Evidence of population bottleneck in Astragalus michauxii (Fabaceae), a narrow endemic of the southeastern United States

Conserv Genet (2014) 15:153–164
DOI 10.1007/s10592-013-0527-2

RESEARCH ARTICLE

Evidence of population bottleneck in *Astragalus michauxii* (Fabaceae), a narrow endemic of the southeastern United States

Wade A. Wall · Norman A. Douglas · William A. Hoffmann · Thomas R. Wentworth · Janet B. Gray · Qiu-Yun Jenny Xiang · Brian K. Knaus · Matthew G. Hohmann

Received: 7 January 2013 / Accepted: 26 August 2013 / Published online: 13 October 2013

Abstract Genetic factors such as decreased genetic diversity and increased homozygosity can have detrimental effects on rare species, and may ultimately limit potential adaptation and exacerbate population declines. The Gulf and Atlantic Coastal Plain physiographic region has the second highest level of endemism in the continental USA, but habitat fragmentation and land use changes have resulted in catastrophic population declines for many species. *Astragalus michauxii* (Fabaceae) is an herbaceous plant endemic to the region that is considered vulnerable to extinction, with populations generally consisting of fewer than 20 individuals. We developed eight polymorphic microsatellites and genotyped 355 individuals from 24 populations. We characterized the population genetic diversity and structure, tested for evidence of past bottlenecks, and identified evidence of contemporary gene flow between populations. The mean ratios of the number of alleles to the allelic range (*M* ratio) across loci for *A. michauxii* populations were well below the threshold of 0.68 identified as indicative of a past genetic bottleneck. Genetic diversity estimates were similar across regions and populations, and comparable to other long-lived perennial species. Within-population genetic variation accounted for 92 % of the total genetic variation found in the species. Finally, there is evidence for contemporary gene flow among the populations in North Carolina. Although genetic factors can threaten rare species, maintaining habitats through prescribed burning, in concert with other interventions such as population augmentation or (re)introduction, are likely most critical to the long term survival of *A. michauxii*.

Keywords *Astragalus* · Bottleneck · Endemism · Genetic diversity · Microsatellites · *Pinus palustris*

Electronic supplementary material The online version of this article (doi:10.1007/s10592-013-0527-2) contains supplementary material, which is available to authorized users.

W. A. Wall (✉) · N. A. Douglas · W. A. Hoffmann · T. R. Wentworth · Q.-Y. J. Xiang
Department of Plant Biology, North Carolina State University, Box 7612, Raleigh, NC 27695, USA
e-mail: Wade.A.Wall@usace.army.mil

W. A. Wall · M. G. Hohmann
US Army Corps of Engineers, Engineer Research and Development Center, P. O. Box 9005, Champaign, IL 61826, USA

N. A. Douglas
Department of Biology, Oberlin College & Conservatory, Science Center K107, 119 Woodland St., Oberlin, OH 44074-1097, USA

J. B. Gray
Endangered Species Branch, Fort Bragg, NC 28310, USA

B. K. Knaus
Department of Botany and Plant Pathology, Oregon State University, 2082 Cordley Hall, Corvallis, OR 97331, USA

Introduction

Land use changes over the last 500 years have led to habitat loss and population isolation for many species, leading to a greater concern about the loss of biodiversity and its effects on the biosphere (Balmford and Bond 2005). In the southeastern United States, many terrestrial and aquatic species are threatened with extinction (Dobson et al. 1997). Many of the rare plant species of the coastal plain are part of the longleaf pine (*Pinus palustris* Mill.) ecosystem, an assemblage of fire-dependent communities

dominated by temperate savanna that covered roughly 37 million hectares from east Texas to Florida to southeastern Virginia (Frost 1993). The longleaf pine ecosystem depends on frequent fires to maintain an open understory and facilitate longleaf pine seedling recruitment. Urbanization, fire suppression, and conversion to agriculture have fragmented the original longleaf pine ecosystem and reduced it to 2 % of its former area (Frost 2006), resulting in population declines for many endemic species (Van Lear et al. 2005).

Anthropogenic habitat fragmentation in the southeastern United States has occurred against a background of past, highly dynamic, climatically driven vegetation change during the transition from the Pleistocene to the Holocene (Webb III 1987; Williams et al. 2004). Thus, possible explanations for the current habitat fragmentation of some southeastern plant populations include, in the short term, changes in land use following European settlement and, in the long term, climatic changes that have occurred in the past 20,000 years. Habitat fragmentation results in increased genetic isolation and smaller plant populations (Young et al. 1996), and with increasing time since fragmentation, genetic differentiation between populations generally increases and genetic diversity within populations decreases (Young et al. 1996). If populations have been isolated since the early Holocene, it is assumed that genetic differentiation between populations would be greater relative to that due to population isolation since European settlement.

Extreme reductions in population size can lead to genetic bottlenecks, which are of conservation concern because of the increased risk of extinction (Frankham 2005). Across taxa, it has been observed that genetic diversity is lower in threatened species (Spielman et al. 2004) and in species with restricted ranges (Hamrick and Godt 1989) relative to common, widespread species. This suggests that genetic factors increase the likelihood of extinction of species by reducing fitness of individuals within small populations (Leimu et al. 2006) and restrict evolutionary potential (Franklin 1980). Thus, assessing and maintaining the genetic variation found within rare species is of concern for both their short- and long-term viability and is one of the cornerstones of conservation genetics (Frankham et al. 2010).

Astragalus michauxii (Kuntze) F.J. Herm. is a rare legume endemic to the Fall-line Sandhills region of North Carolina, South Carolina, and Georgia (USA). Only a few species of *Astragalus* are represented east of the Mississippi. Most members of this genus occur in Asia, western North America, and South America, with the North American species belonging to the aneuploid Neo-Astragalus clade, which began to diversify 4.4 Ma (Scherson et al. 2008). The disjunct species in eastern North America may be relicts from past vegetation assemblages of Pliocene or Pleistocene origin (Noss 2013). In South Carolina and Georgia, all known populations of *A. michauxii* exist in a highly fragmented landscape and are separated from the North Carolina populations by over 200 km. The average population size is less than 20 individuals (Wall et al. 2012) and there is substantial risk that these small, isolated populations may be subject to reduced genetic diversity and the negative effects that ensue.

In this study, we investigated genetic diversity and structure of *A. michauxii* populations across the range of the species using eight microsatellite loci. We estimated the level of genetic variation within and among populations and identified the most likely number of genetic clusters within *A. michauxii*. In addition, we searched for evidence of past genetic bottlenecks and contemporary gene flow. We hypothesized that the small, isolated populations would exhibit strong genetic differentiation between populations and low genetic diversity within populations, as well as evidence of genetic bottlenecks because of recent land use changes following European settlement. By examining the genetic variation of *A. michauxii* within the context of past climatic change and land use history, this study provides useful information for any future conservation or restoration efforts of this species and other rare plants of the Fall-line Sandhills.

Methods

Species

Astragalus michauxii is an herbaceous, long-lived legume that is generally found in upland longleaf pine habitat in what has been characterized as the pine/scrub oak sandhill community (Schafale and Weakley 1990). The largest extant populations are found in the loamy soil variant of this community type that generally have higher pH and more nutrients. These areas are known locally as "pea swales" or "bean dips" because of their high diversity of Fabaceae species (James 2000). The species is largely restricted to the Fall-line Sandhills of North Carolina, South Carolina, Georgia, and Alabama (USA) (Sorrie and Weakley 2001; Peet 2006; NatureServe 2012), an extensive ancient dune system located at the boundary between the Coastal Plain and the Piedmont physiographic regions that is characterized by a rolling topography with excessively-well drained, sandy soils in the interfluvial areas.

Astragalus michauxii is one of the earliest flowering legumes in the Fall-line Sandhills, flowering in early May and producing mature fruits by early July (Radford et al. 1968). The species is most likely an obligate outcrosser,

and flowers are pollinated by a variety of flying insects. Fruit set is commonly quite low relative to the number of flowers; however the reasons for this have not been identified (e.g. an overabundance of self pollen, pollinator limitation, florivory, etc.). Like most legumes, *A. michauxii* produces a seed bank, but seed density appears to be rather low (Weeks 2005); possible factors include low seed set and pre- or post-dispersal seed predation. Seed dispersal is probably highly limited, since the seeds lack any obvious dispersal mechanism. Based on four years of observations across 39 populations, only one recruit was observed across all the monitored populations (Wall et al. 2012). Although census sizes are currently stable due to high adult survivorship, the probability of long-term positive population growth rates may be low because of the lack of recruitment (Wall et al. 2012).

NatureServe has given *A. michauxii* a global status of G3 (NatureServe 2012), indicating that the species is considered vulnerable to extinction due to a restricted range with few populations. In Georgia it has an S2 ranking, indicating that it is present in 20 or fewer populations and considered imperiled. *Astragalus michauxii* has an S3 ranking in North Carolina and South Carolina, indicating that the species is considered vulnerable to extinction with typically 21–100 populations and 3,000–5,000 individuals in total. In South Carolina, *A. michauxii* is only known from two populations. In North Carolina most of the populations are found on several public land holdings, mainly on Fort Bragg Military Reservation and the state-managed Sandhills Gamelands, with varying degrees of isolation between populations.

Sampling and population genetic methods

During June 2009–2010, leaves were collected from individuals located in 22 populations (as defined by the North Carolina Natural Heritage Program) on Fort Bragg and Camp Mackall Military Reservations (North Carolina) and from two populations in Georgia (Burke and Candler Counties) (Fig. 1). Leaf samples were stored in a −80 °C freezer until extraction. DNA was extracted from 355 individuals across the 24 populations, using the CTAB method with minor modifications (Doyle and Doyle 1987). This sampling included 80 % of individuals within the populations. Possible polymorphic microsatellite regions were identified using a recently published protocol (Jennings et al. 2011). Briefly, DNA was first sheared using a Bioruptor sonicator (Diagenode Inc., Denville, NJ, USA) and barcoded Illumina DNA libraries were created (Cronn et al. 2008). Libraries were enriched for microsatellites using hybridization with three probes containing repeated dinucleotide motifs. After hybrid capture, the microsatellite-enriched libraries were quantified using a Nanodrop 1000 spectrophotometer (Thermo Scientific, Wilmington, DE, USA). Libraries were then pooled and paired-end sequenced on one lane of an Illumina Genome Analyzer II. The resulting microreads were then sorted by the 4 bp barcodes and searched for dinucleotide motifs, (located near the center of reads to optimize primer development). After filtering redundant reads, sequences were analyzed using BatchPrimer3 to identify PCR primer sites (You et al. 2008).

Microsatellite-containing sequences were screened using agarose gels, and eight polymorphic loci were identified. 6 μL multiplexed PCR reactions were performed using fluorescently-labeled forward primers (6-FAM, HEX, NED) as follows: 3 μL Qiagen multiplex PCR master mix (Qiagen, Hilden, Germany), 0.6 μL Q-Solution, 0.8 μL H_2O, 0.6 μL multiplexed primer pair mix, and 1.0 μL diluted (1:8 DNA:H_2O) DNA. PCR cycling conditions were 95 °C for 15 min; 45 cycles at 94 °C for 30 s, 58 °C for 1 min 30 s., and 72 °C for 1 min; 60 °C for 30 min. PCR products were genotyped on an ABI 3730

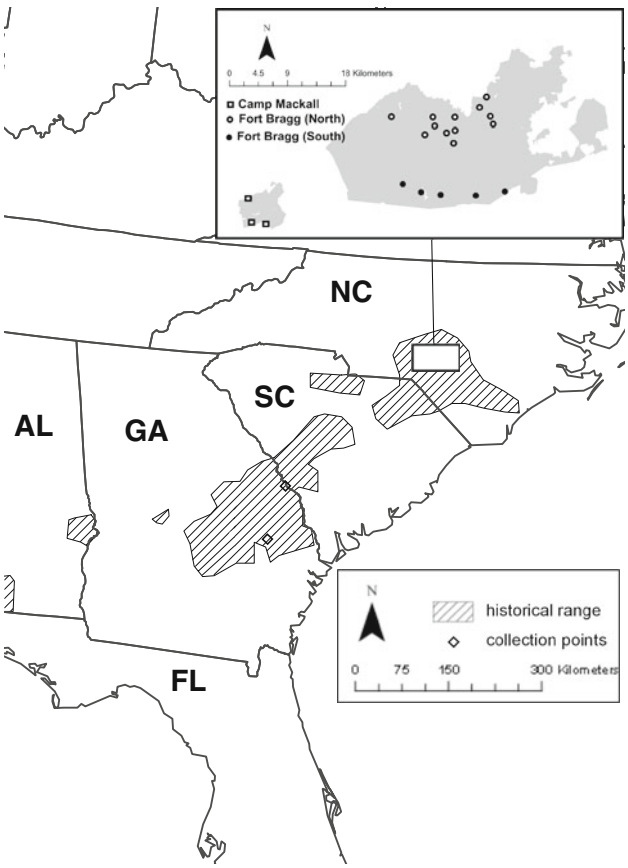

Fig. 1 Historic range and collection sites of *Astragalus michauxii*. Historical range determined based on voucher specimens (UNC Herbarium Flora of the Southeast; http://www.herbarium.unc.edu/seflora). Current range is greatly restricted, with most populations in North Carolina. Survey of Georgia populations only located 13 individuals

sequencer (Applied Biosystems, USA). PCR was performed again for loci that failed to amplify. After a second failed amplification, a locus was marked as missing for that sample. Genotyping was performed using GeneMarker 1.8 (Softgenetics LLC, State College, PA, USA). Exported peak heights were binned using TANDEM (Matschiner and Salzburger 2009). We genotyped 5 % of the individuals twice in order to assess data quality. Consistency across all loci in duplicate samples was 95.5 %.

Genetic structure and diversity

Hardy–Weinberg equilibrium and linkage disequilibrium among loci are assumptions in a number of different genetic analyses. We tested for significant departures from Hardy–Weinberg equilibrium (HWE) in Arlequin 3.5 (Excoffier and Lischer 2010) using a test analogous to Fisher's exact test (Guo and Thompson 1992), whereby the marginal probabilities of the observed allele frequencies are compared to values from simulated data sets explored using a Markov Chain. We performed the Markov Chain with 1,000,000 iterations and a burn-in of 100,000 iterations. We tested for linkage disequilibrium for the eight loci within individual populations in Arlequin 3.5 using a likelihood ratio test (Slatkin and Excoffier 1996), with no assumption of linkage equilibrium. Likelihood was calculated using the expectation–maximization (EM) algorithm (Dempster et al. 1977) to estimate allele frequencies. We performed 10,000 permutations with an initial EM value of 2. Since adjustments for multiple comparisons can make it difficult to identify significance, even when differences exist (Moran 2003), we assessed evidence of linkage disequilibrium using p-values both unadjusted and adjusted by the sequential Bonferroni correction method for multiple comparisons (Holm 1979). We calculated the average number of alleles and absolute number of private alleles (i.e. those found in a single population) for each population. Because populations consisted of varying numbers of individuals, we used a rarefaction method (Kalinowski 2004) for calculating allelic richness and private allelic richness available in HP-Rare 1.1 (Kalinowski 2005). We calculated expected and observed heterozygosity (Nei 1987) using Arlequin 3.5.

We explored the genetic structure of *A. michauxii* populations using the Bayesian clustering algorithm implemented in the software program BAPS 5.2 (Corander et al. 2003, 2004) to infer the number of genetic clusters (K). We used the spatial clustering option, as it has been shown to provide superior results if the genetic data are sparse (Corander et al. 2008). We performed 100 runs in fixed mode for each value of K from one to 24 and assessed admixture within the identified genetic clusters. We also explored the genetic structure of *A. michauxii* using STRUCTURE 2.3.4 under an admixture model with correlated allele frequencies with a burnin period = 100,000, 750,000 MCMC iterations after burnin, and averaged over three runs. STRUCTURE results are available in the supplemental material. We performed an analysis of molecular variance (AMOVA) (Excoffier et al. 1992) as implemented in Arlequin 3.5, to quantify the variance found within the BAPS-identified populations. We also used AMOVA to assess the overall significance of the defined groups. For comparative purposes, we estimated F_{ST} as calculated in Arlequin 3.5; the estimate is identical to the weighted θ (Weir and Cockerham 1984) and is the variance component among groups and the variance component among populations divided by the total variance. Population structure was further visualized graphically by performing a non-metric multi-dimensional scaling (NMDS) ordination using the R (R Development Core Team 2012) package labdsv (Roberts 2010). Pairwise population genetic distances were calculated using Nei's unbiased D (Nei 1978) as implemented in GenAlex 6.5 (Peakall and Smouse 2006). To test for isolation by distance (IBD), we performed a Mantel test using the R package vegan (Oksanen et al. 2009), using the log-transformed geographic distances (to improve normality) and the pairwise population genetic distances (Nei's unbiased D).

Estimating gene flow between populations

To detect possible first generation migrants between populations we used GeneClass2 (Piry et al. 2004), an implementation of the Bayesian method of Rannala and Mountain (1997). For each individual we computed the probability that it was a recent emigrant from another population (i.e. that the identified individual's allele frequencies are more similar to another population than its resident population). We used the simulation algorithm of Paetkau et al. (2004) with the following settings: type I error threshold of 0.01, 1,000 simulated individuals, and all loci included. Gene flow among populations may follow a source-sink metapopulation model, with gene flow from large populations to small populations. To test whether emigration was correlated with population size (i.e. whether larger populations contribute disproportionately to the number of migrants), we performed a logistic regression with emigration as the response variable and population size as the explanatory variable. We tested for significant differences in population size for resident and assigned populations of the identified migrants by simulating 10,000 replicate data sets with replacement and calculating the 95 % confidence interval for mean population size for each group of populations (resident or assigned).

Evidence of genetic bottlenecks across multiple temporal scales

To detect the genetic imprint of recent bottlenecks in *A. michauxii*, we used the software package BOTTLENECK 1.2.02 (Piry et al. 1999), an implementation of the method described by Cornuet and Luikart (1996). We fit two models using the program: the stepwise mutation model (SMM), and a two-phase model (TPM) mutation model with 95 % of the mutations single-step and a variance of 12 (Piry et al. 1999). For both models, 10,000 datasets were simulated with same observed number of alleles and population sample sizes under mutation-drift equilibrium, with the p value calculated as the probability of obtaining the mean expected heterozygosity (H_e) from the observed data based on the distribution of expected mutation-drift heterozygosity (H_{eq}) values. We used a one-tailed Wilcoxon signed rank test to detect significant heterozygote excess in populations because this test is most appropriate when the sample size is less than 30 and the number of loci is less than 10. We only tested for evidence of recent bottlenecks in populations that had >20 gene copies ($N = 10$ for diploid individuals).

To test for bottleneck events that may have occurred over longer time periods (>100 generations), we used the M ratio test (Garza and Williamson 2001) as implemented in Arlequin 3.5. This implementation of the M ratio test excludes monomorphic loci because these can erroneously increase the M ratio. The M ratio is the mean number of alleles in a population divided by the allelic size range. When alleles are lost from a population, the number of alleles decreases at a faster rate than the allelic size range, so small (<0.68) M ratio values are indicative of populations that have gone through a genetic bottleneck at some time in the past. We estimated 95 % confidence intervals by resampling the M ratio estimates for each locus within a population with replacement 10,000 times (Swatdipong et al. 2009). We also calculated a one-sided 95 % confidence interval (M_c) for each population. M_c is a value estimated through 10,000 simulations of a population at equilibrium such that M_c is less than the simulated M values in 95 % of the simulations based on the mean size of non-stepwise mutations and θ. Settings were mean size of non-stepwise mutations = 3.5, θ = 10, and 10,000 iterations, as recommended by Garza and Williamson (2001).

Results

Microsatellite development generated 100 candidate loci, from which we developed eight polymorphic microsatellite loci that amplified consistently and produced no more than two bands in our initial screens (Table 1). Allele frequencies detected significant departures from Hardy–Weinberg expectations in only 19 out of 176 tests at the population level (22 populations * 8 loci); however, our power to detect significant departures was low because of small sample sizes in many of the populations. The northern populations on Fort Bragg had more loci out of HWE compared to populations from other regions. Microsatellite marker AM29 was out of HWE in eight of 22 populations and three of four regions, indicating that this locus may be influenced by null alleles or allelic dropout. After correcting for multiple tests, none of the loci were significantly out of HWE ($p > 0.05$). Linkage disequilibrium (LD) was detected in 91 of 616 tests representing all possible pairwise loci combinations within populations, although only three comparisons were significant after using sequential Bonferroni adjustment. As with testing Hardy–Weinberg equilibrium, our power to detect LD was low because of small sample sizes. Significant evidence of LD was not isolated to pairs of loci, but rather was encountered in all loci compared and was never greater than 30 % (Table S1), suggesting that the observed LD is due to population-level effects such as null alleles, admixture, inbreeding, or genetic drift due to a bottleneck event, rather than physical linkage between loci.

Genetic diversity in *A. michauxii* averaged 10.88 alleles per locus across all populations, with larger populations having a greater number of alleles compared to smaller populations ($R^2 = 0.78$, $p < 0.001$; Table 2). After adjusting for sample size, allelic richness was similar across populations ($R^2 = 0.11$, $p = 0.07$). Allelic richness ranged from 2.43 to 3.80 across populations. As with allele number, the number of private alleles increased with population size ($R^2 = 0.39$, $p = 0.001$), but after rarefaction there was not a significant correlation ($R^2 = 0$, $p > 0.05$).

Genetic clustering indicated the highest posterior probability occurred when the populations were grouped into three clusters of populations. (Table S2 in supplementary material). These clusters combined the Camp Mackall and southern Fort Bragg populations into one cluster, combined the northern and northeastern Fort Bragg populations into another cluster, and separated the two Georgia populations into a distinct cluster (Fig. 2; Fig. S2 in supplemental material includes BAPS and STRUCTURE results for $K = 2$–9). AMOVA results indicated that within population genetic variation accounted for 91.3 % of the total genetic variation (Table 3), with the genetic clusters (K) identified by BAPS 5.2 accounting for 2.9 % of the overall genetic variation. F_{ST} was estimated at 0.08. When populations were grouped according to geographical location, non-metric multidimensional scaling graphically demonstrated the modest population differentiation among *A. michauxii* populations (Fig. 3; Table S3 in supplemental

Table 1 Eight polymorphic loci identified and developed for *Astragalus michauxii*

Primer	Sequence	Repeat	Fragment size (BP)	Alleles
AM_15	F: GTTTCACACTGAGACACAGTTC	GA	24–34 (124–134)	6
	R: AATTCCCAAGTGTAAAAGCTC			
AM_18	F: GAAAACACAAACAAATTCTGG	GA	8–38 (165–195)	13
	R: AGAAAGTCTGTGCTCTCTCATT			
AM_25	F: CAATCCCTAACCTTGAGTTCT	GA	8–36 (107–135)	14
	R: AGCAACGTGGGATAAAAATA			
AM_29	F: AACGGTGTCTGTGTCTATGTC	GT	32–42 (160–170)	6
	R: ATGAAGCGTTTCACATTTTT			
AM_34	F: TGACATACATGCTGAAAGTTG	AG	20–26 (155–161)	4
	R: TTTGGATTCATATAACCACCA			
AM_46	F: GAAAATGGTGGAAAAGGAAT	AG	18–64 (102–148)	22
	R: GTGTAAAAATCGTGCACTTCT			
AM_71	F: AAGATTGTCTAACGATCACCA	GT	187–203 (201 missing)	7
	R:AAAGCCCATGTTTCACTAAAT			
AM_91	F: GGACAAAAGAAGAGGAGAGAG	AG(TACTGG)TG	22–40 (107–125)	10
	R: TAAGTCGAGTTGTTCCAAAGT			

Columns are primer pair name, sequence, repeat, microsatellite range (with fragment size in parentheses), and number of alleles observed over all 22 populations

Table 2 Genetic variation in nineteen *Astragalus michauxii* populations from North Carolina and Georgia, as well as five local and regional geographic groupings, based on eight polymorphic microsatellite loci

Population	Region	N	A	A_R	P	P_R	H_o	H_e	H–W disequilibrium
ASMI053	Camp Mackall	7	3.43	2.61	1.00	0.11	0.55	0.53	None
ASMI054	Camp Mackall	14	4.88	3.23	0.00	0.07	0.59	0.62	AM29
ASMI088	Camp Mackall	5	3.13	2.91	1.00	0.14	0.51	0.55	None
ASMI020	Fort Bragg—North	5	3.13	2.95	0.00	0.12	0.55	0.55	None
ASMI022	Fort Bragg—North	7	4.00	3.42	1.00	0.14	0.60	0.69	AM29, AM46
ASMI023	Fort Bragg—North	27	5.50	3.38	1.00	0.04	0.58	0.66	AM29
ASMI032	Fort Bragg—North	30	6.13	3.48	0.00	0.03	0.62	0.65	None
ASMI034	Fort Bragg—North	24	6.63	3.80	2.00	0.11	0.56	0.71	AM18, AM25, AM29, AM91
ASMI035	Fort Bragg—North	8	4.25	3.36	0.00	0.05	0.58	0.66	AM18, AM25, AM29, AM91
ASMI049	Fort Bragg—North	13	4.50	3.36	0.00	0.00	0.63	0.64	None
ASMI056	Fort Bragg—North	68	7.75	3.63	2.00	0.10	0.62	0.69	AM29, AM34, AM71
ASMI057	Fort Bragg—North	33	5.00	2.93	1.00	0.02	0.47	0.56	AM18, AM46
ASMI061	Fort Bragg—North	6	2.86	2.43	0.00	0.01	0.62	0.55	None
ASMI091	Fort Bragg—North	12	5.00	3.56	0.00	0.07	0.67	0.69	None
ASMI096	Fort Bragg—North	4	3.50	3.50	0.00	0.12	0.69	0.62	None
ASMI030	Fort Bragg—South	9	4.25	3.29	0.00	0.07	0.57	0.62	None
ASMI046	Fort Bragg—South	11	3.88	3.04	0.00	0.02	0.56	0.58	None
ASMI050	Fort Bragg—South	34	5.75	3.19	3.00	0.08	0.58	0.61	AM29
ASMI051	Fort Bragg—South	24	6.00	3.32	1.00	0.10	0.54	0.59	AM29
ASMI097	Fort Bragg—South	5	3.00	2.81	0.00	0.09	0.48	0.53	None
AMBUGA	Georgia	4	3.25	3.25	0.00	0.03	0.56	0.65	None
AMMEGA	Georgia	5	3.63	3.34	0.00	0.12	0.55	0.66	None

Column headings are: *N*, number of individuals; *A*, average number of alleles across loci; A_R, average allelic richness; *P*, number of private alleles; P_R, private allelic richness; H_O, observed heterozygosity; H_e, expected heterozygosity; *H–W disequilibrium*, loci identified as not in Hardy–Weinberg equilibrium

material includes Nei's unbiased *D* pair-wise population distances). The two Georgia populations separated in ordination space from the North Carolina populations, but little separation occurred among the North Carolina populations. Isolation by distance (IBD) results indicated that the genetic distance between populations increased with

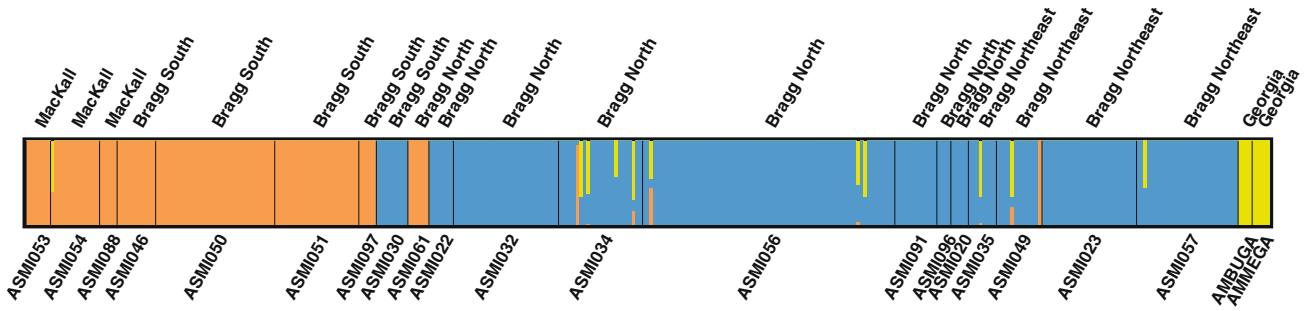

Fig. 2 Population genetic structure for *Astragalus michauxii* as determined by the program BAPS 5.2 using an admixture model with spatial information included and $K = 3$. The highest posterior probability was $K = 3$, with the Camp Mackall and southern Fort Bragg populations clustered and the northern Fort Bragg populations as a second cluster. The Georgia populations form the third cluster

Table 3 Analysis of molecular variance (AMOVA) results for *Astragalus michauxii* populations from North Carolina and Georgia (USA)

Groups (K) correspond to genetic clusters as identified using BAPS 5.2

Source of variation	*d.f.*	Sum of squares	Variance components	Percentage of variation	*p*-value
Among groups (K)	2	39.722	0.07809	2.88	<0.001
Among populations within regions	19	139.067	0.15686	5.78	<0.001
Within populations	688	1704.200	2.47703	91.34	<0.001
Total	709	1882.989	2.71198		

the log of the geographic distance ($r = 0.4553$, $p = 0.003$; Fig. S2 in supplemental material) when all populations were included. However, removal of the Georgia populations indicated that population genetic distance did not increase with the log of the geographic distance for North Carolina populations ($r = 0.05$, $p = 0.36$).

BOTTLENECK results did not indicate evidence of a recent genetic bottleneck in any population under the stepwise mutation model (SMM) or the two-phase mutation model (TPM) based on Wilcoxon's test (Table 4). Contrary to BOTTLENECK results, Critical *M* results indicated a severe bottleneck in *A. michauxii* populations. *M* ratio values across all populations averaged 0.48 and were lower than the 0.68 threshold identified by Garza and Williamson (2001) as indicative of a past genetic bottleneck, with the upper 95 % CI for all estimated *M* ratio values less than the 0.68 threshold in all populations (Fig. 4). The M_c values (90 % SMM) were greater than the observed *M* ratio values in 21 out of 22 populations and the M_c values (80 % SMM) were greater than the observed *M* ratio values in 14 out of 22 populations. Populations with M_c values less than the observed *M* ratio had smaller population sizes (and sample sizes) relative to populations with *M* ratio values less than the M_c values.

For the Fort Bragg and Camp Mackall populations, GeneClass2 results identified 14 putative first-generation migrants out of a total of 346 individuals (Table 5). The average distance between the population where the migrant

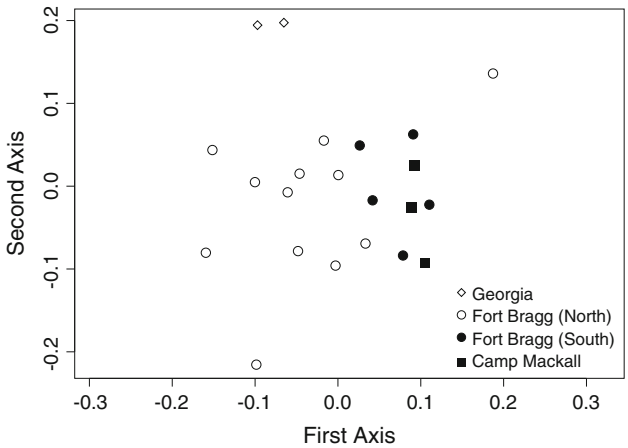

Fig. 3 Non-metric multidimensional scaling ordination of *Astragalus michauxii* population genetic distances based on eight polymorphic microsatellite loci. The Georgia populations appear separate from the North Carolina populations, while the North Carolina populations from Fort Bragg and Camp Mackall are more similar in terms of population genetic distance

was found ("sink") and the source population for individuals identified as migrants was 15.8 km. This distance was not statistically different than the average distance between all sampled North Carolina populations: 15.6 km. There was not a significant correlation between migration and population size, with the population sizes of "sink" populations similar to overall population sizes (18.5 ± 2.6 s.d. vs. 17.3 ± 3.4 respectively; $p > 0.05$).

Table 4 Tests for genetic bottlenecks in *Astragalus michauxii* using BOTTLENECK version 1.2 in populations with >20 gene copies and *M* ratio for all populations as calculated in Arlequin 3.1

Population	Copies	Wilcoxon's test (SMM)	Wilcoxon's test (TPM)	*M* ratio
ASMI020	9.25	–	–	0.43205
ASMI022	13.25	–	–	0.40397
ASMI023	53	0.62891	0.47266	0.49321
ASMI030	18	–	–	0.4812
ASMI032	59.5	0.875	0.76953	0.50957
ASMI034	46	0.84375	0.80859	0.49201
ASMI035	16	–	–	0.49813
ASMI046	22	0.27344	0.27344	0.4777
ASMI049	26	0.15625	0.125	0.48117
ASMI050	67	0.875	0.875	0.456
ASMI051	48	0.99414	0.99414	0.45295
ASMI053	14	–	–	0.50025
ASMI054	27.75	0.76953	0.72656	0.47976
ASMI056	141.5	0.99023	0.98047	0.5195
ASMI057	57.25	0.97266	0.67969	0.46245
ASMI061	12	–	–	0.44602
ASMI088	9.75	–	–	0.43187
ASMI091	24	0.67969	0.47266	0.48862
ASMI096	8	–	–	0.47949
ASMI097	10	–	–	0.50159
AMBUGA	8	–	–	0.5181
AMMEGA	10	–	–	0.49401

Column heading "Copies" is the average number of samples included per locus. Results indicate no recent genetic bottleneck events, but evidence of a severe genetic bottleneck in the more distant past

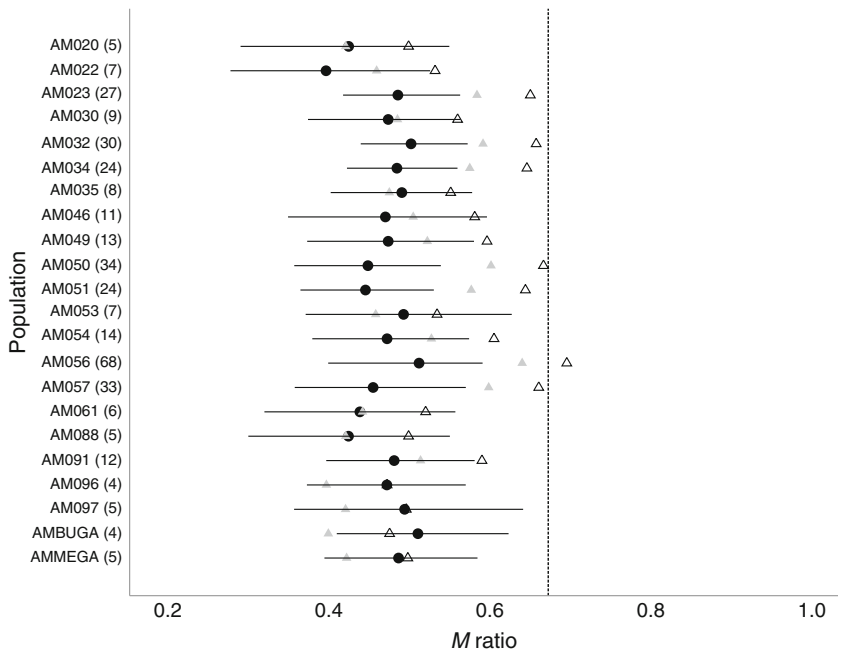

Fig. 4 *M* ratio values (*black circles*) estimated for 22 *Astragalus michauxii* populations across North Carolina and Georgia (USA). *Horizontal lines* represent 95 % CIs, the *vertical line* is the threshold indicative of a past genetic bottleneck, the *open triangles* are the critical M_c (90 % single step mutation model, SSM), and the *gray triangles* are the critical M_c (80 % SMM) (Garza and Williamson 2001). Population sizes are in *parentheses*

Discussion

Habitat fragmentation and degradation has reduced the size and genetic diversity within many plant populations. However, we found that within *A. michauxii* populations, genetic diversity is not lower than other perennial plant species, despite the fact that the longleaf pine ecosystem to which it is endemic has experienced widespread reduction and degradation over the past few centuries. While there are difficulties inherent in comparing microsatellite

diversity across species, mean estimates of expected (H_e) and observed (H_o) heterozygosity for *A. michauxii* populations (0.68 and 0.57 respectively) were comparable to the average H_e (0.61 ± 0.21) and H_o (0.58 ± 0.22) found in a review of plant microsatellite data sets (Nybom 2004).

Populations of *A. michauxii* also exhibited relatively low genetic differentiation, counter to what we expected for a species with varying degrees of population spatial isolation. Within-population genetic variation accounted for 91 % of the total genetic variation; this is not unexpected as other studies have shown that outcrossing, perennial plant species maintain the majority of their genetic diversity within populations (Hamrick and Godt 1996; Nybom 2004). These results are consistent with the genetic structure of three other putative relictual species endemic to the Gulf and Atlantic Coastal Plain Fall-line Sandhills that maintain a large portion of the overall genetic diversity within populations. *Amorpha georgiana* Wilbur (Fabaceae) is a shrub that occurs along river terraces of blackwater rivers in the Fall-line Sandhills and was estimated to contain 89 % of its genetic variation within populations (Straub and Doyle 2009). *Pyxidanthera brevifolia* Wells (Diapensiaceae) occupies xeric upland habitats similar to those occupied by *A. michauxii*, with genetic evidence suggesting within population genetic variation accounted for 90.5 % of the total genetic variation of the species (Wall et al. 2010). Finally, *Lilium pyrophilum* M.W. Skinner and Sorrie (Liliaceae), which occupies relatively more mesic habitats in the Fall-line Sandhills (Douglas et al. 2011) also demonstrates low population differentiation (N. A. Douglas, unpublished data).

The relatively modest genetic differentiation and moderate genetic diversity within *A. michauxii* suggest that population fragmentation occurred following the extensive fragmentation, reduction, and degradation of the longleaf pine ecosystem. Either the populations have not been isolated long enough for genetic drift and mutation to have impacted population differentiation, or gene flow persists despite a fragmented distribution. While the two *A. michauxii* populations in Georgia appear to be distinct from the North Carolina populations based on the NMDS ordination, the genetic differences between the two regions may best be explained by an isolation by distance model (Fig. 3 in supplemental material). If populations had been isolated since the end of Pleistocene, we would expect greater genetic differentiation between populations and/or lower genetic diversity, as has been found in other species (Reisch et al. 2003).

However, these scenarios are not mutually exclusive. If *A. michauxii* is a long-term inhabitant of the Fall-line Sandhills (which has been demonstrated for another endemic taxon, Wall et al. 2010), then it is possible that climatic changes since the Pleistocene, as well as European settlement, have led to the current habitat fragmentation of the Georgia and North Carolina populations. The formation of aeolian river dunes and braided river channels in the Atlantic Coastal Plain during the Late Pleistocene (Ivester et al. 2001; Leigh 2008) suggest an environment with exposed soil and dry, windy climatic conditions (Leigh 2008), not unlike the Great Basin and other ecosystems with a large number of *Astragalus* species (Barneby 1964). The colder, drier conditions of the Pleistocene most likely reduced plant productivity and biomass accumulation, and the region would have been characterized as an open savanna with scattered *Picea* and *Pinus* species and an herbaceous understory (Watts 1980). As climatic conditions became progressively warmer and wetter during the Holocene, it is likely that the biomass and canopy cover increased in many areas. Evidence from extant and historical local populations suggests that *A. michauxii* is sensitive to woody encroachment (North Carolina Natural Heritage Program data). Thus *A. michauxii* may have become restricted to habitats with xeric soil conditions where competition was reduced through increasing fire frequency (upland sites in the Fall-line Sandhills).

These reductions in population size could have led to a genetic bottleneck. Although within population genetic diversity of *A. michauxii* may be comparable to that of other perennial plant species, there is evidence of a past genetic bottleneck based on the ratio of the number of alleles to the allelic range size (*M*-ratio) within the 24 sampled populations. Although the expected mutation-drift heterozygosity (H_{eq}) was not significantly less than H_e in the BOTTLENECK analysis, indicating no evidence of a recent genetic bottleneck, this method may not be as sensitive as evaluating *M* ratio values (Girod et al. 2011). Furthermore, *M* ratio values may be reduced for 100 or more generations, longer than the statistics evaluated in BOTTLENECK. Since *A. michauxii* has an estimated generation time of 9.6 years (Wall, unpublished data), it is not possible to identify whether the bottleneck occurred before or after the anthropogenic fragmentation of the longleaf pine ecosystem using the *M* ratio values alone, as they could be detecting bottlenecks up to 9.6 ka.

For the North Carolina populations, there is evidence of continued interpopulation gene flow (Table 5). Indeed, recent gene flow appears to have occurred broadly across the North Carolina populations, with no evidence of isolation by distance in these populations. However, we must distinguish between statistical migrants, which GeneClass2 identifies based on population gene frequencies, and actual gene flow via pollen or seeds. The results suggest gene flow across an average of 15 km. This is an extreme distance (Greenleaf et al. 2007) for an entomophilous species (Karron 1987; Geer et al. 1995; Crone and Lesica 2004; Becker et al. 2011) with no obvious long distance dispersal

Table 5 *Astragalus michauxii* individuals identified by GeneClass 2 as being the result of possible interpopulation gene flow ($p \geq 0.01$)

Resident population	Assigned population	Individual	Probability	Distance (km)
ASMI022	ASMI032	Ind-36	0.005	5.098
ASMI023	ASMI054	Ind-45	0.006	40.415
ASMI035	ASMI054	Ind-124	0.001	39.628
ASMI049	ASMI034	Ind-137	0.009	6.986
ASMI051	ASMI034	Ind-323	0.003	10.304
ASMI051	ASMI030	Ind-327	0.005	5.379
ASMI053	ASMI050	Ind-1	0.005	26.853
ASMI054	ASMI032	Ind-9	0.01	32.460
ASMI056	ASMI034	Ind-198	0.003	1.956
ASMI056	ASMI046	Ind-188	0.007	13.891
ASMI057	ASMI096	Ind-219	0.001	8.783
ASMI057	ASMI097	Ind-223	0.006	17.144
ASMI057	ASMI035	Ind-238	0.01	2.184
ASMI096	ASMI049	Ind-263	0.01	9159
ASMI097	ASMI057	Ind-342	0	17,144

Resident population refers to the population in which the individual was actually found and assigned population indicates the most likely source of the inferred gene flow (source population). Fourteen individuals across eleven populations were identified

adaptation. Thus, the *A. michauxii* individuals identified as migrants should not necessarily be viewed as actual migrants, but rather as indicative of contemporary gene flow.

Both genetic and demographic factors can affect the long-term viability of plant populations. Our results suggest limited genetic effects of habitat fragmentation and population isolation within *A. michauxii.* However, maintaining future connectivity between populations will be necessary to reduce the negative impacts of future genetic drift and inbreeding. While *A. michauxii* populations may not currently be affected by deleterious genetic processes, demographic analyses suggest that *A. michauxii* may not be maintaining stable populations (Wall et al. 2012), mainly due to limited recruitment of seedlings. This suggests maintaining or increasing population sizes through active management of habitat or more intensive measures, such as augmentation or reintroduction, will likely be critical for the persistence of the species. If active management options are warranted, we recommend that seed source be restricted to the three identified genetic clusters, especially for the Georgia populations. This will maintain the relative genetic distinctiveness of the Georgia populations and the variability identified in the North Carolina populations.

Acknowledgments We thank Fort Bragg Military Reservation for access and support and the Construction Engineering Research Laboratory (US Army Corps of Engineers Cooperative Agreement #W9132T-07-2-0019 to W. Hoffmann) for funding. We also thank Bert Pittman, Sherry Emerine, Jacob Hilton, Kristen Kostelnik, Linda Lee, Xiang Liu, the North Carolina Natural Heritage Program, Tom Patrick, Andrew Walker, the Xiang Lab, and three anonymous reviewers. Assistance with microsatellite development was generously provided by Pacific Northwest Research Station (USDA Forest Service).

References

Balmford A, Bond W (2005) Trends in the state of nature and their implications for human well-being. Ecol Lett 8:1218–1234

Barneby RC (1964) Atlas of North American Astragalus, vol 13. Memoirs of the New York Botanical Garden, NewYork, pp 911–958

Becker T, Voss N, Durka W (2011) Pollen limitation and inbreeding depression in an "old rare" bumblebee-pollinated grassland herb. Plant Biol 13:857–864

Corander J, Waldmann P, Sillanpää MJ (2003) Bayesian analysis of genetic differentiation between populations. Genetics 163: 367–374

Corander J, Waldmann P, Marttinen P, Sillanpää MJ (2004) BAPS 2: enhanced possibilities for the analysis of genetic population structure. Bioinformatics 20:2363–2369

Corander J, Sirén J, Arjas E (2008) Bayesian spatial modeling of genetic population structure. Comput Stat 23:111–129

Cornuet JM, Luikart G (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. Genetics 144:2001–2014

Crone EE, Lesica P (2004) Causes of synchronous flowering in *Astragalus scaphoides*, an iteroparous perennial plant. Ecology 85:1944–1954

Cronn R, Liston A, Parks M et al (2008) Multiplex sequencing of plant chloroplast genomes using Solexa sequencing-by-synthesis technology. Nucleic Acids Res 36:e122

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. J R Stat Soc: Ser B (Methodol) 39:1–38

Dobson AP, Rodriguez JP, Roberts WM, Wilcove DS (1997) Geographic distribution of endangered species in the United States. Science 275:550–553

Douglas NA, Wall WA, Xiang QY et al (2011) Recent vicariance and the origin of the rare, edaphically specialized Sandhills lily, *Lilium pyrophilum* (Liliaceae): evidence from phylogenetic and coalescent analyses. Mol Ecol 20:2901–2915

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull 19:11–15

Excoffier L, Lischer HL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour 10:564–567

Excoffier L, Smouse PE, Quattro J (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131:479–491

Frankham R (2005) Genetics and extinction. Biol Conserv 126: 131–140

Frankham R, Ballou JD, Briscoe DA (2010) Introduction to conservation genetics, 2nd edn. Cambridge University Press, Cambridge

Franklin IR (1980) Evolutionary change in small populations. In: Soule ME, Wilcox BA (eds) Conservation biology: an evolutionary-ecological perspective. Sinauer Associates, Sunderland, pp 135–149

Frost CC (1993) Four centuries of changing landscape patterns in the longleaf pine ecosystem. In: Hermann SM (ed) Proceedings of the tall timbers fire ecology conference. Tall Timbers Research Station, Tallahassee, pp 17–37

Frost CC (2006) History and future of the longleaf pine ecosystem. In: Jose S, Jokela EJ, Miller DL (eds) The longleaf pine ecosystem: ecology, silviculture, and restoration. Springer Science, New York, pp 9–42

Garza JC, Williamson EG (2001) Detection of reduction in population size using data from microsatellite loci. Mol Ecol 10:305–318

Geer SM, Tepedino VJ, Griswold TL, Bowlin WR (1995) Pollinator sharing by three sympatric milkvetches, including the endangered species *Astragalus montii*. Great Basin Nat 55:19–28

Girod C, Vitalis R, Leblois R, Freville H (2011) Inferring population decline and expansion from microsatellite data: a simulation-based evaluation of the msvar method. Genetics 188:165–179

Greenleaf SS, Williams NM, Winfree R, Kremen C (2007) Bee foraging ranges and their relationship to body size. Oecologia 153:589–596

Guo SW, Thompson EA (1992) Performing the exact test of Hardy–Weinberg proportion for multiple alleles. Biometrics 48:361–372

Hamrick JL, Godt MJW (1989) Allozyme diversity in plant species. In: Brown AHD, Clegg MT, Kahler AL, Weir BS (eds) Plant population genetics, breeding and genetic resources. Sinauer Associates, Sunderland, pp 43–63

Hamrick JL, Godt MJW (1996) Effects of life history traits on genetic diversity in plant species. Philos Trans R Soc B: Biol Sci 351: 1291–1298

Holm S (1979) A simple sequentially rejective multiple test procedure. Scand J Stat 6:65–70

Ivester AH, Leigh DS, Godfrey-Smith DI (2001) Chronology of inland eolian dunes on the Coastal Plain of Georgia, USA. Quat Res 55:293–302

James MMR (2000) Legumes in loamy soil communities of the Carolina sandhills: their natural distributions and performance of seeds and seedlings along complex ecological gradients. Master of Science thesis, University of North Carolina

Jennings TN, Knaus BJ, Mullins TD et al (2011) Multiplexed microsatellite recovery using massively parallel sequencing. Mol Ecol Resour 11:1060–1067

Kalinowski ST (2004) Counting alleles with rarefaction: private alleles and hierarchical sampling designs. Conserv Genet 5:539–543

Kalinowski ST (2005) HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic richness. Mol Ecol Notes 5:187–189

Karron JD (1987) The pollination ecology of co-occuring geographically restricted and widespread species of *Astragalus* (Fabaceae). Biol Conserv 39:179–193

Leigh DS (2008) Late quaternary climates and river channels of the Atlantic Coastal Plain, Southeastern USA. Geomorphology 101: 90–108

Leimu R, Mutikainen PIA, Koricheva J, Fischer M (2006) How general are positive relationships between plant population size, fitness and genetic variation? Ecology 94:942–952

Matschiner M, Salzburger W (2009) TANDEM: integrating automated allele binning into genetics and genomics workflows. Bioinformatics 25:1982–1983

Moran MD (2003) Arguments for rejecting the sequential Bonferroni in ecological studies. Oikos 100:403–405

NatureServe (2012) NatureServe Explorer: An online encyclopedia of life (web application), Version 7.1. http://www.natureserve.org/explorer. Accessed 28 May 2012

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. Columbia University Press, New York

Noss RF (2013) Forgotten grasslands of the South: natural history and conservation. Island Press, Washington

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

Oksanen J, Kindt R, Legendre P, et al. (2009) Vegan: community ecology package

Paetkau D, Slade R, Burden M, Estoup A (2004) Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power. Mol Ecol 13:55–65

Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol Ecol Notes 6:288–295

Peet RK (2006) Ecological classification of longleaf pine woodlands. In: Jose S, Jokela EJ, Miller DL (eds) The longleaf pine ecosystem: ecology, silviculture, and restoration. Springer, New York, pp 51–93

Piry S, Luikart G, Cornuet J-M (1999) BOTTLENECK: a computer program for detecting recent reductions in the effective population size using allele frequency data. J Hered 90:502–503

Piry S, Alapetite A, Cornuet JM et al (2004) GENECLASS2: a software for genetic assignment and first-generation migrant detection. J Hered 95:536–539

R Development Core Team (2012) R: a Language and environment for statistical computing. Austria, Vienna

Radford AE, Ahles HE, Bell CR (1968) Manual of the vascular flora of the Carolinas. University of North Carolina Press, Chapel Hill

Rannala B, Mountain JL (1997) Detecting immigration by using multilocus genotypes. Proc Natl Acad Sci 94:9197–9201

Reisch C, Poschlod P, Wingender R (2003) Genetic variation of Saxifraga paniculata Mill. (Saxifragaceae): molecular evidence for glacial relict endemism in central Europe. Biol J Linn Soc 80:11–21

Roberts DW (2010) labdsv: ordination and multivariate analysis for ecology

Schafale MP, Weakley AS (1990) Classification of the natural communities of North Carolina: third approximation, p 321

Scherson RA, Vidal R, Sanderson MJ (2008) Phylogeny, biogeography, and rates of diversification of New World *Astragalus* (Leguminosae) with an emphasis on South American radiations. Am J Bot 95:1030–1039

Slatkin M, Excoffier L (1996) Testing for linkage disequilibrium in genotypic data using the expectation–maximization algorithm. Heredity 76:377–383

Sorrie BA, Weakley AS (2001) Coastal plain vascular plant endemics: phytogeographic patterns. Castanea 66:50–82

Spielman D, Brook BW, Frankham R (2004) Most species are not driven to extinction before genetic factors impact them. Proc Natl Acad Sci USA 101:15261–15264

Straub SCK, Doyle JJ (2009) Conservation genetics of *Amorpha georgiana* (Fabaceae), an endangered legume of the Southeastern United States. Mol Ecol 18:4349–4365

Swatdipong A, Primmer CR, Vasemägi A (2009) Historical and recent genetic bottlenecks in European grayling, *Thymallus thymallus*. Conserv Genet 11:279–292

Van Lear DH, Carroll W, Kapeluck P, Johnson R (2005) History and restoration of the longleaf pine-grassland ecosystem: implications for species at risk. For Ecol Manag 211:150–165

Wall WA, Douglas NA, Xiang QY et al (2010) Evidence for range stasis during the latter Pleistocene for the Atlantic Coastal Plain endemic genus, *Pyxidanthera Michaux*. Mol Ecol 19:4302–4314

Wall WA, Hoffmann WA, Wentworth TR et al (2012) Demographic effects of fire on two endemic plant species in the longleaf pine-wiregrass ecosystem. Plant Ecol 213:1093–1104

Watts WA (1980) Late-quaternary vegetation history at white pond on the inner Coastal Plain of South Carolina. Quat Res 13:187–199

Webb T III (1987) The appearance and disappearance of major vegetational assemblages: long-term vegetational dynamics in eastern North America. Theory and models in vegetation science. Springer, Berlin, pp 177–187

Weeks S (2005) Factors limiting growth in *Astragalus michauxii* (sandhills milk-vetch). Doctoral dissertation, North Carolina State University

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. Evolution 38:1358–1370

Williams JW, Shuman BN, Webb T III et al (2004) Late-quaternary vegetation dynamics in North America: scaling from taxa to biomes. Ecol Monogr 74:309–334

You F, Huo N, Gu Y et al (2008) BatchPrimer3: a high throughput web application for PCR and sequencing primer design. BMC Bioinformatics 9:253

Young A, Boyle T, Brown T (1996) The population genetic consequences of habitat fragmentation for plants. Trends Ecol Evol 11:413–418