GENETIC STRUCTURE OF THE *MECARDONIA ACUMINATA* (PLANTAGINACEAE) COMPLEX IN THE SOUTHEASTERN USA

Adjoa Richardson Ahedor
Department of Biology, Chemistry and Environmental Science
Christopher Newport University
Newport News, VA 23606
Adjoa.ahedor@cnu.edu

Jennifer Messick
Department of Biology
University of Central Oklahoma
Edmond, OK 73034

Wayne J. Elisens
Abigail J. Moore
Department of Microbiology and Plant Biology and Oklahoma Biological Survey
University of Oklahoma
Norman, OK 73019

Keywords: Lamiales, ISSR, Gratioleae, taxonomy

ABSTRACT

Mecardonia acuminata (Plantaginaceae) is found in the southeastern United States and has traditionally been divided into three varieties. A quantitative analysis of morphological data supported the division into the three varieties, although the ranges of the varieties found in that study were not the same as their traditional ranges. Here we use ISSR data to examine the relationships of 238 individuals from 23 locations throughout the range of *M. acuminata*. Although there is genetic structure that is congruent across different analyses, the groups recovered are not consistent with morphology or geography. The results indicate eastern-western distributions of the species with centers of diversity not only in the south but as far north as the Central Basin of the Interior Low Plateaus. The results further suggest ongoing diversification of lineages of *M. acuminata*, or the presence of widespread genes that govern the morphological traits that are traditionally used in delimitating the varieties.

INTRODUCTION

The genus *Mecardonia* Ruiz & Pav. (Plantaginaceae) contains approximately ten species native to North and South America, with its center of diversity in Brazil and Argentina and two species found in North America north of Mexico (Pennell 1946; Rossow 1987; Greppi et al. 2017; Ahedor 2019). Within the Plantaginaceae, *Mecardonia* is in the tribe Gratioleae (Albach et al. 2005; Scatigna et al. 2022), where it is sister to the Central and South American genus *Darcya*, with these two genera together sister to the remainder of the tribe (Scatigna et al. 2022). It is distinguished from the remaining members of the Gratioleae by having a five-lobed, zygomorphic, white or yellow corolla; four anthers with the thecae separated by

the connective; and unequal sepals (Scatigna et al. 2022).

Mecardonia acuminata (Walter) Small has the most northern distribution of any species of *Mecardonia* and is the only one that is restricted to the United States (Ahedor 2019). It is distinguished from all other species of *Mecardonia* by having white corollas that completely lack clavate hairs (Rossow 1987) and distinguished from the other North American species, *M. procumbens* (Mill.) Small, by having erect stems and white corollas with purple veins instead of spreading or prostrate stems and yellow corollas with red veins (Ahedor 2019). It is typically found on loam soil that may be acidic or sub-acidic and is usually in ditches or near streams in pineland or deciduous woodland (Pennell 1935). It ranges from Maryland to Missouri, south to Florida and Texas (Pennell 1922; Rossow 1987; Ahedor 2019). Flowering occurs through the summer followed by formation of capsule fruits throughout the fall (Pennell 1935; Rossow 1987; Wunderlin and Hansen 2003). The South American *M. tenella* (Cham. & Schltdl.) Pennell is pollinated by three types of bees, which collect fragrance, oil, and pollen (Cappellari et al. 2009) and the flowers of *M. acuminata* are also visited by bees (A. Ahedor, personal observations).

Traditionally, *M. acuminata* has been divided into three varieties or subspecies, depending on the treatment (Pennell 1935; Rossow 1987). *Mecardonia acuminata* var. *acuminata* is the most widespread and occurs throughout most of the range of the species. *Mecardonia acuminata* var. *microphylla* (Raf.) Pennell was separated by having shorter pedicels (< 10 mm according to Rossow 1987 or < 12 mm according to Pennell 1935), wider leaves that are less cuneate at the base (Pennell 1935; Rossow 1987), and wider sepals (> 2 mm; Rossow 1987). Its range is traditionally considered to be restricted to the Coastal Plain in Georgia, South Carolina, and Florida (Rossow 1987), predominantly in areas where long-leaf pine grows (Pennell 1935). The third variety, *M. acuminata* var. *peninsularis* Pennell, is traditionally considered to be restricted to peninsular Florida and is distinguished morphologically by having smaller leaves, sepals, and corollas than var. *acuminata* (Pennell 1935); by being branched at the base, instead of being branched only above the base (Pennell 1935; Rossow 1987); and by having pedicels that are ascending instead of spreading (Rossow 1987).

Ahedor and Elisens (2015) performed quantitative morphological analyses to test these hypotheses. They found that, using the classic morphological characters, specimens corresponding to the three varieties could be identified well to the north and (in the case of var. *peninsularis*) west of their traditional ranges. A canonical discriminant function analysis plot showed the three varieties to fall into separate groups. When comparing the morphological characters between varieties, they found that var. *peninsularis* had significantly smaller leaves and a significantly greater proportion of ascending (vs. divaricate) pedicels and basal (versus mid-point or intermediate) branching than the other two varieties. Variety *microphylla* had significantly shorter flowering and fruiting pedicels than the remaining two varieties. The widespread var. *acuminata* did not show any unique traits but could be distinguished by not having the distinctive characteristics of either of the other two varieties (so having larger leaves, divaricate pedicels, mid-point or intermediate branching, and longer flowering and fruiting pedicels; Ahedor and Elisens 2015).

Here we test the morphologically delimited varieties using inter simple sequence repeat (ISSR) data. ISSR are cost-effective and highly reproducible markers (Monfared et al. 2018), which have been useful for genetic studies below the species level to determine population structure (Alansi et al. 2016) and analyze genetic variability of populations (Christopoulos et

al. 2010). They have been successfully employed in the genetic characterization of the varieties of *M. procumbens* found in South America (Pérez de la Torre et al. 2010). We answer the following questions: 1) Do the ISSR data consistently divide *M. acuminata* into well-supported groups? 2) If so, do these groups correspond to the varieties delimited based on morphology? 3) If not, is there a way to delimit new infraspecific taxa that are congruent with both molecular and morphological data, or would it be better to treat *M. acuminata* as a species without infraspecific taxa?

MATERIALS AND METHODS

Sampling Strategy

Herbarium specimens of *M. acuminata* from BRIT, FLAS, GA, MO, and OKL (acronyms following Thiers (updated continuously), were initially examined to choose sampling locations of the three varieties in the southeastern USA. Varieties were identified based on characters reported by Pennell (1935), Rossow (1987), and Ahedor and Elisens (2015). In total, 238 individuals were sampled from 23 locations in seven southeastern states: Alabama, Florida, Georgia, Louisiana, Mississippi, Tennessee, and Texas (Table 1, Figure 1). No individuals were located in Oklahoma as the few habitats present had been disturbed at the time of sampling. Since Oklahoma occurs in the fringes of the range, the species is sparsely distributed compared to other states. The collected plants were also identified to variety based on Pennell (1935), Rossow (1987), and Ahedor and Elisens (2015). At 15 of the locations, the plants could be unambiguously identified as one of the three varieties, while the remaining eight locations contained individuals unambiguously identified as one of the varieties in addition to individuals that were morphologically intermediate between that variety and a second variety. In addition, several of the locations had plants that were morphologically identified as one variety but were in the traditional range of another variety. While the goal was to sample 11 individuals at each location, some locations did not have enough individuals to do that. Leaf tissues were silica-dried and stored in the freezer and voucher specimens were deposited at OKL.

ISSR Amplification and Scoring

DNA was extracted from the leaves using the modified CTAB method of Doyle and Doyle (1987). Fifty ISSR primers obtained from the University of British Columbia (UBC) were screened and seven primers that revealed both intra- and inter-location variability were selected for the study (Table 2). For each individual, ISSR regions were amplified with a single primer at a time via PCR. Total reaction mixtures of 25 μL consisted of 2.0 μL DNA, 1.5 μL of 15 μM primer, 4.0 μL of 1.25 mM dNTPs, 2.0 μL of 5U/μL Taq, 2.5 μL of 50 mM $MgCl_2$, and 1× *Taq* polymerase buffer. The PCR was performed on a MiniCycler (MJ Research Inc., South San Francisco, CA, USA) with 1.5 min at 94°C; 35 cycles of 40 sec at 94°C, 45 sec at 45°C, 1.5 min at 72°C; 40 sec at 94°C, 45 sec at 45°C and 5 min at 72°C (Wolfe and Randle 2001). All experiments included negative control reaction mixtures that had all ingredients except DNA. The PCR products were resolved on a 1.5% agarose gel in 1× TAE, with a 100 bp standard marker ladder loaded alongside to determine the size of the fragments. Gels were stained with ethidium bromide and images were visualized in UV light. Images were captured and analyzed using Kodak Digital Science ID software (Kodak, Rochester, NY, USA). Loci for each of the primers were assigned based on fragment sizes, and the ISSR data were scored as diallelic, 0 (band absent) or 1 (band present).

Table 1 Twenty-three sampling locations and 238 individuals sampled from the distribution range of *Mecardonia acuminata* in southeastern USA. Varieties represent pure varieties and intermediates. All vouchers are deposited at OKL. Location is the standard postal abbreviation for the state, followed by the parish or county within that state. The location code is given in parentheses, if different from the state and county/parish. N is the number of individuals sampled from that population.

Varieties based on morphology	Variety based on geography	Voucher	Location	Latitude (°N)	Longitude (°W)	N
peninsularis	*peninsularis*	Elisens 1061	FL Citrus	28.7295	82.2715	11
peninsularis	*peninsularis*	Elisens 1064	FL Levy	29.4415	82.6365	11
peninsularis	*peninsularis*	Elisens 1141	FL Polk	28.3109	82.0561	11
microphylla	*microphylla*	Elisens 1059	FL Liberty (FL Libe)	30.2043	84.7483	11
microphylla	*microphylla*	Elisens 1058	FL Calhoun (FL Calh)	30.4072	85.1622	11
acuminata	*acuminata*	Ahedor 112	AL Franklin (AL Fran)	34.4820	87.6490	11
acuminata	*acuminata*	Ahedor 113	AL Lawrence (AL Lawr)	34.4880	87.5007	11
acuminata	*acuminata*	Ahedor 101	LA Allen (LA Alln)	30.5185	93.0152	11
acuminata	*acuminata*	Ahedor 102	LA Beauregard (LA Beau)	30.5100	93.2328	10
acuminata	*acuminata*	Ahedor 103	TX Nacadoches (TX Naca)	31.6190	94.6832	11
acuminata	*acuminata*	Ahedor 105	TN Marshall (TN Marsh)	35.6251	86.8105	7
acuminata	*acuminata*	Ahedor 106	TN Maury (TN Maur)	35.5872	86.8975	10
acuminata	*acuminata*	Ahedor 104	TN Rutherford 1 (TN Ruth 1)	35.7394	86.5955	11
acuminata	*acuminata*	Ahedor 108	TN Rutherford 2 (TN Ruth 2)	35.6551	86.4576	7
acuminata	*acuminata*	Ahedor 109	TN Rutherford 3 (TN Ruth 3)	35.8738	86.2844	8
acuminata + peninsularis	*acuminata*	Elisens 1057	AL Covington (AL Covi)	31.1718	86.2908	11
acuminata + peninsularis	*acuminata*	Elisens 1056	MS George (MS Geor)	30.7791	88.7171	11
acuminata + peninsularis	*acuminata*	Ahedor 111	TN Rutherford 4 (TN Ruth 4)	36.0590	86.4847	11
acuminata + peninsularis	*acuminata*	Ahedor 107	TN Bedford (TN Bedf)	35.6772	86.5223	10
acuminata + peninsularis	*acuminata*	Elisens 1053	LA St. Tammany (LA St.Tm)	30.4962	90.1988	11
acuminata + microphylla	*acuminata*	Elisens 1047	LA Winn	31.7532	92.9170	11
peninsularis + microphylla	*acuminata*	Elisens 1066	GA Wilcox (GA Wilc)	31.9488	83.5589	11
peninsularis + microphylla	*acuminata*	Ahedor 110	TN Wilson (TN Wils)	36.0274	86.3673	10

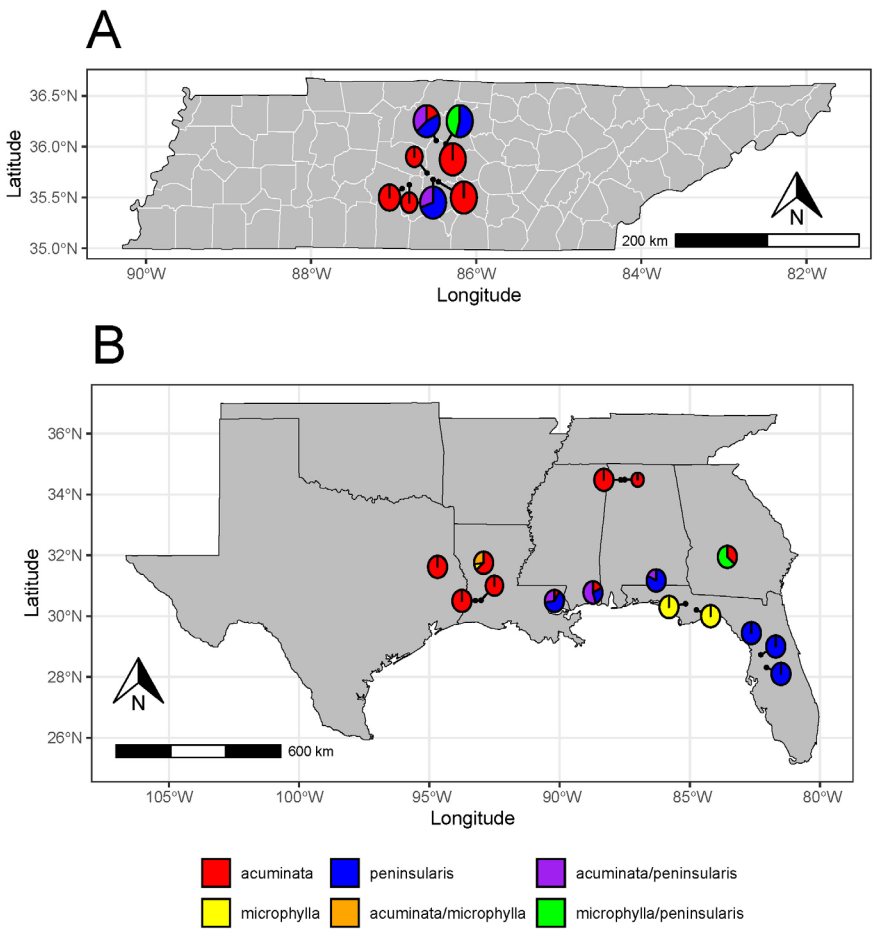

Figure 1 Map of southeastern USA showing sampled locations based on morphology. (A) Sampled locations in Tennessee. (B) Sampled locations in remaining states. Circles with a single color indicate locations with a single variety, while circles with two colors indicate locations with intermediates.

Table 2 Attributes of ISSR primers used to generate markers from 238 individuals sampled for *Mecardonia acuminata*.

Primer	Sequence	Total number of Loci (Grand Total = 94)	Range of Fragment Sizes (bp)	Number of Genotypes per Location
UBC 807	$(AG)_8T$	14	215 – 1400	4 –11
UBC 809	$(AG)_8G$	13	204 – 1500	2 – 9
UBC 812	$(GA)_8A$	15	220 – 1400	2 – 11
UBC 815	$(CT)_8G$	14	230 – 1500	3 – 11
UBC 836	$(AG)_8YA$	18	180 – 2700	3 – 11
UBC 842	$(GA)_8YG$	8	200 – 900	1 – 7
UBC 845	$(CT)_8RG$	12	260 – 1700	3 – 10

Population Genetic Statistics

The level of genetic variation was assessed per primer. The number of genotypes per primer was estimated, and the total number of ISSR loci was estimated for each sampling location. Alleles that occurred in more than half of the locations were considered common alleles, while alleles that only occurred in a single population were considered private alleles. The ISSR data were analyzed using POPGENE version 1.31 (Yeh et al. 1997) to determine allelic diversity, genetic diversity, genetic differentiation, and gene flow. Percentage polymorphism (P) was calculated as the number of polymorphic loci divided by the total number of loci obtained for a primer (Nei 1987). Nei's genetic diversity was used to estimate expected heterozygosity (H_E), and Shannon's index of phenotypic diversity was used to estimate observed heterozygosity (H_O) (Yeh et al. 1997). Levels of genetic differentiation (Nei 1972, 1973) were estimated as: average gene diversity in total (all) locations (H_T), within locations (H_S), and among locations (G_{ST}). G_{ST} measures genetic differentiation among subpopulations (locations) due to combined effects of all evolutionary forces and is ideal for non-model systems (Nei 1973). Gene flow (Nm, where N is the overall sample size and m is the fraction of immigrants per generation) was estimated as $(1/G_{ST})/4G_{ST}$ (Ellstrand and Elam 1993).

To assess isolation by distance, pairwise F_{ST} values were estimated for pairs of locations (in POPGENE). Geographic distances between each location were determined using PASSAGE 2.0 (Rosenberg and Anderson 2011). These two distance matrices were then compared using a Mantel test (Mantel, 1967) in PASSAGE.

Population Structure Analysis

Population structure was analyzed in three different ways. An analysis was performed in *adegenet* version 2.1.5 (Jombart 2008; Jombart and Ahmed 2011; RRID:SCR_000825), also using *ade4* version 1.7-18 (Dray and Dufour 2007) in R version 4.1.1 (R Core Team 2021). For these analyses, individuals were divided into two or three groups. Individuals were initially clustered with the *find.clusters* command and 50 principal components (PCs). A DAPC analysis (Jombart et al. 2010) was then performed on the clustered individuals with the optimal number of PCs as shown with the *optim.a.score* command. Ten analyses each were run with the plants divided into two or three groups. In preliminary analyses, *adegenet* placed individuals with missing data into their own group, despite the fact that these individuals came from several different populations. Therefore, the analyses presented here were run on a reduced dataset of 228 individuals, with the 10 individuals containing missing data removed.

STRUCTURE v.2.3.4 (Pritchard et al. 2000; RRID:SCR_002151) was run using the No Admixture model (Pritchard et al. 2010), following the recommendations for dominant markers such as ISSR (Falush et al. 2007). The number of groups (K) tested ranged from 1 – 23, with 10 runs for each value of K. These analyses were run with correlated allele frequencies, as recommended for closely related groups (Falush et al. 2003, 2007; Pritchard et al. 2010); an inferred value of alpha; and a burn-in of 100,000 generations followed by an analysis of 100,000 generations.

A Q-matrix was analyzed with K = 2 – 23 and visualized using STRUCTURE HARVESTER (Earl and vonHoldt 2012; RRID:SCR_017636). The variance across all iterations of each value of K was then minimized using CLUMPP (Jakobsson and Rosenberg 2007), and the optimal value of K was identified graphically using DISTRUCT (Rosenberg 2004).

Maps were made in R using ggplot2 version 3.4.0 (Wickham 2016;

RRID:SCR_014601), ggspatial version 1.1.5 (Dunnington 2022), maps version 3.4.0 (Deckmyn 2022; RRID:SCR_019296), scatterpie version 0.1.8 (Yu 2022), dplyr version 1.0.7 (Wickham et al. 2022; RRID:SCR_016708), and sf version 1.0-3 (Pebesma 2018).

Splits networks were constructed using SplitsTree6 (Huson and Bryant 2006). Two different analyses were performed, one with all of the individuals and one with the same 228 individuals used in the *adegenet* analyses. In both cases, Hamming Distances were used, and the bootstrap analysis consisted of 100 bootstrap replicates.

RESULTS

Population Genetic Statistics

Ninety-five loci were scored for all seven primers with a range of 8 (UBC 842) to 18 (UBC 836) loci per primer (Table 2). The largest fragment size range scored was for UBC 836, with 180 – 2700 bp, and the smallest range scored was for UBC 842 with 200 – 900 bp. The number of genotypes (unique banding patterns) per primer and sampling location ranged from one (no variation among individuals) to 11 (variation in all individuals).

A mean of 46.5% was obtained for percentage polymorphism (P), the lowest was 32.63% (TN Bedf and TN Maur), and the highest was 58.95% (TN Ruth 2); all these locations were in the Central Basin of the Interior Low Plateaus (Table 3). Mean genetic diversity within location (H_S) was 0.153, and total genetic diversity for all locations (H_T) was 0.239. The average genetic diversity among locations (G_{ST}) was estimated to be 0.361, and the level of gene flow (Nm) was estimated to be 0.887. Observed heterozygosity (H_O) was higher than expected heterozygosity (H_E) for all locations (Table 3), with H_O 0.231±0.034 and H_E 0.153±0.024.

The Mantel Test indicated that there was not a significant correlation between genetic and geographic distance (χ = 234.796, r^2 = 0.008, t = - 0.123, p = 0.90184).

Population Structure

The networks from the SplitsTree analyses were relatively unresolved, with no groups of > 11 individuals that had bootstrap support over 10% in either analysis.

Analyses with *adegenet* showed consistent results when the individuals were divided into two groups (Figure 2). All ten replicate runs showed the same division of individuals, and all but one replicate had one PC as optimal with an identical eigenvalue and proportion of conserved variance. While there was much more variation in optimal number of PCs, eigenvalues, and proportion of conserved variance when the individuals were divided into three groups, there were still only two optimal solutions, each of which was found five times (Figure 3, Figure 4). However, none of the results corresponded to population or geography.

Bayesian clustering using STRUCTURE, and subsequent analyses using CLUMPP and DISTRUCT to estimate ΔK (following Evanno et al. 2005) revealed K = 2 as the best value, with a second optimum at K = 3 (Figure 5). Similar to the *adegenet* results, the STRUCTURE results for all ten replicate runs with K = 2 and K = 3 showed the same grouping of individuals in each run (Figure 6, Figure 7). Although none of these results corresponded to population or geography, the *adegenet* and STRUCTURE analyses in which the individuals were separated into two groups gave congruent results. In addition, both of the *adegenet* results for three groups of individuals and the STRUCTURE results for three groups of individuals gave congruent results for the

populations outside of Tennessee. For the populations in Tennessee, the STRUCTURE results for three groups corresponded to one of the two *adegenet* results.

Table 3 Genetic variability at 95 ISSR loci in 23 sampling locations of *Mecardonia acuminata*. P = Percentage of polymorphic loci. Common loci are loci present in at least 52 % of locations sampled and rare alleles present in 48% of locations. H_E = Nei's genetic diversity (expected heterozygosity), H_O = Shannon's index of phenotypic diversity (observed heterozygosity).

Varieties Based on Morphology	Location	Total Loci	Common Loci	Private alleles	P	H_E (s.d.)	H_O (s.d.)
peninsularis	FL Citrus	68	60	0	52.63	0.164 (0.19)	0.25 (0.28)
	FL Levy	82	68	1	57.89	0.183 (0.19)	0.279 (0.28)
	FL Polk	55	48	0	36.68	0.119 (0.19)	0.177 (0.270)
microphylla	FL Libe	62	55	0	53.68	0.163 (0.19)	0.250 (0.27)
	FL Calh	52	49	0	48.42	0.165 (0.20)	0.244 (0.29)
acuminata	AL Lawr	51	47	0	36.84	0.132 (0.19)	0.197 (0.27)
	AL Fran	57	53	0	44.21	0.128 (0.18)	0.197 (0.26)
	LA Alln	64	54	0	36.84	0.110 (0.18)	0.169 (0.26)
	TN Bedf	53	50	0	32.63	0.125 (0.20)	0.184 (0.28)
	LA Beau	53	50	1	51.58	0.162 (0.20)	0.244 (0.29)
	TN Maur	61	55	0	32.63	0.165 (0.20)	0.247 (0.27)
	TN Ruth1	53	48	0	41.05	0.135 (0.19)	0.205 (0.27)
	TN Ruth 2	65	58	2	58.95	0.190 (0.19)	0.289 (0.28)
	TN Ruth 3	56	48	0	42.21	0.154 (0.20)	0.230 (0.28)
	TX Naca	55	50	0	47.37	0.155 (0.19)	0.234 (0.28)
acuminata + peninsularis	AL Covi	63	53	0	44.21	0.144 (0.19)	0.218 (0.28)
	MS Geor	70	52	6	55.76	0.165 (0.19)	0.254 (0.27)
	TN Ruth 4	70	62	0	56.85	0.168 (0.19)	0.258 (0.27)
	LA St.Tm	68	55	1	48.42	0.170 (0.20)	0.253 (0.29)
acuminata + microphylla	LA Winn	70	55	1	50.53	0.167 (0.12)	0.251 (0.29)
peninsularis + microphylla	GA Wilc	61	55	0	52.63	0.169 (0.19)	0.258 (0.27)
	TN Marsh	55	51	0	44.32	0.141 (0.18)	0.216 (0.27)
	TN Wils	59	51	0	43.16	0.135 (0.19)	0.206 (0.27)
Mean		**61 ± 8.15**	**54 ± 5.05**	**NA**	**46.5 ± 7.97**	**0.153±0.024**	**0.231±0.035**

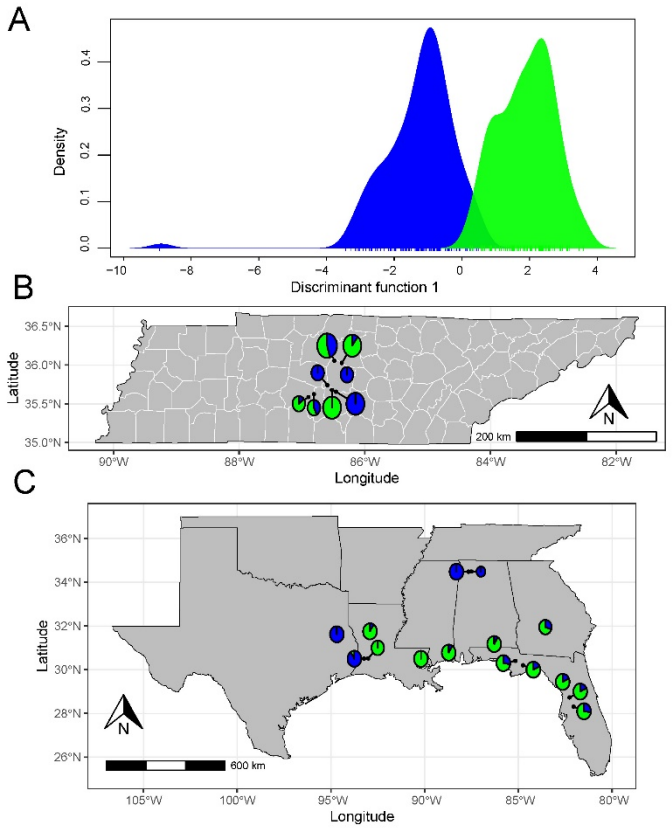

Figure 2 Results from the *adegenet* analysis of the 228 individuals without missing data, in which the plants were divided into two groups. (A) Plot from *adegenet* with the axis explaining 10.6% of the variation. (B) Distribution of the two groups within Tennessee. (C) Distribution of the two groups at the remaining sites.

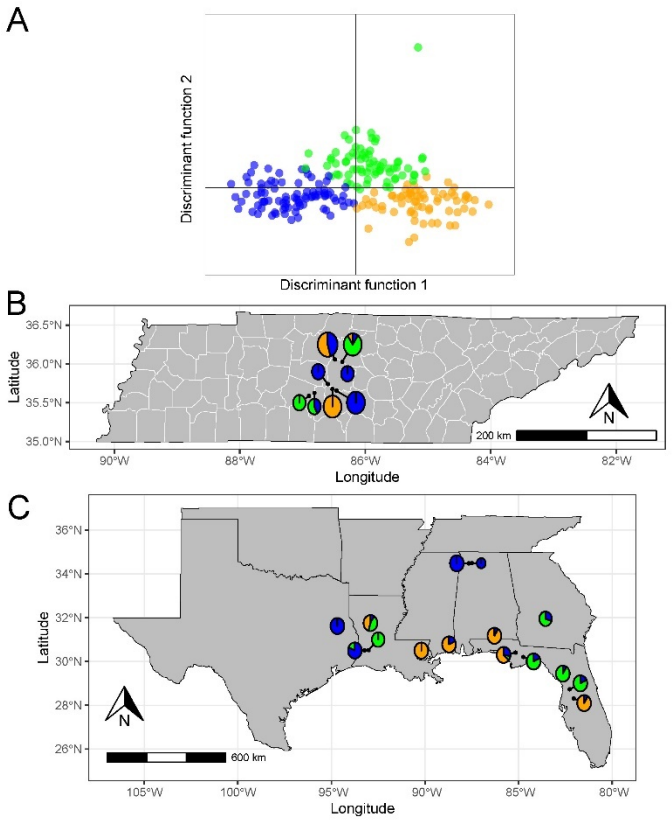

Figure 3 Results from the *adegenet* analysis of the 228 individuals without missing data, in which the plants were divided into three groups, division of individuals found in five of the ten runs. (A) Plot from *adegenet* with the division into groups explaining 33.7% of the variation. (B) Distribution of the three groups within Tennessee. (C) Distribution of the three groups at the remaining sites.

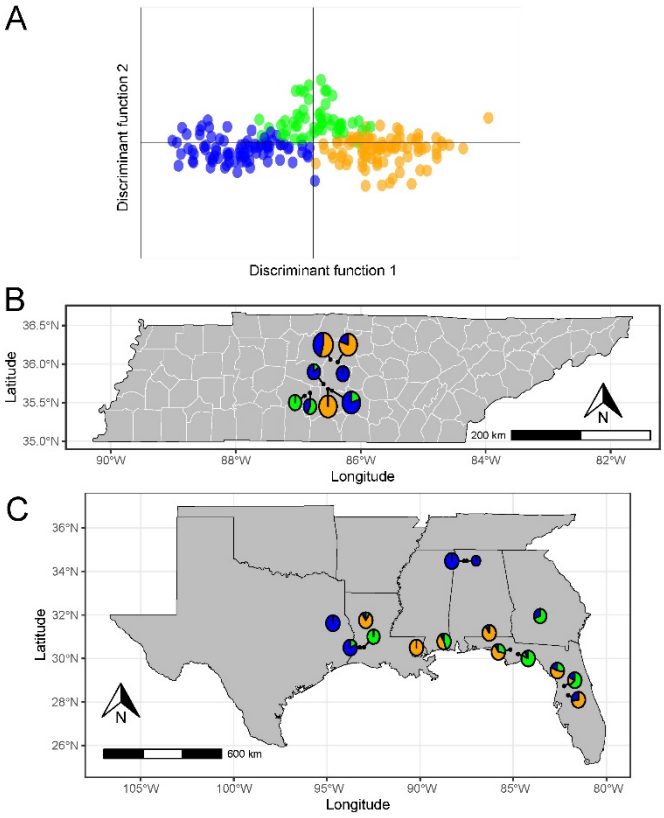

Figure 4 Results from the *adegenet* analysis of the 228 individuals without missing data, in which the plants were divided into three groups, division of individuals found in the remaining five runs. (A) Plot from *adegenet* with the division into groups explaining 37% of the variation. (B) Distribution of the three groups within Tennessee. (C) Distribution of the three groups at the remaining sites.

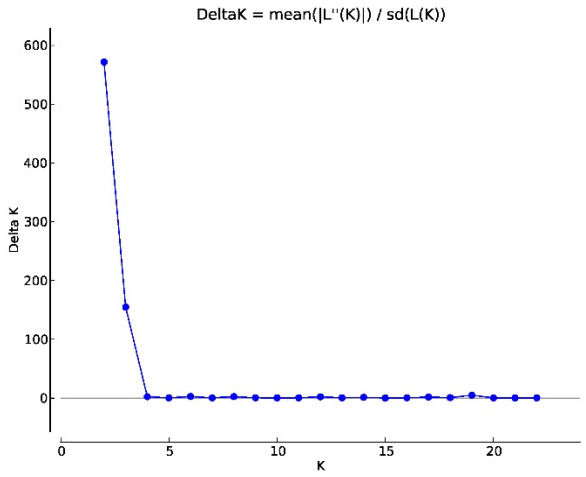

Figure 5 Delta K plot of results from CLUMPP and DISTRUCT analyses for STRUCTURE show two groups is favored for optimal value of K, with three groups being second best.

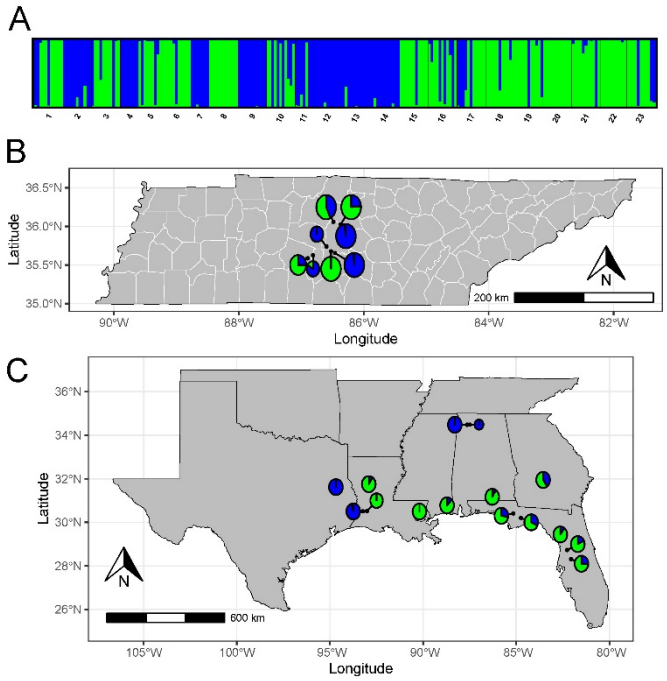

Figure 6 Results from the STRUCTURE analysis of all 238 individuals, in which the plants were divided into two groups. (A) STRUCTURE plot. (B) Distribution of the two groups within Tennessee. (C) Distribution of the two groups at the remaining sites. Populations are numbered as follows: 1: FL Calh, 2: LA Beau, 3: TN Wils, 4: AL Lawr, 5: TN Maur, 6: LA Winn, 7: TN Ruth 1, 8: TN Bedf, 9: Al Fran, 10: GA Wilc, 11: TN Marsh, 12: TX Naca, 13: TN Ruth 3, 14: TN Ruth 2, 15: AL Covi, 16: FL Libe, 17: TN Ruth 4, 18: FL Levy, 19: FL Citrus, 20: LA St. Tim, 21: MS Geor, 22: LA Alln, and 23: FL Polk.

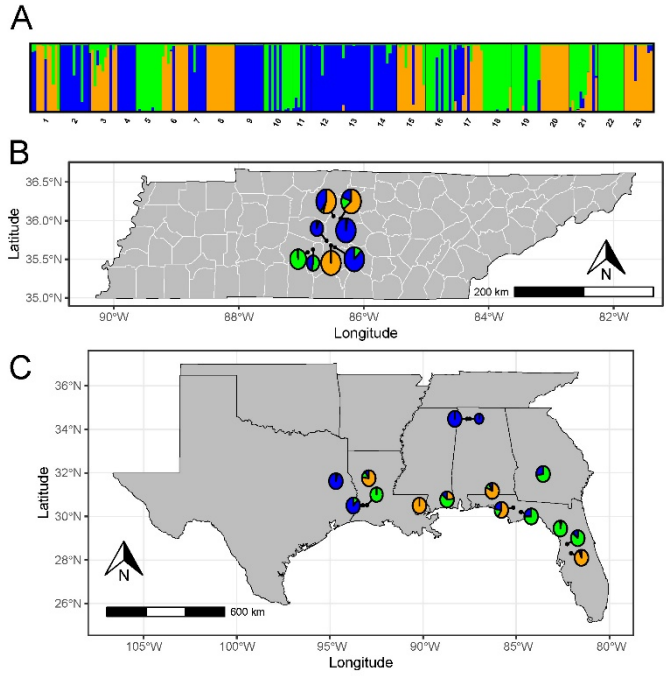

Figure 7 Results from the STRUCTURE analysis of all 238 individuals, in which the plants were divided into three groups. (A) STRUCTURE plot. (B) Distribution of the three groups within Tennessee. (C) Distribution of the three groups at the remaining sites. Populations are numbered as follows: 1: FL Calh, 2: LA Beau, 3: TN Wils, 4: AL Lawr, 5: TN Maur, 6: LA Winn, 7: TN Ruth 1, 8: TN Bedf, 9: Al Fran, 10: GA Wilc, 11: TN Marsh, 12: TX Naca, 13: TN Ruth 3, 14: TN Ruth 2, 15: AL Covi, 16: FL Libe, 17: TN Ruth 4, 18: FL Levy, 19: FL Citrus, 20: LA St. Tim, 21: MS Geor, 22: LA Alln, and 23: FL Polk.

DISCUSSION

Both *adegenet* and STRUCTURE analyses of the genetic data showed patterns that were largely congruent. However, the molecular groups were not congruent with the morphological classification of the populations. The morphological classification showed all groups outside of Georgia and Florida to be either pure var. *acuminata* or var. *acuminata* admixed with one of the other varieties. The three populations in peninsular Florida were var. *peninsularis*, the two populations in the Florida

panhandle were var. *microphylla*, and the single Georgia population was intermediate between those two varieties. In contrast, the molecular results show that the two populations from northern Alabama, one population from Louisiana, the single population from Texas, and three or four of the eight populations in Tennessee form one group (either one of two groups or one of three groups, depending on the analysis). When the plants were divided into three groups, the three populations from peninsular Florida were in two different groups, as were the two populations from the Florida panhandle.

As molecular and morphological data are not congruent in *M. acuminata* and neither type of data recovers groups that are geographically coherent, we cannot support the recognition of infraspecific taxa. It may be that *M. acuminata* is in the process of diversifying, but that the new lineages have not been independent long enough to become distinct with molecular or morphological data. However, it is also possible that the genes that govern the traits that define each of the three morphological varieties are widespread throughout *M. acuminata* as a whole, making the whole species polymorphic for those traits and allowing them to predominate wherever in the range of *M. acuminata* they are most favorable. This interpretation is supported by the large number of intermediate individuals and populations and the fact that these putative intermediates are not only on the edges of the ranges of the proposed varieties, but also well within their core ranges.

The molecular data show evidence of population structure that is recoverable in two of the three analyses (with SplitsTree not showing any significant large-scale patterns). It is not clear what the cause of this signal is. It may be that there is a northern/western and an eastern/southern group, but that a lack of sampling in the northwestern corner of the range obscures this pattern. Similar east-west divisions have been found in a number of other groups from the southeastern United States (e.g., Soltis et al. 2006; Barrow et al. 2017; Myers et al. 2020). Explanations for these patterns include barriers to dispersal formed by various different major rivers in the area (e.g., Soltis et al. 2006; Wallace and Doffitt 2013; Hatmaker et al. 2018; Lyman and Edwards 2022) and recolonization from separate refugia in Texas, the Florida Peninsula, and potentially elsewhere along the Gulf coast (e.g., Barrow et al. 2017; Myers et al. 2020; Naranjo et al. 2023).

However, other groups also show either a complex molecular pattern (e.g., Wallace and Doffitt 2013) or little geographic signal in the data (e.g., Konrade et al. 2019). Given that there have been repeated cycles of glaciation throughout the Pleistocene, there would have been multiple north-south cycles of movement. Thus, we would not necessarily expect the same lineages to always retreat into the same refugia each glacial cycle. Rather, lineages that originated from one refugium could retreat into another refugium during the next glaciation, thus leading to complex patterns of the type we find here. Wallace and Doffitt (2013) hypothesized that in *Trillium*, the combination of a complex Pleistocene history, low dispersal capabilities, and discontinuous habitat patches may have led to its complex genetic patterns, as is likely the case with *M. acuminata*.

AUTHOR CONTRIBUTIONS

ARA designed the study, collected the plants, performed data analysis, and wrote the initial draft of the manuscript. JM performed data analysis. WJE assisted with study design, plant collection, and data analysis. AJM performed data analysis and wrote the final draft of the manuscript. All authors edited the manuscript and approved the final version.

FUNDING

The research was partly funded by the University of Oklahoma Graduate Student Research Grant and the Microbiology and Plant Biology Department of the University of Oklahoma.

ACKNOWLEDGMENTS

The authors thank Brent Berger, Dwayne Estes, and Michael MacRoberts for assistance with fieldwork; Nancy Zehrbach for assistance with ISSR gel analysis and software; and Tina Lee for assistance in the lab. Appreciation also goes to Amy Buthod, Leann Monaghan, Alexa Beemer, and John Ahedor for providing comments on the manuscript. The authors also wish to thank reviewers of this manuscript for their insightful comments.

LITERATURE CITED

Ahedor, A.R. and E. Elisens. 2015. Morphological analyses of the *Mecardonia acuminata* (Plantaginaceae) species complex in the Southeastern USA. *Southeastern Naturalist* 14:173-196.

Ahedor, A.R. 2019. *Mecardonia*. In: Flora of North America Editorial Committee, eds. *Flora of North America North of Mexico*, Vol 17. New York and Oxford: Oxford University Press.

Alansi, S., M. Tarroum, F. Al-Qurainy, S. Khan, and M. Nadeem. 2016. Use of ISSR markers to assess the genetic diversity in wild medicinal *Ziziphus spina-christi* (L.) Willd. collected from different regions of Saudi Arabia. *Biotechnology and Biotechnological Equipment* 30:942-947.

Albach, D.C., H.M. Meudt, and B. Oxelman. 2005. Piecing together the "new" Plantaginaceae. *American Journal of Botany* 92:297-315.

Barrow, L.N., J.A. Soto-Centeno, A.R. Warwick, A.R. Lemmon, and E.M. Lemmon. 2017. Evaluating hypotheses of expansion from refugia through comparative phylogeography of south-eastern Coastal Plain amphibians. *Journal of Biogeography* 44:2692-2705.

Cappellari, S.C., B. Harter-Marques, P. Aumeier, and W. Engels. 2009. *Mecardonia tenella* (Plantaginaceae) attracts oil-, perfume-, and pollen-gathering bees in southern Brazil. *Biotropica* 41: 721-729.

Christopoulos, M.V., D. Rouskas, E. Tsantili, and P.J. Bebeli. 2010. Germplasm diversity and genetic relationships among walnut (*Juglans regia* L.) cultivars and Greek local selections revealed by Inter-Simple Sequence Repeats (ISSR) markers. *Scientia Horticulturae* 125:584-592.

Deckmyn, A. 2022. *Maps: R package for geographical maps*. https://www.rdocumentation.org/packages/maps/versions/3.4.1

Dray, S. and A.B. Dufour. 2007. The ade4 package: implementing the duality diagram for ecologists. *Journal of Statistical Software* 22:1-20.

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

Dunnington, D. 2022. *ggspatial: Spatial Data Framework for ggplot2*. https://paleolimbot.github.io/ggspatial/, https://github.com/paleolimbot/ggspatial

Earl, D.A. and B.M. vonHoldt. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* 4:359-361.

Ellstrand, N.C. and D.R. Elam. 1993. Population genetic consequences of small population size: implications for plant conservation. *Annual Review of Ecology and Systematics* 24:217-242.

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.

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

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

Greppi, J.A., M. de las Mercedes Sosa, and M. Dematteis. 2017. A new polyploid species of *Mecardonia* (Gratioleae, Plantaginaceae) from South America. *Phytotaxa* 303:264-270.

Hatmaker, E.A., M.E. Staton, A.J. Dattilo, D. Hadziabdic, T.A. Rinehart, E.E. Schilling, R.N. Trigiano, and P.A. Wadl. 2018. Population structure and genetic diversity within the endangered species *Pityopsis ruthii* (Asteraceae). *Frontiers in Plant Science* 9:943. doi: 10.3389/fpls.2018.00943

Huson, D.H. and D. Bryant. 2006. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* 23:254-267.

Jakobsson, M. and N.A. Rosenberg. 2007. CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23:1801-1806.

Jombart, T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24:1403-1405.

Jombart, T. and I. Ahmed. 2011. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* 27:3070-3071.

Jombart, T., S. Devillard, and F. Balloux. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics* 11:94. doi: 10.1186/1471-2156-11-94

Konrade, L., J. Shaw, and J. Beck. 2019. A rangewide herbarium-derived dataset indicates high levels of gene flow in black cherry (*Prunus serotina*). *Ecology and Evolution* 9:975-985.

Lyman, R.A. and C.E. Edwards. 2022. Revisiting the comparative phylogeography of unglaciated eastern North America: 15 years of patterns and progress. *Ecology and Evolution* 12:e8827. doi: 10.1002/ece3.8827

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

Monfared, M.A., D. Samsampour, G.R. Sharifi-Sirchi, and F. Sadeghi. 2018. Assessment of genetic diversity in *Salvadora persica* L. based on inter simple sequence repeat (ISSR) genetic marker. *Journal of Genetic Engineering and Biotechnology* 16:661-667.

Myers, E.A., A.D. McKelvy, and F.T. Burbrink. 2020. Biogeographic barriers, Pleistocene refugia, and climatic gradients in the southeastern Nearctic drive diversification in cornsnakes (*Pantherophis guttatus* complex). *Molecular Ecology* 29:797-811.

Naranjo, A.A., R.A. Folk, M.A. Gitzendanner, D.E. Soltis, and P.S. Soltis. 2023. Ancestral area analyses reveal Pleistocene-influenced evolution in a clade of coastal plain endemic plants. *Journal of Biogeography* 50:393-405.

Nei, M. 1972. Genetic distance between populations. *American Naturalist* 106:283-292.

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

Nei, M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* 89:583-590.

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

Pebesma, E. 2018. Simple features for R: standardized support for spatial vector data. *The R Journal* 10:439-446.

Pennell, F.W. 1922. Some overlooked Scrophulariaceae of Rafinesque. *Torreya* 22:77-84.

Pennell, F.W. 1935. The Scrophulariaceae of eastern temperate North America. *The Academy of Natural Sciences of Philadelphia Monograph* 1, 1-650.

Pennell, F.W. 1946. Reconsideration of the *Bacopa-Herpestis* problem of the Scrophulariaceae. *Proceedings of the Academy of Natural Sciences of Philadelphia* 98:83-98.

Pérez de la Torre, M., P. Zirilli, N. Ulrich, L. Setten, and A. Escandón. 2010. Caracterización molecular en el género *Mecardonia* Ruiz & Pav. (Plantaginaceae) utilizando marcadores ISSR. *Revista de la Facultad de Agronomia Universidad Nacional De La Plata* 109:23-30.

Pritchard, J.K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155:945-959.

Pritchard, J.K., X. Wen, and D. Falush. 2010. Documentation for STRUCTURE Software:Version 2.3. https://web.stanford.edu/group/pritchardlab/structure_software/release_versions/v2.3.4/structure_doc.pdf

R Core Team. 2021. R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing. https://www.R-project.org/

Rosenberg, M.S. and C.D. Anderson. 2011. PASSaGE. Pattern Analysis, Spatial Statistics and Geographic Exegesis. Version 2. *Methods in Ecology and Evolution* 2:229-232.

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

Rossow, R.A. 1987. Revisión del género *Mecardonia* (Scrophulariaceae). *Candollea* 42:431-474.

Scatigna A.V., V. Castro Souza, M. de las Mercedes Sosa, G. dalla Colletta, R. Moura Machado, and A. Olmos Simões. 2022. Phylogenetics of Gratioleae (Plantaginaceae): paraphyly of *Stemodia* and its implications for generic circumscriptions, with insights from floral evolution. *Botanical Journal of the Linnean Society* 200:194-217.

Soltis, D.E., A.B. Morris, J.S. McLachlan, P.S. Manos, and P.S. Soltis. 2006. Comparative phylogeography of unglaciated eastern North America. *Molecular Ecology* 15:4261-4293.

Thiers, B.M. (updated continuously). *Index Herbariorum*. https://sweetgum.nybg.org/science/ih/

Wallace, L.E. and C.H. Doffitt. 2013. Genetic structure of the mesic forest-adapted herbs *Trillium cuneatum* and *Trillium stamineum* (Melanthiaceae) in the south-central United States. *Castanea* 78:154-162.

Wickham, H. 2016. *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag.

Wickham, H., R. François, L. Henry, K. Müller, and D. Vaughn. 2022. *dplyr: A Grammar of Data Manipulation*. https://dplyr.tidyverse.org, https://github.com/tidyverse/dplyr

Wolfe, A.D. and C.P. Randle. 2001. Relationships within and among species of the holoparasitic genus *Hyobanche* (Orobanchaceae) inferred from ISSR banding patterns and nucleotide sequences. *Systematic Botany* 26:120-130.

Wunderlin, R.P. and B.F. Hansen. 2003. *Guide to the Vascular Plants of Florida*. Gainesville (FL): University Press of Florida.

Yeh, F.G., R.-C. Yang, B.J.T. Boyle, Z.-H. Ye, and J.X. Mao. 1997. POPGENE, the user friendly software for population genetic analysis. University of Alberta (Canada): Molecular Biology and Biotechnology Centre. https://sites.ualberta.ca/~fyeh/popgene.html (6 June 2018).

Yu, G. 2022. *scatterpie: scatter pie plots*. https://CRAN.R-project.org/package=scatterpie

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in the Dryad Digital Repository at https://doi.org/10.5061/dryad.4b8gthtj3.