DNA MARKER APPROACHES TO SPECIES CONSERVATION AND RESTORATION

C. Echt[1], S. Josserand, V. Hipkins, and B. Crane

[1] USDA Forest Service, Southern Institute of Forest Genetic, Saucier, MS

Maintaining or increasing gene diversity must be an integral part of species conservation and restoration. Because we seldom, if ever, know which particular genes are essential for successful conservation and restoration efforts, we must monitor gene diversity by proxy, that is, by using a suitable sampling of random genetic markers from the genome. A properly selected set of DNA markers can provide accurate data about population inbreeding, gene flow, differentiation, and substructure. When applied to managed populations, DNA markers also can help assess the genetic diversity and level of inbreeding of seed orchards and restored stands. Currently, the DNA markers of choice are microsatellites (SSRs) because of their potential for efficient detection of multiple alleles. By way of example, we analyzed the distribution of longleaf pine SSR allele diversity in natural and managed germplasm sources at different spatial scales across longleaf's natural range. Specifically, we looked for population genetic differences among populations and ecoregions, evidence for inbreeding, maintenance of gene diversity in germplasm collections, and, at the northern extent of the range, localized spatial patterns of relatedness.

MATERIALS and METHODS

The 745 longleaf pine samples from several sources, represented in 17 populations (18 cohorts), were genotyped for 10 SSR marker loci: NZPR0143, PtTX052, PtTX4003, PtTX4058, PtRIP_0984, PtSIFG_0561, PtSIFG_0745, PtSIFG_3147, PtSIFG_4102, and PtSIFG_4218 (Echt et al. 2011a). Samples from Virginia were additionally genotyped for markers PtSIFG_6065 and PtSIFG_6067 (Echt et al. 2011b). We omitted samples that were missing data for more than one marker, yielding 709 samples analyzed. The institutional sources of samples were as follows: clonal archives at Southern Region National Forest System's Seed Orchards (prefix NFS, Table 1), clonal archives at North Carolina Forest Service Nurseries (NCFS), a seedling seed orchard at Berry College (BC), native Virginia trees (VDF), a Virginia Department of Forestry provenance trial established with OP seed from International Forest Company (IFC), and a coastal and piedmont clone mix obtained from NCFS (NCFS_NCmix). Kurt Johnsen and Chris Maier collected samples from the VDF provenance trial; Billy Apperson and Bob Eaton collected samples from native Virginia stands. Records show that trees from which scions or needles were collected, and native Virginia trees, were established before 1930 and therefore are presumed to be naturally regenerated from native germplasm. The same is assumed true for stands from which IFC collected seed.

The NFGEL lab isolated DNA and determined genotypes of sample groups NF, NCFS, and BC. The SIFG lab isolated DNA and determined genotypes of sample groups IFC, NCFS and VDF.

All DNA extractions were from needle tissue. Following PCR amplification, both labs used ABI DNA Analyzers (capillary electrophoresis) for allele fragment detection. We ran control DNAs of known longleaf pine genotypes in both labs to standardize allele assignments, which allowed merging of data sets for analyses.

Gene diversity statistics, N_a, N_e, and H_e, were calculated with GenAlEx v6.501 software (Peakall & Smouse 2012). Estimates of inbreeding, F, for each population were obtained by simultaneously estimating frequencies of null alleles (non-amplifying SSR alleles that can deflate observed heterozygosity) using an individual inbreeding model and a Gibbs sampler with 10,000 iterations, as described by Chybicki & Burczyk (2009) and implemented in their INEst v1 software. F in this context is reported as the probability of alleles at a locus being identical by descent; that is, not Wright's F, but Malecot's. Genetic differentiation among populations or regions was measured by Jost's D_{est} with bootstrap 95% CI calculated by the DEMEtics R-package (Gerlach et al. 2010); hierarchical AMOVA for F_{st} was calculated by GenAlEx using 9999 permutations. Principal component analyses were conducted in GenAlEx using pairwise D_{est} values. All the above metrics were calculated using genotypes from the first set of 10 SSR loci listed above. For the Virginia population samples only, using genotypes from all 12 marker loci, two-dimensional local spatial autocorrelation (2D-LSA) was conducted with GenAlEx and plotted as bubble charts in Excel; sibling groups were identified with their 95% CI by maximum likelihood analysis using ML-RELATE software (Kalinowski et al. 2006); estimations for proportions of seedling progeny contributed by specific parents were conducted using a full-pedigree likelihood method implemented in COLONY.v2 software (Jones & Wang 2010).

RESULTS and DISCUSSION

Marker allele diversity results are summarized in Table 1. For all but one population, genotypes from 25 to 60 individuals were analyzed. This level of sampling is sufficient to estimate population allele frequencies accurately (Hale et al. 2012). Diversity measures for the one exception, a clonal archive of 17 sampled genets representing, a southern Alabama population (NFS_ALs), did not contradict our overall conclusions. We saw no evidence for differences in genetic diversity (N_a, N_e, H_e) among germplasm sources, with but one exception. The exception, a native Virginia provenance (VDF_VAnative), is an interesting and instructive case that we discuss in detail below. (Consequently, this population was not included in our range-wide diversity and differentiation analyses.) Once null alleles were accounted for, we saw no evidence of inbreeding in any population (95% CI of F included zero). We conclude that the various germplasm collections adequately represent natural longleaf pine gene diversity.

Contrary to expectations for populations from different ecozones, we measured slight or non-existent population differentiation across the range. AMOVA showed 1% of genetic variation was between populations, the rest within ($F_{st} = 0.01, p < 0.000$). In pairwise population comparisons, the maximum differentiation in allele frequencies was between the two most geographically distant populations from coastal North Carolina (NCFS_NCmix) and east Texas (NFS_TX) ($D_{\mathrm{est}} = 0.09$, 95% CI = 0.07 – 0.11).

To better assess broad regional evolutionary and biogeographic trends, we pooled samples within the five main geographic regions (Figure 1, Table 1) and then calculate pairwise differences. We noted with interest that the montane longleaf region (Mtn in Figure 1) is evolutionarily closest to the central Gulf (S coast) region. Also of interest is that piedmont longleaf is equally diverged from the east coast and south coast regions; we cannot speculate on its biogeographic origins. The largest, though still relatively small, difference was seen between the east coast and the west regions. Based on their isozyme survey, Schmidtling and Hipkins (1998) hypothesized a Holocene migration of longleaf westward up through south Texas toward the east coast. Our current analyses do not contradict this scenario, but do present the alternative hypothesis of post-glacial migration northward and outward from the central Gulf region. In any event, we conclude that there has been historically high gene flow through longleaf's range; any genetic effects of recent population declines and habitat fragmentation are not yet evident.

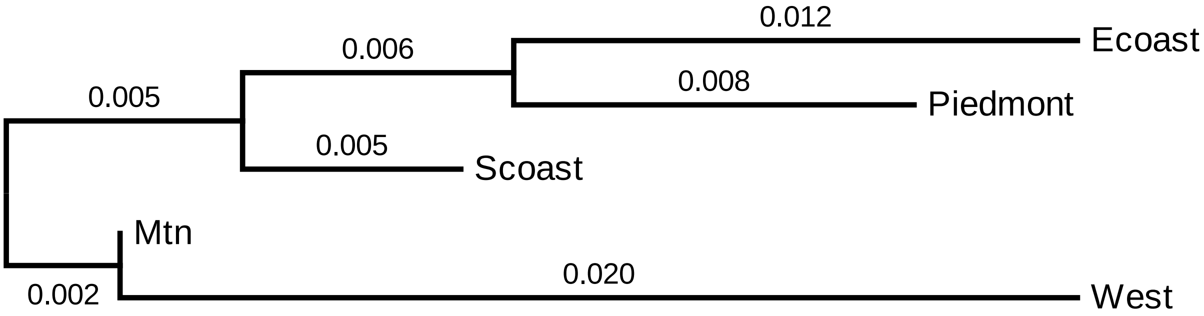
The VDF_VAholland population was not included in the preceding discussion of range-wide analyses because it proved to be a special case. This population derived from seeds collected at a small site of ten trees, 80 to 100 year-old, near Holland, VA. It was used by the Virginia Department of Forestry as the native Virginia representative in a provenance trial. Initial principal coordinate analyses (not shown) demonstrated it to be a genetic outlier. It had considerably lower marker diversity than other populations, but no evidence of population level inbreeding (Table 1). Suspecting consanguineous matings at the site, we genotyped the candidate parent trees and others from the surrounding area (these made up the VDF_VAnative population) and conducted 2D-LSA to look for clusters of relatedness. This spatial analysis identified four related individuals within the Holland site. Subsequent sibship analysis showed that these four trees were a group of full- and half-sibs (one of several possible sibship groups at the site). Further, a separate parentage analysis indicated that up to 40% of the seedlings from the Holland site shared one parent and at least one tree at the site from which seed was collected did not contribute any seedlings to the provenance trial. Therefore, not only were there strong familial relationships among many of the ten potential seed source trees, their progeny had very skewed parental representation. What is particularly notable about the Holland, VA provenance is that, despite its narrow genetic base, it performed as well or better than the other seven south-wide longleaf provenances in the trial (Creighton and Johnsen, personal communication). In conclusion, we have seen that fine-scale spatial analysis of individuals can identify strong genetic relationships not evident in population level analyses. We recommend that local genetic structures should be determined to identify seed sources that can maximize genetic diversity in germplasm conservation and tree improvement programs.

TABLE 1. Summary statistics for longleaf pine population genetic diversity and inbreeding.

Population	Region	n	N_a	N_e	H_e	F (SE)
IFC_SC	E. coast	50	6.8	3.6	0.64	0.013 (0.010)
VDF_VAnative	E. coast	60	6.8	3.4	0.64	0.011 (0.010)
VDF_VAholland	E. coast	51	4.0	2.3	0.53	0.007 (0.008)
NCFS_NCcoast	E. coast	32	6.4	3.7	0.65	0.006 (0.007)
NCFS_NCmix	E. coast	51	6.0	3.9	0.65	0.009 (0.009)
IFC_NC	piedmont	44	6.8	3.6	0.66	0.011 (0.010)
NFS_NC	piedmont	39	6.8	3.7	0.66	0.009 (0.010)
NCFS_AL	montane	26	5.9	3.5	0.65	0.008 (0.009)
BC_GA	montane	30	6.4	3.8	0.67	0.012 (0.012)
IFC_AL	montane	52	7.2	3.8	0.66	0.013 (0.015)
NFS_ALs	S. coast	17	5.6	3.3	0.63	0.007 (0.009)
NFS_FL	S. coast	25	6.5	3.6	0.64	0.008 (0.010)
NFS_MS	S. coast	26	6.3	3.4	0.65	0.010 (0.012)
IFC_FL	S. coast	48	6.5	3.8	0.67	0.025(0.023)
IFC_GA	S. coast	50	7.3	3.5	0.64	0.014 (0.013)
IFC_MS	S. coast	52	6.8	3.7	0.66	0.048 (0.024)
NFS_TX	west	30	6.7	3.9	0.68	0.005 (0.005)
NFS_LA	west	26	6.0	3.4	0.65	0.013 (0.016)

Population name prefix denotes the institutional source of the germplasm, followed by the state were germplasm originated, ending with a suffix further specifying germplasm origin. N_a mean number of alleles per locus (allelic richness), N_e effective allele number, H_e expected heterozygosity, F null-corrected inbreeding coefficient and its standard error.

FIGURE 1. Evolutionary relationships among regional pools of longleaf pine populations. The neighbor-joining algorithm (Saitou & Nei 1987) was used to generate the tree of evolutionary distances, conducted in MEGA5 (Tamura et al. 2011). The optimal tree with the sum of branch length = 0.05 is shown. The tree is drawn to scale, with branch lengths in D_{est} units.



REFERENCES

Chybicki IJ and Burczyk J. 2009. Simultaneous Estimation of Null Alleles and Inbreeding Coefficients. *Journal of Heredity,* 100:106–113.

Echt CS, Saha S, Krutovsky K, Wimalanathan K, Erpelding JE, Liang C, Nelson CD. 2011a. An annotated genetic map of loblolly pine based on microsatellite and cDNA markers. BMC Genetics 2011, 12:17

Echt CS, Saha S, Deemer DL, Nelson CD. 2011b. Microsatellite DNA in genomic survey sequences and UniGenes of loblolly pine. Tree Genet. Genome 7: 773-780.

Gerlach G, Jueterbock A, Kraemer P, Depperman J, and Harmand P. 2010. Calculations of population differentiation based on G_{st} and D: forget G_{st} but not all of statistics! Molecular Ecology 19:3845-3852.

Hale ML, Burg TM, Steeves TE (2012) Sampling for Microsatellite-Based Population Genetic Studies: 25 to 30 Individuals per Population Is Enough to Accurately Estimate Allele Frequencies. PLoS ONE 7(9): e45170.

Jones OR, and Wang J. 2010. COLONY: a Program for Parentage and Sibship Inference from Multilocus Genotype Data. Molecular Ecology Resources 10: 551–555.

Kalinowski ST, Wagner AP, and Taper ML. 2006. Ml-relate: a Computer Program for Maximum Likelihood Estimation of Relatedness and Relationship. Molecular Ecology Notes 6: 576–579.

Peakall R and Smouse PE. 2012. GenAlEx 6.5: Genetic Analysis in Excel. Population Genetic Software for Teaching and Research—an Update. Bioinformatics 28: 2537–2539.

Saitou N and Nei M. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4:406-425.

Schmidtling RC, Hipkins V. (1998) Genetic diversity in longleaf pine (*Pinus palustris*): influence of historical and prehistorical events. Canadian Journal of Forest Research 28:1135-1145.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, and Kumar S. 2011. MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Molecular Biology and Evolution* 28:2731-2739