5-2016

Population genetics and genomics within the genus *Pityopsis*

Elizabeth Anne Hatmaker
University of Tennessee - Knoxville, ehatmake@utk.edu

To the Graduate Council:

I am submitting herewith a thesis written by Elizabeth Anne Hatmaker entitled "Population genetics and genomics within the genus *Pityopsis*." I have examined the final electronic copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science, with a major in Entomology and Plant Pathology.

Margaret E. Staton, Major Professor

We have read this thesis and recommend its acceptance:

Phillip A. Wadl, Edward E. Schilling, Robert N. Trigiano

Accepted for the Council:
Dixie L. Thompson

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

Population genetics and genomics within the genus *Pityopsis*

A Thesis Presented for the
Master of Science
Degree
The University of Tennessee, Knoxville

Elizabeth Anne Hatmaker
May 2016

Acknowledgements

I would like to thank the following people for their help and support through this process. First, Drs. Phillip A. Wadl, Margaret E. Staton, and Đenita Hadziabdic-Guerry for their support, both emotional and academic. Thank you to Dr. Robert Trigiano for hiring me six years ago—your confidence in my work throughout my time at UT has changed my life in so many ways. I thank my committee members (Drs. Phillip A. Wadl, Margaret E. Staton, Edward E. Schilling, and Robert Trigiano) for their expertise and helpful comments toward the completion of this thesis, as well as my additional co-authors Adam Dattilo and Dr. Timothy Rinehart. I would also like to thank Drs. Loran Anderson, Susan Carr, Joan Walker, and Eugene Wofford for their assistance in tissue collection, as well as T.J. Savereno. Sarah Boggess deserves a huge thank you for both her friendship and laboratory expertise, as does Laura Poplawski—the two of you make science more fun. Additionally, I would like to thank Thomas Lane and Nathan Henry for their patience and willingness to share their knowledge of bioinformatics and computer science, as well as Emel Ören and Tyler Edwards for their support in and out of the laboratory. Above all, I would like to thank my parents and family for their constant encouragement and support. This work was supported by USDA Agreement no. 58-6404-1-637 and the J. Frank Schmidt Family Charitable Foundation. Mention of trade names of commercial products in this thesis is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the University of Tennessee or the U. S. Department of Agriculture.

Abstract

Pityopsis (Asteraceae) includes seven species; one species, *P. ruthii*, is federally endangered. The genus exhibits a range of ploidy levels, widespread hybridization among species with overlapping ranges, and interesting adaptive traits such as fire-stimulated flowering. However, taxonomy of *Pityopsis* has remained unresolved. Resolving interspecific relationships can lead to a deeper understanding of the inheritance and hybridization patterns, as well as the evolution of adaptable traits. Our first objective was to examine population structure and gene flow within *Pityopsis ruthii*. Polymorphic microsatellite markers (7 chloroplast and 12 nuclear) were developed and used to examine genetic diversity of 814 *P. ruthii* individuals from 33 discrete locations along the Hiwassee and Ocoee Rivers. A total of 198 alleles were detected with the nuclear loci and 79 alleles with the chloroplast loci. Bayesian cluster analyses of both rivers identified six clusters when the chloroplast microsatellites were used, whereas only two clusters were identified from the nuclear microsatellites. The population structure of *P. ruthii* will allow delineation of conservation units that account for subpopulations along each river. Our second objective was to examine the relationships of the seven species within *Pityopsis* using phylogenetic analyses. The chloroplast genome was sequenced for six species and two varieties. A reference chloroplast genome was assembled *de novo* from *P. falcata*, the species with the highest depth of read coverage. Reads from seven other individuals were then aligned to the *P. falcata* chloroplast genome and an individual genome was assembled for each. To utilize all informative sites for the full length of the chloroplast, a multiple sequence alignment of the eight chloroplast genomes was constructed, and from this, a phylogeny using both the maximum likelihood and maximum parsimony methods. Our findings using the entire chloroplast genome deviate from the results of previous phylogenetic studies of *Pityopsis* and do not support previously defined clades or sections within the genus. Our two objectives add meaningful information about the diversity of *P. ruthii* and the evolutionary history of *Pityopsis*, now available for use by conservationists, molecular ecologists, and evolutionary biologists.

Table of Contents

Chapter 1. An introduction to population genetics, phylgenomics, and *Pityopsis* 1
Chapter 2. Population structure and genetic diversity within the endangered species *Pityopsis ruthii* (Asteraceae) 7
Abstract 9
Introduction 10
Materials and Methods 13
Plant material and microsatellite genotyping 13
Analysis of nuclear data 15
Analysis of chloroplast data 16
Results 17
Genetic diversity in nuclear data 17
Genetic diversity in chloroplast data 19
Discussion 23
Chapter 3. Comparing chloroplast genomes in *Pityopsis* species (Asteraceae) 29
Abstract 31
Introduction 32
Materials and Methods 34
Library construction and sequencing 34
Sequence trimming and alignment 35
Genome assembly and annotation 36
Alignment and comparison 36
Results 37
Chloroplast genome sequencing and assembly 37
Phylogenetic analyses 38
Discussion 39
Comparison of chloroplast genomes 39
Pairwise comparisons and phylogenetic analyses of *Pityopsis* species 40
Chapter 4. Conclusion 43
Literature Cited 46
Appendices 52
Appendix 1. Tables 53
Appendix 2. Figures 69
Vita 80

List of Tables

Table 1. Summary statistics for 12 nuclear microsatellite loci in *Pityopsis ruthii* populations..... 54
Table 2. Genetic variability of 12 nuclear microsatellite loci estimated for 33 subpopulations of *Pityopsis ruthii*..... 55
Table 3. Analysis of molecular variance (AMOVA) from nuclear microsatellite data collected from *Pityopsis ruthii* using Arlequin (version 3.5.1.2) 56
Table 4. Subpopulation pairwise F_{ST} values (below diagonal) and gene flow estimates (above diagonal) for *Pityopsis ruthii* on the Hiwassee River 57
Table 5. Subpopulation pairwise F_{ST} values (below diagonal) and gene flow estimates (above diagonal) for *Pityopsis ruthii* on the Ocoee River 58
Table 6. Characterization of seven chloroplast microsatellite markers from *Pityopsis ruthii* 59
Table 7. Summary statistics by population for the seven *Pityopsis ruthii* chloroplast microsatellite markers 60
Table 8. Analysis of molecular variance (AMOVA) from chloroplast microsatellite data collected from *Pityopsis ruthii* using Arlequin (version 3.5.1.2). 61
Table 9. Wilcoxon tests for recent bottlenecks using the program BOTTLENECK for *Pityopsis ruthii* subpopulations 62
Table 10. Tissue storage method, date, and location of *Pityopsis* individuals used for sequencing. 63
Table 11. Statistics from original sequences for all *Pityopsis* individuals 64
Table 12. Statistics of *Pityopsis* sequences mapped to the *Helianthus annuus* reference..... 65
Table 13. Genes included in the *Pityopsis* chloroplast genomes..... 66
Table 14. Pairwise alignment comparison of eight *Pityopsis* chloroplast genomes. Percent dissimilarity as calculated in MEGA7. 67
Table 15. Estimates of evolutionary divergence between *Pityopsis* cp genomes and *Aster spathulifolius*. Number of base differences per sequence between sequences are shown. Analyses were conducted in MEGA7. 68

List of Figures

Figure 1. Bar plot (top panel) showing Bayesian assignment probabilities of nuclear data for the Hiwassee and Ocoee Rivers subpopulations using the program STRUCTURE for two clusters (*k* = 2). Bar plot (lower panel) showing Bayesian assignment probabilities of chloroplast data for the Hiwassee and Ocoee Rivers subpopulations for six clusters (*k* = 6) 70
Figure 2. Bar plots of the individual Bayesian assignment probabilities of nuclear data for the *Pityopsis ruthii* Hiwassee River subpopulations using the program STRUCTURE for two or four clusters... 71
Figure 3. Bar plots of the Bayesian assignment probabilities of nuclear data for the *Pityopsis ruthii* Ocoee River populations using the program STRUCTURE for three clusters..... 72
Figure 4. Bar plot of individual Bayesian probabilities of chloroplast data for *Pityopsis ruthii* Hiwassee River populations using the program STRUCTURE for four clusters..... 73
Figure 5. Bar plot of the individual Bayesian assignment probabilities of chloroplast data for *Pityopsis ruthii* subpopulations along the Ocoee River using STRUCTURE for two clusters (k = 2)..... 74
Figure 6. Isolation by distance graph of *Pityopsis ruthii* populations using nuclear microsatellites..... 75
Figure 7. Isolation by distance graph of *Pityopsis ruthii* populations using chloroplast microsatellites.... 76
Figure 8. Visualization of *Pityopsis* chloroplast gene map with annotations. Length in kb and region of chloroplast on inside circle 77
Figure 9. Phylogenetic relationships within *Pityopsis* using the maximum parsimony approach. The tree was constructed using the min-mini heuristic model in MEGA7..... 78
Figure 10. Phylogenetic relationships within *Pityopsis* using the maximum likelihood approach. The tree was constructed using the Kimura 2-parameter model in MEGA7. The tree with the highest log likelihood is shown..... 79

Chapter 1. An introduction to population genetics, phylogenomics, and *Pityopsis*

Botany has long embraced the study of genetics. The earliest theories of heredity were derived from Gregor Mendel's pea breeding experiments (Carlson 2004). With new molecular techniques, plant biology and genetics are currently used to understand complex trait inheritance (Nordborg and Weigel 2008), genotype and phenotype associations, evolution of adaptive traits, and population dynamics. Population and phylogenetic studies of plants are particularly useful for endangered species, where conservation and management strategies can utilize genetic information for maintenance and restoration of diversity.

The field of population genetics, though historically a theoretical exercise, has grown recently thanks in large part to the advanced molecular and computational techniques available to scientists (Ouborg et al. 2010, Davey et al. 2011). Genetic diversity may be observed as complex phenotypes with multiple variables that are often difficult to measure (Avolio et al. 2012); the measurement of allele frequencies offers a tractable, quantitative method to understand, interpret, and utilize measures of diversity. Scientists can now detect the alleles of tens to thousands of genes within and among populations. Evolutionary biology provides a framework in which to interpret these allelic changes. For example, population genetics can reveal an excess or lack of diversity within a species (Ouborg et al. 2010). Higher levels of genetic and genotypic diversity within a plant population can lead to increased diversity within the community as a whole (Crutsinger et al. 2006); a community of diverse species can in turn lead to a higher level of ecosystem functioning than a monoculture (Tilman et al. 1997). Alternatively, sampling a population's genetic diversity may reveal inbreeding within the population (Ouborg et al. 2010). This knowledge is beneficial to conservation, as management of threatened species aims to conserve the adaptive ability of the species in the event of environmental changes (Frankham et al. 2009, Ouborg et al. 2010). Though useful in conservation and ecology, estimating genetic diversity and population structure for use in a correlative approach has limitations (Ouborg et al. 2010). Neutral molecular markers do not discern the underlying reasons of a population's decline, which

may be better understood by studying functional genes and pathways. To augment the usage of DNA markers, new molecular approaches have led to the establishment of the new field of eco-genomics, or conservation genomics (Ouborg et al. 2010).

Genomic tools have been widely used to study evolution and systematics, leading to the advent of phylogenomics, a marriage between genomics and evolution (Eisen and Fraser 2003). The reconstruction of phylogenies provides a framework for understanding the evolution of genes and genomes, which can then be applied to conservation and climate change biology (Davis et al. 2010, Hoffmann et al. 2015). Chloroplast DNA, in particular, is useful in understanding the evolutionary history of plants and the events leading to current population structure and patterns of distribution due to the lack of recombination and uniparental inheritance (Byrne 2007). Therefore, phylogenies reconstructed from chloroplast genomes are able to elucidate recent and historical events within the lineage of plant species; however, chloroplasts only track the maternal lineage (Parks et al. 2009).

Population genetics and phylogenomics are useful for studying many groups of plants, including the genus *Pityopsis*, the focus of this thesis. *Pityopsis* is a member of Asteraceae and has been the subject of several phylogenetic studies (Gowe and Brewer 2005, Teoh 2008), but intergeneric relationships for all species and varieties in the genus have not been fully resolved.

Pityopsis was widely considered to be part of *Heterotheca* until 1980, when Semple, Blok, and Heiman distinguished it based on anatomical, morphological, and ecological differences (Semple et al. 1980). Semple and Bowers later (1985) revisited the genus and proposed two distinct sections based on morphological differences in rosette growth and stem leaf traits. Section *Pityopsis* includes *P. falcata, P. flexuosa, P. pinifolia*, and *P. ruthii,* whereas section *Graminifoliae* includes *P. aspera, P. graminifolia*, and *P. oligantha*. Gowe and Brewer (2005) attempted to determine the evolution of fire-based traits using a phylogeny based on morphological data, showing that fire-based traits occurred in section *Graminifoliae* but not *Pityopsis*. Teoh (2008) examined fire-dependent flowering in the genus based on a phylogeny

from molecular data, but not all varieties were included, leaving certain issues of hybridity and polyploidy unresolved, though supporting an allopolyploid lineage of tetraploid *P. graminifolia* var. *latifolia*.

Pityopsis currently includes seven recognized species (Semple 2006), though Weakley (2010) considers the genus to include about 8-13 taxa, including species and varietal rankings. Alternate systems have been proposed, but we are choosing to use Semple (2006) as our taxonomic basis for this study. *Pityopsis falcata*, or the sickle-leaved golden aster, is found in the sandplains of New England. Though it is considered a conservation concern due to the narrow range, it can be quite locally abundant (Vickery 2002, Farnsworth 2007). Similarly, *P. flexuosa* is only found in northern Florida but can attain high local abundances (Gowe and Brewer 2005). It is considered endangered by the state of Florida (USDA 2016).

Pityopsis graminifolia is the most widespread species of the genus, with five varieties: *aequilifolia, graminifolia, latifolia, tenuifolia,* and *tracyi*. Two varieties, *P. graminifolia* var. *aequilifolia* and var. *tracyi*, are found exclusively in Florida. *Pityopsis graminifolia* var. *aequilifolia* is endemic to central Florida and the hexaploid *P*. *graminifolia* var. *tracyi* occurs in the Florida peninsula (Semple 2006). *Pityopsis graminifolia* var. *graminifolia* has a range from eastern Louisiana to the Florida panhandle and north into southeastern North Carolina (Semple 2006). The remaining two varieties, *P*. *graminifolia* var. *tenuifolia* and *P*. *graminifolia* var. *latifolia* have a large range throughout the southeastern U.S., from Oklahoma and Texas to Virginia and North Carolina, and further south into Florida, though *P*. *graminifolia* var. *latifolia* has the widest range of the species and is a tetraploid (Semple 2006).

Often found alongside *P. graminifolia* var. *tenuifolia* in xeric sandhills and long-leaf pine communities of Florida, Georgia, and South Carolina, *P. aspera*, commonly called pineland silkgrass, has been more widely researched than other species (Gowe and Brewer 2005, Gornish 2013) with respect to

fire-influenced flowering and general ecology (Gornish 2013). *Pityopsis aspera* contains two varieties: *P. aspera* var. *aspera* and *P. aspera* var. *adenolepis*. Clewell (1985) proposed that *P. aspera* var. *adenolepis* was a separate species, but Semple's taxonomy places it as a variety of *P. aspera* (Semple and Bowers 1985, Semple 2006).

Pityopsis oligantha, or the large-flowered goldenaster, is similar to *P. graminifolia* and is found from the panhandle of Florida west to Mississippi, but has been reported in both Louisiana and southeastern Texas (Holmes and Singhurst 2012). It occurs in fire-maintained long-leaf pine communities and savannas (Gowe and Brewer 2005) and is considered vulnerable due to habitat destruction and fragmentation and forest management practices. Similarly, *P. pinifolia* also grows in long-leaf pine communities. Known as the sandhill goldenaster, *P. pinifolia* is considered threatened in Georgia (USDA 2016). The species can be found in the sandhills of North Carolina, South Carolina, and Georgia, typically in low abundances in its natural habitat of long-leaf pine communities. However, it can grow abundantly in open spaces (Gowe and Brewer 2005).

Pityopsis ruthii, or Ruth's golden aster, is endangered and endemic to southeastern Tennessee. It grows along short stretches of unshaded phyllite rock outcrops on the Hiwassee and Ocoee Rivers (Bowers 1972) and is at risk for short-term extinction due to altered river flow resulting from damming, which has led to higher competition rates and loss of seed dispersal (Thomson and Schwartz 2006). *Pityopsis ruthii* is also threatened by habitat encroachment from non-related species. Thomson and Schwartz (2006) posit this encroachment is a side effect of damming due to lower water flows which are unable to adequately scour the habitat; high flows scouring the rock keep competition low. However, Moore et al. (2016) hypothesize that drought actually helps maintain *P. ruthii* habitat and prevents encroachment at some sites.

Understanding the ecology and evolutionary history of *Pityopsis* will enable more informed conservation practices within the genus, including management and reintroduction of *P. ruthii* and other

threatened species. A basis of understanding regarding the diversity present in subpopulations and what constitutes a population will inform breeding for augmentation studies and management of the species. Further work is also needed to clarify the species and section divisions and the mechanisms of inheritance of complex traits within *Pityopsis*. Studies of inheritance and physiology of drought-tolerance and fire-stimulated flowering are of particular interest and a phylogenetic tree will address whether these arose in species concurrently or from a single evolutionary event. To this purpose I have completed two objectives: to examine population structure and gene flow within the endangered species *Pityopsis ruthii,* and to collect and analyze molecular phylogenetic data to help understand the relationships of the seven species within *Pityopsis.*

Chapter 2. Population structure and genetic diversity within the endangered species *Pityopsis ruthii* (Asteraceae)

Division of Labor among Co-authors

This chapter was produced in collaboration with my committee members and other scientists and will be published in a peer-reviewed journal. My co-authors include Adam Dattilo, Denita Hadziabdic, Timothy A. Rinehart, Edward E. Schilling, Margaret E. Staton, Robert N. Trigiano, and Phillip A. Wadl.

Drs. Schilling and Trigiano gave intellectual guidance and helped with experimental design, as well as presentation of the results. Dr. Staton helped with writing and interpretation. Dr. Hadziabdic provided analytical guidance. Dr. Wadl helped with sample collection, experimental design, analytical and interpretative methods, and writing. Dr. Rinehart provided technical support. Mr. Dattilo helped with sample collection and experimental design. I designed the experiment, analyzed and interpreted the results, presented data appropriately, and wrote the chapter.

Abstract

Pityopsis ruthii (Ruth's golden aster) is a federally endangered herbaceous perennial endemic to southeastern Tennessee. This Asteraceae species grows along the Hiwassee and Ocoee Rivers and is at risk for short-term extinction. Genetic studies for *P. ruthii* are lacking and are needed to provide novel information to conservationists and researchers in order to facilitate preservation of the species. Genetic variation and gene flow of natural plant populations were evaluated for 814 individuals from 33 discrete locations using 19 polymorphic microsatellites (7 chloroplast and 12 nuclear). A total of 198 alleles were detected with the nuclear loci and 79 alleles with the 7 chloroplast loci. Gene flow was estimated, with the Hiwassee River showing overall higher levels than the Ocoee River locations. Population structure and clustering patterns were examined using Bayesian cluster analyses. Nuclear and chloroplast data grouped individuals into different clusters. From the chloroplast microsatellites, three clusters were identified and all were present in sampling sites at both rivers, indicating a lack of allele fixation along rivers. In contrast, the nuclear markers, revealed two separate clusters, one for each river. When the Hiwassee River locations were analyzed, four clusters were identified for both the chloroplast and nuclear microsatellites, though the individuals clustered differently. Both data sets showed similar clustering among the Ocoee River locations, with two clusters. We recommend *P. ruthii* be managed as four populations within the Hiwassee River habitat and two populations within the Ocoee River habitat. Understanding diversity within populations of *P. ruthii* will impact the current conservation methods and plans by defining subpopulations to ensure effective retention of genetic diversity, especially in augmentation and translocation studies to add diversity to a particular population. The diversity and population structure results also provide a baseline of genetic diversity for population augmentation and future monitoring of the species.

Introduction

Endangered species generally have small or declining populations, and often these populations suffer from inbreeding and erosion of genetic diversity resulting in elevated extinction risks (Frankham 2003). The delineation of conservation units is a critical first step in conservation of a species to ensure that resource managers know where population boundaries lie (Funk et al. 2012), and to monitor and conserve existing genetic diversity. The federally endangered *Pityopsis ruthii* (Small) Small, also known as Ruth's golden aster, is endemic to Polk County in southeastern Tennessee, USA. *Pityopsis ruthii* (2n=2x=18) grows on unshaded phyllite rock boulders in and on the adjacent slopes of the Hiwassee and Ocoee Rivers (Bowers 1972). The riparian habitat of *P. ruthii* is highly dynamic and is typified by seasonally high temperatures, frequent drought, and regular inundating flood flows. One study found that altered river flow due to damming has apparently led to higher competition rates and lower seed dispersal, which has put *P. ruthii* at risk for extinction in the near future (Thomson and Schwartz 2006), though further studies have not supported such claims (Moore et al. 2016).

Two geographically separated populations of *Pityopsis ruthii* remain in the wild, one containing approximately 1,000 individuals along ~3 kilometers of the Ocoee River and a larger population of around 12,000 individuals along ~6.5 kilometers of the Hiwassee River (Moore et al. 2016). In spite of its endangered status, as evidenced by its small population size and narrow geographic range, relatively little research has focused on species recovery. The US Fish and Wildlife Service species recovery plan identifies actions necessary for the delisting of the plant, including defining what constitutes a viable population and developing management protocols that ensure the existence of self-sustaining populations along both the Hiwassee and Ocoee Rivers (USFWS 1992). Knowledge regarding existing population structure and genetic diversity, and the delineation of conservation units for *P. ruthii* are critically needed to facilitate long-term conservation and management efforts.

The population dynamics of *P. ruthii* can be used to inform ecological and conservation issues, both in dealing with conservation of this endangered species, and when addressing larger ecological issues, such as the role of damming on the genetic diversity of riparian species, which can be monitored over time using molecular means outlined in this study. Community ecology has been shown to change with damming, often leading to an increase in non-native plants, which can drastically alter the composition of the riparian ecosystem (Greet et al. 2013). The relatively small system, extensive sampling, and annual census of individuals, combined with the knowledge of all known populations of *P. ruthii* (Moore et al. 2016) provides an ideal scenario to conduct comprehensive population studies as a model for other endangered plant species. *P. ruthii* can therefore serve as a model plant to explore the effects of conservation techniques and river flow on a riparian plant.

Information of genetic structure is essential for understanding the scales over which dispersal, genetic drift, and selection operate in populations (Slatkin 1987). Genetic research determining population diversity can be invaluable when forming and revising an endangered species management plan, as maintaining diversity is critical for conservation (Powell et al. 1996). Fragmentation or disappearance of natural populations can lead to reduced gene flow among populations and thus increase genetic differentiation among populations and genetic structuring due to genetic drift (Hartl et al. 1997, Ouborg et al. 2006). The effects of isolated and fragmented populations on attributes of the genetic structure of *P. ruthii*, in particular genetic erosion, are unknown, especially when habitat degradation is taken into account. Estimating genetic diversity and genetic drift, as well as determining gene flow of the endangered plant species *P. ruthii* using chloroplast and nuclear microsatellites will advance molecular ecology and conservation efforts for the species. Understanding genetic drift and gene flow allows further inferences about the history of the species and delineation of viable populations, as called for in the species recovery plan (USFWS 1992).

Molecular markers provide a useful method to examine diversity. Microsatellite markers in particular are a popular and cost-effective way to measure genetic diversity within populations (Frankham et al. 2009, Abdul-Muneer 2014). The short tandem repeats in non-coding sections of DNA exhibit co-dominant inheritance in nuclear DNA and are highly variable (Wan et al. 2004). This study uses both chloroplast and nuclear microsatellites to examine genetic diversity within *P. ruthii*. Nuclear microsatellite markers were developed and demonstrated to be suitable for determination of genetic diversity in a limited sample size from the Hiwassee and Ocoee River populations of *P. ruthii* (Wadl et al. 2011a). Additionally, microsatellite markers were developed from the chloroplast genome and used as a complement to the nuclear markers to understand the natural history of the species as well as recent ecological changes. The haploid nature of the chloroplast allows identification of loci with a single allele in each individual, creating a nice complement for use with biparentally inherited molecular markers such as nuclear microsatellites. Chloroplast microsatellites (cpSSRs) have uniparental inheritance, a lack of recombination, and a slower mutation rate than nuclear microsatellites (Provan et al. 2001). Chloroplast microsatellites are especially useful for understudied groups such as native plants with a lack of *a priori* knowledge about genetic structure and species of little economic importance (Wheeler et al. 2014). Neutral, non-coding, and easy to develop, microsatellites are suitable when studying small occurrences of native plants under a variety of evolutionary pressures, such as in *P. ruthii*. This study is the first to use highly variable microsatellite loci to examine the genetic diversity and population structure of *P. ruthii*. The results of this study will provide valuable genetic information that can be combined with the development of effective propagation and reintroduction techniques (Wadl et al. 2011b, Wadl et al. 2014), geospatial mapping of all known occurrences, and stressors affecting *P. ruthii* (Moore et al. 2016) to guide conservation and management decisions for the species.

Materials and Methods

Plant material and microsatellite genotyping

In 2010, leaf samples were collected from 814 *Pityopsis ruthii* individuals across 33 discrete geographical sites: 25 sites on the Hiwassee River and 8 sites on the Ocoee River. Subpopulations were established as groups of plants based on well delineated breaks in suitable habitat within the riparian area for each river. The average census counts (2011-2014) for the subpopulations sampled ranged from 15 to 1034 plants for the Hiwassee River and 12 to 491 plants for the Ocoee River (Moore et al. 2016). At the time of sampling, we collected leaf tissue from all individual plants at a known subpopulation if the total number of individuals was less than 50. When subpopulations were greater than 50, we used a random number generator to randomly sample up to 50 individuals. Since sampling occurred in 2010 additional locations have been discovered and further delineations of the locations has occurred based on natural breaks in suitable habitat (Moore et al. 2016), explaining why some locations have fewer than five individuals sampled. Lastly, sampling of individuals is difficult because *P. ruthii* reproduces asexually by rhizomes and sexually via seeds, therefore, distinguishing individual plants is challenging. To ensure sampling of individuals rather than clones, plants occupying the same crevice and occurring less than 15 cm apart were considered an individual, whereas plants more than 15 cm apart in the same crack and plants occupying apparently distinct crevices were considered separate individuals. We further attempted to ensure that a single genetic individual was used for microsatellite genotyping by using DNA isolated from a single leaf for each individual.

Total genomic DNA was isolated from leaf samples using the Qiagen DNeasy Plant Mini Kit (Qiagen, Valencia, CA, USA) according to the manufacturer's protocol. The concentration and quality of the DNA was measured with a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Inc., Wilmington, DE, USA). Twelve polymorphic microsatellite primer pairs described previously for *P. ruthii* (Wadl et al. 2011a) were used to genotype individuals. Additionally, a 400 base pair (bp) genomic DNA

library was developed from next generation sequencing of a single *P. ruthii* genotype using the Ion Torrent Personal Genome Machine sequencing platform (ThermoFisher). The *de novo* assembly yielded chloroplast genome reads that were then assembled and screened for microsatellites using the program Imperfect SSR Finder (Stieneke and Eujayl 2007). Twenty-one loci from sequences were obtained and screened with samples from five locations along the Hiwassee and Ocoee Rivers. Using a 2% agarose gel and the Qiaxcel Advanced Capillary Electrophoresis System (Qiagen), 7 loci were found to be polymorphic among a subset of 66 individuals, 11 from each of 6 sampling sites—2 from the Ocoee River and 4 from the Hiwassee River.

For microsatellite analyses, all 814 *P. ruthii* individuals were amplified using a 10µl polymerase chain reaction with 4 ng genomic DNA, 2.5 mM $MgCl_2$, 1X GeneAmp PCR Buffer (Applied Biosystems, Carlsbad, CA), 0.2 mM dNTPs, 0.25 µM primer (forward and reverse), 0.4 U AmpliTaq Gold DNA polymerase (Applied Biosystems), 5% dimethyl sulfoxide (DMSO), and sterile water. The reaction was run using the following cycling conditions: 94 ° C for 3 min; 35 cycles at 94 ° C for 40 sec, 55 ° C for 40 sec, and 72 ° C for 30 sec; and one cycle at 72 ° C for 1 min for final extension. All individuals were amplified using 12 nuclear microsatellite loci and 7 chloroplast microsatellite loci. Amplicons were visualized using the QIAxcel Advanced Capillary Electrophoresis System (Qiagen) and sized using an internal 25-500 base pair (bp) DNA size marker. Electropherograms were visualized using the software BioCalculator (Qiagen) for nuclear data and the next iteration of the software, ScreenGel version 1.4.0 (Qiagen) for chloroplast data. The raw allelic data was compiled into an Excel worksheet and all nineteen loci were binned using the Excel add-in FLEXIBIN (Amos et al. 2007), which bins raw allele length data into allele size categories using an automated algorithm and reduces false inflation of diversity. The binned data was used in subsequent data analyses. When both datasets were combined for Bayesian analyses, chloroplast data was coded as diploid.

Analysis of nuclear data

The Excel add-in GenAlEx 6.5 (Peakall and Smouse 2001, Peakall and Smouse 2012) was used to estimate genetic diversity among the 814 samples. The mean number of alleles (N_A), effective number of alleles (N_E), observed (H_O) and expected (H_E) heterozygosity, gene flow (N_m), and *F* statistics were calculated across all populations for each locus. A Bayesian analysis was performed using STRUCTURE version 2.3.4 (Pritchard et al. 2000) using the admixture model, which infers whether the individual *i* has inherited a portion of its genetic material from ancestors in population *k.* For measuring different values of *k*, 10 independent replicates were made for each *k* value between 1 and 10 with a burn-in period of 100,000 iterations and 100,000 Markov Chain Monte Carlo (MCMC) repetitions. Estimation of the best *k* value was determined using STRUCTURE Harvester (Earl 2012) following the Evanno et al. (2005) method, which identifies the appropriate number of clusters (*k*) using the ad hoc statistic Δ*k*. This is based on the second order rate of change in the log probability of the data between successive values of *k*. The analysis was performed three times: once with all individuals, once with individuals found in the population located along the Hiwassee River, and once with individuals found in the population located along the Ocoee River. Additionally, subsampling of populations over 25 individuals was performed using a random number generator for further Bayesian analyses, to ensure accuracy despite uneven sampling. Subsampling yielded a smaller dataset of 683 individuals.

The apportionment of genetic variation was determined by an analysis of molecular variance (AMOVA) using GenAlEx 6.5 (Peakall and Smouse 2001, Peakall and Smouse 2012). The significances of variance components for each hierarchical comparison (among populations, among individuals, among individuals within populations) were tested using 9,999 permutations. Additionally, GenAlEX was used for pairwise calculations of F_{ST} and gene flow estimates between populations. To determine the occurrence of isolation by distance (IBD), a Mantel test between the genetic and geographic distances was evaluated using GenAlEx with 9,999 permutations.

BOTTLENECK version 1.2.02 (Piry et al. 1999) was used to determine which populations may have undergone significant reductions in size and to test for allele frequency mode-shifts (i.e. distortion away from the typical L-shaped distribution). We also tested for the presence of an excess of observed heterozygotes by using the Wilcoxon signed rank test to evaluate deviations from 50:50 deficiency/excess (Cornuet and Luikart 1996, Luikart and Cornuet 1998). Heterozygote excess was tested under all three mutation models, infinite alleles (IAM), two-phase (TPM), and the step-wise mutation model (SMM). For TPM we set ps = 0.9 (the frequency of single step mutations) and the variance of those mutations as 12. These are generic values typical for many microsatellite markers (Busch et al. 2007).

Analysis of chloroplast data

The Excel add-in GenAlEx 6.5 was used to calculate several diversity indicators. The haploid data set was combined with geographic coordinates for input into the program. Population differentiation (F_{ST}) was calculated for all samples, and Shannon's diversity index, diversity (*h*), and unbiased diversity (*uh*) were calculated for each population and locus. Nei's gene diversity (Nei 1973) was calculated for populations using GenAlEx as well. A principal coordinates analysis (PCA) based on a covariance matrix was also calculated. A Mantel test using population pairwise geographic distance and genetic distance (Mantel 1967) was performed to determine isolation by distance.

Clustering of the populations was performed using STRUCTURE 2.3.4 (Pritchard et al. 2000). Posterior probabilities were estimated for three different chloroplast data sets: all sampling sites, Hiwassee River sampling sites, and Ocoee River sampling sites. For the Hiwassee River sites, *k* = 1-15. For the Ocoee River sites, *k* = 1-10. For all chloroplast data, *k* = 1-30. An admixture model was assumed for all analyses. The burn-in generation and the MCMC were set to 250,000, with 20 iterations. Delta K, the optimal number of clusters for the sample set, was estimated using the Evanno method (Evanno et al. 2005)through STRUCTURE Harvester, as described previously. We also subsampled the dataset for the

chloroplast data using the same 683 individuals as the subsampled nuclear dataset, with no more than 25 individuals per location. Burn-in generation was 100,000 and MCMC was 50,000 for analysis of subsampled data, with 10 iterations.

A standard AMOVA was calculated using the program ARLEQUIN (Excoffier et al. 2005, Excoffier and Lischer 2010), using a pairwise distance matrix with 9,999 permutations and a threshold of 5% missing data, which excluded 10 individuals. Three hierarchical AMOVA analyses were performed for both the nuclear and the chloroplast data sets. The first analysis included all sampling sites as one hierarchical group, the second analysis included all sampling sites on the Hiwassee River, and the third included all sampling sites from the Ocoee River. Haplotype frequency was also analyzed using the program ARLEQUIN.

Results

Genetic diversity in nuclear data

A total of 814 *Pityopsis ruthii* individuals from 33 discrete locations were genotyped using 12 nuclear and 7 chloroplast microsatellite loci. For the purpose of clarity, each discrete location will be referred to as a subpopulation.

For the 12 polymorphic nuclear loci, 198 alleles were detected (Table 2). All loci demonstrated an overall departure from HWE due to significant heterozygote deficiency when all 814 samples were analyzed together. The number of alleles detected per locus (A) ranged from 9 (PR028) to 24 (PR029), with a mean allelic richness (A_R) of 3.35, ranging from 2.55 (PR002) to 4.26 (PR029). The observed heterozygosity (H_O) was 0.49 and deviated from the expected heterozygosity (H_E) of 0.65. Population differentiation was large (F_{ST} = 0.24) and the inbreeding coefficient was moderate (F_{IS} = 0.22). Average gene flow (N_m) across loci was 0.90 and ranged from 0.45 (PR009) to 1.76 (PR035).

The genetic variability of the 12 microsatellite loci was assessed for each subpopulation and between the two rivers (Table 2). For individuals in the Hiwassee River subpopulations, mean allelic

richness was 3.82 and ranged from 2.77 (H-02-01) to 6.00 (H-12-04), whereas among the Ocoee River individuals, mean allelic richness was 2.77 and ranged from 2.37 (O-04-01) to 3.01 (O-06-01). On the Hiwassee River H_O was 0.53 and ranged from 0.38 (H-03-01) to 0.71 (H-09-02) whereas H_E was 0.64 and ranged from 0.51 (H-02-01) to 0.74 (H-04-01). Across the Ocoee River locations, H_O was 0.36 and ranged from 0.32 (O-03-01) to 0.40 (O-02-01) while H_E was 0.54 and ranged from 0.46 (O-04-01) to 0.59 (O-06-01). The inbreeding coefficient was higher for the Ocoee River individuals (0.35) than the Hiwassee River individuals (0.22) and the range of inbreeding coefficient values was much larger for the Hiwassee River individuals (0.22 to −0.01) compared to the Ocoee River individuals (0.27 to 0.44).

Forty-five private alleles were found in four Ocoee River and 13 Hiwassee River subpopulations, of which H-04-04 (11), O-05-01 (5), H-06-05 (4), H-07-01 (4), and H-11-01 (4) had the most, 5 subpopulations (O-02-03, H-09-03, H-07-03, H-06-04, H-06-02) had 2 private alleles, and 7 subpopulations had 1 private allele (O-04-01, O-01-01, H-06-01, H-12-06, H-05-01, H-04-05, H-01-02). Private alleles were found at 12 loci, with PR002 (7) and PR031 (7) having the most and PR003 and PR035 having 1 each (Table 3). Thirty-four private alleles occurred at a frequency of <0.05, 5 at a frequency between 0.05 and 0.09, 5 at a frequency between 0.10 and 0.20, and 1 at a frequency greater than 0.25 (PR029, allele 245).

The variance components of the AMOVA analyses were highly significant at all hierarchical levels ($P < 0.001$, Table 4). Grouping of all subpopulations together indicated that most (68%) of the variation is explained within individuals, and 14% and 18% of the variation is due to differences among individuals within subpopulations and among subpopulations. Additional AMOVA grouping subpopulations by river found similar results to the grouping of all subpopulations, with a greater percentage of the variation explained within individuals rather that among the individuals with the subpopulation. Nearly 20% greater variation was found within individuals on the Hiwassee River as compared to the Ocoee River.

Populations of *P. ruthii* demonstrated a high level of subpopulation differentiation (Table 3, F_{ST} = 0.24, P <0.001; Table 4, F_{ST} = 0.18, P <0.001). Pairwise comparisons of F_{ST} measures were all significantly different from zero for the Hiwassee River except for H-09-01 and H-09-03, H-09-01 and H-09-02, and H-08-03 and H-08-04 (Table 5, P<0.05), and the comparison for the Ocoee River were all significantly different than zero (Table 6, P<0.05). The greatest differentiation for the Hiwassee River subpopulations was observed between H-02-01 and H-12-04 (F_{ST} =0.29) and H-02-01 and H-03-01 (F_{ST} =0.29) and the lowest differentiation between H-08-04 and H-08-07 (F_{ST} =0.02). For the 300 pairwise subpopulation comparisons of F_{ST} measures for the Hiwassee River, 61 were less than 0.10, 155 were between 0.10 and 0.19, and 81 were greater than 0.20. For the Ocoee River subpopulations the greatest differentiation was observed between O-04-01 and O-03-01 (F_{ST} = 0.33) and O-06-01 and O-04-01 (F_{ST} = 0.32) and the lowest between O-02-01 and O-02-02 (F_{ST} = 0.02), with 21 out of 28 comparisons greater than 0.15. Along the Hiwassee River, gene flow was highest between H-09-01 and H-09-03 and lowest between H-02-02 and H-03-01. Within the Ocoee River subpopulations, gene flow was highest between O-02-01 and O-02-02 and lowest between O-03-01 and O-04-01 (Table 6). Overall, the Hiwassee River subpopulations had lower F_{ST} and much higher gene flow estimates than the Ocoee River subpopulations.

Genetic diversity in chloroplast data

We detected a total of 79 alleles among the seven chloroplast loci, with an average of 11.3 alleles per locus, ranging from a minimum of 5 to a maximum of 17 (Table 7). Loci cpPR002 and cpPR010 showed the highest diversity (h=0.54) whereas cpPR004 (h=0.38) showed the least. Private alleles were identified at every locus, with a total of 15 across all loci and subpopulations (Table 8). Thirteen subpopulations had private alleles: four from the Ocoee River and eleven from the Hiwassee River. In one subpopulation, H-04-04, private alleles were detected at multiple loci, whereas the other subpopulations had a single private allele. Subpopulation H-04-04 had the highest number of private

alleles in both the nuclear and chloroplast data sets. Six other subpopulations showed private alleles at both nuclear and chloroplast loci: O-05-01, O-04-01, O-01-01, H-06-07, H-05-01, and H-04-05. Shannon's Information Index was highest for Ocoee River subpopulation O-02-01 (1.068) and lowest for O-02-03 (0.620), whereas for the Hiwassee River subpopulations it was highest for H-01-06 (1.212) and lowest for H-08-07 (0.278). Diversity and unbiased diversity among the Ocoee River subpopulations were greatest in O-03-01 and lowest in O-04-01, and highest in H-01-06 and lowest in H-12-06 among the Hiwassee River subpopulations.

The haplotype analysis in ARLEQUIN detected 102 unique haplotypes from the Ocoee River subpopulations and 176 unique haplotypes from the Hiwassee River subpopulations, with only five complete haplotypes shared between the two rivers. The diversity within the two rivers was comparable, with the Ocoee River (h = 0.49) showing slightly higher diversity per subpopulation than the Hiwassee River (h = 0.45) despite having fewer individuals (Table 7).

A hierarchical AMOVA with two groups (Table 9), Hiwassee and Ocoee Rivers, revealed that differences between river only explained 5% of variation, whereas differences among populations within groups explained 32% of variation and differences within subpopulations explained 63% of variation (P = 0.03). Among the Hiwassee River subpopulations, 37% of all variation could be explained among subpopulations, whereas differences within subpopulations explained 63% of variation (P < 0.01). Analysis of the Ocoee River subpopulations revealed 22% variation among populations and 78% variation explained by differences within subpopulations (P < 0.01). The amount of variation attributed to differences among subpopulations was 15% higher for the Hiwassee River than for the Ocoee River. The Hiwassee River subpopulations had a higher F_{ST} (0.37) than the Ocoee River subpopulations (0.22).

Population structure

STRUCTURE analysis of nuclear microsatellites found evidence for two distinct clusters when the Hiwassee and Ocoee Rivers were combined, which separated the two rivers (Fig. 1). Analysis of the

Hiwassee River subpopulations with STRUCTURE identified two genetically distinct clusters (Fig. 2). Cluster one (green) is composed of individuals from H-09-03, H-09-02, H-09-01, H-05-01, and a portion of H-04-04 and the other cluster (red) are composed of the remaining subpopulations. To further dissect substructure within the nuclear data, we removed the individuals from H-09-03, H-09-02, H-09-01, H-05-01, and H-04-04 that clustered together in Fig. 2 (green) and analyzed with STRUCTURE. This analysis indicated the presence of three clusters (supplemental information). Based on these results, we selected $\Delta K = 4$ for assignment of individuals to clusters (Fig. 2). Cluster one (yellow) is composed H-11-01, H-08-04, H-06-07, H-06-04, H-06-02, H-06-01, H-04-05, and H-02-01. The second cluster (blue) is composed of H-09-03, H-09-02, H-09-01, and H-05-01. Cluster three (green) is composed of H-08-07, H-08-06, H-08-04, H-08-03, H-07-03, H-07-02, H-07-01, H-06-05, and H-12-04. The fourth cluster (red) is composed of H-12-06, H-04-04, H-03-01, H-01-06, and H-01-02. Although admixture is evident among all Hiwassee River subpopulations, it is higher within H-08-04, H-07-03, H-12-06, H-12-04, H-04-05, and H-04-04. The STRUCTURE analysis identified three clusters for the Ocoee River populations (Fig. 3). Subpopulations O-06-01 and O-03-01 were in one cluster (red), O-05-01 and O-04-01 grouped in another cluster (blue), whereas O-02-03, O-02-02, O-02-01, and O-01-01 clustered into a third group (green). Although admixture is evident among all Ocoee River subpopulations, it is highest within O-06-01, O-03-01, O-02-03, and O-01-01.

The STRUCTURE results using the chloroplast microsatellites differ from the nuclear microsatellites. When all populations are combined, $\Delta K = 6$ (Fig. 1) and admixture is more apparent than in the nuclear data. Clusters one and two include only subpopulations from the Hiwassee River and exhibit very little admixture, whereas clusters three, four, five, and six exhibit a great deal of admixture and are difficult to distinguish. Cluster one (green) is composed of H-07-03, H-07-02, H-06-05, and H-12-06. Cluster two (red) is composed of H-11-01, H-06-01, and portions of H-06-02 and H-06-04. Cluster three (yellow) includes individuals from both rivers, and is composed of O-05-01, O-04-01, H-09-03, H-

09-01, H-12-04, and H-01-02. Clusters four (blue), five (indigo), and six (magenta) exhibit high levels of admixture and split several subpopulations including O-01-01, H-08-07, H-08-06, H-08-04, H-07-01, H-06-07, H-04-04, and H-01-06. Cluster four in the combined analysis using chloroplast data and cluster three in the analysis of Hiwassee alone show strong similarities, as do clusters one of the combined and cluster four of Hiwassee. Subpopulations H-07-03 and H-04-04 show high admixture in both nuclear and chloroplast microsatellites.

Analysis of the Hiwassee River subpopulations using chloroplast microsatellites revealed ΔK = 4 (Fig. 4). Cluster one (green) is composed of H-11-01, H-06-04, H-06-02, and part of H-02-01. Cluster two (yellow) is composed of H-09-03, H-09-01, H-09-02, H-07-03, H-12-04, H-04-04, H-03-01, H-02-01, H-01-06, and H-01-02. Cluster three (red) includes H-08-07, H-08-06, H-08-04, H-08-03, H-06-07, H-05-01, and H-04-05. Cluster four (blue) is composed of H-07-02, H-06-05, and H-12-06, with very little admixture shown. Although admixture is evident among all subpopulations, it is higher within H-07-03, H-06-04, H-06-02, H-04-04, H-03-01, and H-02-01. The STRUCTURE analysis identified two clusters for the Ocoee subpopulations (Fig. 5). Cluster one (green) includes O-06-01, O-03-01, O-02-03, O-02-02, O-02-01, and O-01-01. Cluster two (red) is composed of O-05-01 and O-04-01, which also clustered together in cluster one of the combined analysis. Admixture is low throughout both clusters, with only a few scattered individuals evidencing crossover between clusters. When both chloroplast and nuclear microsatellites were analyzed together, both coded as diploid, two clusters were detected and the subpopulations separated by river (supplemental information).

Additionally, we randomly subsampled all subpopulations over 25 individuals to ensure the accuracy of our STRUCTURE results with more uniform sampling size, as the program has been considered unreliable when uneven sampling occurs (Puechmaille 2016). In the combined data set, we found Δk = 3 and 5 for the Hiwassee River when subsampled and Δk = 6 for the Ocoee River when subsampled (supplemental information), differing from the previous analysis and detecting additional

clusters. However, subpopulations clustered similarly between the subsampled and original data, with individuals in smaller populations showing higher levels of admixture but the main clusters remaining the same.

The pairwise correlation between geographic and genetic distance to determine isolation by distance using nuclear data shows a clear and significant ($P < 0.01$) separation the Hiwassee and Ocoee Rivers (Fig. 6A). The separation is also noticeable between rivers when using the chloroplast data from both rivers (Fig. 7A). Isolation by distance is less apparent when the Hiwassee and Ocoee River subpopulations are separated, though a significant positive relationship is apparent between geographic distance (km) and genetic distance in both data sets (Fig. 6B and 6C; Fig. 7B and 7C).

Wilcoxon tests to detect bottlenecks in *P. ruthii* subpopulations showed that cluster one and cluster four of the Hiwassee River (Fig. 2, k = 4) had significant signs of a recent genetic bottleneck using all three models (Table 10). The infinite allele model showed likely recent bottlenecks and loss of genetic diversity in all of the Hiwassee River clusters, as well as two of the Ocoee River subpopulations ($P < 0.05$). BOTTLENECK did not detect a bottleneck in the third cluster of the Ocoee River (Fig. 3) using any of the models.

Discussion

In general, *Pityopsis ruthii* is characterized by high levels of variation at nuclear microsatellite loci (Tables 2 and 3) and moderate levels of variation for chloroplast microsatellite loci (Tables 7 and 8). Expected heterozygosity is considerably higher ($H_E = 0.63$) than that found in endemic cliff dwelling perennial species (*Opisthopappus longilobus*; $H_E = 0.20$ and *O. taihangensis*; $H_E = 0.14$) from China (Guo et al. 2013). Within the Hiwassee River populations, slightly higher inbreeding coefficients (F_{IS}) were found in the peripheral populations and suggest a higher degree of inbreeding in these populations compared to the central populations, though analyses of chloroplast microsatellites did not show a lack of diversity in those same populations.

Genetic differentiation (F_{ST}) of the Ocoee River populations is higher (0.19) than Hiwassee River populations (0.15) using nuclear data and similar those found in *O. taihangensis* (Guo et al. 2013). However, F_{ST} calculated from chloroplast data is higher for both the Ocoee River populations (0.22) and the Hiwassee River populations (0.37). Chloroplast microsatellites often show high F_{ST} values, and though we found greater genetic differentiation in chloroplast microsatellites than nuclear, our F_{ST} = 0.37 is much smaller compared to that found in *Begonia* species (F_{ST} = 0.73)(Twyford et al. 2013). However, the seven loci had a higher number of alleles per locus and allelic frequency than loci used to study *Chrysanthemum indicum* and *C. lavandulifolium* (Yang et al. 2006), and equivalent to or greater than chloroplast loci used for *Begonia nelumbiifolia* and *B. heracleifolia* (Twyford et al. 2013). Regarding the proportion of diversity among rivers, the F_{ST} values for both rivers indicate high genetic differentiation among subpopulations. Subopulations with greater average pairwise differences have more genetically variable individuals than populations with lower average pairwise differences. Wright (1931) suggested when gene flow >1, genetic differentiation among populations due to genetic drift can be prevented. Reduced gene flow can be expected to increase inbreeding within populations. However, in general, gene flow estimates greater than 0.5 indicate that migration is adequate to prevent genetic divergence of populations due to drift (Slatkin 1987). All subpopulations along both rivers have gene flow estimates greater than 0.5, and are therefore not in immediate danger of genetic drift causing divergence among populations. Pairwise comparisons of gene flow estimates for subpopulations on the Ocoee River are in general much lower than those subpopulations on the Hiwassee River, as expected due to the much lower number of individuals.

We found significant (P>0.001) F_{IS} values for all subpopulations regardless of river indicating an excess of homozygosity, which indicates inbreeding. Inbreeding in the case of *P. ruthii* could be attributed to mating among relatives, which could lead to lower seed viability or seedling vigor.

Recent bottlenecks were detected in multiple clusters, indicating a loss of genetic diversity, which could impact adaptation. Cluster three of the Ocoee River from the nuclear SSR data (Fig. 3; O-02-03, O-02-02, O-02-01, and O-01-01) showed no signs of a bottleneck, perhaps due to the inclusion of four sampling sites, two of which (O-01-01 and O-02-03) show high admixture with the nuclear data. Coupled with the higher levels of gene flow among these subpopulations, high admixture in the two subpopulations furthest upstream, and the possibility of higher water flows at these upstream subpopulations, the lack of bottlenecking may indicate a founder effect. The four sampling sites that make up cluster three are located upstream from the other sites along the Ocoee, allowing little gene flow from the downstream subpopulations, but also showing little gene flow from cluster three to other locations, resulting in bottlenecks downstream. Considering the geographical proximity to one another, the out-crossing breeding system, and habitat continuity we should expect genetic exchange among populations through pollen or seed dispersal. However, asynchronous flowering within subpopulations could compound inbreeding and explain the high gene diversity (H_E) and high inbreeding coefficient (F_{IS}) observed. Variation of flowering times is not uncommon; flowering starts as early as late July to early August and continues until late October to early November. Mating of individuals from each group could lead to increased inbreeding, explaining the inbreeding coefficients within subpopulations.

Additionally, drought-tolerance appears to be a factor in maintaining diversity for some subpopulations of *P. ruthii*. Subpopulation H-04-04 has the highest number of private alleles for both the nuclear and chloroplast loci, and is considered a drought-maintained subpopulation due to the lack of other drought-tolerant species to compete for resources (Moore et al. 2016). This drought-tolerance at H-04-04 seems to allow diversification from other sites, leading to higher number of private alleles and low F_{ST} values (Table 5). High admixture is apparent in this subpopulation in both the nuclear and chloroplast datasets, which could indicate recruitment to this site from other populations, since conditions are more favorable for *P. ruthii*.

Fewer alleles and lower overall diversity was seen in chloroplast microsatellites when compared to nuclear microsatellites. This was expected due to the lack of recombination and smaller effective population size of chloroplast loci. The clear admixture between rivers in the STRUCTURE results from the chloroplast data is not present in the nuclear data. Coupled with the AMOVA results indicating a lack of variance between the Ocoee and Hiwassee River subpopulations, the differences between chloroplast and nuclear data are evident. One possibility is that the two populations of *P. ruthii* were once a single, widespread population that fragmented due to unknown causes, leading to higher levels of genetic drift in the genome and lower levels in the more conserved plastome. If true, it is not unreasonable to expect a more widespread range of *P. ruthii* possible again in the future.

Considering that this study is a single snapshot of the genetic variation for these subpopulations, determining whether genetic variation is increasing, decreasing, or stable is difficult. Pollen and seed dispersal are the main mechanisms for natural gene flow. In order to evaluate the results of our study, we need to take into consideration what is known about pollen and seed dispersal of *P ruthii*. Seed distribution is thought to be adapted for water dispersal or rolling around on the rock substrate until a seed is lodged into a suitable crevice or blow into the water by the wind (Clebsch and Sloan 1993). Germination of seedlings in wild populations has been observed and the mortality was higher than 90% after 1 year (Clebsch and Sloan 1993). Further studies are needed to determine if seedlings at *P. ruthii* subpopulations with high inbreeding coefficients are suffering from inbreeding depression or if seedling recruitment is limited.

Dams on the Hiwassee and Ocoee Rivers have altered the hydrology of both rivers where *P. ruthii* occur, though it is not clear what effect damming of the river and augmented flows have on seed recruitment. The lack of information on seed recruitment and habitat loss, coupled with high mortality of seedlings within natural populations pose challenges to developing strategies to protect sustainability of these populations. Another scenario is that the detected gene flow levels at least in part reflect

natural gene flow. The fringe subpopulations at the edges of the Hiwassee River and all of the subpopulations on the Ocoee River may reflect this as they are more isolated than other subpopulations. These populations may have somewhat restricted gene flow being situated too far from most other populations for effective pollen or seed exchange, making them more distinct.

Currently, each discrete subpopulation is managed as a separate population (Adam Dattilo, personal communication), though our study shows no evidence for such fragmentation. Rather, we posit that multiple locations have similar genetics and therefore can be managed as one larger population. This is especially evident when viewing data from the nuclear microsatellite markers. Managing the species using a framework with four populations along the Hiwassee River and three along the Ocoee River, as defined by nuclear data clustering (Fig. 2 and 3), will allow researchers to use plants in larger placement areas for augmentation, reintroduction, and/or translocation studies to add diversity to a particular population.

Molecular markers are commonly used in maintenance of germplasm collections, and chloroplast microsatellites in particular have been used to much advantage in several species (Balas et al. 2014). A germplasm core collection should include the majority of diversity without excessive redundancy, which we can now access using the frequency of private alleles detected using the microsatellites outlined in this study. The North Carolina Botanical Garden currently curates the *P. ruthii* germplasm accessions (Michael Kunz, personal communication).

Ongoing pollination studies and reintroduction efforts add to the effort base of knowledge on the ecology of this endangered species (Wadl et al. 2014, Moore et al. 2016). With the habitat topography of *P. ruthii*, surrounded by high ridges and ineffective seed dispersal mechanisms, the species may not be able to migrate with warming climates. Additionally, the rivers show differing levels of population expansion, with subpopulations along the Ocoee River exhibiting greater numbers of flowers per plant and a lower level of competition with other herbaceous and woody plants (Moore et

al. 2016). Moore et al. (2016) hypothesize that cyclical drought also plays a role in maintaining the rocky habitat necessary to support subpopulations of *P. ruthii*. Along with understanding the ecology of the species, a viable method of introducing diversity into the natural habitat is necessary. Cultivation of plants is possible through both stem cuttings and tissue culture, providing methods for reintroduction studies; it is feasible to grow *P. ruthii in vitro* and transplant seamlessly into the natural habitat (Wadl et al. 2014).

Coupled with the molecular markers in this diversity study and the information now available on population structure, preventing further loss of diversity and protecting *Pityopsis ruthii* is possible with adequate management and augmentation studies that take into account the genetically distinct populations. Additionally, the genus *Pityopsis* is useful in studying inheritance of adaptive traits such as fire-dependent flowering and drought tolerance due to the wide range of habitats of different species, as well as the presence of certain traits in one species of the genus and its concurrent absence in another species (Gowe and Brewer 2005). Studies of the genus could also provide insight into polyploidy, evolutionary history, and interspecific hybridization between *P. ruthii* and *P. graminifolia* var. *latifolia*, which often grow close together (Moore et al. 2016). Tracking demography within populations as well as further work on seed dispersal mechanisms and breeding success would be useful in understanding and protecting this endangered species. Our work in identifying conservation units as genetically distinct populations for this species and understanding the underlying genetics of the species will inform conservation practices in the future, as well as further study into the entire genus *Pityopsis*, and has provided a relatively easy and cost-effective way to follow genetic diversity in existing, augmented, or reintroduced populations over time.

Chapter 3. Comparing chloroplast genomes in *Pityopsis* species (Asteraceae)

Division of Labor among Co-authors

This chapter was produced in collaboration with my committee members and other scientists and will be published in a peer-reviewed journal. My co-authors include Timothy A. Rinehart, Thomas S. Lane, Edward E. Schilling, Margaret E. Staton, Robert N. Trigiano, and Phillip A. Wadl.

Drs. Schilling and Trigiano gave intellectual guidance and helped with experimental design. Dr. Staton helped with analysis of data, as well as writing and interpretation. Dr. Wadl guided and helped with analytical, technical, and interpretative methods. Dr. Rinehart provided immense technical support. Mr. Lane provided analytical guidance. I designed the experiment, analyzed and interpreted the results, presented data appropriately, and wrote the chapter.

Abstract

Pityopsis includes several regionally and one federally endangered species of herbaceous perennials. Although four species are highly localized, three species are found throughout the eastern United States and the range of one extends into Mexico. Morphological studies have separated the genus into two distinct clades, but there have been few molecular studies and intergeneric relationships have not been fully resolved or understood. For this study, six species and four varieties were collected from the wild or obtained from herbaria vouchers, and the whole chloroplast genome was sequenced. A reference chloroplast genome was assembled *de novo* from the species with the highest depth of read coverage, *Pityopsis falcata*. Reads from the other individuals were then aligned to the *P. falcata* reference genome and an individual reference genome was assembled for each. To utilize all informative sites from the full length of the chloroplast, a multiple sequence alignment of the eight chloroplast genomes was constructed, and from this, a phylogeny using the maximum likelihood method. Using the entire chloroplast genomes we found no evidence for clades or taxonomic sections that have been previously proposed within the genus. This study will help inform breeding and conservation practices, as well as general knowledge of evolutionary history, hybridization, and speciation within *Pityopsis*.

Introduction

One of the largest and most wide-spread plant families, Asteraceae (Compositae), contains over 20,000 species distributed on all continents but Antarctica (Panero and Funk 2008). Polyploidization has been well-documented in Asteraceae, with all tribes showing evidence of duplication of the basal Compositae genome (Barker et al. 2008). Such polyploidization events, both ancient and recent, drive speciation (Vamosi and Dickinson 2006). The genus *Pityopsis* is a member of Asteraceae, in tribe Astereae, with a wide range of ploidy levels across species and a large range throughout southeastern North America. *Pityopsis* has been the subject of several phylogenetic studies (Gowe and Brewer 2005, Teoh 2008), but intrageneric relationships for all species and varieties in the genus have not been fully resolved. The genus includes many polyploid varieties and several interesting traits such as fire-stimulated flowering and drought-tolerance, and understanding species relationships will allow inferences about the evolution of such traits.

Phylogenetic studies are conducted to clarify taxonomic relationships and classification (Wan et al. 2004). They have proved useful for understanding plant-pathogen interactions (Gilbert and Webb 2007) and community ecology (Vamosi et al. 2009). Additionally, phylogenetic studies can translate to predictions of phenological response and adaptation in related species, especially adaptation in regard to climate change (Hoffmann et al. 2015). Phylogenies have additional use in studies focused on evolutionary history (Heuertz et al. 2006, Byrne 2007). *Pityopsis* is an excellent candidate for such analysis as the genus includes species with and without such traits as fire-adaptive flowering, as well as species with varying ploidy levels (Teoh 2008). In *Pityopsis,* species distinctions are not well understood and require further resolution, which has been difficult due to the differing ploidy levels in the genus. For example, in *P. graminifolia* alone there are three ploidy levels: diploid (*P. graminifolia* var. *graminifolia*), tetraploid (var. *latifolia*), and hexaploid (var. *aequilifolia*), present in different varieties of the species (Semple 2006). The wide range of ploidy levels creates difficulties in analyses using

biparental nuclear markers. However, with a well-supported phylogeny based on molecular markers, *Pityopsis* could be used to examine the evolution of adaptive traits and the role of hybridity in the evolution of polyploidy.

Nuclear microsatellites have been developed for two different *Pityopsis* species and chloroplast microsatellites have been developed for one species (Wadl et al. 2011a, Boggess et al. 2014). However, whole chloroplast (cp) genomes are lacking for all species in the genus. With the availability of next-generation sequencing, phylogenetic studies using entire cp genomes is becoming more reliable and common, especially for closely related species (Parks et al. 2009). Chloroplast genome sequences have become a convenient way to find repetitive sequences and single nucleotide polymorphisms (SNPs) that could be used for further ecological and evolutionary studies, as well as clarifying taxonomy in general with muddled history (Huang et al. 2014). Many similar studies have been conducted on phylogenetic relationships within economically important plants, such as wheat, rice, and maize (Matsuoka et al. 2002), strawberry (Njuguna et al. 2013), and cotton (Xu et al. 2012). Using cp genomes to analyze the species relationships within *Pityopsis* allows further studies regarding past polyploid events using a simplified system due to the haploid nature of chloroplasts, though only the maternal line is revealed in the case of species arising from hybridization events.

Pityopsis includes seven species: *P. aspera* (Shuttlew. ex Small) Small, *P. falcata* (Pursh) Small, *P. flexuosa* (Nash) Small, *P. graminifolia* (Michx.) Nutt., *P. oligantha* (Chapm. ex Torr. & Gray) Small, *P. pinifolia* (Ell.) Nutt., and *P. ruthii* (Small) Small (Semple 2006). Both *P. aspera* and *P. graminifolia* have multiple varieties, some of which have previously been recognized as separate species (Clewell 1985). *Pityopsis* is endemic to the eastern United States, and though *P. graminifolia* and *P. aspera* have a large range, other species in the genus are more localized, such as *P. ruthii* and *P. flexuosa.* All species are perennial and have yellow inflorescences, as indicated by the common name for plants in the genus, goldenaster (Semple and Bowers 1987).

The division of *Pityopsis* into sections remains unclear. According to Semple and Bowers (1985), the genus is divided into two sections: section *Pityopsis* with *P. falcata, P. flexuosa, P. pinifolia,* and *P. ruthii,* and section *Graminifoliae* with *P. aspera, P. graminifolia,* and *P. oligantha*. However, Gowe and Brewer (2005) posited that the genus had two clades, Falcata, which includes *P. falcata*, *P. flexuosa*, *P. graminifolia*, *P. pinifolia*, and *P. oligantha*, and Aspera, which includes *P. aspera*, *P. adenolepis,* and *P. oligantha*. Their phylogeny was constructed based on fire-dependent flowering and other morphological traits. In contrast, a molecular study utilized sequences from chloroplast and nuclear regions of all seven species and concluded that two new clades should be named: Ruthii and Flexuosa (Teoh 2008). Clade Ruthii includes *P. falcata, P. pinifolia, P. ruthii*, and *P. graminifolia* var. *latifolia.* Splitting the species *P. graminifolia*, clade Flexuosa includes *P. graminifolia* var. *aequilifolia*, *P. graminifolia* var. *tenuifolia*, and *P. graminifolia* var. *graminifolia*, as well as *P. aspera, P. adenolepis*, and *P. oligantha*. Both the 2005 and the 2008 studies include *Pityopsis adenolepis* as a separate species as per Clewell (1985), though it is currently considered a variety of *P. aspera* (Semple and Bowers 1985). With little to no consensus between morphological and molecular studies, an in depth study of the taxonomy of the genus is warranted. In this study, eight *Pityopsis* chloroplast genomes were assembled, compared to other Asteraceae chloroplast genomes, and used to construct a phylogenetic tree, which did not support previous divisions of the genus into clades previously proposed.

Materials and Methods

Library construction and sequencing

Leaf tissue of seven species of *Pityopsis* was collected from the southeastern United States (Table 10). Samples from *Pityopsis aspera* var. *aspera* and *P. pinifolia* were obtained from herbarium vouchers Kral 56861, Bowers 45553, and Bowers & Wofford 71562, stored at the University of Tennessee Herbarium (TENN). Leaf tissue from plants maintained in a greenhouse at the University of Tennessee was collected for *P. graminifolia* var. *tracyi*. This study used tissue collected in 2010 and 2013

and kept at -80 ^{o} C from *P. ruthii* and *P. graminifolia* var. *latifolia*, respectively. *Pityopsis aspera* var. *adenolepis, P. graminifolia* var. *tenuifolia*, and *P. graminifolia* var. *graminifolia* were collected from South Carolina in 2014. For *P. oligantha*, *P. graminifolia* var. *aequilifolia*, and *P. flexuosa*, tissue was collected in 2015 from Florida. Vouchers are available at the Florida State University Herbarium (FSU) for *P. oligantha* and *P. flexuosa* (Anderson 28905 and Anderson 28533, respectively).

Total genomic DNA was isolated using a DNeasy Plant Mini Kit (Qiagen, Valencia, CA) following manufacturer's protocol. *Pityopsis ruthii* was sequenced using the Ion Torrent Personal Genome Machine (ThermoFisher). A 400 bp library was constructed and single end reads were output. Genomic DNA of the other six species was cleaned and concentrated using the Zymo Genomic DNA Clean and Concentrator Kit (Zymo Research Corp., Irvine, CA). The libraries were prepared using the Nextera DNA Library Preparation Kit (Illumina, San Diego, CA). DNA was fragmented using transposase-mediated tagmentation and paired end sequenced using dual indexes. The Illumina MiSeq version 3 sequencing platform (Illumina, San Diego, CA) was used to sequence the DNA. Three libraries were pooled for three runs, and four pooled on one run.

Sequence trimming and alignment

The sequence quality of all sequences was checked using FastQC (Andrews 2010) for kmer content, GC content, and average length of reads. Adaptors and low quality ends were trimmed using Trimmomatic v. 0.35 (Bolger et al. 2014). After trimming, quality was assessed again using FastQC, which showed that overall quality improved in all individuals. Using the program Bowtie2 (Langmead and Salzberg 2012), the data from all individuals was aligned against the chloroplast genome of *Helianthus annuus*, which was downloaded from NCBI (GenBank: DQ383815.1; downloaded November, 2015). *Pityopsis falcata* had the highest number of mapped reads and was selected for *de novo* assembly of a reference cp genome.

Genome assembly and annotation

Mapped reads from *Pityopsis falcata* were used to create a reference cp genome using the program ABySS v 1.5.2 (Simpson et al. 2009), which is designed for short, paired-end reads. Gaps within the draft genome were closed using GapFiller v. 11 (Boetzer and Pirovano 2012), which uses an automated strategy to fill gaps within scaffolds. Gaps between contigs were closed using all *P. falcata* reads and default parameters, with the exception of the minimum number of bp allowed to overlap (150, default is 50), the percentage of reads that should have single nucleotide extension in order to close a gap (0.3, default is 0.7), and the maximum difference between gap size and number of nucleotides closed within the gap (150, default is 50). The reference genome from *P. falcata* was annotated using DOGMA (Wyman et al. 2004), which is specific to organelle genomes and also identifies tRNAs and rRNAs. The annotations were manually reviewed and edited using the JBrowse (Skinner et al. 2009) plug-in Web Apollo (Lee et al. 2013). Visualization of the genome annotation was created using the program GenomeVx (Conant and Wolfe 2008).

Alignment and comparison

Using CLC Genomics Workbench (Qiagen, Redwood City, CA, USA), reads from each species and variety were mapped to the *P. falcata* reference with a linear gap cost and length fraction of 0.5, and consensus sequences were extracted to serve as complete cp genomes of their respective species. The cp genomes were then aligned using the default settings of a gap open cost of ten and a gap extension cost of one. A maximum likelihood phylogeny was reconstructed in MEGA7 (Kumar et al. 1994, Kumar et al. 2016) using the Kimura 2-parameter method. Bootstrap analysis was conducted using 400 replicates. A maximum parsimony tree was reconstructed using the min-mini heuristic model in MEGA7 with bootstrap values calculated using 400 replicates. Using MEGA7 (Kumar et al. 1994, Kumar et al. 2016), pairwise differences were calculated between all eight *Pityopsis* cp genomes and the *Aster spathulifolius*

outgroup. Single nucleotide variants (SNVs) between cp genomes and the *P. falcata* reference were called using SAMtools 1.2.2 (Li et al. 2009).

Results

Chloroplast genome sequencing and assembly

Using the Illumina MiSeq sequencing platform, we sequenced total genomic DNA and assembled cp genomes of five species and three varieties from *Pityopsis*, and using the Ion Torrent sequencing platform, we sequenced one species of *Pityopsis* to assemble a cp genome. All individuals sequenced using the Illumina platform yielded paired end reads, while *P. ruthii* had single end reads. Illumina paired-end sequencing produced from 3,451,455 (*P. oligantha*) to 33,339,900 (*P. graminifolia* var. *aequilifolia*) reads per individual (Table 11). Of these reads, 6,571 (*P. graminifolia* var. *aequilifolia*) to 1,407,627 (*P. graminifolia* var. *tracyi*) reads mapped to the *Helianthus annuus* reference cp genome, with 5-189 x coverage (Table 3). The single species sequenced using the Ion Torrent platform, *P. ruthii,* had the highest percentage of mapped reads, with 169X coverage (Table 12). *P. falcata* had the highest number of basepairs mapped to the *Helianthus* reference.

The reference *Pityopsis* cp genome is a single, circular chromosome, with a large single copy (LSC), small single copy (SSC), and two inverted repeat regions (IR) (Figure 8). The *P. falcata* reference was 145,335 bp in length; the LSC was 79,227 bp in length, the SSC was 18,174 bp in length, and the two IRs were 23,966 bp in length. 112 genes were identified, of which 26 were transfer RNA (tRNA) genes, 4 were ribosomal RNA (rRNA) genes, and 82 were protein-coding genes (Table 13). All the rRNA genes were found in the IR regions.

Upon running the entire *P. falcata* reference cp genome through NCBI BLAST, *Aster spathulifolius* (GenBank: KF279514.1) was identified as the organism with the most similar cp genome. Pairwise bp differences were calculated for all eight cp genomes and the closely related outgroup, *Aster spathulifolius.* When compared to the *P. falcata* reference, *P. ruthii* had the highest differentiation with

369 SNVs, while *P. oligantha* had the lowest, with 85. We included a single inverted repeat in the MEGA7 alignment and pairwise analyses. The number of SNVs ranged from 42 (*P. graminifolia* var. *latifolia* vs. *P. graminifolia* var. *tracyi*) to 358 (*P. flexuosa* vs. *P. graminifolia* var. *tracyi*)(Table 15). *P. flexuosa* showed the highest number of SNVs overall, with a total of 2,263 single base mutations from other *Pityopsis* cp genomes. *P. flexuosa* also had the highest number of SNVs when compared to *Aster spathulifolius. P. graminifolia* var. *tracyi* had the fewest SNVs overall, with only 912 single base mutations when compared to all other *Pityopsis* cp genomes (Table 15). *P. graminifolia* var. *tracyi* also had the fewest mutations when compared to *Aster spathulifolius* (2173), though *P. graminifolia* var. *latifolia* displayed only one SNV more (with 2174).

Phylogenetic analyses

When the full cp genomes of all *Pityopsis* individuals were aligned, pairwise comparison showed *P. flexuosa* to be the most distinct from all other species and varieties. The least number of mutations between sequences was observed between *P. graminifolia* var. *aequilifolia* and *P. falcata*; the largest differentiation was seen between *P. flexuosa* and *P. ruthii* (Table 14). All sequences exhibited over 99% similarity with one another in the pairwise comparison (Table 14). The complete cp genome of *Aster spathulifolius* (GenBank: KF279514.1; downloaded February 2016) was used as the outgroup for maximum likelihood (ML) and maximum parsimony (MP) phylogenetic reconstructions. Bootstrap analyses were conducted using 400 replicates. All bootstrap values (BS) were over 40% for the MP tree (Fig. 9), though branching within the ML tree was not as well supported (Fig. 10). The placement of *P. graminifolia* var. *latifolia* was not well supported on either the ML or the MP tree (BS = 41% for both). Both the ML and MP phylogenies showed maximum support (BS = 100) for the separate branching of *P. aspera* var. *adenolepis*, *P. flexuosa*, and *P. oligantha* from the rest of the cp genomes. Separation between *P. graminifolia* var. *aequilifolia*, var. *tracyi, P. falcata,* and *P. ruthii* was moderately supported

using the maximum parsimony approach (BS > 50%), but was less reliable using the maximum likelihood approach (BS < 40%).

Discussion

Comparison of chloroplast genomes

All eight complete *Pityopsis* cp genomes displayed attributes common among angiosperm cp genomes, with quadripartite structure including the LSC, SSC, and a pair of inverted repeats (IRa and IRb). No inversions or genome rearrangements were apparent in the *Pityopsis* cp genome when compared to other Asteraceae species. The length of the *Pityopsis* cp genome (145,355) was 4,000-5,000 bp shorter than seen in other Asteraceae species such as *Aster spathulifolius* (Choi and Park 2015) and *Jacobaea vulgaris* (Doorduin et al. 2011). The length of the two IRs in *Pityopsis* (23,966) was shorter than that of *J. vulgaris* or *A. spathulifolius*, as was the LSC (79,227), but the SSC was larger than found in *A. spathifolius* and within 100 bp of that found in *J. vulgaris* (Doorduin et al. 2011, Huang et al. 2014).

Asteraceae cp genomes contain approximately 114 genes according to Wang et al. (2015). The number of genes identified from the *Pityopsis* cp genome, 112, is slightly lower, though consistent with other Asteraceae species such as *Chrysanthemum* × *morifolium* (Wang et al. 2015). When including genes duplicated in the IRs, 131 genes were identified. This is not the largest number identified in the Asteraceae; it is lower than the number of genes (including duplicates) found in *J. vulgaris* by four genes (Doorduin et al. 2011). Functional groups of genes (Table 13) were all appropriately represented in the *Pityopsis* cp genomes as compared to those of *A. spathulifolius* (Choi and Park 2015). The *Pityopsis* cp genome included 26 tRNA genes, slightly lower than the number found in *A. spathulifolius* (29) and *J. vulgaris* (29). Within Asteraceae, 29 tRNA genes per cp genome is average (Timme et al. 2007, Wang et al. 2015). When compared to other Asteraceae species, the *Pityopsis* cp genome was missing the transfer RNA genes *trnH-GUG*, *trn-T-UGU*, and *trnG-UUC* as well as the protein-coding gene *psbG*. The number of rRNA genes found in the IR of *Pityopsis* is consistent with the number found in several other

Asteraceae species, including *A. spathulifolius* (Choi and Park 2015), *H. annuus* (Timme et al. 2007), *J. vulgaris* (Doorduin et al. 2011), and *Lactuca sativa* (Timme et al. 2007). Four rRNA genes in each IR is typical of Asteraceae (Wang et al. 2015). The *ycf1* and *ndhH* genes did not overlap, consistent with *Helianthus annuus* and other species within Heliantheae, rather than overlapping as seen in Astereae species such as *A. spathulifolius* (Choi and Park 2015). Additionally, the *ycf15* gene was present in *Pityopsis* cp genomes, a phenomenon that distinguishes *Helianthus annuus* from *Chrysanthemum indicum*, *C.* × *morifolium*, and *Guizotia abyssinica*, in which *ycf15* is absent (Wang et al. 2015).

Pairwise comparisons and phylogenetic analyses of *Pityopsis* species

Maximum likelihood and maximum parsimony phylogenetic trees were reconstructed for eight species or varieties within *Pityopsis* (Fig. 9 and 10). The close relationship between *P. flexuosa* and the varieties of *P. graminifolia* seen in by Teoh (2008) was not evident when the whole cp genome was used for phylogeny construction. Our findings are not consistent with the sections of *Graminifoliae* and *Pityopsis* proposed by Semple and Bowers (1985).

The maximum likelihood tree (Fig. 10) and the maximum parsimony phylogeny (Fig. 9) provide strong support for distinction of *P. aspera* var. *adenolepis*, *P. flexuosa*, and *P. oligantha* from other *Pityopsis* species, though support for other clades was low in the maximum parsimony tree. The relationship between *P. aspera* var. *adenolepis* and *P. oligantha* in both phylogenies lends some credence to the Aspera clade from the morphological phylogeny of Gowe and Brewer (2005), though the placement of *P. flexuosa* is not consistent with their findings. The close relationship between *P. oligantha* and *P. flexuosa* is worth noting, as the range of *P. oligantha* overlaps with the smaller habitat of *P. flexuosa*. These two species show a moderate percentage of similarity in their cp genomes as compared to other species in the genus (Table 14) and are differentiated by 273 SNVs. *P. ruthii* was more closely related to all varieties of *P. graminifolia* than seen by Teoh (2008). Our findings were inconsistent with Teoh's clade Ruthii.

Interestingly, *P. falcata* grouped closely with the three varieties of *P. graminifolia*, though the outgrouping of *P. graminifolia* var. *aequilifolia* was only moderately supported (BS > 50%). The ploidy levels of *P. graminifolia* var. *aequilifolia*, *P. graminifolia* var. *latifolia*, and *P. graminifolia* var. *tracyi* are diploid (2n = 18), tetraploid (2n = 36), and hexaploid (2n = 54), respectively (Semple 2006). A study using 12 nuclear microsatellites developed from *P. graminifolia* var. *latifolia* showed cross-transferability of 7 loci to *P. ruthii* and 6 loci to *P. falcata* (Boggess et al. 2014), supporting the close relationship between the three species evidenced by the ML phylogeny. The *P. falcata* cp differs from *P. graminifolia* var. *tracyi*, *latifolia*, and *aequilifolia* by 49, 63, and 70 SNV sites, respectively (Table 15). Varieties of *P. graminifolia* have few sequence differences: *P. graminifolia* var. *latifolia* and *P. graminifolia* var. *tracyi* differ from each other at only 42 sites, whereas *P. graminifolia* var. *aequilifolia* is slightly more polymorphic, differing from *P. graminifolia* var. *latifolia* and *tracyi*, at 65 and 61 sites, respectively (Table 15). These close relationships within *P. graminifolia* uphold the varieties as part of the same species. The lack of differentiation between *P. graminifolia* var. *aequilifolia*, var. *tracyi*, and *P. falcata* is supported by the branching of both phylogenetic trees derived from their alignment, and encourages further study into the relationships between the two species.

Pityopsis ruthii and *P. graminifolia* var. *latifolia* grow alongside one another throughout the range of *P. ruthii,* and successful controlled hybridization has been confirmed using microsatellites (Boggess et al. 2014). Their placement in the two phylogenic trees differs, but neither tree indicates that they are closely related. In both trees, *P. ruthii* is more closely related to *P. falcata* and *P. graminifiolia* var*. aequilifolia*, which do not share habitat with the endangered species. The Flexuosa and Ruthii clade system proposed by Teoh (2008) separated the *P. graminifolia* varieties and is inconsistent with our findings, especially in respect to the placement of *P. flexuosa.* Our two phylogenetic trees offer only slight support to the morphologically distinct Aspera and Falcata clades proposed by Gowe and Brewer

(2005), though this distinction within the genus are not well supported based on either our ML or MP phylogenies.

Our phylogenies exhibit a striking difference in tree topology than in previous molecular studies, which placed both *P. falcata* and *P. ruthii* separately from all varieties of *P. graminifolia* (Teoh 2008). Pairwise sequence comparison of the genomes showed the highest number of differences and lowest percent similarity between *P. ruthii* and *P. graminifolia* var. *tracyi*, which is surprising given their close placement in the phylogeny. As chloroplasts are not directly affected by nuclear genome duplication events, nuclear studies may be necessary to elucidate the relationships between varieties of *P. graminifolia*. It is unknown whether *Pityopsis* polyploids are auto- or allopolyploids, though there is some evidence that allopolyploidy is the mechanism of genome duplication in *P. graminifolia* var. *latifolia*, and that polyploid varieties within *Pityopsis* may be allopolyploid hybrids of other species and varieties (Teoh 2008).

The cp genomes from six species of *Pityopsis* will provide information to future researchers interested in the genus. Using the entire cp genome, a well-supported clade of *P. aspera* var. *adenolepis*, *P. flexuosa*, and *P. oligantha* was seen, while other relationships remained unresolved. The variation between chloroplast genomes of *Pityopsis* species provide a mechanism of distinguishing between species and varieties which should be examined in future studies, as well as for understanding diversity within the genus. We have developed whole chloroplast genomes that will allow further study of individual species as well, opening possibilities for future work in chloroplast transcriptomics, furthering knowledge of variable regions within the chloroplast, and providing information for future studies of *Pityopsis* and Asteraceae.

Chapter 4. Conclusion

Our study into *Pityopsis* has used multiple approaches to elucidate the relationships within the genus as a whole and in the endangered species *Pityopsis ruthii*. We have accomplished our goals with our two objectives, which were to study the genetic diversity and population structure within *P. ruthii* and to examine taxonomy within the genus using phylogenetic analyses. For the first objective, we defined conservation units within populations of *Pityopsis ruthii* that inform conservation practices and ensure adequate management of viable populations as deemed necessary in the species recovery plan (USFWS 1992). We have also provided cost-effective tools in the form of microsatellite markers, which will allow scientists to track genetic diversity within these populations over time. Based on our estimates of gene flow within *P. ruthii*, the species is not currently in danger of divergence among populations caused by genetic drift. Reintroduction and augmentation efforts will benefit from the baseline genetic diversity information and provide a reliable way to monitor changes within populations and subpopulations. Chloroplast microsatellites have been used to study hybridity and evolution within polyploid complexes (Yang et al. 2006), as have nuclear microsatellites (Ferriol et al. 2014), and both can be used in future studies to clarify the relationship between *P. ruthii* and *P. graminifolia* var. *latifolia*. Our study will inform further efforts to understand the genetics, ecology, and physiology of *P. ruthii*, and our methods and the results add to the information available on the study and protection of endangered plant species.

For the second objective, we sequenced and assembled chloroplast genomes from six species of *Pityopsis*, which were used to construct two phylogenetic trees. The two trees showed a consensus regarding the close relationship between *P. aspera* var. *adenolepis*, *P. flexuosa*, and *P. oligantha* and separation of these species from others within the genus. Otherwise, our findings did not support previous taxonomical divisions of the genus. Assembling whole chloroplast genomes enabled variant calling and pairwise comparison between species, providing further tools to study variation within the genus. Species within *Pityopsis* have been included in ecological studies of fire-dependent flowering

(Heuberger and Putz 2003, Gowe and Brewer 2005), sandplains communities (Farnsworth 2007), and sandhill pine communities (Provencher et al. 2001, Sharma et al. 2012). Understanding relationships within *Pityopsis* will also allow researchers to better understand and compare studies using *P. aspera* and *P. graminifolia*, as both have multiple varieties and are often found in sandhill pine and fire-maintained communities. We have provided a foundation by developing genomic resources to further study community ecology within such communities, as well as supplement physiological studies of fire-dependent flowering. Additionally, availability of chloroplast genomes encourages future work on chloroplast transcriptomics, and physiology of photosynthesis and energy production within species of *Pityopsis.* Knowledge of variable regions within *Pityopsis* chloroplasts adds to the knowledge of Asteraceae as a whole, and provides valuable information for future studies of *Pityopsis*.

Though *Pityopsis* is currently an understudied genus, it has much to offer in the way of discovery. Understanding the evolutionary history and current diversity of *Pityopsis* will enable more thorough studies of the genus, and our work has contributed to that aim. Physiology of the genus, especially in regard to drought-tolerance, will be of interest and could illuminate mechanisms of survival of *P. ruthii*. Further work on hybridity in the genus, particularly between *P. ruthii* and *P. graminifolia* var. *latifolia* is warranted. As always, additional knowledge waits for those who take the time to explore, and researchers will not be lacking in future directions.

Literature Cited

2016. PLANTS Database. National Plant Data Team, Greensboro, NC 27401-4901 USA., USDA, NRCS.
Abdul-Muneer, P. 2014. Application of microsatellite markers in conservation genetics and fisheries management: recent advances in population structure analysis and conservation strategies. Genetics research international **2014**.
Amos, W., J. Hoffman, A. Frodsham, L. Zhang, S. Best, and A. Hill. 2007. Automated binning of microsatellite alleles: problems and solutions. Molecular Ecology Notes **7**:10-14.
Andrews, S. 2010. FastQC: A quality control tool for high throughput sequence data. Reference Source.
Avolio, M. L., J. M. Beaulieu, E. Y. Lo, and M. D. Smith. 2012. Measuring genetic diversity in ecological studies. Plant Ecology **213**:1105-1115.
Balas, F., M. Osuna, G. Domínguez, F. Pérez-Gragera, and M. López-Corrales. 2014. Ex situ conservation of underutilised fruit tree species: establishment of a core collection for Ficus carica L. using microsatellite markers (SSRs). Tree Genetics & Genomes **10**:703-710.
Barker, M. S., N. C. Kane, M. Matvienko, A. Kozik, R. W. Michelmore, S. J. Knapp, and L. H. Rieseberg. 2008. Multiple paleopolyploidizations during the evolution of the Compositae reveal parallel patterns of duplicate gene retention after millions of years. Molecular Biology and Evolution **25**:2445-2455.
Boetzer, M., and W. Pirovano. 2012. Toward almost closed genomes with GapFiller. Genome Biol **13**:R56.
Boggess, S. L., P. A. Wadl, D. Hadziabdic, B. E. Scheffler, A. S. Windham, W. E. Klingeman, and R. N. Trigiano. 2014. Characterization of 12 polymorphic microsatellite loci of Pityopsis graminifolia var. latifolia. Conservation Genetics Resources **6**:1043-1045.
Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics:btu170.
Bowers, F. D. 1972. The existence of Heterotheca ruthii (Compositae). Castanea:130-132.
Busch, J. D., P. M. Waser, and J. A. DeWOODY. 2007. Recent demographic bottlenecks are not accompanied by a genetic signature in banner-tailed kangaroo rats (Dipodomys spectabilis). Molecular ecology **16**:2450-2462.
Byrne, M. 2007. Phylogeography provides an evolutionary context for the conservation of a diverse and ancient flora. Australian Journal of Botany **55**:316-325.
Carlson, E. A. 2004. Mendel's legacy. The Annotated and Illustrated Double Helix.
Choi, K. S., and S. Park. 2015. The complete chloroplast genome sequence of Aster spathulifolius (Asteraceae); genomic features and relationship with Asteraceae. Gene **572**:214-221.
Clebsch, E., and A. Sloan. 1993. Final Report, Contract Between the University of Tennessee and the Tennessee Department of Environment and Conservation for the study of various aspects of the ecology and life history of the endangered plant species Ruth's Golden-Aster (Pityopsis ruthii). Unpublished report prepared for the Tennessee Department of Environment and Conservation, Nashville, Tennessee.
Clewell, A. F. 1985. Guide to the vascular plants of the Florida panhandle. Tallahassee: Florida State University Press 605p.-illus., maps, keys.. En Icones, Maps. Geog **3**.
Conant, G. C., and K. H. Wolfe. 2008. GenomeVx: simple web-based creation of editable circular chromosome maps. Bioinformatics **24**:861-862.
Cornuet, J. M., and G. Luikart. 1996. Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. Genetics **144**:2001-2014.
Crutsinger, G. M., M. D. Collins, J. A. Fordyce, Z. Gompert, C. C. Nice, and N. J. Sanders. 2006. Plant genotypic diversity predicts community structure and governs an ecosystem process. Science **313**:966-968.

Davey, J. W., P. A. Hohenlohe, P. D. Etter, J. Q. Boone, J. M. Catchen, and M. L. Blaxter. 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nature Reviews Genetics **12**:499-510.

Davis, C. C., C. G. Willis, R. B. Primack, and A. J. Miller-Rushing. 2010. The importance of phylogeny to the study of phenological response to global climate change. Philos Trans R Soc Lond B Biol Sci **365**:3201-3213.

Doorduin, L., B. Gravendeel, Y. Lammers, Y. Ariyurek, T. Chin-A-Woeng, and K. Vrieling. 2011. The complete chloroplast genome of 17 individuals of pest species Jacobaea vulgaris: SNPs, microsatellites and barcoding markers for population and phylogenetic studies. DNA research:dsr002.

Earl, D. A. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation genetics resources **4**:359-361.

Eisen, J. A., and C. M. Fraser. 2003. Phylogenomics: intersection of evolution and genomics. Science **300**:1706.

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

Excoffier, L., G. Laval, and S. Schneider. 2005. Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evolutionary bioinformatics online **1**:47.

Excoffier, L., and H. E. Lischer. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Molecular ecology resources **10**:564-567.

Farnsworth, E. J. 2007. Plant life history traits of rare versus frequent plant taxa of sandplains: implications for research and management trials. Biological Conservation **136**:44-52.

Ferriol, M., H. Merle, and A. Garmendia. 2014. Microsatellite evidence for low genetic diversity and reproductive isolation in tetraploid Centaurea seridis (Asteraceae) coexisting with diploid Centaurea aspera and triploid hybrids in contact zones. Botanical Journal of the Linnean Society **176**:82-98.

Frankham, R. 2003. Genetics and conservation biology. Comptes Rendus Biologies **326**:22-29.

Frankham, R., D. A. Briscoe, and J. D. Ballou. 2009. Introduction to conservation genetics. 2nd edition. Cambridge University Press.

Funk, W. C., J. K. McKay, P. A. Hohenlohe, and F. W. Allendorf. 2012. Harnessing genomics for delineating conservation units. Trends in Ecology & Evolution **27**:489-496.

Gilbert, G. S., and C. O. Webb. 2007. Phylogenetic signal in plant pathogen–host range. Proceedings of the National Academy of Sciences **104**:4979-4983.

Gornish, E. S. 2013. Effects of density and fire on the vital rates and population growth of a perennial goldenaster. AoB PLANTS **5**:plt041-plt041.

Gowe, A. K., and J. S. Brewer. 2005. The evolution of fire-dependent flowering in goldenasters (Pityopsis spp.)1. The Journal of the Torrey Botanical Society **132**:384-400.

Greet, J., R. Cousens, and J. Webb. 2013. More exotic and fewer native plant species: riverine vegetation patterns associated with altered seasonal flow patterns. River Research and Applications **29**:686-706.

Guo, R., L. Zhou, H. Zhao, and F. Chen. 2013. High Genetic Diversity and Insignificant Interspecific Differentiation in Opisthopappus Shih, an Endangered Cliff Genus Endemic to the Taihang Mountains of China. The Scientific World Journal **2013**.

Hartl, D. L., A. G. Clark, and A. G. Clark. 1997. Principles of population genetics. Sinauer associates Sunderland.

Heuberger, K. A., and F. E. Putz. 2003. Fire in the suburbs: ecological impacts of prescribed fire in small remnants of longleaf pine (Pinus palustris) sandhill. Restoration Ecology **11**:72-81.

Heuertz, M., S. Carnevale, S. Fineschi, F. Sebastiani, J. Hausman, L. Paule, and G. Vendramin. 2006. Chloroplast DNA phylogeography of European ashes, Fraxinus sp.(Oleaceae): roles of hybridization and life history traits. Molecular ecology **15**:2131-2140.

Hoffmann, A., P. Griffin, S. Dillon, R. Catullo, R. Rane, M. Byrne, R. Jordan, J. Oakeshott, A. Weeks, and L. Joseph. 2015. A framework for incorporating evolutionary genomics into biodiversity conservation and management. Climate Change Responses **2**:1.

Holmes, W. C., and J. R. Singhurst. 2012. Pityopsis oligantha (Asteraceae) new to Texas. Phytoneruon **110**:1-4.

Huang, H., C. Shi, Y. Liu, S.-Y. Mao, and L.-Z. Gao. 2014. Thirteen Camellia chloroplast genome sequences determined by high-throughput sequencing: genome structure and phylogenetic relationships. BMC evolutionary biology **14**:1.

Kumar, S., G. Stecher, and K. Tamura. 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Molecular Biology and Evolution:msw054.

Kumar, S., K. Tamura, and M. Nei. 1994. Mega. Bioinformatics **10**:189-191.

Langmead, B., and S. L. Salzberg. 2012. Fast gapped-read alignment with Bowtie 2. Nature methods **9**:357-359.

Lee, E., G. A. Helt, J. T. Reese, M. C. Munoz-Torres, C. P. Childers, R. M. Buels, L. Stein, I. H. Holmes, C. G. Elsik, and S. E. Lewis. 2013. Web Apollo: a web-based genomic annotation editing platform. Genome Biol **14**:R93.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin. 2009. The sequence alignment/map format and SAMtools. Bioinformatics **25**:2078-2079.

Luikart, G., and J. M. Cornuet. 1998. Empirical evaluation of a test for identifying recently bottlenecked populations from allele frequency data. Conservation Biology **12**:228-237.

Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. Cancer research **27**:209-220.

Matsuoka, Y., Y. Yamazaki, Y. Ogihara, and K. Tsunewaki. 2002. Whole chloroplast genome comparison of rice, maize, and wheat: implications for chloroplast gene diversification and phylogeny of cereals. Molecular Biology and Evolution **19**:2084-2091.

Moore, P. A., P. A. Wadl, J. A. Skinner, R. N. Trigiano, E. C. Bernard, W. E. Klingeman, and A. J. Dattilo. 2016. Current knowledge, threats, and future efforts to sustain populations of Pityopsis ruthii (Asteraceae), an endangered southern Appalachian species 1. The Journal of the Torrey Botanical Society **143**:117-134.

Nei, M. 1973. Analysis of gene diversity in subdivided populations. Proceedings of the National Academy of Sciences **70**:3321-3323.

Njuguna, W., A. Liston, R. Cronn, T.-L. Ashman, and N. Bassil. 2013. Insights into phylogeny, sex function and age of Fragaria based on whole chloroplast genome sequencing. Molecular Phylogenetics and Evolution **66**:17-29.

Nordborg, M., and D. Weigel. 2008. Next-generation genetics in plants. Nature **456**:720-723.

Ouborg, N., P. Vergeer, and C. Mix. 2006. The rough edges of the conservation genetics paradigm for plants. Journal of Ecology **94**:1233-1248.

Ouborg, N. J., C. Pertoldi, V. Loeschcke, R. K. Bijlsma, and P. W. Hedrick. 2010. Conservation genetics in transition to conservation genomics. Trends in genetics **26**:177-187.

Panero, J. L., and V. A. Funk. 2008. The value of sampling anomalous taxa in phylogenetic studies: Major clades of the Asteraceae revealed. Molecular Phylogenetics and Evolution **47**:757-782.

Parks, M., R. Cronn, and A. Liston. 2009. Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. BMC biology **7**:1.

Peakall, R., and P. Smouse. 2001. GenAlEx V5: genetic analysis in Excel. Population genetic software for teaching and research. Australian National University, Canberra, Australia.

Peakall, R., and P. E. Smouse. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. Bioinformatics **28**:2537-2539.
Piry, S., G. Luikart, and J.-M. Cornuet. 1999. BOTTLENECK: a program for detecting recent effective population size reductions from allele data frequencies. Montpellier, France.
Powell, W., G. C. Machray, and J. Provan. 1996. Polymorphism revealed by simple sequence repeats. Trends in plant science **1**:215-222.
Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics **155**:945-959.
Provan, J., W. Powell, and P. M. Hollingsworth. 2001. Chloroplast microsatellites: new tools for studies in plant ecology and evolution. Trends in Ecology & Evolution **16**:142-147.
Provencher, L., B. J. Herring, D. R. Gordon, H. L. Rodgers, K. E. Galley, G. W. Tanner, J. L. Hardesty, and L. A. Brennan. 2001. Effects of hardwood reduction techniques on longleaf pine sandhill vegetation in northwest Florida. Restoration Ecology **9**:13-27.
Puechmaille, S. J. 2016. The program STRUCTURE does not reliably recover the correct population structure when sampling is uneven: sub-sampling and new estimators alleviate the problem. Molecular ecology resources.
Semple, J. C. 2006. Pityopsis. Pages 222-228 Flora of North America. Oxford University Press, New York.
Semple, J. C., V. C. Blok, and P. Heiman. 1980. Morphological, anatomical, habit, and habitat differences among the goldenaster genera Chrysopsis, Heterotheca, and Pityopsis (Compositae-Astereae). Canadian Journal of Botany **58**:147-163.
Semple, J. C., and F. D. Bowers. 1985. A revision of the goldenaster genus Pityopsis Nutt.(Compositae: Astereae). University of Waterloo biology series.
Semple, J. C., and F. D. Bowers. 1987. Cytogeography of Pityopsis Nutt., the grass-leaved goldenasters (Compositae: Astereae). Rhodora:381-389.
Sharma, A., S. Jose, K. K. Bohn, and M. G. Andreu. 2012. Effects of reproduction methods and overstory species composition on understory light availability in longleaf pine–slash pine ecosystems. Forest Ecology and Management **284**:23-33.
Simpson, J. T., K. Wong, S. D. Jackman, J. E. Schein, S. J. Jones, and I. Birol. 2009. ABySS: a parallel assembler for short read sequence data. Genome research **19**:1117-1123.
Skinner, M. E., A. V. Uzilov, L. D. Stein, C. J. Mungall, and I. H. Holmes. 2009. JBrowse: a next-generation genome browser. Genome research **19**:1630-1638.
Slatkin, M. 1987. Gene flow and the geographic structure of natural populations. Science **236**:787-792.
Stieneke, D., and I. Eujayl. 2007. Imperfect SSR Finder.
Teoh, V.-H. 2008. Phylogeny, Hybridization and the Evolution of Fire-stimulated Flowering Within the Grass-leaved Goldenasters (Pityopsis, Asteraceae). University of Mississippi.
Thomson, D. M., and M. W. Schwartz. 2006. Using population count data to assess the effects of changing river flow on an endangered riparian plant. Conservation Biology **20**:1132-1142.
Tilman, D., J. Knops, D. Wedin, P. Reich, M. Ritchie, and E. Siemann. 1997. The influence of functional diversity and composition on ecosystem processes. Science **277**:1300-1302.
Timme, R. E., J. V. Kuehl, J. L. Boore, and R. K. Jansen. 2007. A comparative analysis of the Lactuca and Helianthus (Asteraceae) plastid genomes: identification of divergent regions and categorization of shared repeats. American Journal of Botany **94**:302-312.
Twyford, A. D., C. A. Kidner, N. Harrison, and R. A. Ennos. 2013. Population history and seed dispersal in widespread Central American Begonia species (Begoniaceae) inferred from plastome-derived microsatellite markers. Botanical Journal of the Linnean Society **171**:260-276.
USFWS. 1992. Ruth's golden aster recovery plan. Pages 1-33 *in* U. S. F. a. W. Service, editor. U. S. Fish and Wildlife Service, Atlanta, GA.

Vamosi, J. C., and T. A. Dickinson. 2006. Polyploidy and diversification: a phylogenetic investigation in Rosaceae. International Journal of Plant Sciences **167**:349-358.
Vamosi, S., S. Heard, J. Vamosi, and C. Webb. 2009. Emerging patterns in the comparative analysis of phylogenetic community structure. Molecular ecology **18**:572-592.
Vickery, P. D. 2002. Effects of the size of prescribed fire on insect predation of northern blazing star, a rare grassland perennial. Conservation Biology **16**:413-421.
Wadl, P. A., A. J. Dattilo, B. E. Scheffler, and R. N. Trigiano. 2011a. Development of microsatellite loci for the endangered species Pityopsis ruthii (Asteraceae). American Journal of Botany **98**:e342-e345.
Wadl, P. A., A. J. Dattilo, L. M. Vito, and R. N. Trigiano. 2011b. Shoot organogenesis and plant regeneration in Pityopsis ruthii. Plant Cell, Tissue and Organ Culture (PCTOC) **106**:513-516.
Wadl, P. A., T. A. Rinehart, A. J. Dattilo, M. Pistrang, L. M. Vito, R. Milstead, and R. N. Trigiano. 2014. Propagation for the Conservation of Pityopsis ruthii, an Endangered Species from the Southeastern United States. HortScience **49**:194-200.
Wan, Q. H., H. Wu, T. Fujihara, and S. G. Fang. 2004. Which genetic marker for which conservation genetics issue? Electrophoresis **25**:2165-2176.
Wang, M., L. Cui, K. Feng, P. Deng, X. Du, F. Wan, S. Weining, and X. Nie. 2015. Comparative analysis of Asteraceae chloroplast genomes: Structural organization, RNA editing and evolution. Plant Molecular Biology Reporter **33**:1526-1538.
Weakley, A. S. 2010. Flora of the southern and mid-Atlantic states. University of North Carolina Herbarium, North Carolina Botanical Garden, University of North Carolina, Chapel Hill.
Wheeler, G. L., H. E. Dorman, A. Buchanan, L. Challagundla, and L. E. Wallace. 2014. A review of the prevalence, utility, and caveats of using chloroplast simple sequence repeats for studies of plant biology. Applications in plant sciences **2**.
Wright, S. 1931. Evolution in Mendelian populations. Genetics **16**:97-159.
Wyman, S. K., R. K. Jansen, and J. L. Boore. 2004. Automatic annotation of organellar genomes with DOGMA. Bioinformatics **20**:3252-3255.
Xu, Q., G. Xiong, P. Li, F. He, Y. Huang, K. Wang, Z. Li, and J. Hua. 2012. Analysis of complete nucleotide sequences of 12 Gossypium chloroplast genomes: origin and evolution of allotetraploids. PLoS One **7**:e37128.
Yang, W., B. J. Glover, G. Y. Rao, and J. Yang. 2006. Molecular evidence for multiple polyploidization and lineage recombination in the Chrysanthemum indicum polyploid complex (Asteraceae). New phytologist **171**:875-886.

Appendices

Appendix 1. Tables

Table 1. Summary statistics for 12 nuclear microsatellite loci in *Pityopsis ruthii* subpopulations.

Locus	Repeat motif	*A*	A_R	H_O	H_E	F_{IS}	F_{IT}	F_{ST}	N_m
PR002*	$(TG)_9$	15	2.55	0.50	0.52	0.02	0.25	0.24	0.78
PR003*	$(TG)_{14}$	9	2.83	0.42	0.59	0.26	0.46	0.27	0.69
PR005*	$(CT)_7$	16	3.08	0.76	0.63	-0.23	0.11	0.27	0.66
PR006*	$(CA)_{16}$	17	3.52	0.33	0.67	0.48	0.59	0.21	0.94
PR009*	$(GT)_{11}$	16	3.33	0.43	0.62	0.27	0.53	0.36	0.45
PR020*	$(GT)_8$	13	3.45	0.48	0.68	0.26	0.39	0.18	1.16
PR027*	$(GTGTC)_5$	18	3.30	0.50	0.62	0.17	0.35	0.23	0.86
PR028*	$(GT)_{10}$	9	2.70	0.28	0.54	0.45	0.59	0.26	0.70
PR029*	$(GT)_3A(GT)_8$	24	4.26	0.66	0.77	0.10	0.26	0.17	1.24
PR030*	$(AC)_{12}$	21	4.16	0.53	0.77	0.28	0.41	0.18	1.15
PR031*	$(GT)_9AA(GT)_5$	19	2.94	0.25	0.57	0.54	0.70	0.35	0.47
PR035*	$(GT)_5A(TG)_7(AG)_{15}$	16	4.10	0.71	0.76	0.04	0.16	0.12	1.76
Mean		16.08	3.35	0.49	0.65	0.22	0.40	0.24	0.90

A= number of alleles, A_R= allelic richness, H_O= observed heterozygosity, H_E= expected heterozygosity, F_{IS}= inbreeding coefficient relative to the subpopulation, F_{IT}= inbreeding coefficient relative to total number of individuals, F_{ST}= fixation index, N_m= estimated gene flow

Table 2. Genetic variability of 12 nuclear microsatellite loci estimated for 33 subpopulations of *Pityopsis ruthii*

River	Subpopulation	Sample size	A_R	H_O	H_E	F_{IS}	Private alleles
Ocoee	O-06-01	49	3.01	0.38	0.59	0.37	-
	O-05-01	50	2.96	0.34	0.58	0.42	5
	O-04-01	22	2.37	0.35	0.46	0.27	1
	O-03-01	8	2.45	0.32	0.49	0.41	-
	O-02-03	26	2.74	0.39	0.54	0.29	2
	O-02-02	18	2.78	0.35	0.51	0.34	-
	O-02-01	24	2.90	0.40	0.55	0.29	-
	O-01-01	35	2.94	0.33	0.57	0.44	1
Hiwassee	H-11-01	29	3.68	0.51	0.68	0.27	4
	H-09-03	20	4.02	0.66	0.72	0.11	2
	H-09-01	14	3.94	0.60	0.69	0.17	-
	H-09-02	16	3.68	0.71	0.68	-0.01	-
	H-08-07	6	3.67	0.57	0.65	0.21	-
	H-08-06	9	3.65	0.50	0.65	0.29	-
	H-08-04	15	3.63	0.54	0.67	0.22	-
	H-08-03	4	3.17	0.50	0.54	0.21	-
	H-07-03	58	3.96	0.59	0.72	0.20	2
	H-07-02	29	3.94	0.61	0.73	0.17	-
	H-07-01	4	3.08	0.42	0.55	0.37	4
	H-06-07	16	3.30	0.49	0.61	0.23	-
	H-06-05	50	5.91	0.53	0.60	0.13	4
	H-06-04	28	3.40	0.49	0.65	0.26	2
	H-06-02	33	3.73	0.57	0.69	0.19	2
	H-06-01	14	3.61	0.55	0.67	0.21	1
	H-12-06	11	5.83	0.54	0.60	0.16	1
	H-12-04	11	6.00	0.50	0.60	0.22	-
	H-05-01	50	3.74	0.63	0.70	0.10	1
	H-04-05	31	3.39	0.47	0.65	0.29	1
	H-04-04	44	4.08	0.40	0.74	0.46	11
	H-03-01	15	3.06	0.38	0.53	0.33	-
	H-02-01	25	2.77	0.41	0.51	0.23	-
	H-01-06	21	3.19	0.46	0.61	0.27	-
	H-01-02	29	3.10	0.49	0.55	0.12	1

A_R= allelic richness, H_O= observed heterozygosity, H_E= expected heterozygosity, F_{IS}= inbreeding coefficient relative to the subpopulation

Table 3. Analysis of molecular variance (AMOVA) from nuclear microsatellite data collected from *Pityopsis ruthii* using Arlequin (version 3.5.1.2).

	d.f.	Sum of squares	Variance component	% of variation	*P* value
A. Variance partition					
Among populations	32	1086.36	0.63	17.90	<0.001
Among individuals within populations	781	2638.26	0.50	14.41	<0.001
Within individuals	814	1928.50	2.37	67.69	<0.001
Total	1627	5653.12	3.50	100	
Fixation indices: F_{IS} = 0.18, F_{ST} = 0.18, F_{IT} = 0.32					
B. Variance partition					
Among populations	24	548.58	0.42	12.18	<0.001
Among individuals within populations	557	1937.06	0.43	12.29	<0.001
Within individuals	582	1527.00	2.62	75.53	<0.001
Total	1163	4012.64	3.47	100	
Fixation indices: F_{IS} = 0.14, F_{ST} = 0.12, F_{IT} = 0.25					
C. Variance partition					
Among populations	7	271.02	0.63	20.69	<0.001
Among individuals within populations	224	701.20	0.70	22.84	<0.001
Within individuals	232	401.50	1.73	56.47	<0.001
Total	463	1373.72	3.07	100	
Fixation indices: F_{IS} = 0.28, F_{ST} = 0.21, F_{IT} = 0.44					

A = The first analysis included all sampling sites as one hierarchical group; B = The second hierarchical analysis included all sampling sites on the Hiwassee River; C = The final analysis included all sampling sites on the Ocoee River; F_{ST} = variance among subpopulations relative to the total variance; F_{IS} = inbreeding coefficient of individuals relative to the population; F_{IT} = variance in the total population; F_{CT} = variance among groups relative to the total variance; F_{SC} = variance among subpopulations within groups

Table 4. Subpopulation pairwise F_{ST} values (below diagonal) and gene flow estimates (above diagonal) for *Pityopsis ruthii* on the Hiwassee River.

	H-11-01	H-09-03	H-09-01	H-09-02	H-08-07	H-08-06	H-08-04	H-08-03	H-07-03	H-07-02	H-07-01	H-06-07	H-06-05	H-06-04	H-06-02	H-06-01	H-12-06	H-12-04	H-05-01	H-04-05	H-04-04	H-03-01	H-02-01	H-01-06	H-01-02
H-11-01	0.00	1.58	1.59	1.32	8.01	1.68	4.11	2.01	2.63	1.83	1.80	1.59	1.23	1.94	2.31	2.22	0.94	0.96	1.20	2.01	1.69	0.88	1.40	1.10	0.81
H-09-03	0.14	0.00	92.63	6.02	3.00	1.17	1.52	1.11	1.71	1.33	1.13	1.20	0.85	1.34	1.31	1.46	0.90	0.87	3.06	1.35	2.18	1.11	0.81	1.37	0.90
H-09-01	0.14	0.00^{NS}	0.00	15.20	2.82	1.14	1.48	1.03	1.65	1.36	1.12	1.18	0.83	1.32	1.35	1.43	0.91	0.88	2.42	1.40	2.10	1.00	0.81	1.22	0.81
H-09-02	0.16	0.04	0.02^{NS}	0.00	2.36	0.99	1.21	0.96	1.45	1.17	0.99	0.86	0.72	1.16	1.17	1.15	0.78	0.76	2.77	1.03	1.68	0.84	0.66	1.00	0.68
H-08-07	0.03	0.08	0.08	0.10	0.00	4.75	16.06	5.48	4.05	3.17	4.32	1.53	1.31	1.87	2.16	2.46	1.17	1.13	2.14	2.10	2.95	1.07	1.01	1.51	0.82
H-08-06	0.13	0.18	0.18	0.20	0.05	0.00	3.58	6.14	3.16	2.34	1.97	0.97	1.23	1.54	1.50	1.60	1.35	0.97	0.92	1.15	1.97	0.75	0.66	1.36	0.96
H-08-04	0.06	0.14	0.14	0.17	0.02	0.07	0.00	6.48	3.72	2.88	2.45	1.82	1.45	3.34	3.21	3.62	1.22	1.06	1.26	1.98	2.39	0.86	1.08	1.37	0.91
H-08-03	0.11	0.18	0.20	0.21	0.04	0.04	0.04^{NS}	0.00	3.79	2.56	2.58	1.02	1.61	1.63	1.48	1.38	1.08	0.94	1.04	1.15	2.19	0.66	0.66	1.02	0.76
H-07-03	0.09	0.13	0.13	0.15	0.06	0.07	0.06	0.06	0.00	4.06	3.72	2.04	2.92	3.35	2.99	3.31	2.06	2.52	1.40	2.87	3.82	1.15	1.52	3.00	1.62
H-07-02	0.12	0.16	0.16	0.18	0.07	0.10	0.08	0.09	0.06	0.00	3.13	1.77	1.30	2.16	2.47	2.05	1.26	1.20	1.33	1.99	2.58	0.89	1.08	1.33	1.07
H-07-01	0.12	0.18	0.18	0.20	0.05	0.11	0.09	0.09	0.06	0.07	0.00	1.61	1.27	1.91	2.08	1.99	1.01	0.85	1.23	2.07	3.17	0.74	1.01	1.20	0.80
H-06-07	0.14	0.17	0.17	0.23	0.14	0.20	0.12	0.20	0.11	0.12	0.13	0.00	1.02	2.35	3.28	2.97	0.76	0.85	1.13	2.72	2.50	0.81	1.75	1.04	1.02
H-06-05	0.17	0.23	0.23	0.26	0.16	0.17	0.15	0.13	0.08	0.16	0.16	0.20	0.00	1.34	1.47	1.41	1.97	2.48	0.77	1.29	1.51	0.68	0.84	1.05	0.75
H-06-04	0.11	0.16	0.16	0.18	0.12	0.14	0.07	0.13	0.07	0.10	0.12	0.10	0.16	0.00	3.83	4.92	1.01	0.95	1.09	2.05	2.31	0.90	1.20	1.20	0.86
H-06-02	0.10	0.16	0.16	0.18	0.10	0.14	0.07	0.14	0.08	0.09	0.11	0.07	0.14	0.06	0.00	12.08	1.17	1.27	1.35	3.84	3.14	0.91	1.67	1.39	1.07
H-06-01	0.10	0.15	0.15	0.18	0.09	0.14	0.06	0.15	0.07	0.11	0.11	0.08	0.15	0.05	0.02	0.00	1.13	1.18	1.31	3.85	3.42	0.94	1.65	1.44	0.89
H-12-06	0.21	0.22	0.22	0.24	0.18	0.16	0.17	0.19	0.11	0.17	0.20	0.25	0.11	0.20	0.18	0.18	0.00	2.50	0.82	1.08	1.65	0.66	0.63	1.14	0.93
H-12-04	0.21	0.22	0.22	0.25	0.18	0.21	0.19	0.21	0.09	0.17	0.23	0.23	0.09	0.21	0.16	0.17	0.09	0.00	0.85	1.32	1.81	0.67	0.84	1.53	0.98
H-05-01	0.17	0.08	0.09	0.08	0.10	0.21	0.17	0.19	0.15	0.16	0.17	0.18	0.24	0.19	0.16	0.16	0.23	0.23	0.00	1.38	1.93	0.88	0.81	1.30	0.89
H-04-05	0.11	0.16	0.15	0.20	0.11	0.18	0.11	0.18	0.08	0.11	0.11	0.08	0.16	0.11	0.06	0.06	0.19	0.16	0.15	0.00	3.51	1.04	3.06	2.05	1.25
H-04-04	0.13	0.10	0.11	0.13	0.08	0.11	0.09	0.10	0.06	0.09	0.07	0.09	0.14	0.10	0.07	0.07	0.13	0.12	0.11	0.07	0.00	2.01	1.39	3.14	2.12
H-03-01	0.22	0.18	0.20	0.23	0.19	0.25	0.23	0.27	0.18	0.22	0.25	0.24	0.27	0.22	0.21	0.21	0.28	0.27	0.22	0.19	0.11	0.00	0.60	0.96	0.71
H-02-01	0.15	0.24	0.24	0.27	0.20	0.27	0.19	0.27	0.14	0.19	0.20	0.12	0.23	0.17	0.13	0.13	0.29	0.23	0.24	0.08	0.15	0.29	0.00	0.97	0.80
H-01-06	0.19	0.15	0.17	0.20	0.14	0.16	0.15	0.20	0.08	0.16	0.17	0.19	0.19	0.17	0.15	0.15	0.18	0.14	0.16	0.11	0.07	0.21	0.20	0.00	3.67
H-01-02	0.24	0.22	0.24	0.27	0.23	0.21	0.22	0.25	0.13	0.19	0.24	0.20	0.25	0.23	0.19	0.22	0.21	0.20	0.22	0.17	0.11	0.26	0.24	0.06	0.00

Table 5. Subpopulation pairwise F_{ST} values (below diagonal) and gene flow estimates (above diagonal) for *Pityopsis ruthii* on the Ocoee River.

	O-06-01	O-05-01	O-04-01	O-03-01	O-02-03	O-02-02	O-02-01	O-01-01
O-06-01	0.00	0.89	0.53	1.64	1.07	0.94	1.00	0.94
O-05-01	0.22	0.00	1.33	0.90	1.92	1.52	1.16	1.05
O-04-01	0.32	0.16	0.00	0.50	0.91	0.72	0.64	0.68
O-03-01	0.13	0.22	0.33	0.00	1.21	0.94	0.90	0.98
O-02-03	0.19	0.12	0.22	0.17	0.00	2.19	1.76	1.23
O-02-02	0.21	0.14	0.26	0.21	0.10	0.00	5.65	1.51
O-02-01	0.20	0.18	0.28	0.22	0.12	0.04	0.00	1.21
O-01-01	0.21	0.19	0.27	0.20	0.17	0.14	0.17	0.00

Table 6. Characterization of seven chloroplast microsatellite markers from *Pityopsis ruthii*

Locus	Primer sequences (5'-3')	Repeat motif	N	N_a	N_e	I	h	uh
cpPR002	F: ACTCACTAAGCCGGGATCACT	$(T)_9$	17	3.46	2.48	0.96	0.54	0.59
	R: GGAACCGGGGAAAGTATACAG							
cpPR004	F: ACCGATCCTTGTTTACCAACC	$(GAA)_3$	5	2.61	1.79	0.64	0.38	0.40
	R: TCTCGAGAAACAAGTGGGCTA							
cpPR005	F: ATTCGGCAGATTTTGATTCCT	$(T)_{12}$	5	3.00	1.90	0.76	0.43	0.47
	R: AAAACCCCTTCCCAAACTGTA							
cpPR006	F: ATTGAATTGGGTCCAGGAATC	$(T)_8$	12	3.21	2.35	0.85	0.49	0.52
	R: GCAATGAGATCGTTAAATGGAA							
cpPR010	F: AATGGACGATTCCATCGATTA	$(AG)_4$	16	3.91	2.52	1.00	0.54	0.60
	R: TGAACAAACTCGACAAATGG							
cpPR011	F: CAAAATTTCTTGATTCCCATACA	$(CAG)_3$	15	3.27	2.04	0.75	0.40	0.44
	R: TTTAGGCAGAATACCATCACCT							
cpPR019	F: GCGTATTGATTTGACCCCATA	$(A)_9$	8	3.12	2.07	0.77	0.44	0.47
	R: TTGCGAAAACTTCTGGATAGG							

Number of alleles (N), allele frequency (N_a), number of effective alleles (N_e), Shannon's information index (I), diversity (h), and unbiased diversity (uh)

Table 7. Mean number of alleles (N), allele frequency (N_a), number of effective alleles (N_e), Shannon's information index (I), diversity (h), and unbiased diversity (uh) by population for the seven *Pityopsis ruthii* chloroplast microsatellite markers.

River	Population	Sample size	N_a	N_e	I	h	uh	Private alleles
Ocoee	O-06-01	46	4.000	2.416	1.021	0.574	0.588	--
	O-05-01	50	3.714	2.052	0.840	0.464	0.473	1
	O-04-01	31	3.143	1.730	0.621	0.323	0.334	1
	O-03-01	8	3.286	2.749	1.046	0.609	0.714	--
	O-02-03	25	2.571	1.794	0.620	0.365	0.381	--
	O-02-02	17	3.429	2.475	0.980	0.554	0.591	--
	O-02-01	25	4.000	2.527	1.068	0.578	0.605	1
	O-01-01	33	3.571	1.840	0.794	0.438	0.453	1
Hiwassee	H-11-01	30	3.000	2.192	0.809	0.480	0.501	--
	H-09-03	19	3.286	2.149	0.856	0.476	0.509	--
	H-09-01	16	3.000	1.941	0.764	0.440	0.511	--
	H-09-02	14	3.714	2.976	1.101	0.597	0.650	1
	H-08-07	4	1.429	1.371	0.278	0.196	0.262	--
	H-08-06	8	2.143	1.464	0.468	0.272	0.315	--
	H-08-04	16	2.429	1.567	0.512	0.293	0.318	--
	H-08-03	4	2.000	1.695	0.561	0.367	0.500	--
	H-07-03	57	5.000	2.773	1.193	0.624	0.637	--
	H-07-02	29	3.286	2.357	0.875	0.493	0.517	1
	H-07-01	3	1.571	1.457	0.364	0.254	0.381	--
	H-06-07	17	2.857	2.020	0.760	0.444	0.489	--
	H-06-05	49	2.857	2.037	0.697	0.397	0.413	1
	H-06-04	28	3.714	2.690	1.086	0.609	0.634	--
	H-06-02	33	3.000	2.297	0.813	0.476	0.492	--
	H-06-01	15	3.286	2.349	0.977	0.571	0.616	--
	H-12-06	11	1.714	1.440	0.296	0.168	0.184	1
	H-12-04	11	2.714	1.968	0.719	0.415	0.472	1
	H-05-01	50	2.714	1.869	0.700	0.408	0.418	1
	H-04-05	31	2.857	1.687	0.581	0.319	0.332	1
	H-04-04	44	5.857	3.187	1.339	0.654	0.676	3
	H-03-01	15	3.714	2.390	0.977	0.521	0.585	--
	H-02-01	25	3.857	2.716	1.060	0.587	0.616	1
	H-01-06	20	4.286	2.954	1.212	0.644	0.686	--
	H-01-02	30	4.429	2.336	1.024	0.550	0.573	--

Allele frequency (N_a), number of effective alleles (N_e), Shannon's information index (I), diversity (h), and unbiased diversity (uh)

Table 8. Analysis of molecular variance (AMOVA) from chloroplast microsatellite data collected from *Pityopsis ruthii* using Arlequin (version 3.5.1.2).

	d.f.	Sum of squares	Variance component	% of variation	*P* value
A. Variance partition					
Among groups	1	45.89	0.080 Va	4.66	< 0.01
Among populations within groups	31	433.23	0.542 Vb	32.01	< 0.01
Within populations	771	827.30	1.073 Vc	63.33	0.03
Total	803	1306.41	1.694	100	
Fixation indices: F_{SC} = 0.33, F_{ST} = 0.37, F_{CT} = 0.05					
B. Variance partition					
Among populations	23	362.00	0.632 Va	37.12	< 0.01
Within populations	542	579.93	1.070 Vb	62.88	< 0.01
Total	565	941.94	1.702	100	
Fixation indices: F_{ST} = 0.37					
C. Variance partition					
Among populations	7	67.34	0.299 Va	21.56	< 0.01
Within populations	227	246.70	1.087 Vb	78.44	<0.01
Total	234	314.04	1.386	100	
Fixation indices: F_{ST} = 0.22					

A = The first analysis included all sampling sites as one hierarchical group; B = The second hierarchical analysis included all sampling sites on the Hiwassee River; C = The final analysis included all sampling sites on the Ocoee River; F_{ST} = variance among subpopulations relative to the total variance; F_{IS} = inbreeding coefficient of individuals relative to the population; F_{IT} = variance in the total population; F_{CT} = variance among groups relative to the total variance; F_{SC} = variance among subpopulations within groups

Table 9. Wilcoxon tests for recent bottlenecks using the program BOTTLENECK for *Pityopsis ruthii* subpopulations. The *P* values are reported for the infinite allele model (IAM), two-phase mutational model (TPM), and stepwise mutational model (SMM).

River	Structure cluster	IAM	TPM	SMM	Distribution
Hiwassee	1 (yellow)	0.001	0.008	0.001	Normal
	2 (blue)	<0.001	0.910	0.064	Normal
	3 (green)	<0.001	0.093	0.001	Normal
	4 (red)	0.034	0.001	<0.001	Normal
Ocoee	1 (red)	0.008	0.470	0.077	Normal
	2 (blue)	0.017	0.380	0.850	Normal
	3 (green)	0.340	0.569	0.151	Normal

Table 10. State of tissue after collection, date, and location of *Pityopsis* individuals used for sequencing.

Species	Type of tissue	Year collected	Location
P. aspera var. *adenolepis*	Dried	2014	South Carolina
P. falcata	Dried	2010	Rhode Island
P. flexuosa	Dried	2015	Florida
P. graminifolia var. *aequilifolia*	Dried	2015	Florida
P. graminifolia var. *latifolia*	Frozen	2013	Tennessee
P. graminifolia var. *tracyi*	Fresh	2014	Florida
P. oligantha	Dried	2015	Florida
P. ruthii	Frozen	2010	Tennessee

Table 11. Statistics from original sequences for all *Pityopsis* individuals

Species	Avg. contig length (bp)	No. reads	Total bp	% GC
P. aspera var. *adenolepis*	231.80	3780160	876251906	0.50
P. falcata	235.94	4448332	1049529236	0.36
P. flexuosa	218.68	5245587	1149603092	0.51
P. graminifolia var. *aequilifolia*	170.16	33339900	567294195	0.35
P. graminifolia var. *latifolia*	231.77	3553331	823569125	0.35
P. graminifolia var. *tracyi*	230.27	4950830	1140021069	0.36
P. oligantha	232.32	3451455	801833869	0.38
P. ruthii	288.92	3609012	1042705488	0.35

P. ruthii was sequenced using the Ion Torrent platform. All other species and varieties were sequenced using the Illumina MiSeq platform.

Table 12. Statistics of *Pityopsis* sequences mapped to the *Helianthus annuus* reference using Bowtie2

Species	No. mapped reads	No. bp mapped	% mapped	Coverage
P. aspera var. *adenolepis*	22554	4109187	0.49	27.39x
P. falcata	134967	28365271	2.55	189.10x
P. flexuosa	23800	5682784	0.40	37.89x
P. graminifolia var. *aequilifolia*	6571	762368	0.29	5.08x
P. graminifolia var. *latifolia*	55981	12950819	1.36	86.34x
P. graminifolia var. *tracyi*	140727	27133639	2.29	180.89x
P. oligantha	12546	2306733	0.30	15.38x
P. ruthii	113690	25458733	3.76	169.72x

P. ruthii was sequenced using the Ion Torrent platform. All other species and varieties were sequenced using the Illumina MiSeq platform.

Table 13. Genes included in the *Pityopsis* chloroplast genomes

Category	Group of genes	Name of genes
Self replication	Large subunit of ribosomal proteins	rpl2, 14, 16, 20, 22, 23, 32, 33, 36
	Small subunit of ribosomal proteins	rps2, 3, 4, 7, 8, 11, 12, 14, 15, 16, 18, 19
	DNA dependent RNA polymerase	rpoA, B, C1, C2
	rRNA genes	rrn4.5, 4, 16, 23
	tRNA genes	trnA-UGC, C-GCA, D-GUC, E-UUC, F-GAA, fM-CAU, G-GCC, G-UCC, I-CAU, I-GAU, K-UUU, L-CAA, L-UAA, L-UAG, M-CAU, N-GUU, P-UGG, Q-UUG, R-ACG, S-GCU, S-GGA, V-GAC, V-UAC, W-CCA, Y-GUA
Photosynthesis	Photosystem I	psaA, B, C, I, J
	Photosystem II	psbA, B, C, D, E, F, H, I, J, K, L, M, N, T, Z
	NADH oxidoreductase	ndhA, B, C, D, E, F, G, H, I, J, K
	Cytochrome b6/f complex	petA, B, D, G, L, N
	ATP synthase	atpA, B, E, F, H, I
	Rubisco	rbcL
Other genes	Translational initiation factor	infA
	Maturase	matK
	Protease	clpP
	Envelop membrane protein	cemA
	Subunit Acetyl-CoA-carboxylase	accD
	c-type cytochrom synthesis gene	ccsA
Unknown gene	Conserved Open Reading Frames	orf188, ycf1, ycf2, ycf3, ycf4, ycf15, ycf68

Table 14. Pairwise alignment comparison of eight *Pityopsis* chloroplast genomes. Percent dissimilarity as calculated in MEGA7.

	P. aspera var. *adenolepis*	*P. falcata*	*P. flexuosa*	*P. graminifolia* var. *aequilifolia*	*P. graminifolia* var. *latifolia*	*P. graminifolia* var.*tracyi*	*P. oligantha*	*P. ruthii*
P. aspera var. *adenolepis*								
P. falcata	0.002							
P. flexuosa	0.002	0.003						
P. graminifolia var. *aequilifolia*	0.002	0.001	0.003					
P. graminifolia var. *latifolia*	0.002	0.001	0.003	0.001				
P. graminifolia var. *tracyi*	0.001	0.000	0.003	0.001	0.000			
P. oligantha	0.001	0.002	0.002	0.002	0.002	0.000		
P. ruthii	0.002	0.001	0.003	0.001	0.001	0.001	0.002	

Table 15. Estimates of evolutionary divergence between *Pityopsis* cp genomes and *Aster spathulifolius*. Number of base differences per sequence from between sequences are shown. All positions containing gaps and missing data were eliminated. Analyses were conducted in MEGA7.

	P. aspera var. *adenolepis*	*P. falcata*	*P. flexuosa*	*P. graminifolia* var. *aequilifolia*	*P. graminifolia* var. *latifolia*	*P. graminifolia* var.*tracyi*	*P. oligantha*	*P. ruthii*
P. aspera var. *adenolepis*								
P. falcata	202							
P. flexuosa	234	370						
P. graminifolia var. *aequilifolia*	186	70	347					
P. graminifolia var. *latifolia*	181	63	340	65				
P. graminifolia var. *tracyi*	161	49	358	61	42			
P. oligantha	147	209	237	198	205	183		
P. ruthii	207	79	377	83	70	58	220	
A. spathulifolius	2,314	2,202	24,78	2,204	2,174	2,173	2,341	2,202

Appendix 2. Figures

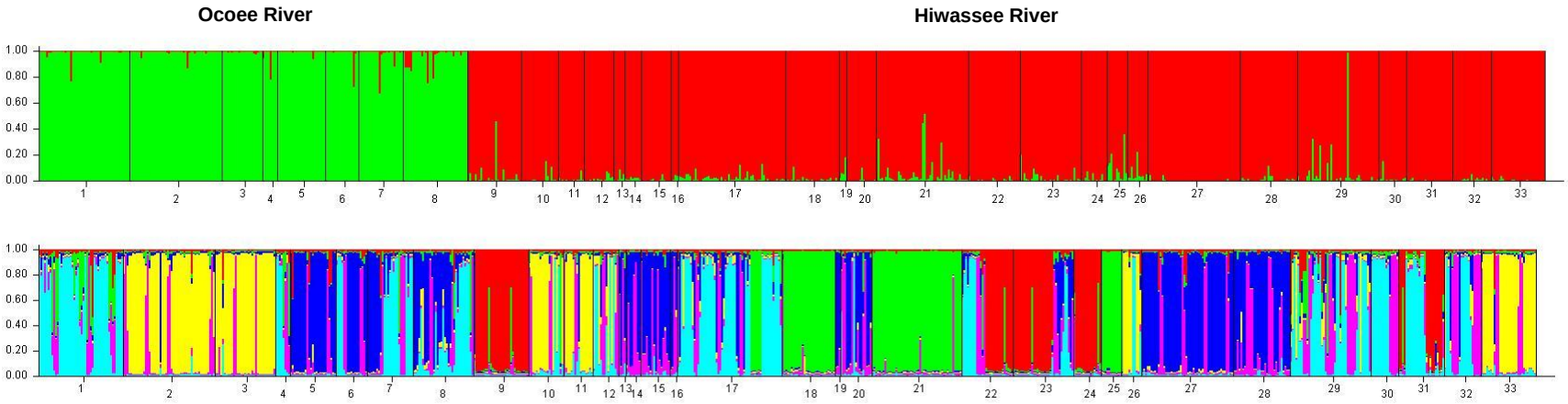

Figure 1. Bar plot (top panel) showing Bayesian assignment probabilities of nuclear data for the Hiwassee and Ocoee Rivers subpopulations using the program STRUCTURE for two clusters (*k* = 2). Bar plot (lower panel) showing Bayesian assignment probabilities of chloroplast data for the Hiwassee and Ocoee Rivers subpopulations using the program STRUCTURE for six clusters (*k* = 6). The x-axis indicates the individuals sampled and the y-axis indicates the assignment probability of individuals to each of the two clusters. Each vertical line represents and individual's probability of belonging to one of *k* clusters (represented by different colors) or a combination of if ancestry is mixed.

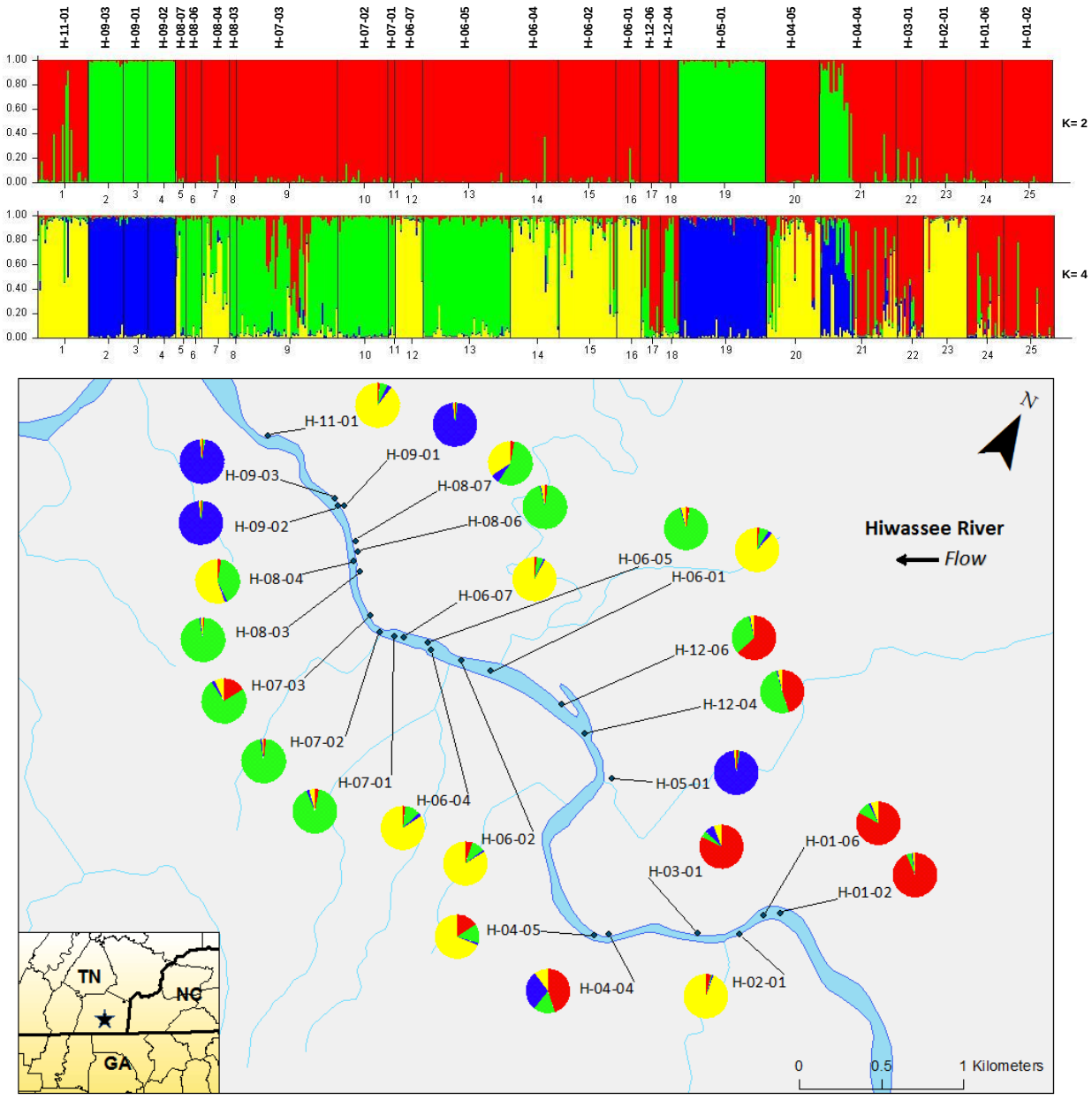

Figure 2. Bar plots of the individual Bayesian assignment probabilities of nuclear data for the *Pityopsis ruthii* Hiwassee River subpopulations using the program STRUCTURE for two or four clusters.

Each vertical line represents and individual's probability of belonging to one of k clusters (represented by different colors) or a combination if ancestry is mixed. Map of the sampled populations. Pie charts correspond to the population assignment for the four genetic groups defined by the Bayesian assignment of Structure.

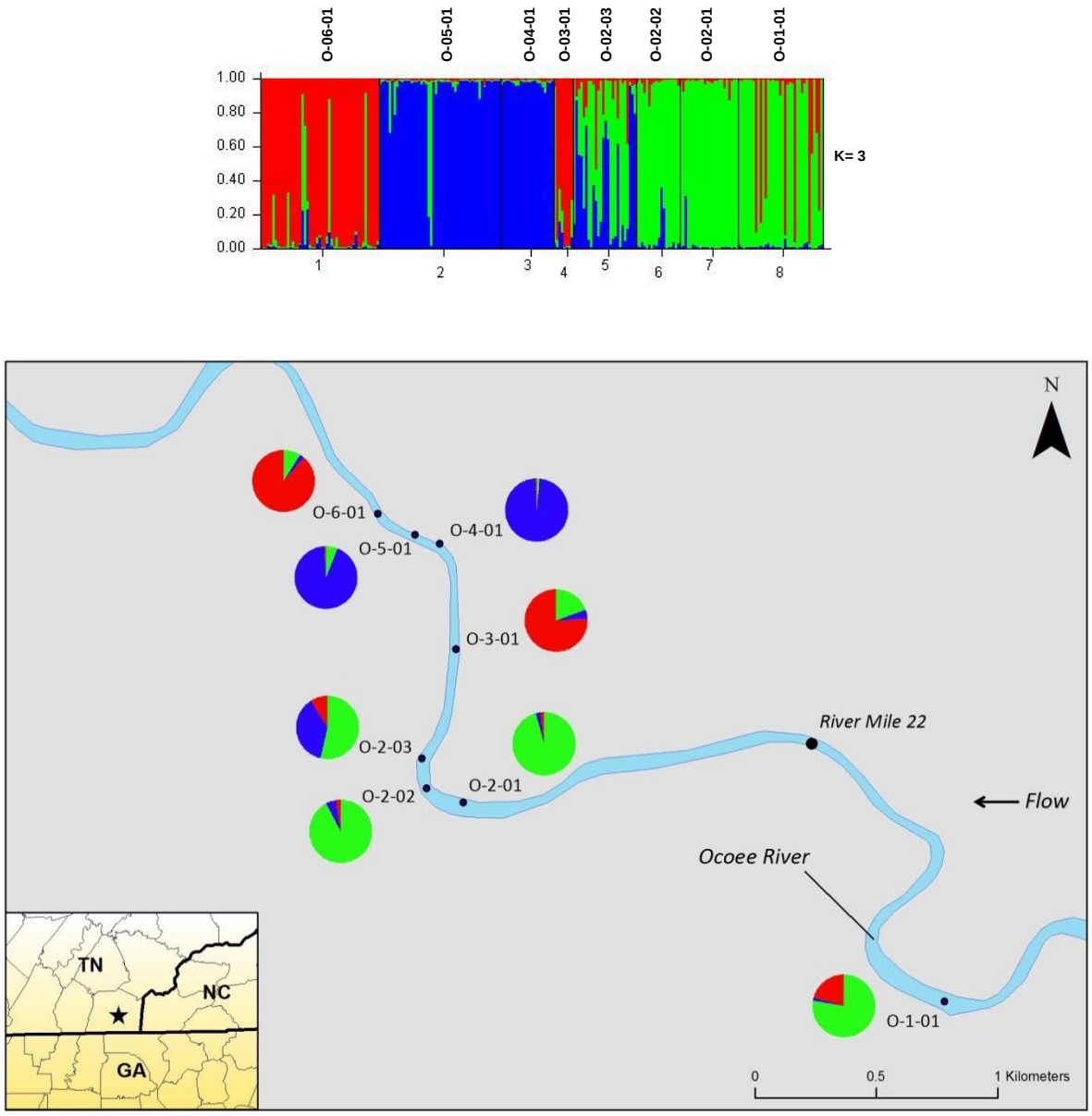

Figure 3. Bar plots of the Bayesian assignment probabilities of nuclear data for the *Pityopsis ruthii* Ocoee River populations using the program STRUCTURE for three clusters.

Each vertical line represents and individual's probability of belonging to one of k clusters (represented by different colors) or a combination of if ancestry is mixed. Map of the sampled populations. Pie charts correspond to the population assignment for the three genetic groups defined by the Bayesian assignment of Structure.

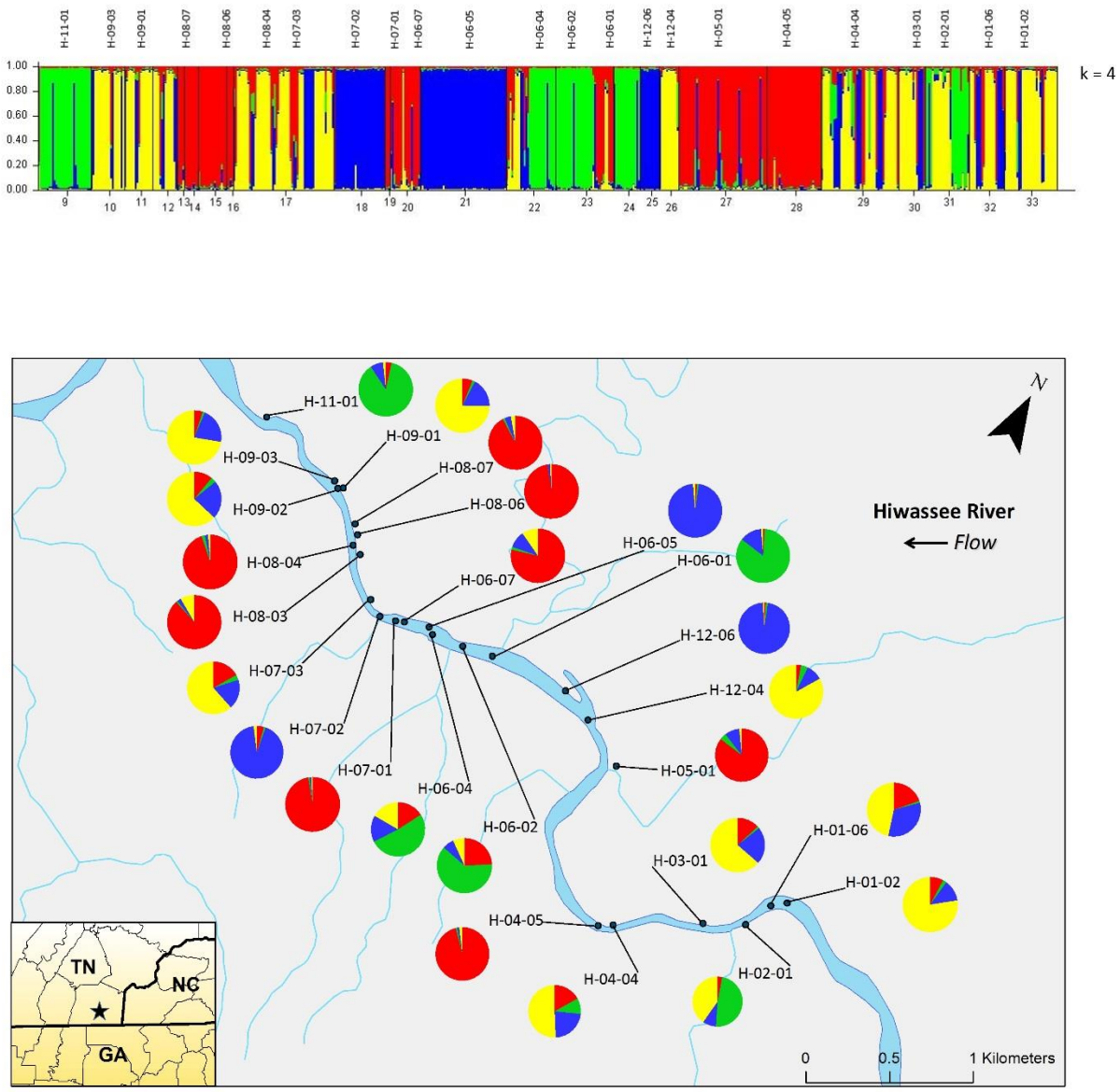

Figure 4. Bar plot of individual Bayesian probabilities of chloroplast data for *Pityopsis ruthii* Hiwassee River populations using the program STRUCTURE for four clusters.

Each vertical line represents and individual's probability of belonging to one of k clusters (represented by different colors) or a combination of if ancestry is mixed. Map of the sampled populations. Pie charts correspond to the population assignment for the three genetic groups defined by the Bayesian assignment of Structure.

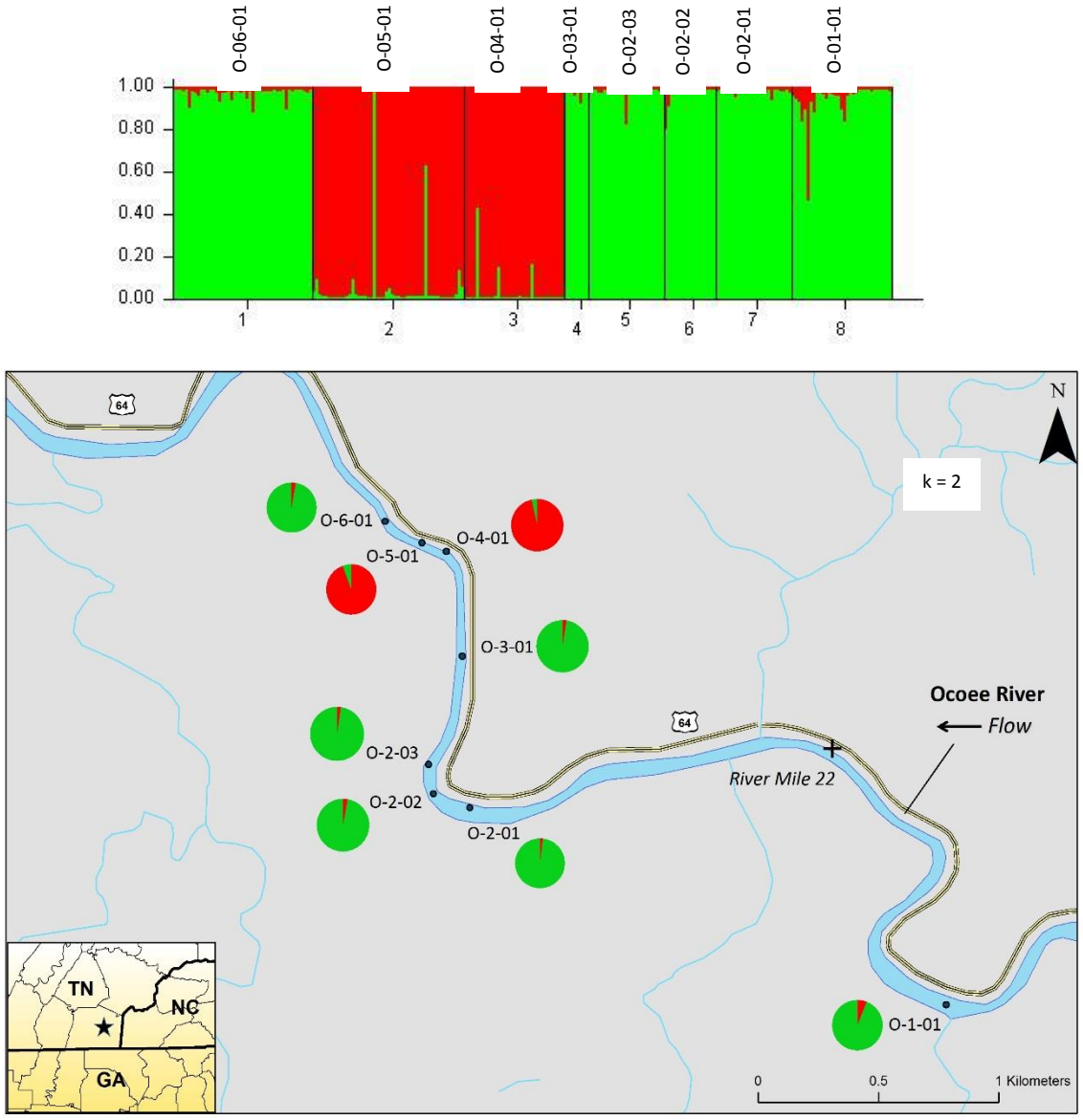

Figure 5. Bar plot of the individual Bayesian assignment probabilities of chloroplast data for *Pityopsis ruthii* subpopulations along the Ocoee River using STRUCTURE for two clusters (k = 2).

Each vertical line represents and individual's probability of belonging to one of *k* clusters (represented by different colors) or a combination of if ancestry is mixed. Pie charts correspond to the population assignment for the three genetic groups defined by the Bayesian assignment of Structure.

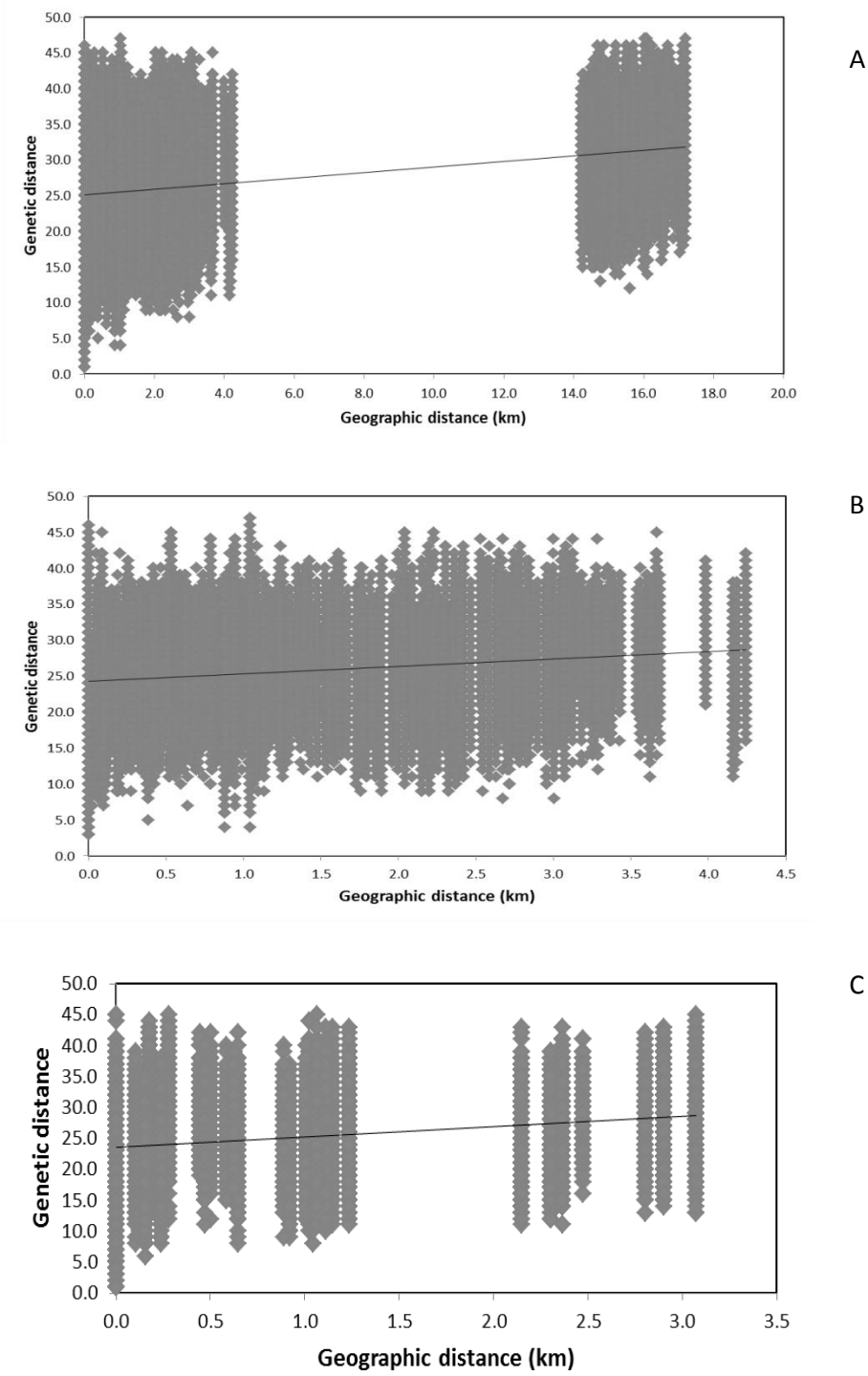

Figure 6. Isolation by distance graph of *Pityopsis ruthii* populations using nuclear microsatellites.

A. Correlation between pairwise genetic distance values and geographic distance for all populations. B. Correlation between pairwise genetic distance values and geographic distance for Hiwassee River populations. C. Correlation between pairwise genetic distance values and geographic distance for Ocoee River populations.

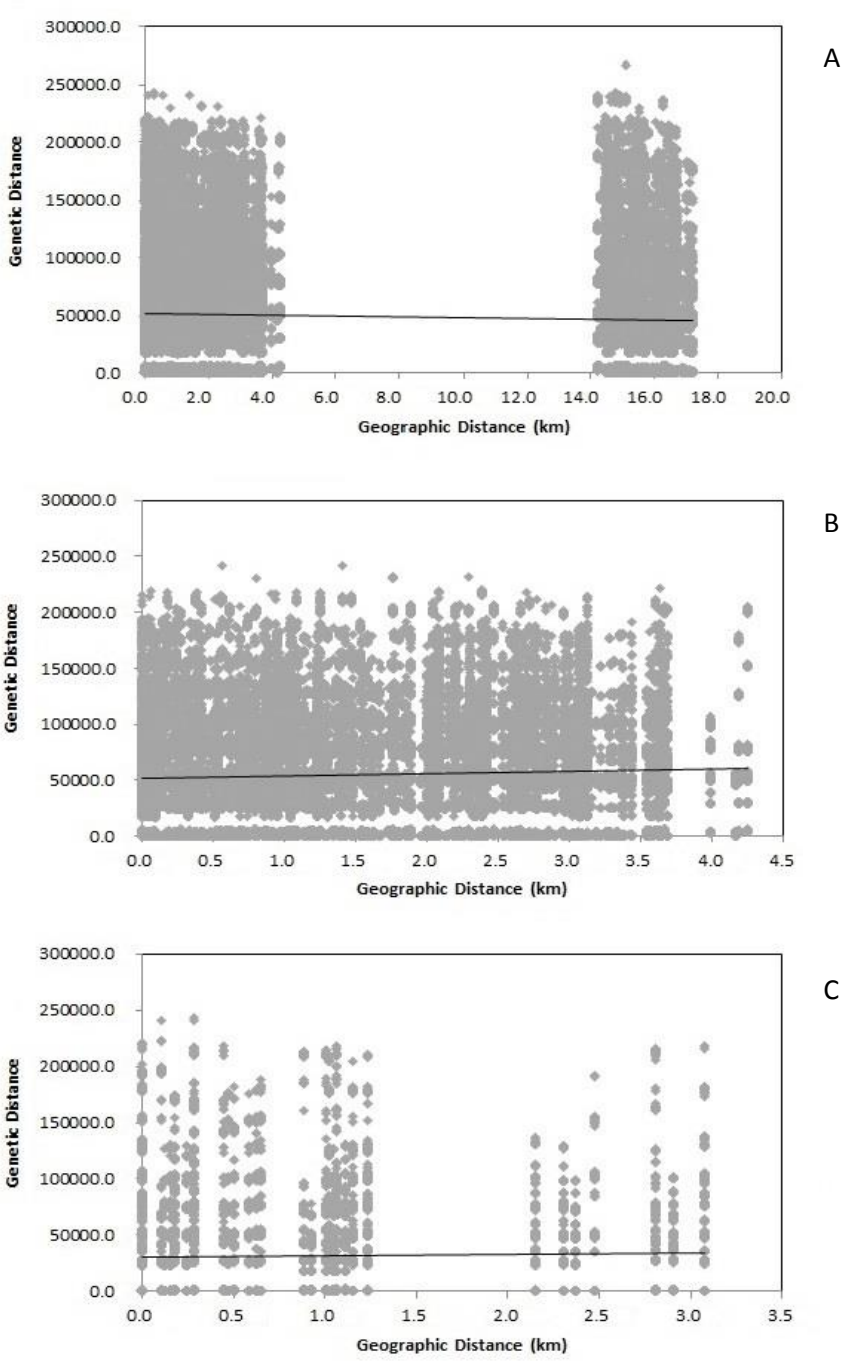

Figure 7. Isolation by distance graph of *Pityopsis ruthii* populations using chloroplast microsatellites.

A. Correlation between pairwise genetic distance values and geographic distance for all locations; B. Correlation between pairwise genetic distance values and geographic distance for locations along the Hiwassee River; C. Correlation between pairwise genetic distance and geographic distance for locations along the Ocoee River.

Figure 8. Visualization of *Pityopsis* chloroplast gene map with annotations. Length in kb and region of chloroplast on inside circle. Genes on the outer circle are color coded based on the gene name. Genes on the inside of the outer circle are minus (-) strand and genes on the outside of the outer circle are plus (+) strand.

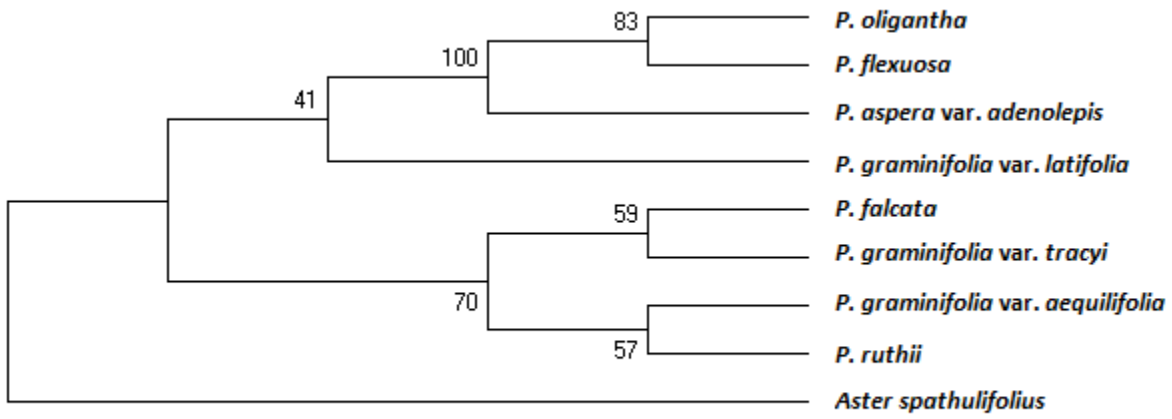

Figure 9. Phylogenetic relationships within *Pityopsis* using the maximum parsimony approach. The tree was constructed using the min-mini heuristic model in MEGA7. Numbers above or below nodes are bootstrap values calculated from 400 replicates (> 40%).

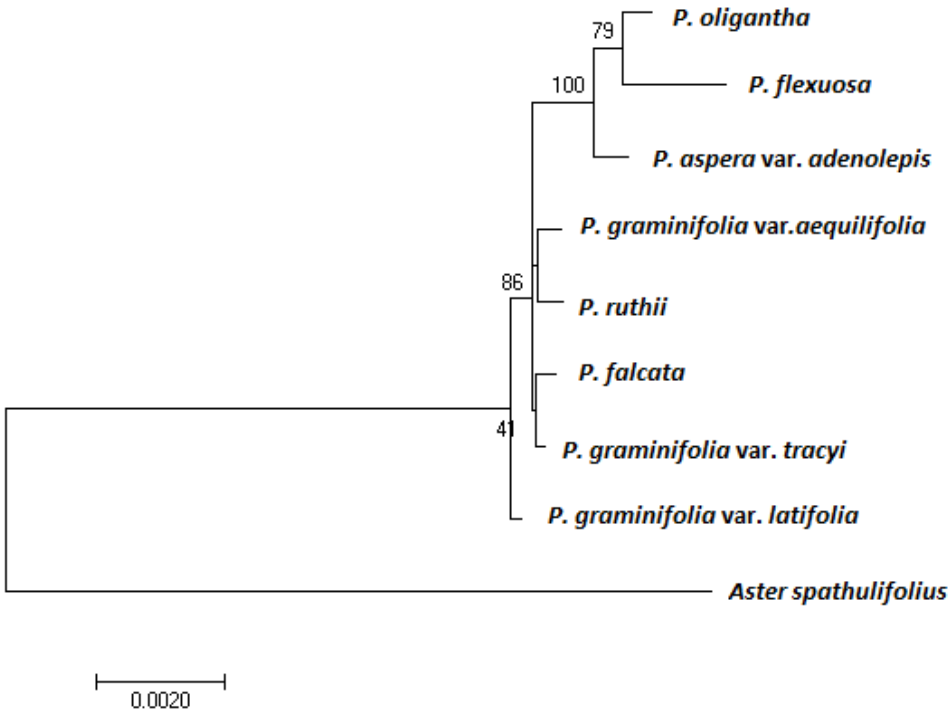

Figure 10. Phylogenetic relationships within *Pityopsis* using the maximum likelihood approach. The tree was constructed using the Kimura 2-parameter model in MEGA7. The tree with the highest log likelihood is shown. Numbers above or below nodes are percentages of trees that exhibited particular branching from 400 replicates (> 40%).

Vita

Elizabeth Anne Hatmaker was born in Knoxville, TN to parents John and Joan Hatmaker. She attended Karns High School in Knoxville, TN, graduating in 2009. She later attended the University of Tennessee, Knoxville, where she majored in Ecology and Evolutionary Biology and graduated in 2014. After graduating, she worked as a research associate under Dr. Robert Trigiano until fall of 2014, when she accepted a graduate research assistantship under Dr. Phillip Wadl and later studied under Dr. Margaret Staton. Elizabeth graduated with a Master of Science in Entomology and Plant Pathology, with a concentration in Bioinformatics and Genomics, in May of 2016.