Conserv Genet (2011) 12:1015–1026
DOI 10.1007/s10592-011-0204-2

RESEARCH ARTICLE

Patterns of differentiation among endangered pondberry populations

Craig S. Echt · Dennis Demeer · Danny Gustafson

Received: 10 September 2010 / Accepted: 17 March 2011 / Published online: 1 April 2011

Abstract Pondberry, *Lindera melissifolia*, is an endangered and partially clonally reproducing shrub species found in isolated populations that inhabit seasonally wet depressions in forested areas of the lower Mississippi River alluvial valley and southeastern regions of the United States. With eleven microsatellite loci, we quantified population genetic differentiation and diversity among 450 genets in 10 locations distributed across pondberry's range. We used estimates of F_{st} and Jost's D_{est} to measure genetic differences between populations and between geographic regions. The largest pairwise regional difference was found between eastern and western regional population groups ($F_{st} = 0.23$, $D_{est} = 0.67$), with the northern-most population groups in each region exhibiting larger divergence from each other than the southern-most population groups. Genetic diversity was lowest in the Sand Pond Conservation Area ($A_e = 1.9$, $H_e = 0.36$), which was the northern-most pondberry population, and highest in the Francis Marion National Forest ($A_e = 4.1$, $H_e = 0.69$), although we identified only 17 genets in that admixed population. Following adjustments for estimated null allele frequencies, we identified heterozygote excess in four eastern populations and found no evidence for inbreeding in any population. The observed patterns of differentiation indicate a phylogeography that exhibits an Appalachian Mountain discontinuity coupled with northward migrations along the Southern Atlantic Coastal Plain and into the Mississippi Alluvial Plain. The genetic consequences of this proposed phylogeographical structure may affect selection of germplasm sources for population reestablishment programs across pondberry's range.

Keywords Clonal plant · Heterozygote excess · Phylogeography · Nuclear SSR markers

Introduction

Lindera melissifolia (Walt.) Blume, known as pondberry, is an uncommon, dioecious, rhizomatous shrub of the Lauraceae. It exists in sparsely distributed and highly localized populations across seven states in southeastern United States (U.S. Fish and Wildlife Service 2010). Pondberry occupies microhabitats of seasonally flooded depressions and sinks in wooded areas, where it characteristically grows in discrete colony patches of sparsely branched deciduous woody stems with aromatic leaves (Wright 1990; U.S. Fish and Wildlife Service 1993; Godt and Hamrick 1996; Shotz 2005). Seed production is common although seedling establishment is rare, vegetative reproduction predominates in the form of single-sex colonies, and population sex ratios tend to be biased toward males, from 7:1 to 19:1 (Wright 1994; Devall et al. 2001; Hawkins et al. 2007). The species is endemic to three ecoregions of the continental United States: the Middle Atlantic Coastal Plain and southern portions of the Southeastern Plains, both of which are characterized by broadleaf-coniferous evergreen forests, and the Mississippi Alluvial Plain, which is

Electronic supplementary material The online version of this article (doi:10.1007/s10592-011-0204-2) contains supplementary material, which is available to authorized users.

C. S. Echt (✉) · D. Demeer
Southern Institute of Forest Genetics, Southern Research Station, USDA Forest Service, 23332 Success Road, Saucier, MS 39574, USA
e-mail: cecht@fs.fed.us

D. Gustafson
Department of Biology, The Citadel, 171 Moultrie Street, Charleston, SC 29409, USA

characterized by riverine bottomland forests (Environmental Protection Agency 2010).

Pondberry is listed as federally endangered (U.S. Fish and Wildlife Service 1986) and has a global status of G2—Imperiled (NatureServe 2010). The 1993 U.S. Fish and Wildlife Service Pondberry Recovery Plan specifies that permanent protection of 25 self-sustaining populations, distributed throughout the historic range, is required to remove pondberry's officially endangered status (U.S. Fish and Wildlife Service 1993). Determining what constitutes a "self-sustaining population" is one of the tasks of the Recovery Plan. That determination is contingent on defining the biological and ecological requirements of the species, as well as on quantifying genetic variation within and between colonies and populations (U.S. Fish and Wildlife Service 1993). The Recovery Plan also calls for protecting pondberry genetic resources by establishing seed banks, developing cultivation methods, and planting new or reintroduced populations and colonies with seeds and planting stock from nurseries.

Although there are a number of recent studies on pondberry's reproductive biology, growth and ecology (Devall et al. 2001; Smith et al. 2004; Aleric and Kirkman 2005; Connor et al. 2007; Hawkins et al. 2009a, b), there has been only one study on the distribution of its genetic variation (Godt and Hamrick 1996). In that study, 27 isozyme loci were used to genotype 720 stems from 15 sample sites across pondberry's range. Godt and Hamrick (1996) reported limited gene flow, a dominant role of genetic drift in determining pondberry genetic structure, and gene diversity that is much lower than found for the average woody angiosperm. They identified only 66 unique multilocus genotypes (genets) across all populations and acknowledged that the low allelic diversity of the isozyme markers likely contributed to the low observed genotypic diversity observed within populations (Godt and Hamrick 1996).

The objectives of the current study were to use highly polymorphic neutral markers to define genetically distinct populations, quantify the genetic differences among them, and provide information on pondberry's phylogeographic structure. Our intent was to identify populations that are most diverged and most closely related, those that may be most genetically at-risk, and those that could serve as potential germplasm sources for conservation and reestablishment programs.

Materials and methods

Study sites and sampling

Sample site information is listed in Table 1 and mapped in Fig. 1. In the years 2004, 2005 and 2006 we sampled mature plants from across the full extent of pondberry's range, including the 7 states and 12 of the 44 counties where it had been reported (U.S. Fish and Wildlife Service 2010). At sites where pondberry habitat covered 1 km^2 or less (Table 1), we sampled from all observed colonies. We regarded a colony as a cluster of stems separated from other clusters or patches of pondberry growth. At the larger sites we sampled from areas of highest pondberry densities. Shotz (2005) reported the discovery of pondberry in Alabama at two stands separated by 0.75 km and our AlCovin collection included both of these stands. At all sites we sampled from disjunct single stems and from multiple stems from within obvious colony structures, except in Arkansas where we sampled only a single stem from each colony. In areas where colonies were not spatially distinct, we randomly sampled stems at a spacing of at least 10 m. From each sampled stem, we collected a leaf into a labeled paper envelope and covered the envelopes with dehydrated silica gel granules for field storage and transport to the laboratory. In the laboratory, the envelopes were sealed in airtight plastic bags and stored in freezers until used for DNA isolation. DNA isolation was from 10 to 20 mg of desiccated leaf per well using the DNeasy 96 Plant Kit according to the manufacturer's instructions (Qiagen, USA). We measured purified DNA concentrations by Hoechst dye fluorometry.

Genotyping

We genotyped samples with eleven dinucleotide motif SSR (microsatellite) markers: LmSI001, LmSI002, LmSI004, LmSI013, LmSI027, LmSI035, LmSI047, LmSI049, LmSI049b, LmSI050, and LmSI061 (Echt et al. 2006). Marker LmSI049b was not previously reported, however, we characterized it as an unlinked marker that PCR co-amplified with the LmSI049 PCR primer pair. Marker amplification by PCR, SSR allele separation by capillary electrophoresis and allele size determination were as described by Echt et al. (2006) using 100 nM of each PCR primer, not the 100 μM as reported. All allele assignments were independently verified by having a second person examine the allele chromatogram profiles. We used the program MICRO-CHECKER to check for genotyping errors that are common to microsatellite marker analysis (van Oosterhout et al. 2004).

Individual genotypes with an indeterminate allele composition were recorded as missing data for that locus-sample pair. If a sample had missing data for two or more loci, then we removed it from subsequent analyses. From a set of samples having identical multilocus genotypes, we retained only one for population genetic analysis. If a sample's multilocus genotype differed only by a 2-bp allele at one locus, then it was considered to have been from a

Table 1 Location and sampling information for 14 pondberry sites

Site code	Area name	No. stems sampled	No. Genets identified	Site area, km^2	Site alias[a]	County, State	Lat.[b]	Long.[b]
AlCovin	Covington	161	7	0.25	–	Covington, Alabama	31.15	−86.29
ArSFSLm	SFSLWMA middle	56	55	1.0	–	Poinsett, Arkansas	35.69	−90.43
ArSFSLn	SFSLWMA north	27	27	1.0	–	Craighead, Arkansas	35.70	−90.42
ArSFSLs	SFSLWMA south	28	28	1.0	–	Poinsett, Arkansas	35.68	−90.44
ArSLSP	Sand Pond Conservation Area	24	20	0.80	AR12; H	Clay, Arkansas	36.49	−90.61
GaTayl	Butler	60	26	0.10	–	Taylor, Georgia	32.57	−84.26
MsBoliv	Shelby	30	13	0.50	–	Bolivar, Mississippi	33.70	−90.93
MsDNFc	DNF Colby	495	114	4.8	COLB; E	Sharkey, Mississippi	32.87	−90.72
MsDNFm	DNF mid-Forest region	134	23	130	–	Sharkey, Mississippi	32.74	−90.75
MsDNFn	DNF north	224	38	5.0	RED; E	Sharkey, Mississippi	32.91	−90.70
NcCumb	Big Pond Bay	105	33	0.25	–	Cumberland, North Carolina	34.91	−78.58
NcSamp	Pondberry Bay Preserve	134	33	0.25	NC1; A	Sampson, North Carolina	34.98	−78.47
ScFMNF	Francis Marion National Forest	291	17	36	FM; B	Berkeley, South Carolina	33.04	−79.77
ScMCAS	Marine Corps Air Station	107	16	0.12	–	Beaufort, South Carolina	32.47	−80.72

SFSLWMA St. Francis Sunken Lands Wildlife Management Area, *DNF* Delta National Forest, *DNF north* Delta National Forest Red Gum Research Natural Area and area northward

[a] Locale code used by Godt and Hamrick (1996)

[b] Latitude and longitude in decimal degrees, WGS84 datum; coordinates approximate midpoint of collection area

multiply sampled genet and was excluded from subsequent analyses of population genetic parameters. The rationale for excluding these samples was that a single step-wise allele difference would most likely have resulted from either genotyping error or recent SSR somatic mutation. We use here Scrosati's (2001) definition of genet, and extend it to include an individual not represented by multiple ramets, that is, any sexually produced multilocus genotype. We estimated the probability of identity $P_{(ID)}$, that is, the probability that two individuals drawn at random from a subpopulation will have the same multilocus genotype, using the method of Paetkau et al. (1998) as implemented by the software program GENALEX 6.1 (Peakall and Smouse 2006).

Estimates of null allele frequencies and null-adjusted visible SSR allele frequencies were obtained by a maximum likelihood method using the EM algorithm (Dempster et al. 1977), as implemented in Option 8.1 of the software program GENEPOP 4.0.7 (Rousset 2007). We modified the input data file for use with Option 8.1 to include a dummy population of a single sample with all locus genotypes coded as 9999, which is the explicit Genepop code for null homozygotes. This modification suppressed the program's default interpretation of the largest allele integer in a subpopulation as a null allele, and ensured correct allele counts and frequency estimates for each subpopulation.

Population genetic analyses

We used Bayesian model-based clustering to infer assignment of individuals to genetic clusters, correlate clusters to sample locations, and identify admixed populations, as implemented by STRUCTURE ver. 2.3 (Pritchard et al. 2000; Hubisz et al. 2009). For all STRUCTURE simulation runs we used the admixture model and the independent allele frequencies model, either with or without the location prior model, and set all other run parameters to their default values.

We obtained neighbor-joining (NJ) trees for "net nucleotide distances" among inferred genetic clusters (i.e., ancestral populations), as provided by the default STRUCTURE output. We also used STRUCTURE to draw NJ trees for independently estimated pairwise population F_{st} and D_{est} values. These NJ trees were obtained by editing the distance matrix in the allele-frequency divergence section of a STRUCTURE results file template. The ΔK rates of change of Ln P(D) (estimated log probability of data) for K inferred clusters were calculated as described by Evanno et al. (2005) and implemented by CORRSIEVE (Campana et al. 2010). For display of STRUCTURE Q plots, we used CLUMPP to standardize cluster labels among STRUCTURE runs (Jakobsson and Rosenberg 2007). DISTRUCT was used to generate color-coded bar graphs from CLUMPP output and grey-scale-coded bar graphs from STRUCTURE output (Rosenberg 2004).

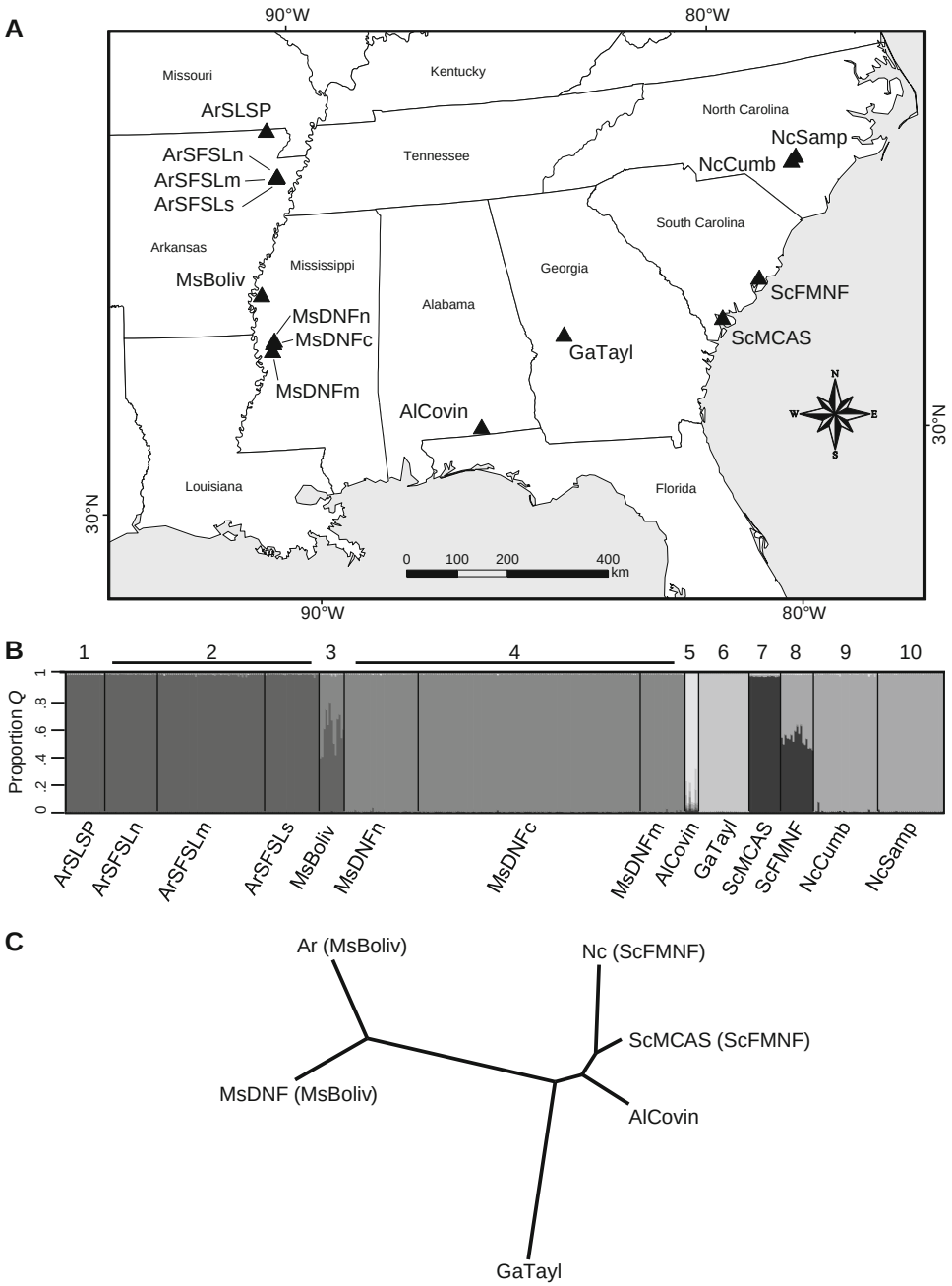

Fig. 1 Pondberry sample site locations and Bayesian model-based assignment of individuals to inferred genetic clusters. **a** Map of the southeastern United States showing locations of 14 pondberry sample sites (*triangles*). **b** STRUCTURE *Q* plot bar graph of 450 individuals for $K = 6$ with the location prior model; each shade of grey corresponds to one of the six inferred clusters where each vertical grey-scale bar represents an individual; vertical black bars separate sets of individuals between the sites named below the graph; location numbers used for the location prior and their association with sites are above the graph; the scale on the left is the proportion of estimated cluster composition, *Q*, within each individual. **c** NJ tree derived from the net nucleotide distances among the six inferred clusters; branches are labeled with the site or geographic area associated with each cluster; parentheses denote sites admixed with two clusters

We calculated multilocus means for H_e expected heterozygosity (Nei 1973), A_e effective allele number, where $A_e = 1/(1 - H_e)$ (Kimura and Crow 1964; Jost 2008, Eq. 6), and F fixation index, where $F = (H_e - H_o)/H_e$ (Nei 1977), as implemented by GENALEX ver. 6.41 (Peakall and Smouse 2006). The following analyses were also performed in GENALEX: principal coordinate analysis (PCoA) of population genetic distance matrices using the non-standardized covariance approach (Orlóci 1978), estimation of *F*st by AMOVA (Peakall et al. 1995), and Mantel tests for correlation of matrices for genetic and geographic distances (Mantel 1967; Smouse et al. 1986; Smouse and Long 1992). All AMOVA and Mantel tests were run with 999 permutations. With the aid of spreadsheet functions we calculated the multilocus mean F in each population from null-adjusted estimated allele frequencies.

In addition to F_{st}, we measured pairwise population differentiation as estimates of Jost's D, where $D = [(H_T - H_S)/(1 - H_S)]\,[n - (n - 1)]$, H_T and H_S are Nei's (1973) average total heterozygosity and subpopulation heterozygosity, respectively, for individual loci and n is the number

of subpopulations (Jost 2008, Eq. 11). When subpopulations share all alleles in equal proportions, $D = 0$ and when they share no alleles, $D = 1$, regardless of the magnitude of heterozygosity. The estimator of D, D_{est} (Jost 2008, Eq. 12), was obtained by replacing H_T and H_S with their respective unbiased estimators (Nei and Chesser 1983), as implemented in the software program Smogd (Crawford 2010), which also provided bootstrap 95% confidence intervals (1000 replications) for D_{est}. We also obtained D_{est} using the software program Spade (Chao and Shen 2009) for verification purposes. We used arithmetic means across loci for multilocus expectations of D_{est}.

Results

Genotype summary

From 11 SSR loci, we amplified 174 alleles (mean = 15.8, SD = 8.20, range = 4–36). Loci LmSI027 and LmSI049b were monoallelic in all three populations in the St. Francis Sunken Lands Wildlife Management Area (ArSFSL) and locus LmSI004 was monoallelic in the MsDNFm samples. Among the 14 sites, the median probability of identity, $P_{(ID)}$ (Waits et al. 2001), for an 11-locus genotype was 7.2×10^{-9}, with a range from 6.0×10^{-5} in ArSLSP to 2.7×10^{-11} in ScFMNF. Based on the low $P_{(ID)}$ values, we concluded that our 11 SSR markers would together identify every genet in each sample set, thus providing a level of individual genetic identification greater than from isozyme markers (Godt and Hamrick 1996). Guided by our genotype quality criteria, we identified 450 genets, from which 11 loci provided data for 4,942 individual locus genotypes out of a possible 4,950. These genotype data are available from Online Resources, (Supplemental Table S1). We regarded the eight missing genotypes as technical failures during PCR and did not code them specifically as null allele homozygotes in subsequent analyses.

Among all sites within ArSFSL and within the Delta National Forest (MsDNF), we identified 110 and 175 genets, respectively. In contrast, within the Francis Marion National Forest (ScFMNF), which also was an area of continuous forest habitat, we identified only 17 genets among the four largest known stands. At the disjunct Alabama Covington (AlCovin) site, we found only seven genotypes among 161 individual stems sampled from across extent of the habitat. Shotz (2005) reported that one of two stands at AlCovin site contained several hundred stems. When we sampled 41 stems from across the breadth of that stand, we detected only one genotype, which indicated that the entire shallow depression was monoclonal. We observed at other sites colonies that were monogenic, or composed of a dominant genet with low frequencies of other genotypes, and will discuss those results in a separate publication. At each of the remaining disjunct sites, which were all 0.5 km^2 or less in habitable area, we identified 13 to 33 genets (Table 1). It is likely that from these sites (AlCovin, GaTayl., MsBoliv, NcCumb, NcSamp, and ScMCAS) we collected a majority fraction of genets present because the habitable areas were relatively small. Assuming this is true, then the small sample sizes of these sites approached actual population sizes.

We estimated a null allele to be present in at least one locus in each of the 10 populations identified in subsequent analyses described in the following section (Appendix Table 3). Estimated null allele frequencies ranged from 0.00 to 0.26 per locus per population, from 0.00 to 0.08 per locus, and from 0.00 to 0.04 per population. The mean null allele frequency was 0.02 over all loci and populations.

Population structure

To initially see whether individuals were genetically correlated with sample sites, we ran STRUCTURE simulations for 1 through 12 inferred genetic clusters ($K = 1 \ldots 12$). These initial simulations were run without the location prior model and with a burn-in of 2000 replications followed by 8000 replications for each of 20 runs per K. The highest mean ΔK (Evanno et al. 2005) among the runs was for $K = 2$, which unambiguously assigned all individuals from the western region (Arkansas and Mississippi) into one cluster and all individuals from the eastern region (Alabama, Georgia, South Carolina, and North Carolina) into the other. As K increased, individual assignments remained highly correlated with increasingly specific associations to sample sites. After $K = 2$, the next highest ΔK, and the most likely mean estimated log probability of data, Ln P(D), was seen at $K = 6$. However, the trend of association between cluster and specific sites continued up to $K = 8$, beyond which no additional population structure was consistently delineated and additional clusters resulted in merely increasing admixture of minor cluster components within individuals and sites. Among the $K = 8$ runs, as well as at lower K values, we found sets of multimodal solutions that provided limited combinations of specific sites split among different clusters. Estimated Q plots in Supplemental Fig. S1 show representative modes. What these modalities demonstrated was that even though there was clear association between genetic cluster and sample location, there was no "correct" estimation for a single K that clearly defined current pondberry population structure. Multiple runs across the range of K values also provided evidence that the three sample sites within the ArSFSL comprised a genetically distinct set of individuals, as did the three sites within MsDNF, thus suggesting

10 genetically distinct populations (ArSLSP, ArSFSL, MsBoliv, MsDNF, AlCovin, GaTayl, ScMCAS, ScFMNF, NcCumb, and NcSamp). All runs above $K = 3$ revealed that the MsDNF location was an admixture of two genetic clusters that were geographically dispersed among individuals within MsDNF. Local genetic structuring of pondberry populations will be addressed elsewhere.

Having established that clustering outcomes were correlated with sampling locations, we ran 24 STRUCTURE simulation runs for $K = 6$ using the location prior model (Hubisz et al. 2009) with the 10 population locations and with an empirically determined burn-in of 100,000 replications followed by 400,000 replications. We performed these runs specifically to explore broad genetic relationships among inferred clusters. As with the initial runs, the location prior $K = 6$ runs provided a limited set of modal solutions. The estimated Q plot in Fig. 1b shows the most common mode. A NJ tree of cluster genetic distances indicated a correspondence between genetic and geographic distances that suggested phylogeographical structure (Fig. 1c). NJ trees derived from the other modal solutions and from other K values did not contradict this general structure.

As noted by Pritchard et al. (2000), K clusters may not necessarily correspond to "real" populations. We performed additional analyses to further explore genetic relationships among individual sites and regions and discern possible phylogeographic patterns among them. Pairwise estimated F_{st} and D_{est} were determined for our fourteen sample sites and principal coordinate analyses (PCoA) were conducted with the resulting distance matrices (Appendix Table 4). As seen in Fig. 2a, the first principal coordinate axis widely separated the eight sites in the western region from the six sites in the eastern region, supporting the STRUCTURE runs for $K = 2$. To measure divergence between the western and eastern regions, we pooled samples within each and estimated pairwise F_{st} as 0.23 (95% CI = 0.17–0.30, AMOVA $P = 0.001$) and D_{est} as 0.67 (95% CI = 0.63–0.71).

When plotted against the second principal coordinate axis, the 14 sites clustered into five regional population groups: Ar (including ArSLSP, ArSFSLn, ArSFSLm and ArSFSLs), MsBoliv, MsDNF, SE (including AlCovin, GaTayl and ScMCAS) and NE (including ScFMNF, NcCumb, and NcSamp) (Fig. 2a). Divergence among population groups, especially the SE and NE groups, was

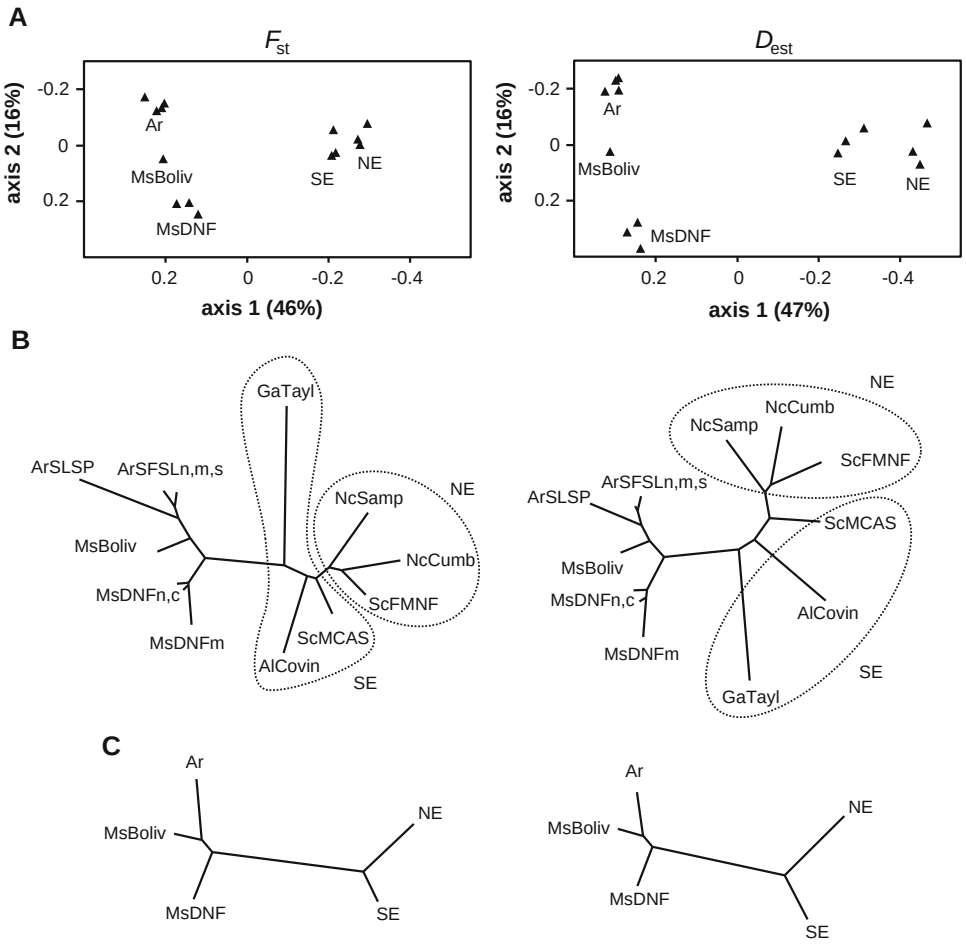

Fig. 2 Analyses of F_{st} (*left column*) and Jost's D_{est} (*right column*) matrices for 14 pondberry sites. **a** Scatter plots of site coordinates (*triangles*) for the first two eigenvector axes from principal coordinate analysis (PCoA); group labels, defined in the main text, are next to each cluster of sites; axis percentage value is the proportion of total PCoA variation represented by the axis, **b** NJ trees for the same site data, dotted lines demark the SE and NE population groups identified by PCoA, **c** NJ trees for pooled samples from the five population groups identified by PCoA

more evident with D_{est} than with F_{st}. In NJ trees drawn from the F_{st} and D_{est} matrices, the main western and eastern divergence was obvious, but the secondary associations among sites within regions were obscured, though still discernable (Fig. 2b). A prominent feature of the NJ trees not seen in the two-dimensional PCoA plots was the extended branch length for the GaTayl site. That site, however, was discriminated from all other sites along the third principal coordinate axis, which was evident in three-dimensional PCoA plots (available upon request). The majority of information was present in the first two axes that contained ~62% of total variation among six axes.

When we pooled samples and generated F_{st} and D_{est} matrices for the five population groups (Table 2), then the resulting NJ trees for each measure did provide obvious concordance between genetic and geographic population structures (Fig. 2c). We reversed the axes in Fig. 2a from the original PCoA output to aid in visualizing the concordance with map locations (Fig. 1a).

Table 2 Pairwise measures of genetic differences for sample pools among five pondberry population groups identified by principal coordinate analysis, with SSR locus averages of Jost's D_{est} above the diagonal, estimated F_{st} (AMOVA) below the diagonal

	AR	MsBoliv	MsDNF	SE	NE
AR	–	0.197	0.317	0.637	0.777
MsBoliv	0.095	–	0.205	0.676	0.759
MsDNF	0.147	0.089	–	0.632	0.766
SE	0.264	0.259	0.251	–	0.389
NE	0.309	0.291	0.288	0.109	–

Based on results from STRUCTURE, F_{st} and D_{est} analyses, we identified 10 populations among the 14 sample sites and pooled samples from within the ArSFSL and MsDNF locations for subsequent analyses. This pooling was done notwithstanding subpopulation differences within these continuously forested locations that were evident from F_{st} and D_{est} results (Appendix Table 4). Guided by the observed west-east division, we conducted separate Mantel tests to assess isolation by distance among the four populations in the western regions (ArSLSP, ArSFSL, MsBoliv, MsDNF) and the six populations in the eastern region (AlCovin, GaTayl, ScMCAS, ScFMNF, NcCumb, and NcSamp). We found a positive correlation between differentiation and geographic distance in the western region ($R = 0.86$, $P = 0.03$) and somewhat weaker evidence for a positive correlation in the eastern region ($R = 0.51$, $P = 0.05$) (Fig. 3).

Genetic diversity and heterozygosity

We used the effective number of alleles, A_e, as the main measure of genetic diversity because it provides an accurate comparison of relative diversity among populations (Jost 2008). For eleven SSR loci, the overall A_e for the

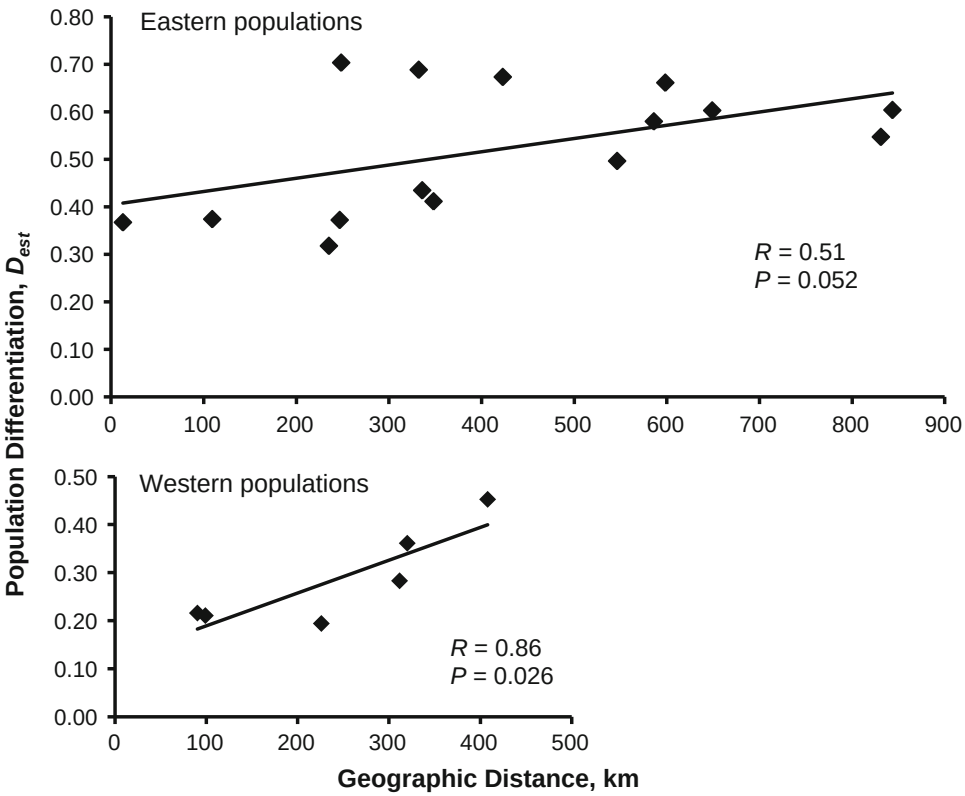

Fig. 3 Evidence for isolation-by-distance among pondberry populations within geographic regions. Scatter plots of coordinates (*diamonds*) and least squares regression (*line*) for pairwise D_{est} and geographic distances among six eastern populations (*upper plot*) and four western populations (*lower plot*); Mantel test results are inset

species was 5.78. A_e in the eastern region was 3.15 and in the western region, 2.91. The respective expected heterozygosity values, H_e, were 0.76, 0.61, and 0.54. We did not find sufficient evidence to conclude that there was greater diversity in the eastern region, at least among these eleven SSR loci (mean A_e difference = 0.24, 95% CI = −0.86 to 1.35, $t = 0.49$, $df = 20$, one-tailed $P = 0.31$). The population with the lowest diversity in the eastern region was GaTayl ($A_e = 2.3$, $H_e = 0.47$), while ArSLSP ($A_e = 1.9$, $H_e = 0.36$) had the lowest in the western region (Fig. 4). Highest diversities in the eastern and western regions were in the Francis Marion National Forest (ScFMNF $A_e = 4.1$, $H_e = 0.69$) and in the Delta National Forest (MsDNF $A_e = 3.8$, $H_e = 0.62$). We did not normalize allele diversity measures among populations based on varying sample sizes (i.e., rarefaction) or apply unbiased estimators because the underlying statistical assumptions of sampling from large populations would not have been valid in those cases, mentioned previously, where small sample sizes of genets likely approximated actual population sizes.

Calculating diversity values that included null allele frequency estimates did not provide any significant differences from values that excluded them. Notable differences were observed, however, for some populations when we applied null frequency estimates to the fixation index, F (Fig. 4). Because F-statistics are heavily influenced by observed proportions of heterozygous genotypes, we thought it prudent to rely on null-corrected F estimates and mitigate possible bias that could be introduced by misclassifying cryptic null heterozygotes as homozygotes. Accordingly, we observed negative F, that is, a preponderance of heterozygous genotypes in excess of Hardy–Weinberg proportions, for all populations except ArSFSL ($F = 0.01$, 95% CI = −0.01 to 0.03) (Fig. 4). While the null-adjusted F for ArSFSL did not statistically differ from zero, the value without null-adjustment did and was significantly positive ($F = 0.08$, 95% CI = 0.01 to 0.08, HWE U-test $P = 0.0002$). The other notable consequence of excluding null allele adjustment was the increased width of confidence intervals for F, particularly in ArSLSP, MsBoliv, AlCovin, and NcSamp (Fig. 4). This increased uncertainly resulted from the uneven distribution of locus-specific null allele frequencies among populations (Appendix Table 3).

All eastern populations had moderate evidence for heterozygote excess (90% CI < 0), and four of the six, GaTayl, ScMCAS, NcCumb, and NcSamp, had stronger evidence (95% CI < 0). Although three western populations had negative F, all four western populations lacked statistical evidence to conclude a departure from random mating (95% CI included zero). Consequently, for no pondberry population did we find evidence to suggest inbreeding.

Discussion

Patterns of divergence and phylogeographic structure

We found the most differentiation among pondberry populations between the eastern and western regions, where the northern-most population groups exhibited more differentiation between regions than the southern-most population groups (Table 2). This pattern indicates that southern pondberry populations had a more recent common ancestor than the northern populations. It further suggests that divergent northward migration into the Mississippi Alluvial Plain to the west and along the Southern Atlantic Coastal Plain to the east preceded the west-east split that eventually occurred among the ancestral southern populations. This general phylogeographical structure of a west-east divide is consistent with the Appalachian Mountain discontinuity hypothesized for other plant species in eastern North America (Soltis et al. 2006). While it is possible

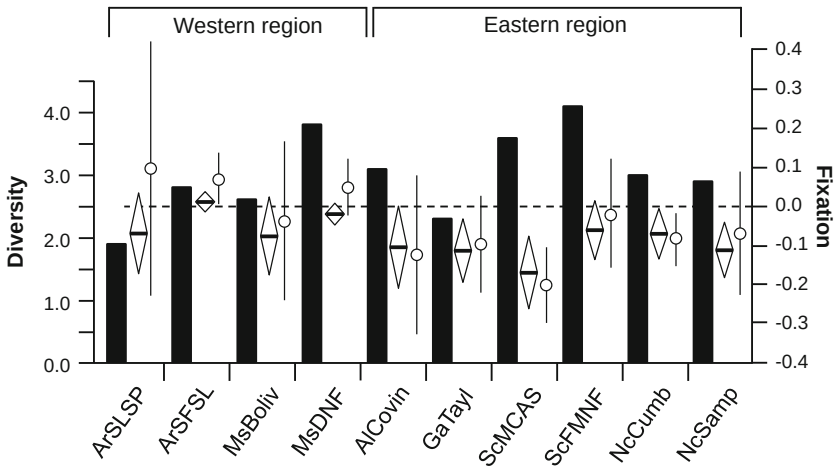

Fig. 4 Comparisons among ten pondberry populations, grouped geographic region, of genetic diversity A_e (*solid vertical bars*), and fixation index F including null alleles (*horizontal bars*), and F excluding null alleles (*circles*); *open diamonds* and *vertical lines* are 95% CI of the respective F multilocus means; *dashed line*, $F = 0$

that pondberry's phylogeographical structure originated from Pleistocene barriers to gene flow, our data do not provide any clue about the times that migrations and divergences occurred. Alternatively, and as cautioned by Soltis et al. (2006), pseudocongruence, that is, differences in marker mutation rates or varying divergence times among ancestral populations, may have been involved. Given pondberry's restriction to microhabitats of seasonally flooded wet depressions, we may gain additional insights to its phylogeographical patterns by ecological niche modeling, as was recently done for American beech (Morris et al. 2010).

There has been much divergence among populations within regions since the west-east split of the species range and divergence has been greatest in the eastern region. Although the average geographic distance between sampled populations was greater in the eastern region (423 km) than in the western region (229 km), differentiation remained consistently higher at comparable pairwise distances in the eastern region (Fig. 3). Therefore, geographic distance alone cannot account for the variation in differentiation seen in each region. This variation likely resulted from greater isolation and genetic drift throughout the eastern region, as evident in the pronounced differentiation that we observed even between neighboring sample sites, e.g., NcCumb and NcSamp (Appendix Table 4). We will consider ecological factors that affect regional differentiation and gene flow in future studies.

Heterozygote excess

From the H_e and H_o values of fifteen pondberry populations reported by Godt and Hamrick (1996), we can see that F is negative for eight of them (AR11, AR12, COLB, FM1, FM3, FM4, RED, SUN), indicating a heterozygote excess among the isozyme loci they studies and supporting our findings with microsatellite loci. A heterozygote excess among microsatellite loci also has been reported in other partially clonal plant species (Stoeckel et al. 2006; Meerow et al. 2007; Rasmussen and Kollmann 2008). Heterozygote excess is expected to occur in clonal populations (Birky 1996; Balloux et al. 2003), in populations that have recently experienced severe reduction in size, i.e., a bottleneck (Cornuet and Luikart 1996), and in small dioecious populations (Robertson 1965; Rasmussen 1979; Falconer 1989; Pudovkin et al. 1996; Balloux 2004). A biased sex ratio of parents can further increase heterozygote excess (Balloux 2004). Given that pondberry is dioecious, frequently found in small isolated populations, largely clonal, and has populations that are often dominated by single-sex colonies, it would be difficult to separate the specific factors responsible for the varying degrees of heterozygote excess that we observed. However, the ArSFSL and MsDNF populations had the largest number of genets and had the highest (null-adjusted) F values (Fig. 4). This suggests that small population size, which is indicated in most populations, may have a major role in determining heterozygote excess in pondberry.

Implications for management

Although the isolated Alabama population, AlCovin, had relatively favorable measures of genetic diversity and no detectable inbreeding (Fig. 4), with only seven genets found it appears to be the one most at risk for unrecoverable loss of genotypic diversity through drift. This risk becomes more pronounced given that there are no known natural long-range dispersal mechanisms for pondberry that could contribute to gene flow from other populations. Two other populations in the eastern region, ScMCAS and ScFMNF, are also notable for the few number of genets that we found (16 and 17, respectively) among the large number of stems sampled (107 and 291, respectively). The population most at risk for continued loss of diversity in the western region appears to be ArSLSP. It is disjunct, at the northern extent of pondberry's range, had the lowest allelic diversity, and had two of eleven SSR loci at fixation. Establishing sufficient numbers of novel, sexually reproductive, genets at these sites would increase diversity in the breeding pool and potentially avoid long-term population decline. Our diversity measures, however, are for neutral genetic markers and as such can only serve as proxies for adaptive genetic diversity. Adaptive diversity can only be directly measured as quantitative variation in the as-yet-unknown phenotypic traits that determine long-term reproductive success, plant health, and colony establishment.

Based on the wide genetic differentiation that we observed between the eastern and western regions, a conservative approach to stocking pondberry nurseries for use in recovery programs would be to restrict germplasm transfers to within a region. Doing so would preserve the natural pattern of diversity established by pondberry's phylogenetic history that we have proposed. Any further restrictions of germplasm transfers within a region would depend on adaptive considerations for specific ecotypes, for which the current genetic analyses provide no information. Although maintenance of heterozygosity is a general goal of conservation genetics, in the case of pondberry excess heterozygosity may be a symptom of very small effective population sizes. The low numbers of genets that we sampled from most populations is a cause for concern because it limits the range of adaptive variation that can be provided through sexual generation of new genotypes. While clonal growth and high heterozygosity may provide a degree of tolerance to extinction events even

in small populations, the ultimate consequence of predominantly clonal reproduction in forest plants appears to be suppression of sexual reproduction leading to sexual extinction (Honnay and Bossuyt 2005). If pondberry nurseries were established as seed nurseries to engender crossing among a wide variety of regional genotypes, then they could provide genotypically diverse seed sources needed for successful pondberry restoration programs. Propagating vegetative cuttings from existing genets for restocking populations in restoration programs would not increase overall genet diversity, which may be a concern in areas where genet diversity is already limited.

Acknowledgments The Army Corps of Engineers—Vicksburg provided partial funding of this research. The U.S. Fish and Wildlife Service provided collection permits. We thank Kevin Potter for generating the ArcGIS map graphics and thank him and James Roberds for providing helpful comments on an early draft of the manuscript. We also appreciate the help with sample collections provided by Kris Connor, Margaret Devall, Shea Hammond, Tracy Hawkins, Thomas Kubisiak, Nathan Schiff, Stephanie Skojak, and Ralph Pearce.

Appendix

See Tables 3 and 4.

Table 3 Estimated null allele frequencies of 11 SSR loci in 10 pondberry populations

Locus	ArSLSP	ArSFSL	MsBoliv	MsDNF	AlCovin	GaTayl	ScMCAS	ScFMNF	NcCumb	NcSamp	Mean
LmSI001	0	0.09	0	0.07	0	0	0	0	0	0	0.02
LmSI002	0	0	0	0.01	0	0	0	0	0	0	0
LmSI004	0.15	0	0	0	0.12	0	0	0	0	0	0.03
LmSI013	0	0.03	0	0	0.05	0	0	0	0	0	0.01
LmSI027	0.01	0	0.02	0	0	0	0.02	0	0	0	0.01
LmSI035	0	0.02	0	0.02	0.14	0.07	0	0	0	0	0.03
LmSI047	0.16	0.1	0.26	0.06	0	0.09	0	0.07	0.01	0.02	0.08
LmSI049	0	0.05	0	0.01	0	0	0.04	0.12	0	0.15	0.04
LmSI049b	0.01	0	0.03	0	0	0	0.13	0	0.05	0	0.02
LmSI050	0.07	0.02	0.02	0.13	0	0	0	0	0	0	0.02
LmSI061	0	0.03	0.09	0	0	0	0	0	0	0	0.01
Mean	0.04	0.03	0.04	0.03	0.03	0.01	0.02	0.02	0.01	0.02	0.02

Table 4 Pairwise measures of genetic differences among 14 pondberry sites, with SSR locus averages of Jost's D_{est} above the diagonal, F_{st} (AMOVA) below the diagonal

Site	ArSLSP	ArSFSLn	ArSFSLm	ArSFSLs	MsBoliv	MsDNFn	MsDNFc	MsDNFm	AlCovin	GaTayl	ScMCAS	ScFMNF	NcCumb	NcSamp
ArSLSP		0.24	0.23	0.24	0.30	0.46	0.43	0.55	0.78	0.79	0.76	0.90	0.84	0.90
ArSFSLn	0.17		0.05	0.09	0.22	0.34	0.37	0.45	0.73	0.79	0.61	0.82	0.77	0.79
ArSFSLm	0.17	0.03**		0.03	0.21	0.29	0.33	0.41	0.65	0.74	0.59	0.79	0.76	0.79
ArSFSLs	0.18	0.05	0.01*		0.21	0.33	0.35	0.41	0.72	0.78	0.67	0.83	0.83	0.79
MsBoliv	0.22	0.11	0.11	0.11		0.26	0.19	0.33	0.78	0.77	0.70	0.81	0.81	0.75
MsDNFn	0.25	0.17	0.15	0.16	0.11		0.08	0.17	0.68	0.70	0.65	0.74	0.79	0.74
MsDNFc	0.22	0.17	0.16	0.16	0.09	0.03		0.13	0.72	0.75	0.66	0.74	0.84	0.77
MsDNFm	0.31	0.21	0.20	0.20	0.15	0.07	0.06		0.72	0.79	0.71	0.77	0.87	0.82
AlCovin	0.47	0.32	0.31	0.34	0.35	0.28	0.31	0.30		0.70	0.50	0.60	0.55	0.60
GaTayl	0.50	0.41	0.39	0.41	0.42	0.36	0.37	0.40	0.36		0.69	0.67	0.58	0.66
ScMCAS	0.39	0.26	0.27	0.28	0.30	0.27	0.29	0.28	0.17	0.33		0.37	0.43	0.41
ScFMNF	0.44	0.33	0.33	0.34	0.32	0.28	0.31	0.29	0.19	0.31	0.13		0.32	0.37
NcCumb	0.44	0.35	0.35	0.37	0.37	0.34	0.36	0.36	0.22	0.29	0.17	0.12		0.37
NcSamp	0.45	0.35	0.35	0.36	0.35	0.31	0.34	0.33	0.24	0.33	0.16	0.14	0.18	

* AMOVA $P = 0.010$

** AMOVA $P = 0.002$

For all other F_{st}, AMOVA $P = 0.001$

References

Aleric KM, Kirkman K (2005) Growth and photosynthetic responses of the federally endangered shrub, *Lindera melissifolia* (Lauraceae), to varied light environments. Am J Bot 92:682–689

Balloux F (2004) Heterozygote excess in small populations and the heterozygote-excess effective population size. Evolution 58: 1900–1981

Balloux F, Lehman L, de Meeûs T (2003) The population genetics of clonal and partially clonal diploids. Genetics 164:1635–1644

Birky CW Jr (1996) Heterozygosity, heteromorphy, and phylogenetic trees in asexual eukaryotes. Genetics 144:427–437

Campana MG, Hunt HV, Jones H, White J (2010) CorrSieve: software for summarizing and evaluatin Structure output. Mol Ecol Resour 11:349–352

Chao A, Shen T-J (2009) SPADE (Species Prediction and Diversity Estimation), URL http://chao.stat.nuhu.edi.tw/softwareCI.html. Accessed 2 July 2010

Connor K, Schaefer G, Donahoo J, Devall M, Gardiner E, Hawkins T, Wilson DA, Schiff N, Hamel P, Leininger T (2007) Development, fatty acid composition, and storage of drupes and seeds from the endangered pondberry (*Lindera melissifolia*). Biol Conserv 137:489–496

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

Crawford NG (2010) SMOGD: software for the measurement of genetic diversity. Mol Ecol Resour 10:556–557

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

Devall M, Schiff N, Boyette S (2001) Ecology and reproductive biology of the endangered pondberry, *Lindera melissifolia* (Walt) Blume. Natural Areas Journal 21:250–258. URL http://www.treesearch.fs.fed.us/pubs/2750. Accessed 27 Aug 2010

Echt CS, Demeer D, Kubisiak T, Nelson CD (2006) Microsatellites for *Lindera* species. Mol Ecol Notes 6:1171–1173

Environmental Protection Agency (2010) Ecoregion maps http://www.epa.gov/wed/pages/ecoregions.htm. Accessed 6 Sept 2010

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

Falconer DS (1989) Introduction to quantitative genetics, 3rd edn. Longman Scientific and Technical, Wiley, New York

Godt MW, Hamrick JL (1996) Allozyme diversity in the endangered shrub *Lindera melissifolia* (Lauraceae) and its widespread congener *Lindera benzoin*. Can J For Res 26:2080–2087

Hawkins TS, Schiff NM, Gardiner ES, Leininger TD, Devall MS, Wilson DA, Hamel PB, McCown DD, Connor K (2007) Micropropagatin of the endangered shrub pondberry (*Lindera melissifolia* [Walt.] Blume). HortSci 42:407–409

Hawkins TS, Skojac DA Jr, Lockhart BR, Leininger TD, Devall MS, Schiff NM (2009a) Bottomland forests in the lower Mississippi Alluvial Valley associated with the endangered *Lindera melissifolia*. Castanea 74:105–113

Hawkins TS, Schiff NM, Leininger TD, Gardiner ES, Devall MS, Hamel PB, Wilson DA, Connor KF (2009b) Growth and intraspecific competitive abilities of the dioecious *Lindera melissifolia* (Lauraceae) in varied flooding regimes. J Torrey Bot Soc 136:91–101

Honnay O, Bossuyt B (2005) Prolonged clonal growth: escape route or route to extinction? Oikos 108:427–432

Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. Mol Ecol Resour 9:1322–1332

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23:1801–1806

Jost L (2008) G_{ST} and its relatives do not measure differentiation. Mol Ecol 17:4015–4026. doi:10.1111/j.1365-294X.2008.03887.x

Kimura M, Crow J (1964) The number of alleles that can be maintained in a finite population. Genetics 49:725–738

Mantel N (1967) The detection of disease clustering and a generalized regression approach. Cancer Res 27:209–220

Meerow AW, Gideon M, Kuhn DN, Motamayor JC, Nakamura K (2007) Genetic structure and gene flow among south Florida populations of *Iris hexagona* Walt. (Iridaceae) assessed with 19 microsatellite DNA loci. Int J Plant Sci 168:1291–1309

Morris AM, Graham CH, Soltis DE, Soltis PS (2010) Reassessment of phylogeographical structure in an eastern North American tree using Monmonier's algorithm and ecological niche modelling. J Biogeogr 37:167–1667

NatureServe (2010) NatureServe Explorer: an online encyclopedia of life [web application]. http://www.natureserve.org/explorer/. Accessed 6 Sept 2010

Nei N (1973) Analysis of gene diversity in subdivided populations. Proc Nat Acad Sci USA 70:3321–3323

Nei M (1977) F-statistics and analysis of gene diversity in subdivided populations. Ann Hum Genet 41:225–233

Nei M, Chesser RK (1983) Estimation of fixation indices and gene diversities. Ann Hum Genet 47:253–259

Orlóci L (1978) Multivariate analysis in vegetation research, 2nd edn. Springer, The Hague

Paetkau D, Waits LP, Clarkson PL, Craighead L, Vyse E, Ward R, Strobeck C (1998) Variation in genetic diversity across the range of North American brown bears. Cons Biol 12:418–429

Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol Ecol Notes 6:288–295. URL www.anu.edu.au/BoZo/GenAlEx/genalex_download.php Accessed 2 July 2010

Peakall R, Smouse PE, Huff DR (1995) Evolutionary implications of allozyme and RAPD variation in diploid populations of dioecious buffalograss *Buchloë dactyloides*. Mol Ecol 4:135–147

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959

Pudovkin AI, Zaykin DV, Hedgecodk D (1996) On the potential for estimating the effective number of breeders from heterozygote-excess in progeny. Genetics 144:383–397

Rasmussen DI (1979) Sibling clusters and gene frequencies. Am Nat 113:948–951. URL http://www.jstor.org/stable/2460316. Accessed 3 Sept 2010

Rasmussen KK, Kollmann J (2008) Low genetic diversity in small peripheral populations of a rare European tree (*Sorbus torminalis*) dominated by clonal reproduction. Conserv Genet 9:1533–1539

Robertson A (1965) The interpretation of genotypic ratios in domestic animal populations. Anim Prod 7:319–324

Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. Mol Ecol Notes 4:137–138

Rousset F (2007) GENEPOP'007: a complete reimplementation of the GENEPOP software for Windows and Linux. Mol Ecol Resour 8:103–106

Scrosati R (2001) An updated definition of genet applicable to clonal seaweeds, bryophytes, and vascular plants. Basic Appl Ecol 3:97–99

Shotz A (2005) Alabama. Castanea 70:317. doi:10.2179/0008-7475

Smith CD III, Hamel PB, Devall MS, Schiff NM (2004) Hermit thrush is the first observed dispersal agent for pondberry (*Lindera melissifolia*). Castanea 69:1–8

Smouse PE, Long JC (1992) Matrix correlation analysis in anthropology and genetics. Yearbook Phys Anthropol 35:187–213

Smouse PE, Long JC, Sokal RR (1986) Multiple regression and correlation extensions of the Mantel test of matrix correspondence. Syst Zool 35:623–627

Soltis DE, Morris AB, McLachlan JS, Manos PS, Soltis PS (2006) Comparative phylogeography of unglaciated eastern North America. Mol Ecol 15:4261–4293

Stoeckel S, Grange J, Fernández-Manjarres JF, Bilger I, Frascaria-Lacoste N, Mariette S (2006) Heterozygote excess in a self-incompatible and partially clonal forest tree species—*Prunus avium* L. Mol Ecol 15:2109–2118

U.S. Fish and Wildlife Service (1986) Endangered and threatened wildlife and plants: determination of endangered status for *Lindera melissifolia*. Fed Regis 51(147):27495–27500. http://ecos.fws.gov/docs/federal_register/fr1169.pdf. Accessed 23 July 2010

U.S. Fish and Wildlife Service (1993) Recovery Plan for Pondberry (*Lindera rnelissifolia*). U.S. Fish and Wildlife Service. Atlanta, Georgia. 56 pp. http://ecos.fws.gov/docs/recovery_plan/930923a.pdf. Accessed 23 July 2010

U.S. Fish and Wildlife Service (2010) Endangered species program: an online database of federally endangered species in the United States [web application]. http://www.fws.gov/endangered/. Accessed 6 Sept 2010

van Oosterhout D, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors. Mol Ecol Notes 4:535–538

Waits LP, Luikart G, Taberlet P (2001) Estimating the probability of identity among genotypes in natural populations: cautions and guidelines. Mol Ecol 10:249–256

Wright RD (1990) Species biology of *Lindera melissifolia* (Walt.) Blume in northeast Arkansas. In: Mitchell RS, Scheviac CJ, Leopold DJ eds. Ecosystem Management: rare species and significant habitats. Proceedings of the 15th Annual Natural Areas Conference, New York State Mus Bull 471:176–179. http://nysl.nysed.gov/Archimages/76089.PDF. Accessed 2 Aug 2010

Wright RD (1994) Sex ratio and success, an assessment of *Lindera melissifolia* in Arkansas. Proc Arkansas Acad Sci 48:230–233