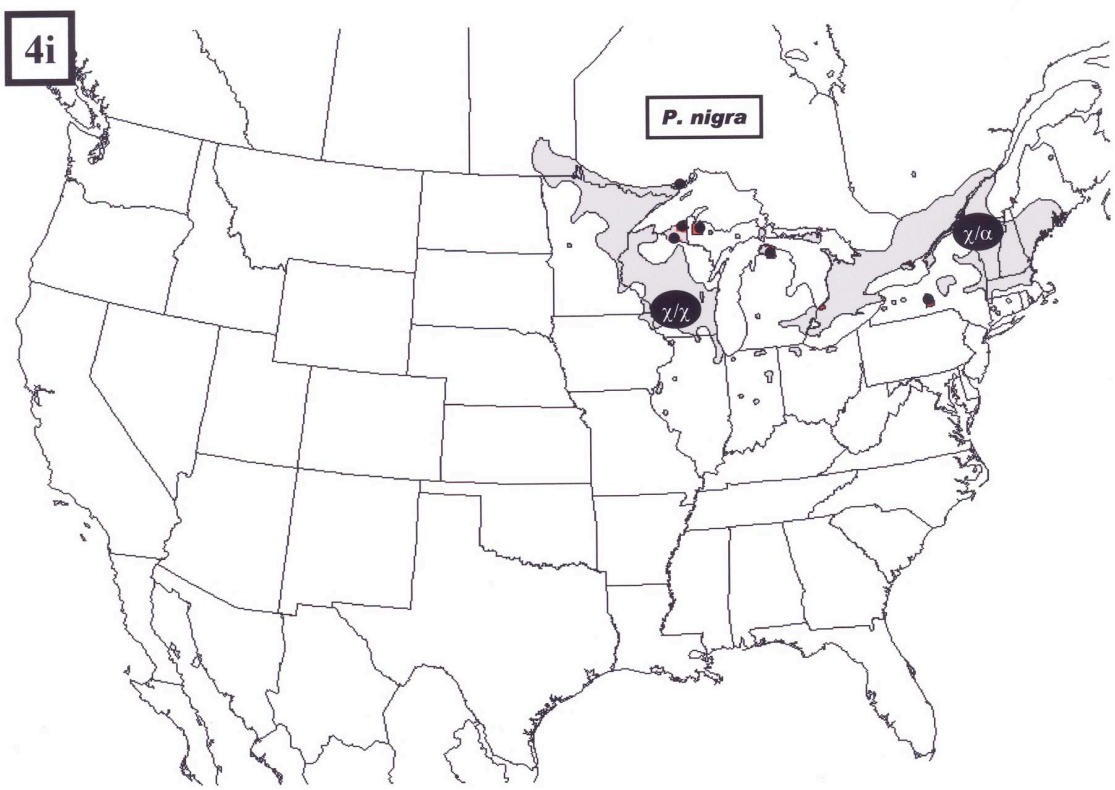


Fig. 4. continued.

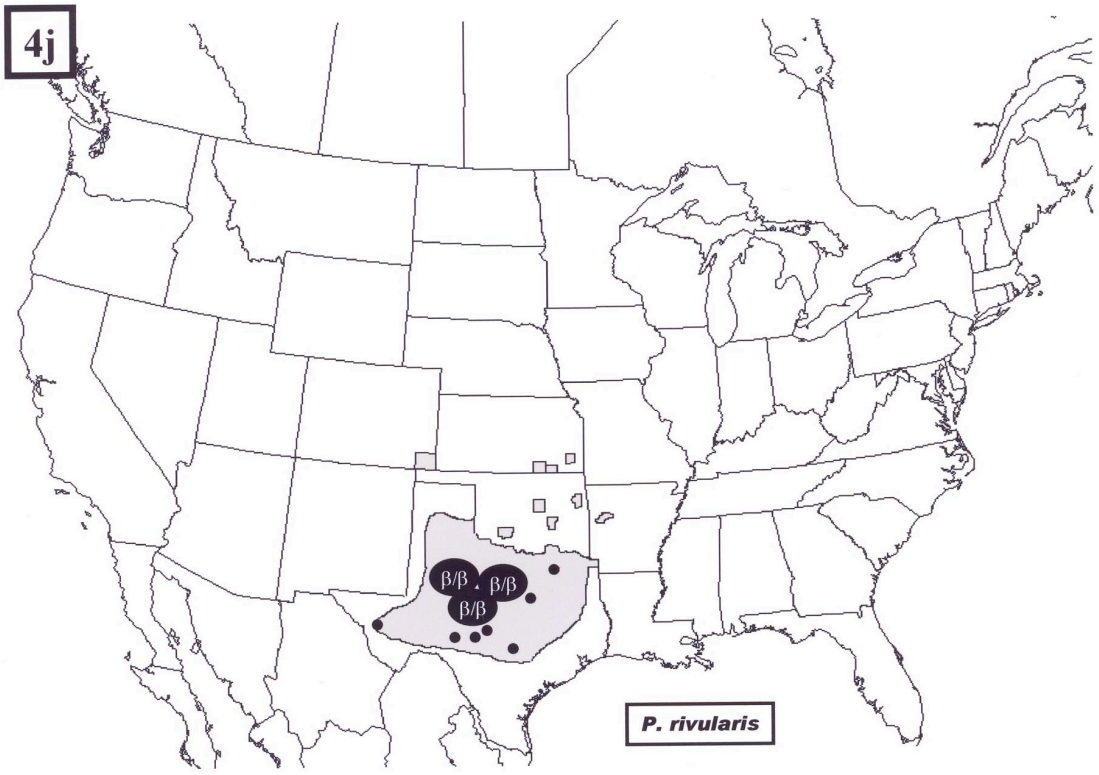


Fig. 4. continued.

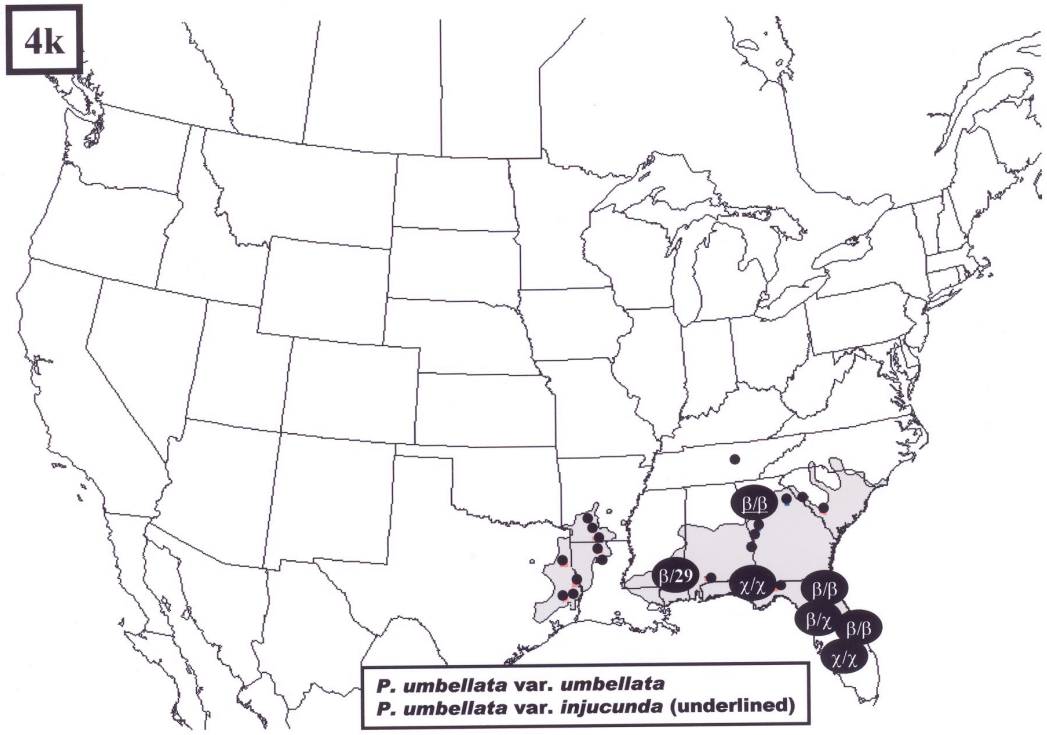


Fig. 4. continued.

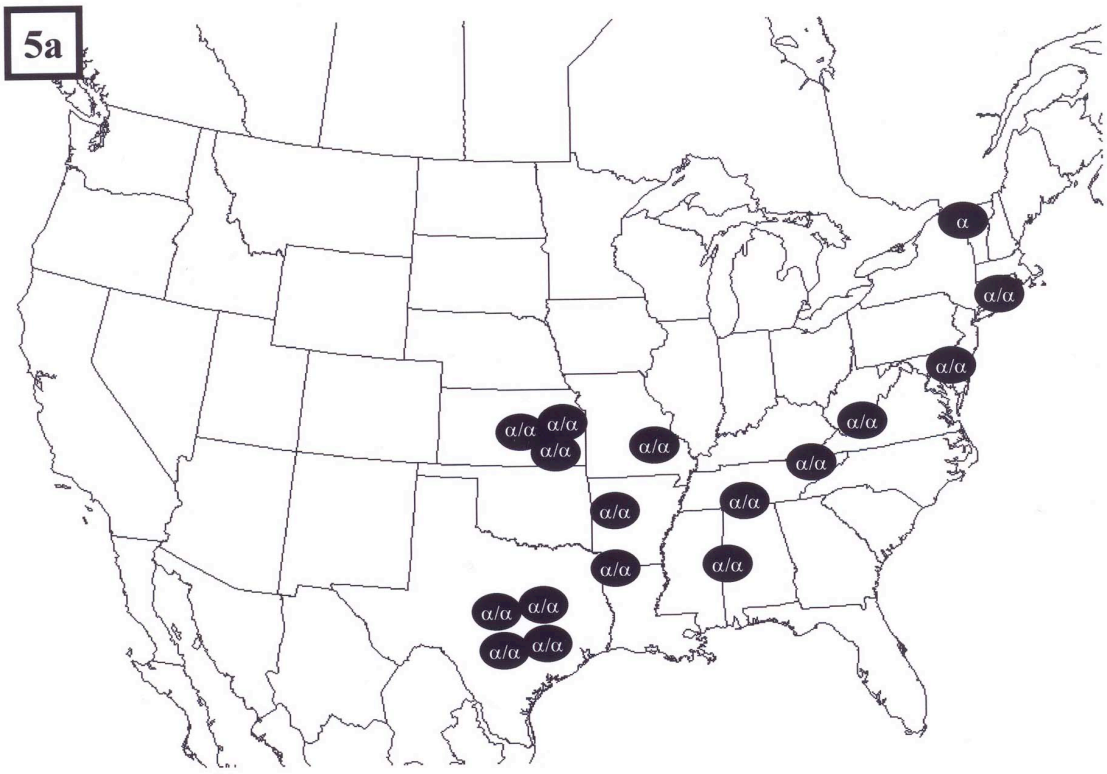


Fig. 5. (a) to (d). The distribution of the three primary (, , and) and two possible recombinant (- and 29) haplotypes observed in this study.

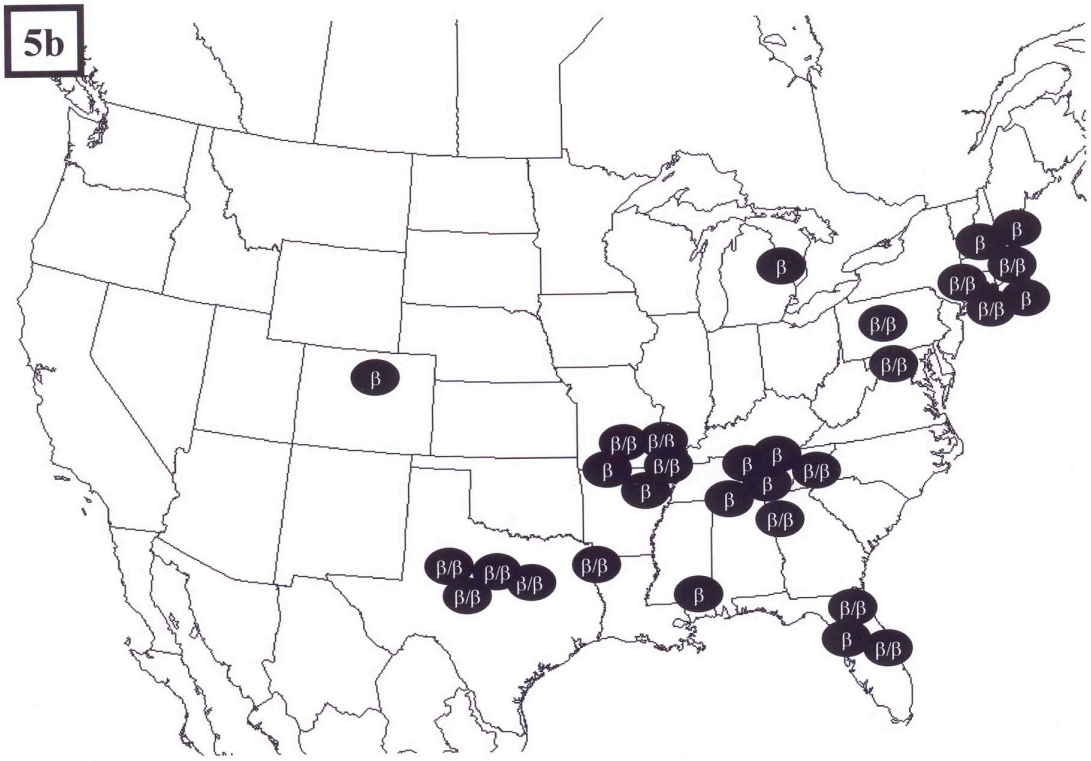


Fig. 5. continued.

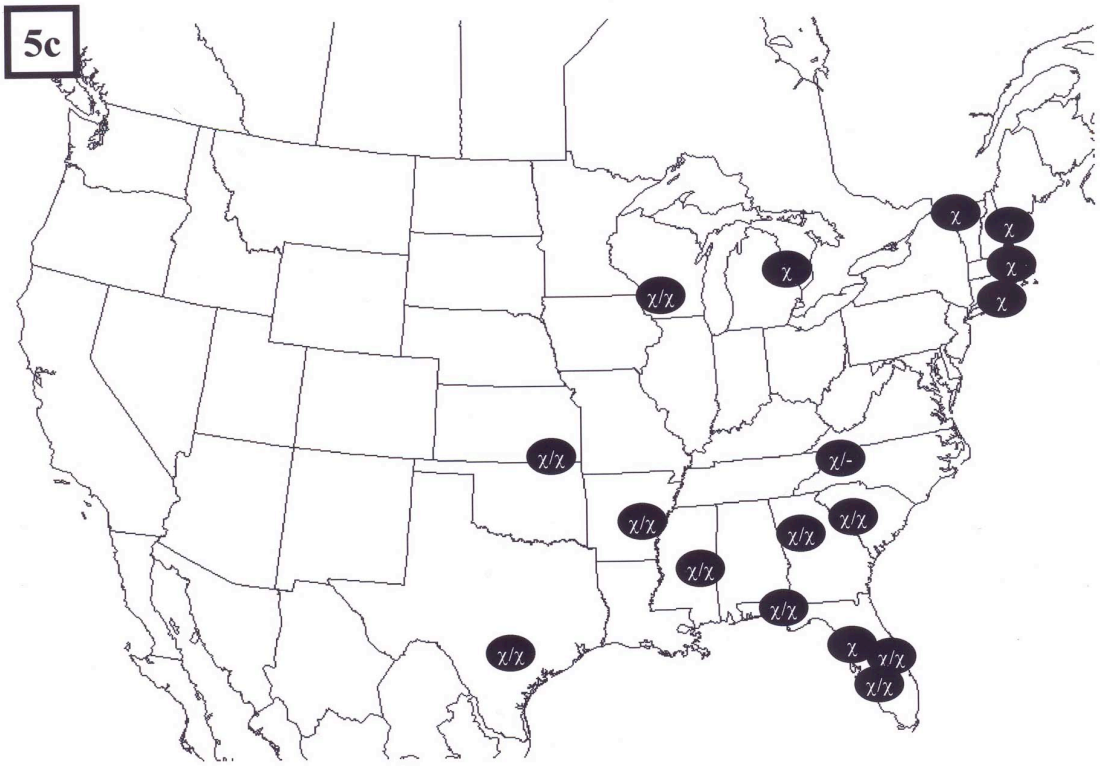


Fig. 5. continued.

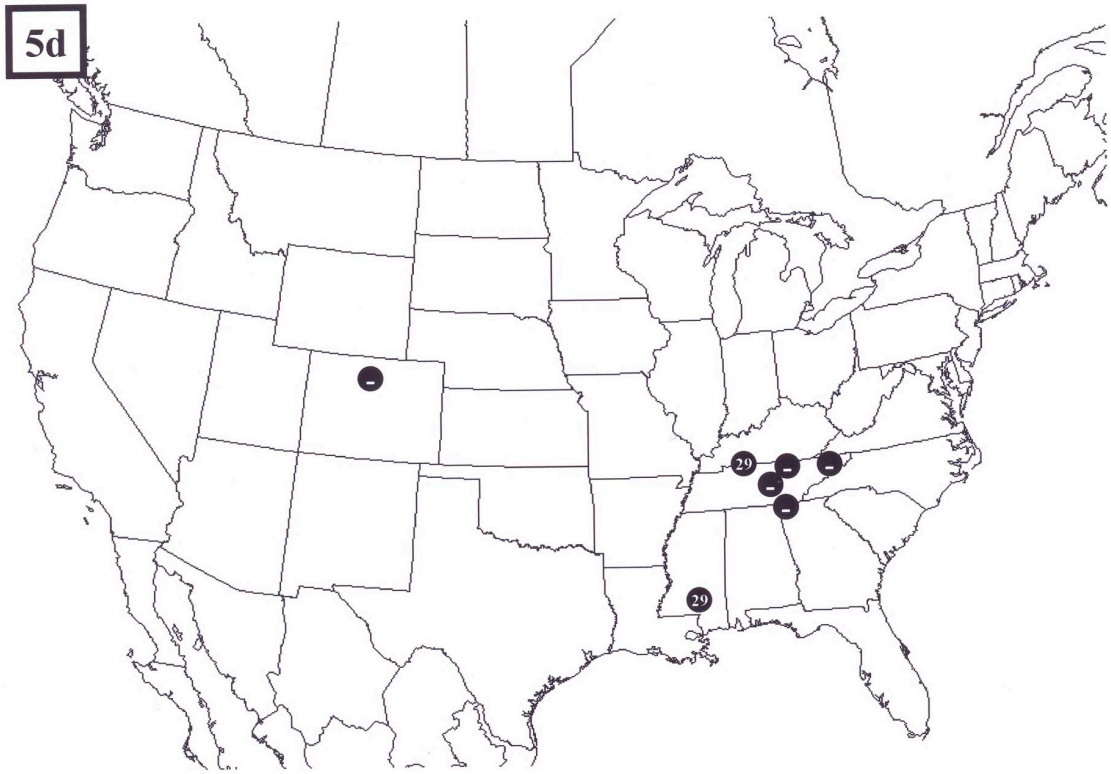


Fig. 5. continued.

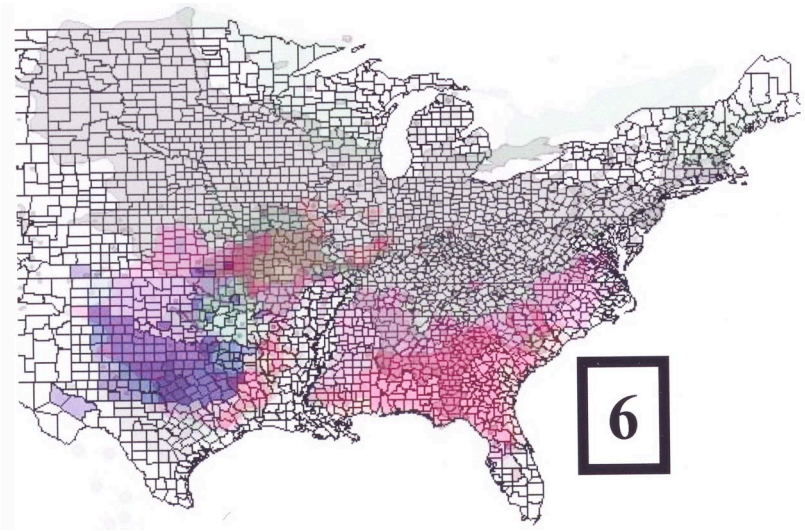


Fig. 6. Superposition of the ranges of all North American plum taxa showing that the ranges of the majority of the named taxa overlap in a region from northeast Texas through Oklahoma and Arkansas to Missouri and west Tennessee.

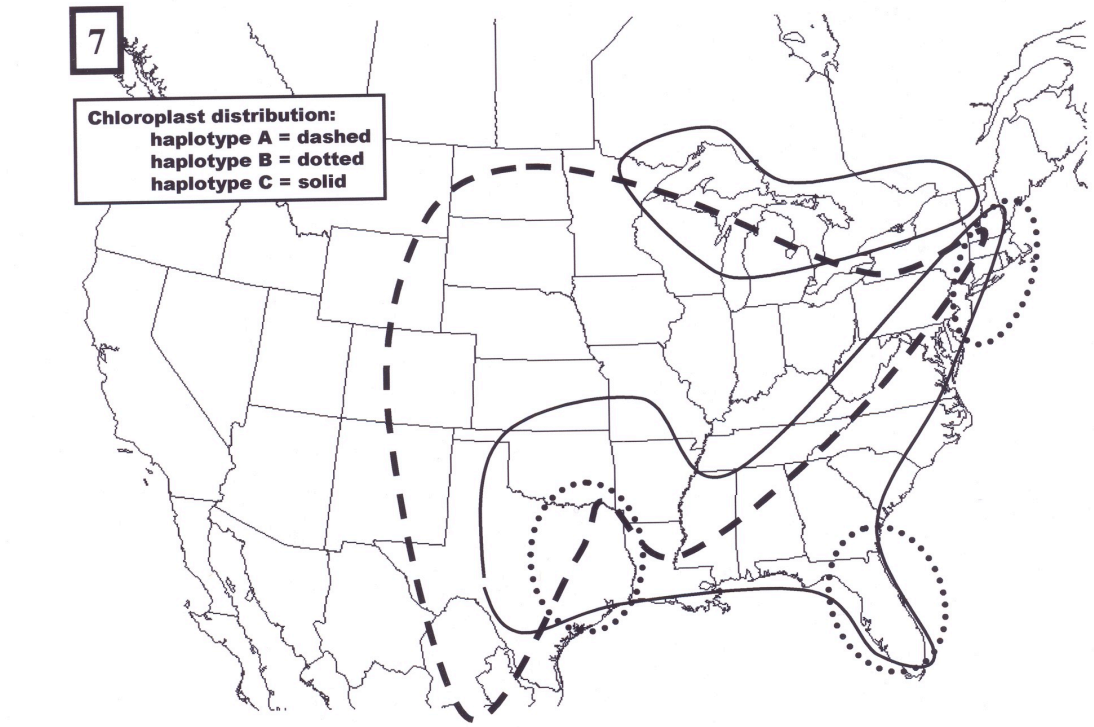


Fig. 7. Distribution of the primary chloroplast haplotypes. Recreation of results from Shaw and Small (submitted [Part 4]).

Part 6
General Conclusions and Future Directions

The initial goal of this dissertation was to evaluate the taxonomic / phylogenetic relationships of a group of taxa with long-standing taxonomic problems, the North American plums (*Prunus* subgenus *Prunus* section *Prunocerasus*; Rosaceae), using tools from the field of molecular systematics. The importance of this course of study includes but is not limited to: (1) better understanding an infamously difficult group of taxa for academic reasons, (2) accumulating data to better understand North American plum taxa so that conservation efforts on taxa of conservation concern can be better focused (e.g., *P. geniculata* is Federally Endangered), (3) provide an evolutionary history for a group of taxa of economic importance (many North American plum species have been and are potential food sources), (4) testing the limits of tools currently used in the field of plant molecular phylogenetics as the North American plum syngameon posits a problem at the boundary between molecular phylogenetics and population genetics, and (5) because the North American plums range throughout the eastern USA and *P. subcordata* is disjunct to the western USA, this study adds to a body of knowledge surrounding longstanding biogeographic questions of North America, namely the Atlantic and Gulf Coastal Plain disjunction to the Great Lakes region and the eastern North America-western North America disjunction.

This research dissertation is the accumulation of information from four original research papers and a National Science Foundation Dissertation Improvement Grant. All papers have either been published, submitted, or will be submitted to the *American Journal of Botany*, Parts 3, 4, and 5 are original papers concerning the North American plums.

The first published paper (Part 2: The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis) addresses the potential phylogenetic utility of 21 noncoding chloroplast DNA regions that had been used in the field of plant molecular systematics. Before the start of this investigation of the North American plums little was known about the relative rates of evolution among different noncoding chloroplast DNA — a primary source of data for plant molecular systematic studies. A few key papers had provided the molecular systematics community with universal primer pairs for a few noncoding regions that have dominated the field. While these few regions have provided adequate information to resolve species relationships in some taxa, they often provide little resolution at low taxonomic levels. The focus of this part of this dissertation was to determine whether there is any predictable rate heterogeneity among 21 noncoding cpDNA regions identified as phylogenetically useful at low levels. The results of this study clearly show that predictable rate heterogeneity exists among noncoding cpDNA regions. An additional important finding of this work is that a preliminary three-species survey can be used to determine the relative utility of a given region before implementing a full-scale sequencing project.

The second published paper (Part 3: Addressing the "hardest puzzle in American pomology:" Phylogeny of *Prunus* section *Prunocerasus*; Rosaceae based on seven noncoding chloroplast DNA regions) was focused on testing the monophyly of section *Prunocerasus*. To test the hypothesis of monophyly for the North American plums we sequenced the cpDNA *trnG* and *rpL16* introns and the *trnH-psbA* and *trnS-trnG* intergenic spacers for at least two representatives of each of the five subgenera in *Prunus*. Additionally we sampled heavily among *Prunus* subgenus *Prunus* sections *Prunus* and

Armeniaca, and *Prunus* subgenus *Amygdalus* because these groups are putatively most closely related to *Prunocerasus*. Once monophyly of section *Prunocerasus* was shown we added the sequences of *trnL* and *rpS16* introns and the *trnL-trnF* spacer in an attempt to increase resolution within the section. The species of section *Prunocerasus* showed an initial split with *P. subcordata*, the only species from western North America, sister to the rest of the group. The remaining species fell into three primary clades. Within each of the three primary clades there was little phylogenetic resolution. Three main conclusions can be drawn from this paper. First, phylogenetic analysis of cpDNA sequences demonstrated the monophyleticism of *Prunus* subgenus *Prunus* section *Prunocerasus*. Second, this data set provides a maternally-inherited framework on which we can further develop a hypothesis of relationships for the species within *Prunocerasus.* Third, this paper highlights the continued need, especially at low taxonomic levels, to assay the relative utility of multiple cpDNA regions before investigation instead of just using commonly employed tools.

The third paper (Part 4: Chloroplast DNA phylogeny and phylogeography of the North American plums [*Prunus* subgenus *Prunus* section *Prunocerasus*; Rosaceae]) employs a cpDNA-based phylogeographic approach to looking beyond the taxa as terminal evolutionary points and focuses more on the maternally inherited evolutionary history of the group in a geographic context. Here we determined if the three primary cpDNA haplotypes found in Part 4 are confined to the taxa in which they were initially observed. Four main conclusions can be drawn from this study. The first concerns the para- and polyphyly of chloro-haplotypes observed in 12 of 17 of the North American plum taxa. Second, four accessions of the Texas Peachbush, *P. texana*, all contain the same haplotype (T), which is strongly supported as being embedded within section *Prunocerasus* species. Third, the data presented in this study are consistent with a biogeographic hypothesis that the ancestors of section *Prunocerasus* entered North America via the Bering land bridge, migrated south and east, and subsequently radiated from Texas throughout the eastern North America (although the data do not refute an eastern migration route). Last, and perhaps the most important general conclusion, this study highlights the need for more thorough sampling in phylogenetic investigations because most phylogenetic studies include one or a few individual(s) per taxon; therefore, in closely related groups, the true phylogeny might not be reflected but rather the obscuring effects of hybridization and introgression or lineage sorting from ancestral polymorphisms.

The fourth paper (Part 5: Concordant phylogeographic patterns of the chloroplast and nuclear genomes in the North American plum syngameon [*Prunus* subgenus *Prunus* Section *Prunocerasus*; Rosaceae]) employs nDNA sequence data from the *s6pdh* intron one to reevaluate the phylogenetic hypothesis put forth in Part 5 using an independent data set. Two main conclusions can be drawn from this study. First, the results presented here coupled with the results of Part 4 clearly demonstrate that most section *Prunocerasus* taxa are not monophyletic with respect to either their chloroplasts or the nuclear encoded *s6pdh* gene. Although some nuclear and chloroplast haplotypes are more common in some taxa, few are exclusive. In fact, total evidence strongly supports the hypothesis of the North American plums behaving as a syngameon, or hybridizing species group. Second, *P. texana*, currently classified as a peach in subgenus *Amygdalus*,

was positioned with strong support within the North American plums. This position is in agreement with the results of Part 4.

In summary, genetic structure within the North American plums was evaluated using the independent chloroplast and nuclear genomes to gain insights into this taxonomically troublesome group. The fruits of this course of study do not ultimately solve the "hardest puzzle in American pomology," but they do provide a reason as to why the North American plums are the "hardest puzzle in American pomology," or are such a taxonomically difficult group. This study demonstrates that the North American plums are a series of reticulating morphogeographical taxa that model a syngameon as defined by Grant (1981): the most inclusive interbreeding population system in a hybridizing species group whose members behave like well-isolated biological species on its outer boundary, but differs in having a more complex internal structure where many entities come in close contact and hybridize.

Natural extensions of this work include further study of the Federally Endangered *P. geniculata*. This study is warranted because only one accession of this morphologically unique plum was acquired during this study. Because this taxon is restricted in range to the sand-pine scrub community of the Lake Wales Ridge in central Florida it is open to many interesting evolutionary questions beyond whether gene flow between it and the other section *Prunocerasus* taxa is occurring. For example, is genetic diversity correlated to burn frequency of this scrub habitat? Another interesting taxon that deserves further study is *P. texana*. Before this investigation, Texas Peachbush had been allied with the peaches in subgenus *Amygdalus*. The results put forth here not only challenge the taxonomic position of this Texas endemic but it also raise questions as to the definitions of the subgenera (peaches are generally believed to be pubescent, as is *P. texana*, and plums glabrous).

Other courses of future study will certainly employ insights gained from the North American plums to address other specific evolutionary histories within the plethora of subgenera and sections within this large, complex, and economically important genus. Currently, I am in the early phases of a collaborative project to more broadly focus on the evolutionary history of *Prunus*. This will certainly provide a backbone for further interpretation of the evolutionary and biogeographic history of *Prunus* and further illuminate the biogeographic inferences made in this study on the North American plums.

VITA

Joey has lived a good life. He was born to handsome parents on July 23, 1970 in Smithtown, NY and lived there in a suburban setting until the age of 10. From 10 to 15 years of age he lived in Boxboro, MA and Birmingham, AL. He and his large and extended family moved to Chattanooga, TN in 1986.

After coasting through high school, he did construction work for a while installing heating and air conditioning units, apprenticing as a commercial electrician, remodeling houses, and when new construction fell on hard times in the early 1990's (Reaganomics?), he began work in an oppressive La-Z-Boy chair factory in Dayton, TN. There he felt his happy life slipping away to ennui of factory work. So, he began to attend Chattanooga State Technical Community College in 1992 as a part-time student where he reacquainted himself with the life sciences that he thought were cool as a kid. After graduating from the University of Tennessee, Chattanooga in 1998 with a degree in General Biology (Magna Cum Laude), Joey was awarded a graduate assistantship at the University of Tennessee, Knoxville, Tennessee and earned a Master of Science in Botany in May, 2001. After completing his M.S. degree, Joey was again awarded a graduate assistantship at the University of Tennessee, Knoxville and earned the Doctor of Philosophy degree in Botany. As of the spring of 2005, Joey has accepted an Assistant Professorship in the Department of Biology and Environmental Science at his *alma mater*, the University of Tennessee, Chattanooga.

Currently, he and M have Aidan, Girl, and Truckie…