2022

Determinants of Population Genetic Structure in *Chamaecrista fasciculata* (Michx.) Greene (Fabaceae) in the Southeastern United States

Mahboubeh Hosseinalizadeh Nobarinezhad
Old Dominion University, mnobarin@odu.edu

Lisa E. Wallace
Old Dominion University, lewallac@odu.edu

Original Publication Citation

Nobarinezhad, M. H., & Wallace, L. E. (2022). Determinants of population genetic structure in *Chamaecrista fasciculata* (Michx.) Greene (Fabaceae) in the southeastern United States. *Castanea*, 87(1), 129-146.

CASTANEA 87(1): 129–146 May

Determinants of Population Genetic Structure in *Chamaecrista fasciculata* (Michx.) Greene (Fabaceae) in the Southeastern United States

Mahboubeh Hosseinalizadeh Nobarinezhad[1] and Lisa E. Wallace[2,*]
[1]Department of Biological Sciences, Mississippi State University, Mississippi State, MS 39762
[2]Department of Biological Sciences, Old Dominion University, Norfolk, VA 23529

ABSTRACT

Chamaecrista fasciculata is a widely distributed, phenotypically variable species in the eastern U.S. Whereas studies have demonstrated genetic structure and local adaptation in northern areas of its distribution, there has been no comparison of genetic variability among populations at the southern extent where phenotypic variation is more complex. We characterized genetic variation at 14 microsatellite loci for populations in Mississippi and Alabama and compared this to variation in a phenotypic trait, leaf pubescence. Geographic distance, climatic variables, and elevation were evaluated as factors to explain the observed patterns of genetic diversity. A significant amount of variation (19%) resided among populations, but most variation (68%) was among individuals. Assignment of individuals into genetic groups suggests two primary clusters, but these groups are not concordant with known geographical or ecological breaks, nor phenotypic variants. Genetic structure at a regional scale can be characterized as isolation by distance, while environmental factors may play a secondary role in limiting gene flow at local scales. Mean population F_{ST} is strongly associated with allelic diversity and heterozygosity, suggesting that genetic drift influences population variation. Despite the presence of genetic and phenotypic variation in southern populations of *C. fasciculata*, the lack of concordant patterns between these types of variation indicate that they are not driven by the same factors. This study demonstrates how local factors differentially influence the maintenance of intraspecific variation and suggest the southern distributional range is an active area of evolution for *C. fasciculata*.

Key words: *Chamaecrista fasciculata*, climatic variation, microsatellite markers, genetic structure, phenotypic variation.

INTRODUCTION

Species occupying wide distributions that encompass climatic variation are expected to exhibit geographically structured genetic and phenotypic variation among populations. Such structure could arise from contemporary conditions involving adaptive responses to abiotic factors or from historical factors that have isolated populations, for example during glacial-interglacial cycles. Abiotic factors, such as day length, growing season period, precipitation, and temperature, are latitudinally structured; because of adaptive responses to these abiotic variations, plant populations often demonstrate latitudinal variation in their traits. Many studies have documented latitudinal clines in traits such as life history, growth, and phenology (Weber and Schmid 1998; Kollmann and Bañuelos 2004; Maron et al. 2004, 2007; Friedman et al. 2008; Montague et al. 2008), and these can also be associated with genetic variation (Glato et al. 2017). In contrast, differentiation due to historical isolation of populations may not lead to geographic structure that coincides with latitudinal variation in abiotic factors because of the stochastic effects of genetic drift.

*email address: lewallac@odu.edu
Received 21 May 2021; Accepted 4 May 2022

Dispersal, the movement of seeds or diaspores away from the parent plant, is a key life-history stage in plants, and persistence, migration and seedling recruitment are all affected by seed dispersal distances (Howe and Smallwood 1982; Hyatt et al. 2003; Levin et al. 2003). When dispersal between populations is geographically restricted, the accumulation of genetic differences produces a pattern influenced by distance (Slatkin 1993). Under this scenario, spatial genetic correlations are relatively high for pairs of adjacent or nearby individuals and generally decrease as the distance among individuals and populations increases (Sokal and Wartenberg 1983), leading to the pattern known as isolation by distance (IBD). Population structure can also reflect local responses to variation in environmental conditions, and when genetic differentiation increases with differences in abiotic factors, it is known as isolation by environment (IBE). Isolation by environment is more important for local adaptation than isolation by distance (Tiffin and Ross-Ibarra 2014), although it can be difficult to distinguish the impacts of IBD from IBE that might lead to locally adapted populations (Nadeau et al. 2016). Examining patterns of intraspecific structure across spatial scales can provide insight about the roles that genetic drift and environmentally mediated natural selection play in structuring genotypic and phenotypic variation in plants.

Chamaecrista fasciculata (Michx.) Greene., also known as Partridge pea, is a self-compatible, outcrossing, annual, native legume that is widely distributed in the eastern United States from Minnesota to Mississippi and from the east coast of the United States to New Mexico (USDA NRCS 2021). Partridge pea is an important species in many ecosystems because it is a colonizing species that provides cover, nectar, and pollen for animals (USDA NRCS 2021) and fixes nitrogen in symbiosis with rhizobia bacteria. As a frequent species in open landscapes of the eastern United States, it is an ideal system in which to study how genetic drift and environmentally mediated selection can shape a species' genetic structure.

This species exhibits intraspecific phenotypic variation in stem and leaflet pubescence, anther color, growth habit, and stem architecture; these traits may be geographically structured, leading several authors to name intraspecific forms (Pullen 1963; Weakley 2020). The frequency of plants with pubescent leaflets appears to increase at lower latitudes in the southern U.S. (Isely 1975; Nobarinezhad and Wallace, personal observation). Wallace et al. (2020) reported that distinct phenotypes are maintained in a common garden, but we have been unable to identify genetic differences at nrDNA ITS or plastid loci (Wallace, unpublished data), associated with phenotypic variation, which continues to obscure intraspecific taxonomy of *Chamaecrista fasciculata,* especially in the southern United States.

In some areas of its distribution researchers have found evidence of limited gene flow (Fenster 1991) and local adaptation (Etterson 2004), but in other studies there is little evidence of local adaption (Abdala-Roberts and Marquis 2007), or it has been found only at large spatial scales (Galloway and Fenster 2000). These mixed findings suggest that population structure and the response of populations to local factors varies across the distribution of *Chamaecrista fasciculata.* No genetic studies of this species have included populations from the southern United States. Given that phenotypic variation for the species appears to be more complex in the southern United States compared to other areas of its distribution (Pullen 1963; Weakley 2020), it is important to consider population structure and the potential for unique factors that may influence its ecology at the southern range limit.

We conducted this study to better understand how genetic diversity is structured among populations of *Chamaecrista fasciculata* in the southeastern United States and to evaluate if phenotypically distinct populations are genetically distinct at microsatellite markers, which are more rapidly evolving than nrITS or the plastid genome. Additionally, we tested the hypothesis that populations are genetically differentiated at multiple spatial scales. Metapopulation dynamics have been suggested to be important for this species in other areas of its distribution (Galloway and Fenster 2000). Thus, we predicted population genetic structure would follow an isolation by distance (IBD) pattern at local scales because of a low propensity for gene dispersal due to gravity-dispersed seeds and bee pollination, which has been shown to result in most pollen from a source plant being deposited

on immediate neighbors (Cresswell et al. 1995). Under IBD, distance predicts differentiation due to dispersal limitation and drift, irrespective of environmental differences (Wright 1943). Across a larger spatial expanse of several latitudinal degrees, we expected environmental differences to play a larger role in the differentiation of populations. Thus, at this scale, variation in climatic variables is expected to be associated with genetic differentiation of populations if gene flow is reduced over long distances or populations are evolving towards local conditions as suggested by geographically structured phenotypic variation (Pullen 1963; Wallace et al. 2020).

MATERIALS AND METHODS

Sampling and Data Collection

In total, 38 populations of *Chamaecrista fasciculata* across Mississippi and southern Alabama, were sampled (Table 1, Figure 1). Since this species is not clonal and seeds are gravity dispersed from the maternal plant, individuals at each collection site were haphazardly sampled at least 2 meters apart to reduce the likelihood of sampling related individuals. GPS coordinates were recorded for each sampling site, and voucher specimens have been deposited in the Mississippi State University Herbarium (MISSA; Table 1). We focused on leaf pubescence as a phenotypically variable trait because of its potential adaptive role (Housman et al. 2002; Hanley et al. 2007). To assess pubescence, we measured trichome density on leaves from 2–8 pressed plants per population. Two leaves from the upper part of each plant were randomly selected and two leaflets from the middle part of each selected leaf were removed and observed using a stereo microscope with 7x magnification (ca. 6 mm^2 area). Three photos were taken per leaflet from the bottom, middle and upper part of each leaflet using a Canon PowerShot G9X camera. These photos were uploaded into ImageJ 1.x (Schneider et al. 2012), with resolution of 6000 × 4000-pixel per photo. The number of trichomes was counted manually for the bottom, middle and upper part of each leaflet in an area of 2285 × 2181 pixels (ca. 1.25 mm^2 area), and a mean value from the three portions was recorded as the number of trichomes per leaflet. A mean value from all studied leaflets per plant was determined, and these were used to calculate a population mean number of trichomes per leaflet. Populations were considered glabrous if no leaflets from any individual contained trichomes, mixed if some plants were pubescent and others glabrous, or pubescent if all plants had pubescent leaves. Leaf pubescence was evaluated relative to genetic structure by considering whether populations with pubescent leaf phenotypes were genetically divergent from populations with glabrous leaf phenotypes and if populations with a mixture of pubescent and glabrous phenotypes exhibited greater genetic admixture than purely pubescent or purely glabrous populations.

Leaf samples for use in genetic analyses were kept on ice in the field and then stored in silica gel or at –80°C prior to DNA extraction. Genomic DNA was extracted using a CTAB method (Dellaporta et al. 1983, modified by Schnable lab 2014) or the SYNERGY 2.0 Plant DNA extraction kit with modification (OPS Diagnostics, Lebanon, New Jersey). DNA samples were standardized to 10 ng/μL for use in PCR. Each sampled plant was genotyped at 14 trinucleotide microsatellite loci (Nobarinezhad 2020) that were developed from a transcriptome library of *Chamaecrista fasciculata* (Singer et al. 2009). The 14 loci were amplified using multiplex PCR with the Kapa Multiplex PCR kit (Roche Sequencing and Life Science, Wilmington, Massachusetts) and fluorescent labeled primers (Nobarinezhad 2020). For each sample, three reactions with five, four, and five loci per reaction respectively, were performed in a final volume of 10 μl in the presence of 10 ng of template DNA, 100 μmole of each of the reverse and tagged fluorescent label primers and 10 μmole of tagged forward primer using a KAPA 2G Fast Multiplex PCR kit. The tag in the 5′ forward primer matched the sequence of the fluorescent labeled primer (Culley et al. 2013). The thermal cycler program used to amplify loci included 3 min at 95°C, 30 cycles of 15 s at 95°C, 30 s at 60°C, and 30 s at 72°C, and a final extension step of 1 min at 72°C. Amplified products were genotyped at the Arizona State University DNA Lab with LIZ 600 size standard, and individual alleles were sized using GeneMarker software (SoftGenetics, State College, Pennsylvania). Microsatellite genotypes are deposited in Dryad (**https://doi.org/10.5061/dryad.zw3r2289c**).

Table 1. Population codes, sample size, genetic diversity measures, location, and vouchers for studied populations of *Chamaecrista fasciculata*.

Pop Code	N	N_a	%P	H_O	H_E	F_{IS}	GPS Coordinates	Pubescence[a]	Voucher[b]
A20	24	5.21	100	0.59	0.59	0	33°57′22.21″N 88°42′36.79″W	G	MISSA034317
A40	25	5.71	100	0.5	0.56	0.13	33°30′38.88″N 88°44′17.23″W	G	MISSA033001
A43	11	4.71	92.9	0.53	0.56	0.02	33°54′15.33″N 88°50′47.18″W	G	MISSA033003
A44	11	3.86	92.9	0.39	0.54	0.27	33°28′38.27″N 88°42′38.59″W	G	MISSA033004
A48	18	4.79	100	0.57	0.61	0.04	33°34′39.39″N 89°16′48.61″W	G	MISSA033005
A49	24	5.5	100	0.62	0.63	0.03	33°41′56.11″N 89°21′02.88″W	G	MISSA033006
A50	11	4.64	100	0.55	0.58	0.01	33°46′49.40″N 89°25′43.06″W	G	MISSA033007
A52	24	6.29	100	0.56	0.59	0.04	33°55′37.38″N 89°36′35.02″W	G	MISSA033008
A57	33	6.07	100	0.51	0.61	0.15	33°06′21.70″N 90°01′13.51″W	G	MISSA033011
A59	23	4.93	100	0.51	0.58	0.08	33°17′11.83″N 90°13′23.66″W	G	MISSA033013
A61	11	4.71	100	0.54	0.6	0.09	33°28′22.51″N 90°08′15.71″W	G	MISSA033015
A69	18	3.93	100	0.4	0.49	0.18	33°16′22.51″N 88°47′22.27″W	G	MISSA033023
A70	24	4.86	100	0.45	0.52	0.1	33°30′07.55″N 89°03′45.47″W	M(28.7)	MISSA034318
A71	24	5.57	100	0.53	0.6	0.1	33°21′34.91″N 89°30′42.13″W	G	MISSA034320
A72	24	5.14	100	0.52	0.58	0.08	33°23′23.74″N 89°31′13.67″W	G	MISSA034334
A73	24	5.93	100	0.63	0.67	0.05	33°46′55.52″N 90°00′56.16″W	G	MISSA034322
A74	23	4.79	100	0.46	0.55	0.17	33°23′23.67″N 89°31′54.12″W	G	MISSA034319
A75	23	4.79	100	0.55	0.63	0.11	33°46′41.84″N 90°03′30.65″W	G	MISSA034321
A76	23	3.5	100	0.39	0.45	0.09	33°28′37.07″N 88°48′47.01″W	G	MISSA034335
Mean-North	**20.95**	**5.0**	**99.25**	**0.52**	**0.58**	**0.09**			
A63	11	3.71	100	0.45	0.46	-0.02	31°50′45.24″N 88°41′13.81″W	P(37.7)	MISSA033017
A64	12	2.57	100	0.4	0.4	-0.01	31°42′19.33″N 89°02′12.22″W	P(33.2)	MISSA033018
A66	10	3.86	100	0.58	0.56	-0.05	31°56′53.84″N 89°17′31.41″W	P(32)	MISSA033020
A67	19	3.93	100	0.47	0.5	0.02	32°01′29.09″N 89°24′18.32″W	P(27.7)	MISSA033021

Table 1. ***continued***

Pop Code	N	N_a	%P	H_O	H_E	F_{IS}	GPS Coordinates	Pubescence[a]	Voucher[b]
A68	12	4.14	92.9	0.42	0.53	0.19	32°15′18.21″N 89°24′34.73″W	G	MISSA033022
A82	24	3.79	92.9	0.38	0.46	0.2	31°32′42.36″N 87°30′23.72″W	M(13.7)	MISSA034339
A83	24	5.43	100	0.5	0.53	0.09	32°11′03.19″N 88°19′06.52″W	M(18.1)	MISSA034328
A90	24	2.07	57.1	0.11	0.21	0.35	31°30′35.71″N 87°27′13.93″W	P(19.5)	MISSA034324
A91	24	4.07	78.6	0.38	0.44	0.11	31°44′09.74″N 87°57′31.39″W	P(23.7)	MISSA034327
A80	24	4.57	100	0.27	0.5	0.39	30°48′56.55″N 88°56′47.90″W	M(27.7)	MISSA034340
A81	24	3.43	78.6	0.21	0.35	0.44	30°46′19.77″N 88°54′52.77″W	M(36.3)	MISSA034342
A84	12	2.29	64.3	0.09	0.19	0.41	31°01′35.79″N 89°11′33.93″W	P(31)	MISSA034341
A85	24	4.07	100	0.46	0.52	0.09	31°03′15.69″N 89°11′27.78″W	P(22.3)	MISSA034338
A86	24	4.36	100	0.54	0.54	-0.03	31°01′44.43″N 89°10′55.59″W	M(30.8)	MISSA034336
A87	24	4.79	100	0.49	0.53	0.15	30°50′54.02″N 88°59′31.95″W	M(22.4)	MISSA034337
A88	24	3.07	85.7	0.34	0.39	0.13	30°29′36.16″N 88°20′14.64″W	M(29.5)	MISSA034323
A89	24	3.21	85.7	0.27	0.45	0.33	30°44′25.83″N 87°52′50.41″W	P(28.7)	MISSA034326
A92	18	1.64	35.7	0.18	0.19	0.01	30°35′40.38″N 88°56′03.76″W	G	MISSA034325
Mean-South/Central	**18**	**1.64**	**35.7**	**0.18**	**0.19**	**0.01**			
Species–Mean	**20.4**	**4.32**	**93.4**	**0.44**	**0.51**	**0.12**			
Species–Total	**726.5**	**10.5**	**100**	**0.45**	**0.64**	**0.29**			

N=number of individuals sampled; Na=mean number of alleles per locus; %P=percentage of polymorphic loci; HO =observed heterozygosity; HE=expected heterozygosity; FIS=inbreeding coefficient.

[a]G=all leaves glabrous in the population, M=both glabrous and pubescent leaves noted in the population, P=all leaves pubescent in the population; for mixed and pubescent populations the mean number of trichomes per leaf are indicated.

[b]Vouchers are deposited at the Mississippi State University Herbarium (MISSA).

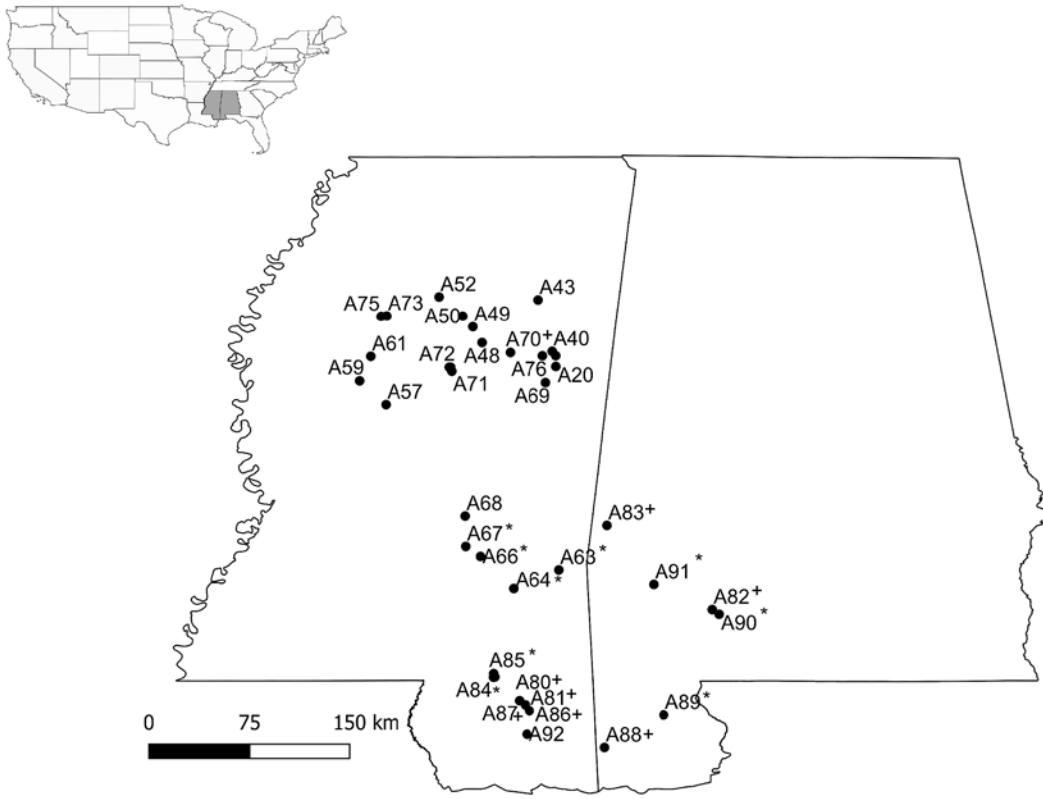

Figure 1. Sampling locations of *Chamaecrista fasciculata* in Mississippi and Alabama. Populations containing pubescent leaf phenotypes are indicated by * and those with a mixture of pubescent and glabrous forms by +. Population names follow those in Table 1.

Data analysis

The number of unique multilocus genotypes for each population was calculated using the Poppr package (Kamvar et al. 2014) using RStudio statistical software v. 1.1.456 (RStudio Team 2018, R Core Team 2020). The possibility of null alleles in each locus and population was checked using the program FreeNA (Chapuis and Estoup 2007). Genetic diversity within populations was assessed as number of alleles per locus (N_a), observed heterozygosity (H_o), expected heterozygosity (H_e), percent polymorphic loci (%P) using GenAlEx version 6.503 (Peakall and Smouse 2012). Within each population each locus was tested for departure from Hardy-Weinberg expectations through permutations of alleles among individuals and statistical significance was assessed using a *p*-value of 0.00009, which is adjusted using the sequential Bonferroni correction for multiple comparisons (Holm 1979). Genotypic linkage disequilibrium was measured for each pair of loci in each population and tested through Fisher's exact test using GENEPOP version 3.2 (Raymond and Rousset 1995) and applying a Bonferroni-type corrected *p*-value of 0.0005.

An analysis of molecular variance (AMOVA) (Excoffier et al. 1992) was conducted using GenAlEx version 6.503 (Peakall and Smouse 2012), which allows the hierarchical partitioning of genetic variance into three components: among populations, among individuals within populations, and within individuals. Statistical significance of AMOVA was assessed by 9,999 permutations. To determine the impact of genetic drift on population genetic variability, we compared mean population F_{ST} with N_a and H_E in linear regression analyses using SPSS v. 27 (IBM Corp 2020). Mean population F_{ST} was

calculated using GESTE v.2 (Foll and Gaggiotti 2006). Genetic data were input as allele frequencies and analyses were run using default settings, reversible jump MCMC method, and 10 pilot runs of a length of 5000 as burn-in prior to drawing samples from a chain of 50,000 in length, separated by a thinning interval of 50.

Two methods were used to evaluate population structuring and de novo assignment of individuals to genetic groups. The Bayesian statistical framework implemented in the program STRUCTURE version 2.3.4 (Pritchard et al. 2000) was run using an admixture model with correlated allele frequencies, a 'burn-in' period of 50,000 MCMC replicates, a sampling period of 100,000 replicates, and 10 iterations of each *K* value, from one to six. These values were selected based on initial runs of five iterations each of *K*=1 to *K*=25 under similar run parameters, which indicated low probability of a *K* value greater than two. For the final runs, multiple posterior probability values (log likelihood (lnL) values) for the 10 iterations of each *K* value were generated, and the most likely number of clusters was determined using STRUCTURE HARVESTER (Earl and vonHoldt 2012) and the Delta K-method of Evanno et al. (2005). CLUMPP (Jakobsson and Rosenberg 2007) was used to aggregate individual assignment probabilities from each of the 10 iterations for the chosen *K* value. STRUCTURE PLOT (Ramasamy et al. 2014) was used to generate plots of individual assignment from the CLUMPP output file. We also tested for population structure using discriminant analysis of principal components (DAPC) (Jombart et al. 2010) because it does not assume adherence to a particular population model as STRUCTURE does. This analysis was performed on allelic data using the adegenet package (Jombart 2008; Jombart and Ahmed 2011) in R (R Core Team 2020). We conducted a de novo assignment of individuals to clusters first by using a *K*-means clustering approach implemented in the find.clusters() function. The optimal number of groups (*K*) was determined based on comparison of successive values of the Bayesian Information Criterion (BIC) for *K*=1 to 37 (i.e., diffNgroup criterion). The resulting groups were then used in DAPC first retaining 60 Principal Components (PCs). Then, an optimal number of PCs to retain was determined using the optim.a.score() command and the DAPC was conducted again with the optimal number of PCs. The compoplot() function was used to display assignment probabilities of each individual to the optimal number of clusters.

We tested for evidence of IBD and IBE over multiple spatial scales (i.e., all populations, northern populations, and south-central populations) using distance-based RDA (dbRDA) (Legendre and Anderson 1999). The northern populations occur in a distinct climate from south-central populations and rarely have pubescent leaves. Assessment of these population groups allowed us to evaluate the relative impacts of distance and environment on genetic structure at local scales. For each analysis a matrix of pairwise population genetic distances (i.e., $F_{ST}/(1\text{-}F_{ST})$) (Rousset 1997) was created using GENEPOP v. 3.2 (Raymond and Rousset 1995; Rousset 2008). The GPS coordinates for each population were used to generate a pairwise population geographic distance matrix using the distGeo() function in the geosphere package v. 1.5 (Hijmans et al. 2021) in R (R Core Team 2020). This matrix was then used in a principal coordinates of neighbor matrices (PCNM; Borcard and Legendre 2002; Borcard et al. 2004) with default threshold values to generate a set of independent variables reflecting spatial arrangement of the populations. PCNM axes were forward selected using the double-stopping criterion of Blanchet et al. (2008) which considers both an alpha level and the adjusted R^2 value relative to the full model including all PCNM axes when selecting significant axes. We used an alpha value of 0.10 in forward selection to allow for greater variation across both coarse and fine geographic scales to be reflected in the chosen PCNM axes used in dbRDA. The resulting axes were tested as predictors of the response variable, genetic distance, in dbRDA. To test for IBE we considered precipitation, temperature, and elevation as potential predictors of population genetic structure. Data were sampled for each site from 30-second resolution layers for 11 variables associated with temperature, eight variables associated with precipitation, and elevation using the WorldClim data set (Fick and Hijmans 2017). Using R (R Core Team 2020), we conducted principal component analysis (PCA) for the temperature and precipitation variables separately to generate sets of independent variables. Loadings on the first two principal component axes were used in

dbRDA against genetic distances. In each analysis of all, northern, south-central, and south-central cluster 1 populations, the first two PCA's accounted for >80% of the observed variation. Elevation was tested directly against genetic distance in dbRDA. Each of the environmental variables was assessed independently and conditional on the PCNM axes selected through forward selection as outlined above. The dbRDA analyses were conducted using the dbRDA() function of the package vegan (Oksanen et al. 2020) with a constant added to correct for negative eigenvalues as suggested by Legendre and Anderson (1999). An analysis of variance was used to evaluate significance of each dbRDA model.

RESULTS

For this study, 756 individuals and a mean 20.4 individuals per population were genotyped. Duplicate multilocus genotypes were detected in four populations (i.e., two each in populations 9 and 29, six in population 37 and 12 in population 38). Each duplicated genotype was removed from populations 9, 29 and 37, while population 38 was removed from the data set because of low sample size of non-duplicated genotypes. Possible null alleles were detected at six loci and in seven of the 37 populations (i.e., Cf6895 in populations 34 and 37, Cf10002 in population 37, Cf9980 in population 35, Cf5782 in population 25, Cf3119 in population 25 and Cf8757 in population 26). We accounted for null alleles at these loci when the predicted null allele frequency was greater than 0.2, as suggested by Dakin and Avise (2004).

At the species level, a mean of 726.5 individuals per locus were considered in estimating genetic diversity; all loci were polymorphic, the number of alleles per locus was 10.5, observed and expected heterozygosity were 0.45 and 0.64, respectively, and the inbreeding coefficient was 0.29. At the population level, the number of alleles per locus ranged from 1.64 to 6.29, with a mean value of 4.32 (Table 1). Polymorphic loci among all populations ranged from 35% to 100%, with a mean of 93.4%. The mean polymorphic loci across all populations was 93.44%. Observed heterozygosity ranged from 0.09 to 0.63 and expected heterozygosity ranged from 0.19 to 0.67 (Table 1). Excepting four populations with negative values, the inbreeding coefficient (F_{IS}) was consistently positive, indicating heterozygote deficiency across these populations (Table 1). After applying Bonferroni correction, 61 out of 518 (11.77%) locus-by-population comparisons among all populations deviated significantly from Hardy-Weinberg Equilibrium (HWE), and 43 out of 3367 (1.27%) tests of linkage disequilibrium among all populations were significant.

The AMOVA indicated significant differentiation among populations (F_{ST}=0.193; p=0.001), with the observed variation residing among individuals (68%), among populations (19%) and among individuals within populations (13%). Mean population F_{ST} was negatively associated with population allelic diversity (r^2=0.829, p<0.001) and expected heterozygosity (r^2=0.873, p<0.001; Figure 2).

The STRUCTURE analysis suggested the presence of two genetic clusters (K=2) based on the highest delta K value. The probability of membership of each population to each of the two genetic clusters is shown in Figure 3. Most populations are grouped into cluster 1, and most individuals are assigned to this cluster with greater than 95% probability. Cluster 2 includes five populations and evidence of admixture with cluster 1. The populations of cluster 2 are not geographically structured nor phenotypically cohesive, but several of these populations have lower allelic diversity and heterozygosity compared to populations in cluster 1.

The best solution of groups determined by the find.clusters() function was K=2. In the DAPC 14 PCs were retained as the optimal number identified by the optim.a.score() command. Assignment of individuals to each of the two clusters by DAPC resulted in extremely similar patterns as indicated by STRUCTURE (Figure 3). Furthermore, these groups are shown to be largely non-overlapping in their composition (Figure 4). Neither STRUCTURE nor DAPC indicated that these clusters correspond to the leaf phenotypic groups or distinguished northern populations from south-central populations, although it is the latter populations that primarily make up the second cluster in both analyses. Additionally, populations containing a mixture of glabrous and pubescent leaf phenotypes

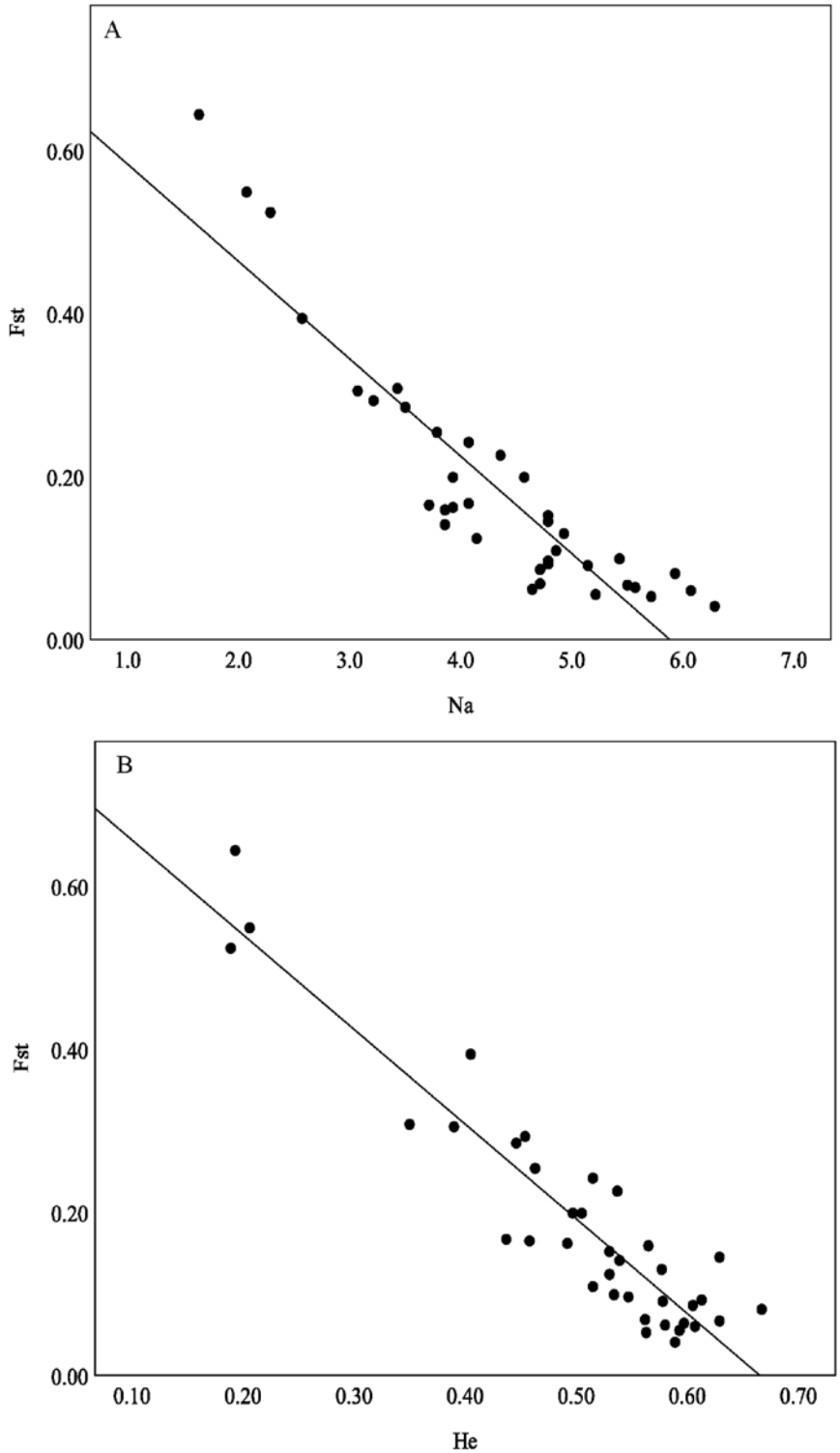

Figure 2. Association between mean population FST and genetic diversity. Panel A indicates Na (r^2=0.829, p<0.001); and panel B indicates He (r^2=0.873, p<0.001).

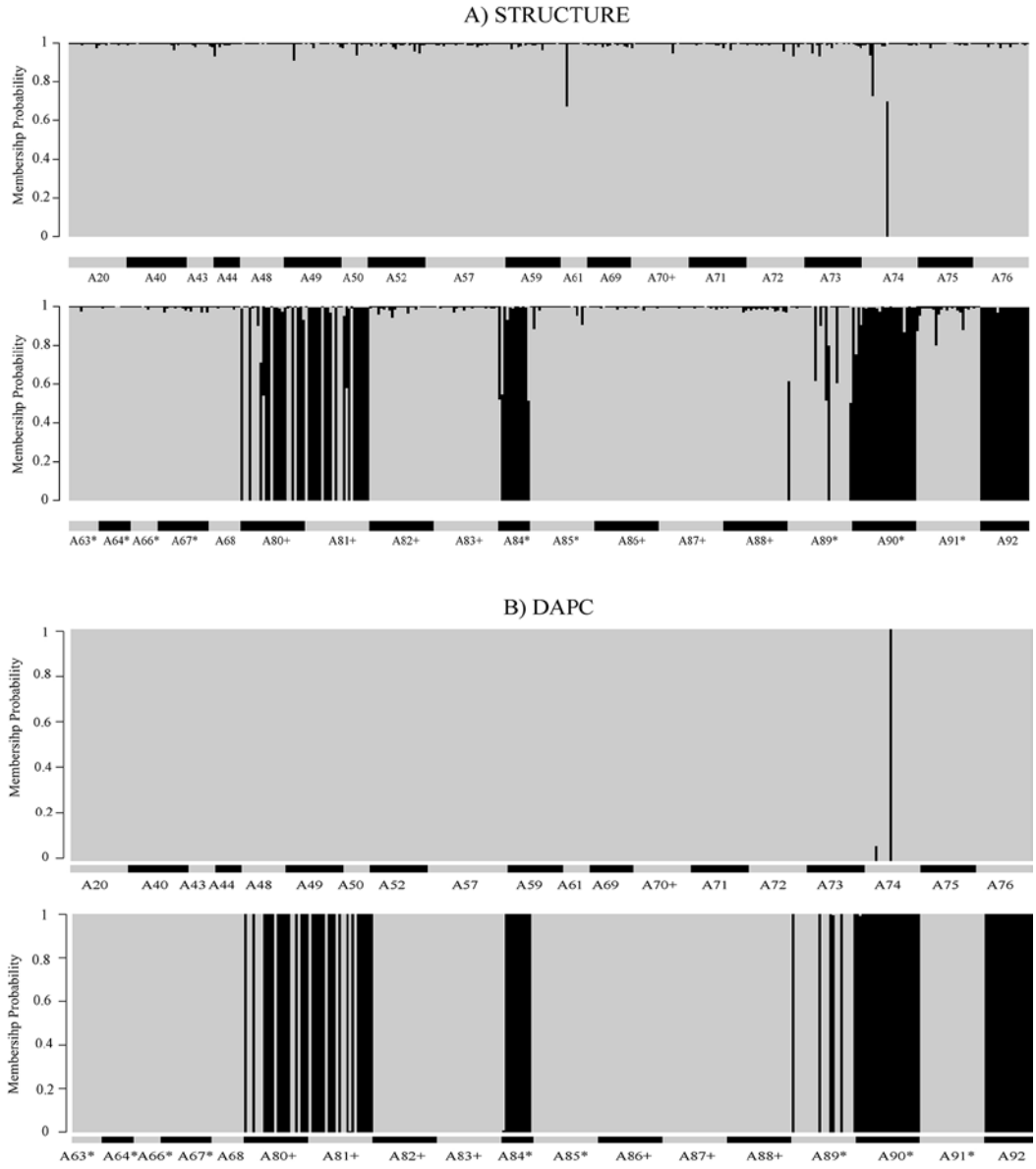

Figure 3. A. Assignment of individuals in each of two clusters (gray or black) according to STRUCTURE. **B.** Assignment of individuals in each of two clusters by DAPC. In both panels, northern populations are shown in the upper plot and south-central populations in the lower plot; populations containing pubescent leaf phenotypes are indicated by * and those with a mixture of pubescent and glabrous phenotypes are indicated by +. Population names follow those in Figure 1 and Table 1.

did not exhibit higher levels of genetic admixture within individuals compared to those with a greater consistency of either glabrous or pubescent phenotypes (Figure 3).

PCA based on climatic variables indicated variation in precipitation and temperature among the sample sites (Figure 5). The first two axes in each analysis accounted for >98% of the observed variation among populations. In both analyses most of the variation is explained in the first dimension, which is driven by temperature seasonality or annual precipitation and precipitation in the

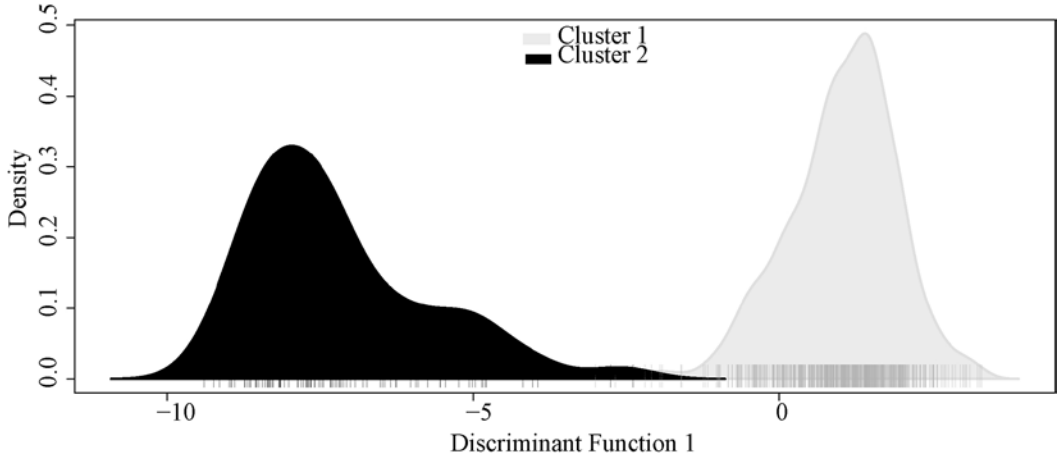

Figure 4. Discriminant analysis of principal components (DAPC) for allelic variation across the 27 populations of *Chamaecrista fasciculata*. *K*-means analysis identified two groups as the most likely solution. DAPC of these groups demonstrates that there is very little overlap between them along the discriminant function shown. See Figure 3 for the individual assignment of each individual to these clusters.

warmest quarter. The 95% confidence interval encompasses greater variation among populations in the south-central areas of Mississippi and Alabama than in northern Mississippi for both climate datasets.

The dbRDA analyses indicated that geographic distance and precipitation were significant predictors of genetic structure across all populations, as well as among northern populations (Table 2). For northern populations elevation was also found to be a significant predictor. Nevertheless, in conditional tests that include geographic distance as a covariate no environmental variables were significant for all populations or those in the north. For the south-central populations neither geographic distance nor any environmental variable was a significant predictor of genetic structure. When the populations of cluster 2 (i.e., as identified by STRUCTURE) were excluded from the analysis of south-central populations, then temperature and precipitation were both significant predictors and this persisted in conditional tests that factored in geographic distance.

DISCUSSION

This survey of populations of *Chamaecrista fasciculata* distributed across its southeastern edge revealed genetic variability within populations that is consistent with a primarily outcrossing breeding system. Fixation indices averaged 0.12 across all populations surveyed, as much as 68% of observed genetic variation occurred among individuals, and mean population polymorphism was 93%. Within-population genetic diversity was variable among populations (mean H_e = 0.19 to 0.67), but the mean value (0.51) was lower than mean H_e based on microsatellite loci in other widespread (0.62) or outcrossing species (0.65) reviewed by Nybom (2004). However, the mean H_e in this study was higher than for other annual species (0.46) or gravity-dispersed species (0.47) (Nybom 2004). Our results are consistent with previous studies of populations from other areas in demonstrating that *C. fasciculata* harbors moderately high genetic variability at neutral loci (Fenster et al. 2003, Bueno et al. 2019).

Leaf pubescence has been reported for *Chamaecrista fasciculata* from coastal and non-coastal areas at its southern edge, although our observations are consistent with Pullen (1963) and Isely (1975) who suggested that pubescence increases towards the Gulf Coast. Excepting one northern population, plants containing pubescent leaves occurred in central and coastal Mississippi and Alabama (Figure 1), areas that are climatically different than northern Mississippi (Figure 5). Given

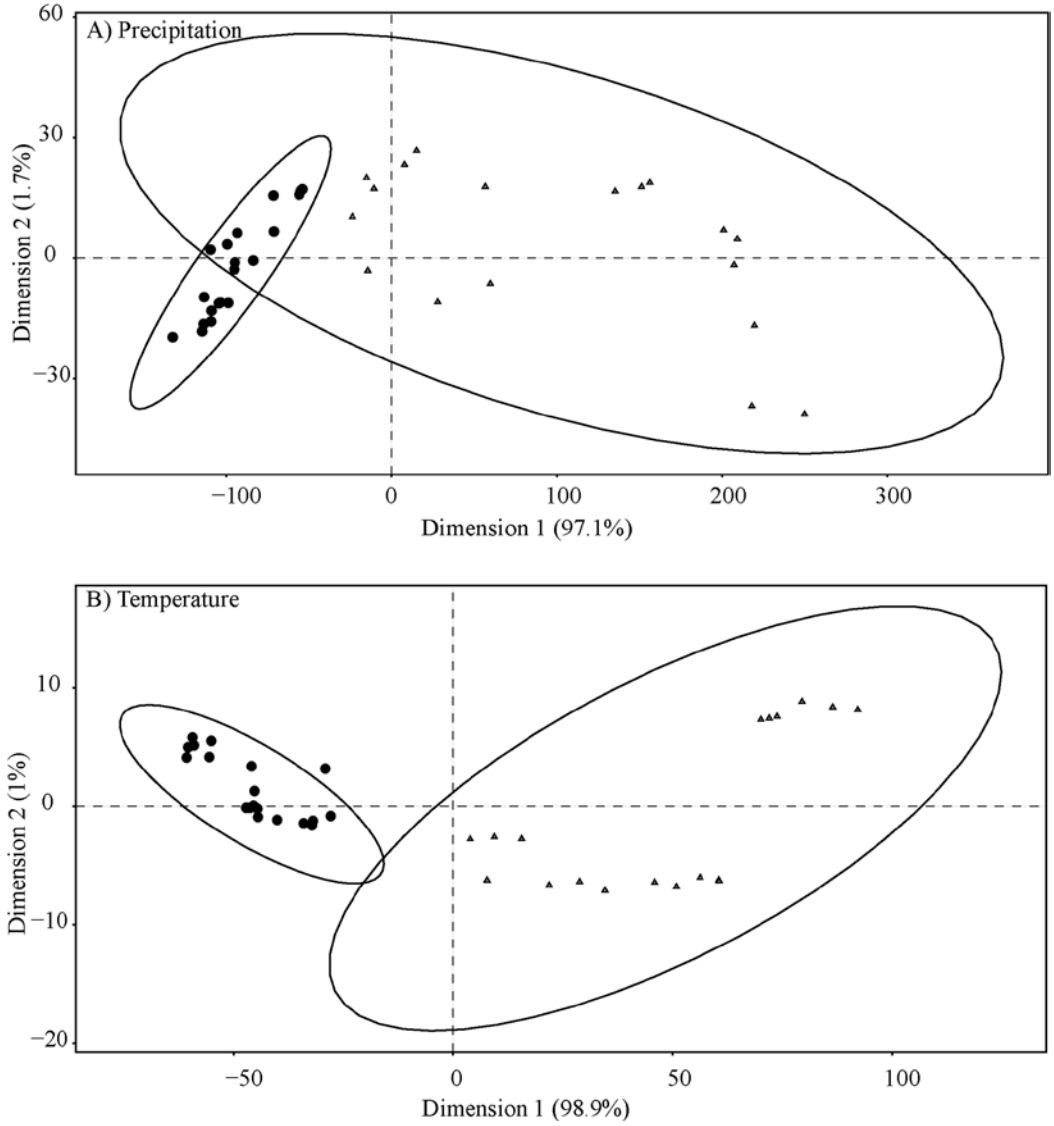

Figure 5. Principal component analysis based on variation in **A.** eight precipitation and **B.** 11 temperature variables among sample locations. Ellipses indicate 95% CI around pre-defined groups of northern (black circles) vs. south-central (gray triangles) populations. In panel A the first two dimensions explain 98.8% of the observed variation, primarily due to annual precipitation and precipitation in the warmest quarter primarily. In panel B the first two dimensions explain 99.9% of the observed variation, primarily due to temperature seasonality. Climate data are derived from the WorldClim dataset (Fick and Hijmans 2017).

that leaf pubescence appears to be heritable in populations from Mississippi (Wallace et al. 2020), we expected that these phenotypically distinct populations may also be genetically differentiated because of limited long-distance gene flow or if phenotypic variation reflects local adaptation as reported in other areas where *C. fasciculata* occurs (Etterson 2004; Lau et al. 2019). Additionally, variation in environmental factors is frequently a significant predictor of variation in floral traits of *C. fasciculata* (Frazee and Marquis 1994; Abdala-Roberts and Marquis 2007). Collectively, these studies point to the presence of spatial variation in phenotypic traits in *C. fasciculata*, which seems

Table 2. Influence of geographic distance and several environmental predictor variables on genetic structure of *Chamaecrista fasciculata* populations based on dbRDA analyses. Marginal tests included individual sets of each variable, whereas conditional tests included the PCNM vectors based on geographic distance as covariates. The column labeled "%var" indicates the percentage of genetic structure that is explained by each predictor variable. *P*-values <0.05 are indicated in bold. Temperature and precipitation variables reflect values along the first two axes of a principal component analysis using WorldClim variables (Fick and Hijmans 2017).

Variable	Marginal tests			Conditional tests		
	F	*P*	% Var	F	*P*	% Var
			All populations			
Distance[a]	4.28	**0.01**	41	NA	NA	NA
Elevation	0.09	0.76	0	0.91	0.79	3.0
Temperature	1.76	0.21	9.0	0.07	0.79	1.0
Precipitation	3.38	**0.04**	17.0	1.23	0.35	8.0
			North			
Distance[a]	5.91	<**0.01**	79.0	NA	NA	NA
Elevation	5.64	<**0.01**	25.0	1.34	0.35	12.0
Temperature	2.35	0.08	23.0	1.16	0.39	21.0
Precipitation	3.35	**0.01**	30.0	2.03	0.15	31.0
			South-Central			
Distance[a]	2.61	0.12	26.0	NA	NA	NA
Elevation	0.28	0.70	2.0	0.83	0.48	6.0
Temperature	1.36	0.30	15.0	0.68	0.61	9.0
Precipitation	2.32	0.16	24.0	2.66	0.09	29.0
			South-Central Cluster 1			
Distance[a]	0.46	0.99	4	NA	NA	NA
Elevation	0.90	0.51	8.0	1.77	0.15	15.0
Temperature	1.99	**0.03**	28.0	2.14	**0.02**	32.0
Precipitation	3.58	<**0.01**	42.0	3.46	<**0.01**	43.0

[a]The number of PCNM axes retained for RDA was 5 for all populations, 7 for northern populations, 7 for south-central populations, and 1 for south-central populations of cluster 1.

to be correlated with environmental variability. In adverse environments, trichomes are beneficial because they influence water balance, protect photosynthetic machinery, and play a role in thermoregulation (Hauser 2014). Thus, the pubescent nature of plants growing at lower latitudes, where the temperature is higher, might reflect the protective nature of trichomes to increase reflectance and reduce the heat load.

Despite our expectations, we did not find evidence that phenotypically distinct populations are genetically unique. Genetic diversity has not consistently been associated with phenotypic diversity (Butlin and Tregenza 1998; Reed and Frankham 2001; Villellas et al. 2014), likely reflecting the impact of different processes on these types of variation. Whereas phenotypic variation is expected to be shaped strongly by environmental factors, genetic variation can reflect demographic and geographic history of populations, as well as environmental barriers to gene flow. In this study, the strongest predictor of genetic differentiation across all populations was geographic distance, rather than environmental variables. We also found evidence that genetic drift may be driving population differentiation as higher mean F_{ST} is strongly associated with lower allelic diversity and lower heterozygosity (Figure 2). Limited gene flow may have enabled the maintenance of both pubescent and glabrous forms in south-central populations as well as the unique genetic variants of cluster 2 in *Chamaecrista fasciculata*.

We expected that patterns of genetic structure would vary by spatial scale and reflect differing influences of geographic distance and environmental variation on gene flow and population persistence.

This hypothesis is supported by these results, which also suggest differential contributions of environmental factors to geographic structure in different areas. We found a strong pattern of isolation by distance when considering all populations and those in northern Mississippi, but not for populations in the southern-central area. Precipitation was also significant in the dbRDA when considering all populations, but this effect disappeared when geographic distance was included as a covariate. The importance of geographic distance to genetic structure is consistent with the small neighborhood area estimated by Fenster (1991) for *Chamaecrista fasciculata* at only 2.4 m, driven by greater limitations on seed than pollen dispersal. The absence of environmental predictors at this scale likely occurs because seeds and pollen are moving only between sites within similar climatic bands. For the northern populations, the overwhelming pattern is one of isolation by distance as well, but elevation and climate may influence population structure too as indicated by the significant results in marginal tests of these variables. That any of these variables would influence genetic structure is surprising given that the sample sites occur within relatively small latitudinal and elevational (i.e., 36–154 m) bands. Populations are distributed across this area, but we have noted variation in phenotypes between populations in the Mississippi River Delta and those in the central and eastern areas of northern Mississippi. As the Delta region has been heavily influenced by human activity, the associated variation in environmental variables may simply covary with other factors that more directly influence the persistence of populations. Both Etterson (2004) and Lau et al. (2019) reported the potential for local adaptation in *C. fasciculata* at small spatial scales, which suggests that metapopulation dynamics may be particularly important to consider for this species.

The south-central populations occur in a more variable climatic space (Figure 5) and are more phenotypically variable compared to northern populations (Table 1). Thus, environmental variables might be expected to have equal importance to geographic distance in generating genetic structure at local scales. The dbRDA results indicated that neither geographic distance nor any of the environmental variables were significant predictors of genetic distance between populations. The uniqueness of the five populations assigned to cluster 2 appears to influence this result because when these populations are excluded, temperature and precipitation become significant predictors in marginal tests and this pattern is maintained when accounting for geographic distance in the conditional tests.

The Southeast has long been described as an area harboring substantial phenotypic variation for *Chamaecrista fasciculata* (Pullen 1963, Isely 1975, Weakley 2020). Additionally, Bueno et al. (2019) also found that a germplasm sample (USDA PI638976) from southern Mississippi has a unique AFLP profile relative to other Mississippi collections. We find that genetic variation is no different, and in this region *C. fasciculata* may continue to actively evolve. Even when considering the limited distance over which gene flow occurs, our analyses suggest that other factors influence population diversity and connectivity in the southeast. For example, populations of cluster 2 are not geographically isolated from those in cluster 1, yet they are not exchanging genes. This may be a historical effect reflecting periods of isolation during glacial cycles. Additionally, the lower observed genetic diversity and greater differentiation of populations, particularly at lower latitudes, may reflect edge effects, which have been reported at the northern and western edges of *C. fasciculata* in the U.S. (Stanton-Geddes et al. 2013). Given that populations of *C. fasciculata* are often short-lived (Stanton-Geddes et al. 2013, L. Wallace, personal observation), more frequent extinction and recolonization at the range edge could constrain genetic variation (Galloway and Fenster 2000).

Overall, our study demonstrates the role that contemporary factors, including geographic distance and climate, can play in generating patterns of spatial genetic variation in *C. fasciculata*. Further sampling of populations along the Gulf Coast seems important for full understanding the contemporary structure of *C. fasciculata* populations and their evolutionary history and would add to the complex story of plant evolution in the Southeast, a hotspot of biodiversity.

ACKNOWLEDGMENTS

The authors thank Dr. Susan Singer for permission to use *Chamaecrista fasciculata* transcriptome

data set to develop the microsatellite markers, the Biology Faculty Fund of Mississippi State University for funding for this project, and the J. Robert Stiffler Endowment at Old Dominion University for support to M.H.N.

LITERATURE CITED

Abdala-Roberts, L. and R.J. Marquis. 2007. Test of local adaptation to biotic interactions and soil abiotic conditions in the ant tended *Chamaecrista fasciculata* (Fabaceae). Oecologia 154:315–326.

Blanchet, F.G., P. Legendre, and D. Borcard. 2008. Forward selection of explanatory variables. Ecology 89:2623–2632.

Borcard, D. and P. Legendre. 2002. All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. Ecol. Model. 153:51–68.

Borcard, D., P. Legendre, C. Avois-Jacquet, and H. Tuomisto. 2004. Dissecting the spatial structure of ecological data at multiple scales. Ecology. 85:1826–1832.

Bueno, E., T. Kisha, S.L. Maki,. E.J.B. von Wettberg and S. Singer. 2019. Genetic diversity of *Chamaecrista fasciculata* (Fabaceae) from the USDA germplasm collection. BMC Res. Notes 12:117 (doi:10.1186/s13104-019-4152-0).

Butlin, R.K. and T. Tregenza. 1998. Levels of genetic polymorphism: marker loci versus quantitative traits. Philos. T. R. Soc. B 353:187–198.

Chapuis, M.P. and A. Estoup. 2007. Microsatellite null alleles and estimation of population differentiation. Molec. Biol. Evol. 24:621–631.

Cresswell, J.E., A.P. Bassom, S.A. Bell, S.J. Collins, and T.B. Kelly. 1995. Predicted pollen dispersal by honeybees and bumblebees foraging on oil-seed rape: a comparison of three models. Funct. Ecol. 6:829–841.

Culley, T.M., T.I. Stamper, R.L. Stokes, J.R. Brzyski, N.A. Hardiman, M.R. Klooster, and B.J. Merritt. 2013. An efficient technique for primer development and application that integrates fluorescent labeling and multiplex PCR. Applic. Pl. Sci. 1:1300027 (doi:10.3732/apps.1300027).

Dakin, E.E. and J.C. Avise. 2004. Microsatellite null alleles in parentage analysis. Heredity 93:504–509.

Dellaporta, S.L., J. Wood, and J.B. Hicks. 1983. A plant DNA mini preparation: version II. Pl. Molec. Biol. Reporter 1:19–21.

Earl, D.A. and B.M. vonHoldt. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method, version 2.3.4. Conservation Genet. Resources 4:359–361.

Etterson, J. 2004. Evolutionary potential of *Chamaecrista fasciculata* in relation to climate change. I. Clinal patterns of selection along an environmental gradient in the Great Plains. Evolution 58:1446–1458.

Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molec. Ecol. 14:2611–2620.

Excoffier, L., P.E. Smouse, and J.M. Quattro. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131:479–91.

Fenster, C.B. 1991. Gene flow in *Chamaecrista fasciculata* (Leguminosae) I. Gene dispersal. Evolution 45:398–409.

Fenster, C.B., X. Vekemans, and O.J. Hardy. 2003. Quantifying gene flow from spatial genetic structure data in a metapopulation of *Chamaecrista fasciculata* (Leguminosae). Evolution 57:995–1007.

Fick, S.E. and R.J. Hijmans. 2017. WorldClim 2: new 1km spatial resolution climate surfaces for global land areas. Int. J. Climatology 37:4302–4315.

Foll, M. and O.E. Gaggiotti. 2006. Identifying the environmental factors that determine the genetic structure of populations. Genetics 174:875–891.

Frazee, J.E. and R.J. Marquis. 1994. Environmental contribution to floral trait variation in *Chamaecrista fasciculata* (Fabaceae: Caesalpinoideae). Amer. J. Bot. 81:206–215.

Friedman, J.H., J.E. Roelle, J.F. Gaskin, A.E. Pepper, and J.R. Manhart. 2008. Latitudinal variation in cold hardiness in introduced *Tamarix* and native *Populus*. Evol. Applic. 1:598–607.

Galloway, L.F. and C.B. Fenster. 2000. Population differentiation in an annual legume: local adaptation. Evolution 54:1173–1181.

Glato, K., A. Aidam, N.A. Kane, D. Bassirou, M. Couderc, L. Zekraoui, N. Scarcelli, A. Barnaud, and Y. Vigouroux. 2017. Structure of sweet potato (*Ipomoea batatas*) diversity in West Africa covaries with a climatic gradient. PLoS One 12(5):e0177697 (**https://doi.org/10.1371/journal.pone.0177697**).

Hanley, M.E., B.B. Lamont, M.M. Fairbanks, and C.M. Rafferty. 2007. Plant structural traits and their role in antiherbivore defense. Pl. Ecol. Evol. Syst. 8:157–178.

Hauser, M.T. 2014. Molecular basis of natural variation and environmental control of trichome patterning. Front. Plant Sci. 320:5 (**https://doi.org/10.3389/fpls.2014.00320**).

Hijmans, R.J., C. Karney, E. Williams, and C. Vennes. 2021. geosphere: Spherical Trigonometry. R package version 1.5-14. (**https://CRAN.R-project.org/package=geosphere**, 14 April 2022).

Holm, S. 1979. A simple sequentially rejective multiple test procedure. Scand. J. Statist. 6:65–70.

Housman, D.C., M.V. Price, and R.A. Redak. 2002. Architecture of coastal and desert *Encelia farinosa* (Asteraceae): consequences of plastic and heritable variation in leaf characters. Amer. J. Bot. 89:1303–1310.

Howe, H.F. and J. Smallwood. 1982. Ecology of seed dispersal. Annual Rev. Ecol. Syst. 13:201–228.

Hyatt, L.A., M.S. Rosenberg, T.G. Howard, G. Bole, W. Fang, J. Anastasia, K. Brown, R. Grella, K. Hinman, J.P. Kurdziel, and J. Gurevitch. 2003. The distance dependence prediction of the Janzen-Connell hypothesis: a meta-analysis. Oikos 103:590–602.

IBM Corp. 2020. IBM SPSS Statistics for Windows (Version 27). IBM Corp., Armonk, New York.

Isely, D. 1975. Leguminosae of the United States: II. Subfamily Caesalpinioideae. Mem. New York Bot. Gard. 25:1–228.

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

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

Kamvar Z.N., J.F. Tabima, and N.J. Grünwald. 2014. Poppr: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. Peer J 2:e281. (doi.org/10.7717/peerj.281).

Kollmann, J. and M.J. Banuelos. 2004. Latitudinal trends in growth and phenology of the invasive alien plant *Impatiens glandulifera* (Balsaminaceae). Divers. Distrib. 10:377–385.

Lau, J.A., S.M. Magnoli, C.R. Zirbel, and L.A. Brudvig, 2019. The limits to adaptation in restored ecosystems and how management can help. Ann. Missouri Bot. Gard. 104:441–454.

Legendre, P. and M.J. Anderson. 1999. Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments. Ecol. Monogr. 69:1–24.

Levin, S.A., H.C. Muller-Landau, R. Nathan, and J. Chave. 2003. The ecology and evolution of seed dispersal: a theoretical perspective. Annual Rev. Ecol. Evol. Syst. 34:575–604.

Maron, J.L., M. Vila, R. Bommarco, S. Elmendorf, and P. Beardsley. 2004. Rapid evolution of an invasive plant. Ecol. Monogr. 74:261–280.

Maron, J.L., S. Elmendorf, and M. Vila. 2007. Contrasting plant physiological adaptation to climate in the native and introduced range. Evolution 61:1912–1924.

Montague, J.L., S.C.H. Barrett, and C.G. Eckert. 2008. Re-establishment of clinal variation in flowering time among introduced populations of purple loosestrife (*Lythrum salicaria*, Lythraceae). J. Evol. Biol. 21:234–45.

Nadeau, S., P.G. Meirmans, S.N. Aitken, K. Ritland, and N. Isabel. 2016. The challenge of separating signatures of local adaptation from those of isolation by distance and colonization history: The case of two white pines. Ecol. and Evol. 6:8649–8664.

Nobarinezhad, M.H. 2020. Genetic structure of the host plant, *Chamaecrista fasciculata*, and its microbial partners. Ph.D. dissertation, Mississippi State University, Mississippi State, Mississippi.

Nybom, H. 2004. Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. Molec. Ecol. 13:1143–1155.

Oksanen, J., F.G. Blanchet, M. Friendly, R. Kindt, P. Legendre, D. McGlinn, P.R. Minchin, R.B. O'Hara, G.L. Simpson, P. Solymos, M.H.H. Stevens, E. Szoecs, and H. Wagner. 2020. vegan: Community Ecology Package. R package version 2.5-7 (**https://CRAN.R-project.org/package=vegan/**, 14 April 2022).

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.

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

Pullen, T.M. 1963. The *Cassia fasciculata* complex (Leguminosae) in the United States. Ph.D. dissertation, The University of Georgia, Athens, Georgia.

Ramasamy, R.K., S. Ramasamy, B.B. Bindroo, and V.G. Naik. 2014. STRUCTURE PLOT: a program for drawing elegant STRUCTURE bar plots in user friendly interface. SpringerPlus 3:431 (**https://doi.org/10.1186/2193-1801-3-431**).

Raymond, M. and F. Rousset. 1995. GENEPOP: population genetics software for exact tests and ecumenicism, version 1.2. J. Heredity 86:248–249.

Reed, D.H. and R. Frankham. 2001. How closely correlated are molecular and quantitative measures of genetic variation? A meta-analysis. Evolution 55:1095–1103.

Rousset, F. 1997. Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. Genetics 145:1219–1228.

Rousset, F. 2008. Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. Molec. Ecol. Resources 8:103–106.

R Core Team. 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

RStudio Team. 2018. RStudio: Integrated development environment for R (Version 1.1.456). RStudio, PBC, Boston, Massachusetts.

Schnable, J. 2014. Dellaporta DNA extraction lightly edited by the Schnable Lab, Iowa State University. (**https://schnablelab.plantgenomics.iastate.edu/docs/resources/protocols/pdf/Dellaporta_DNA_Extraction.2014.08.20.pdf**.

Schneider, C.A., W.S. Rasband, and K.W. Eliceiri. 2012. NIH Image to ImageJ: 25 years of image analysis. Nat. Methods 9:671–675.

Singer, S., J. Doyle, G. May, S. Cannon, S. Maki, and D. Ilut. 2009. Exploring *Chamaecrista fasciculata* genomics data. (**http://serc.carleton.edu/exploring_genomics/chamaecrista/chamaecrista_tr.html**, 14 April2022).

Slatkin, M. 1993. Isolation by distance in equilibrium and non-equilibrium populations. Evolution 47:264–279.

Sokal, R.R. and D. Wartenberg. 1983. A test of spatial autocorrelation using an isolation-by-distance model. Genetics 105:219–237.

Stanton-Geddes, J., R.G. Shaw, and P. Tiffin. 2013. Insights from population genetics for range limits of a widely distributed native plant. Amer. J. Bot. 100:744–753.

Tiffin, P. and J. Ross-Ibarra. 2014. Advances and limits of using population genetics to understand local adaptation. Trends in Ecol. and Evol. 29:673–680.

USDA, NRCS. 2021. The PLANTS Database. National Plant Data Team, Greensboro, NC USA (**http://plants.sc.egov.usda.gov**, 14 April 2022)..

Villellas, J., R. Berjano, A. Terrab, and M.B. Garcia. 2014. Divergence between phenotypic and genetic variation within populations of a common herb across Europe. Ecosphere 5:1–14.

Wallace, L.E., M. Hosseinalizadeh-Nobarinezhad, and R. Coltharp. 2020. Phenotypic variation of partridge pea, *Chamaecrista fasciculata*, persists in a common garden. Castanea 85:93–107.

Weakley, A.S. 2020. Flora of the southeastern United States. University of North Carolina Herbarium, North Carolina Botanical Garden. **https://ncbg.unc.edu/research/unc-herbarium/floras/**, 14 April 2022).

Weber, E. and B. Schmid. 1998. Latitudinal population differentiation in two species of *Solidago* (Asteraceae) introduced into Europe. Amer. J. Bot. 85:1110–1121.

Wright, S. 1943. Isolation by distance. Genetics 28:114–138.