Theor Appl Genet (1995) 90:584–594

C. G. Williams · J. L. Hamrick · P. O. Lewis

Multiple-population versus hierarchical conifer breeding programs: a comparison of genetic diversity levels

Received: 10 August 1994 / Accepted: 30 September 1994

Abstract Advanced-generation domestication programs for forest-tree species has raised some concerns about the maintenance of genetic diversity in forest-tree breeding programs. Genetic diversity in natural stands was compared with two genetic conservation options for a third-generation elite *Pinus taeda* breeding population. The breeding population was subdivided either on the basis of geographic origin and selection goals (multiple-population or MPBS option) or stratified according to genetic value (hierarchical or HOPE option). Most allelic diversity in the natural stands of loblolly pine is present in the domesticated breeding populations. This was true at the aggregate level for both multiple-population (MPBS) and the hierarchical (HOPE) populations. Individual subpopulations within each option had less genetic diversity but it did not decline as generations of improvement increased. Genetic differentiation within the subdivided breeding populations ranged from 1 to 5%, genetic variability is within each subpopulation rather than among subpopulations for both MPBS (> 95%) and the HOPE approaches (> 98%). Nei's G_{st} estimates for among-population differentiation were biased upwards relative to estimates of θ from Weir and Cockerham (1984).

Key words Multiple-population breeding (MPBS) Hierarchical open-ended breeding (HOPE) Genetic conservation · Conifers

Introduction

Genetic diversity bears on the long-term evolutionary success of a species. On a shorter time scale, maintaining genetic diversity is a worthwhile goal of plant domestication programs. Few forest-tree breeding programs have reached the stage of domestication where loss of genetic diversity is an issue; most have undergone less than four cycles of improvement. Recent use of small, elite breeding populations managed for short-term genetic gain could accelerate the loss of genetic diversity in domesticated populations of *Pinus taeda* L. Elite populations should be prone to allele loss through stringent selection, increased inbreeding and genetic drift due to small population sizes. Thus, this is an opportune time to address the question of whether the genetic diversity present in natural stands of forest trees is adequately represented in advanced-generation cultivated populations.

Background

Loblolly pine (*Pinus taeda* L.) is a monoecious, wind-pollinated pioneer species which naturally regenerates as even-aged stands along the Atlantic coastal and in piedmont regions throughout the southern United States. Longevity and long-distance movement of seed and pollen predisposes loblolly pine to extremely high levels of genetic diversity within populations (Hamrick et al. 1992). Only 11% of the total genetic diversity is apportioned among populations across the entire species range (Hamrick, unpublished data).

Plantation culture of loblolly pine dates to the 1930s when the first geographic variation studies for loblolly pine were planted. Forest-tree breeding in the southeastern U.S. was initiated in the 1950s when quantitative genetics theory was applied to the silvicultural improvement of the species. Currently, recurrent breeding populations are subdivided to control the rate of inbreeding, to permit selection for different sets of traits, and to emphasize improvement of the best selections. These small, elite populations isolated through pedigree control are an extreme departure from large, outcrossing natural populations with long-distance pollen and

Communicated by P. M. A. Tigerstedt

C. G. Williams (✉) · J. L. Hamrick · P. O. Lewis
Department of Genetics, North Carolina State University, Box 7614
Raleigh North Carolina 27695-7614, USA

seed dispersal. Genetic diversity criteria should reflect this shift in population size and gene flow. Genetic diversity is often measured using selectively neutral markers. With neutral loci, changes in gene frequency are attributed to genetic sampling or random drift. In subdivided populations, genetic differentiation should occur rapidly among populations as a result of genetic drift (Wright 1931; 1943; 1977 pp. 443–473). Drift shifts allelic frequencies randomly and results in a loss of alleles from one generation to the next.

The smaller the population, the greater the probability of randomly fixing alleles. For example, average heterozygosity (*He*) measured across all loci will decline with a population bottleneck. This can be predicted as $He_t = [1 - 1/(2Ne)]\ He_{t-1}$ where Ne is the variance-effective population size and t refers to the generation (Nei et al. 1975). For example, if the variance-effective population size of a small elite population stand is 5 and average heterozygosity of the previous generation (He_{t-1}) is 0.20 then average heterozygosity will fall to 90% of its original value, i.e., $He_t = 0.18$.

Random fixation of high-frequency neutral alleles will also increase as population size is reduced from one generation to the next. Random allele loss is problematic for advanced-generation elite populations, because allele loss will compromise long-term selection response.

There are two general solutions to the random loss of alleles. First, unrelated germplasm can be periodically infused into an elite population. This solution is hierarchical because the breeding population is subdivided according to the genetic value of the germplasm. Hierarchical schemes have been collectively defined as the HOPE genetic conservation strategy (Eriksson et al. 1994). Second, a breeding population can be divided into replicates or subpopulations which represent different sources or selection criteria. In this case, the population is not subdivided on the basis of genetic value but on co-ancestry. Random allele loss occurs in any one replicate but, on the average, gene frequencies remain fairly constant from one generation to the next. This solution, based on Wright's shifting balance theory, was "the only practical way of bringing about a rapid advance...through subdivision of the population into isolated and hence differentiating small groups, among which selection may be practice" (Wright 1931). It serves as a basis for sub-lining and multiple-population breeding strategies (MPBS) in forest-tree breeding (Burdon and Namkoong 1983; Cotterill 1984; Namkoong 1984; Lowe and van Buijtenen 1986).

Predictions for genetic diversity levels using the HOPE breeding strategy

The hierarchical population structure is used in some maize breeding programs (Kannenberg 1984; Kidd 1992) and for some cooperative forest-tree breeding programs (i.e., Cotterill 1984; McKeand and Bridgwater 1992). HOPE refers to the Hierarchical Open-Ended Breeding System designed at the University of Guelph for maize (Kannenberg 1984). HOPE programs are characterized by a tradeoff between short-term gain and the maintenance of long-term genetic diversity. Each program is composed of a hierarchy of breeding populations with successively higher performance level, culminating in an elite population from which selections are extracted for use in commercial production.

In the example of a three-tiered HOPE program, the entry level of the hierarchy would serve as the gene pool or genetic conservation archive. It functions as a low-cost, dynamic population to conserve alleles which are present in the early generations of domestication. The next level in the hierarchy is bred for long-term germplasm enhancement, where exotic or novel gene complexes are introduced into elite lines. The upper level of the hierarchy is an elite population which would be rapidly cycled to maximize genetic gain per unit time.

In this study, we use a two-tiered HOPE strategy for forest trees designed to improve a single elite population and infuse unrelated selections every other generation (Mahalovich 1989). These enrichment selections would come from a main subpopulation composed of progeny-tested parental selections and their purpose is to maintain genetic diversity and slow the rate of inbreeding within the elite population (Mahalovich 1989, pp. 79–81). Replacing half of the elite selections with enrichment selections is a conservative option with respect to short-term gain; this contrasts with the option to develop a closed elite population with no infusion. Both options make efficient use of resources and can be rapidly cycled to extract maximum short-term genetic gain.

The two-tiered HOPE alternative is expected to have lower levels of genetic diversity than the multiple-population or MPBS counterpart. The HOPE enrichment subpopulation is expected to maintain more genetic diversity than its elite counterpart, relative to natural stands. The third-generation elite subpopulation should have reduced genetic diversity relative to the first-generation main population and natural stands.

Predictions for genetic diversity levels using a multiple-population (MPBS) strategy

The multiple-population breeding strategy (MPBS) combines short-term gain, germplasm enhancement and genetic conservation objectives into the same breeding population (Namkoong 1984; Barnes 1984; Eriksson et al. 1994). This contrasts with the HOPE model where short-term gain, germplasm enhancement and genetic conservation objectives are met with separate subpopulations.

Multiple-population breeding systems are applicable to species for which large numbers of alleles are still available. MPBS is used in public crop breeding for germplasm enhancement (Bramel-Cox and Cox 1988) and in both government and private forestry programs

(Barnes 1984). Under MPBS, small subpopulations can be established from different seed sources and, within each population, separate selection criteria are used. In practice, a MPBS plan tends to have a continuum of subpopulations. Some are selected for diverse product goals while others are replicates, sharing a common regional origin and product goal (Burdon and Namkoong 1983).

The genetic diversity level for the aggregate MPBS breeding population is expected to be higher than the HOPE alternative and as high as the natural stands. The rationale is that MBPS accommodates improvement of novel or exotic germplasm without dilution of well-tested elite selections. Also, MPBS is based on a larger number of smaller subpopulations and thus more rapidly fixes alternate alleles through genetic drift. Among the six subpopulations used in this study, infusion of rare alleles from an exotic subpopulation (HD) or from the additions of germplasm from national forests (STDNC, STDSC) and research programs (GGC) were expected. As in the HOPE alternative, elite, third-generation forward selections (ELA, ELB) are expected to be less genetically diverse than their first-generation progenitors.

The objectives of this study were to test the following hypotheses: (1) that genetic diversity in the aggregate MPBS breeding population is expected to be higher than the HOPE alternative and is as high as the natural stands and (2) that elite, third-generation offspring selections are less genetically diverse than their first-generation progenitors.

Materials and methods

Population origins for HOPE and MPBS programs

Genetic improvement programs for *Pinus taeda* were started four decades ago. Selections were made in natural stands, then branches were removed from each selection and grafted for breeding. Biparental pedigree control has been used since the inception of loblolly-pine breeding programs by using isolation bags to ensure the paternal success of a single, designated pollen parent. Offspring (forward) selections from tests grown from the resulting seeds. Parental (backward) selections are also made on the basis of offspring performance although these offspring may be either wind- or control-pollinated.

The HOPE and MPBS breeding populations are composed of subpopulations which have each been maintained separately by different state, federal, international and industrial breeders. Coancestry levels between any two subpopulations are generally zero but a few selections are related as half-sibs, half first-cousins or grandparent-grandchild ($F = 0.0625$–0.25). To-date, there have been two cycles of generation improvement since the selections in natural stands. Selections in natural stands are referred to as first-generation selections and their progeny are second-generation selections. Third-generation selections are progeny of second-generation matings. No matings in this study are made between generations. The origins of all subpopulations are generalized as follows.

ELITE, ELA, ELB: The elite population is descended from 34 phenotypic selections made in natural stands in coastal Virginia, North Carolina and South Carolina on industrial land holdings. The ELITE population in the HOPE aggregate is composed of 50 forward, third-generation selections. The MPBS subpopulations EL-A and EL-B are replicates which have minimum shared co-ancestry. ELA and ELB are comprised of a few backward second-generation parental selections as well as third-generation forward selections. Selection was made on the basis of early growth in short-term tests (Williams and Lambeth 1992).

HD: This MPBS subpopulation represents exotic loblolly-pine germplasm bred in southern Africa and adapted to eastern North Carolina. These second-generation forward selections are offspring from selections made in Zimbabwe plantations which were planted with seed collected from putative loblolly-pine plantations in South Africa. Selections were made on the basis of half-rotation sawlog quality, early disease screening and adaptability.

MAIN, STDNC, STDSC: These subpopulations represent the best first-generation parental selections from eastern North Carolina and South Carolina, respectively. Half of these selections come from national-forest breeding programs in these regions. For MPBS, the STDNC or STDSC subpopulations are divided by state of origin (North or South Carolina) and for HOPE, they are combined across state boundaries to comprise the MAIN collection. Second-generation parental selections were also added to MAIN to provide a high-value enrichment population.

GGC: Six selections in this MPBS subpopulation were drawn from North Carolina State University-Industry Cooperative research trials where the best first-generation parental selection were intermated (McKeand et al. 1988). These parental selections originated from natural-stand selections made in eastern Virginia, North Carolina and South Carolina on industrial land holdings. Their offspring were tested and selected. The remaining three selections were forward selections from an eastern North Carolina plantation-enrichment population. Selection was based on growth in conventional tests.

Sample collection

For selected loblolly pine, three branch samples per genotype were placed in plastic bags and shipped on ice to prevent protein denaturing. Samples were stored at 4 °C until protein extraction. All genotypes were sampled within the U.S. Atlantic seaboard eastward from 32° N, 80° 30′ W to 38° N, 77° 30′ W.

For natural stands, a single branch was taken from each of 48 randomly selected trees sampled in natural stands at each of three U.S. Atlantic locations: Andrews, South Carolina; Washington, North Carolina, and Windsor, Virginia. A total of 132 advanced-generation selections were sampled for MPBS and HOPE. Three replicate branch samples were assayed; if there was a discrepancy then another set of samples were taken from the genotype and re-analyzed.

The advanced-generation selections were allocated to two different breeding-conservation programs prior to electrophoretic analyses (Tables 1 A, B). First-, second- and third-generation selections are divided into three mutually exclusive breeding-conservation strategies. The multiple-population strategy consisted of six unrelated subpopulations (MPBS). The second, a hierarchical (HOPE) strategy, consists of two subpopulations, an elite population and a main subpopulation of unrelated backward selections. There is additional overlap between the options; selections in the main are represented in several MPBS subpopulations also.

Electrophoresis

Meristem and needle tissue were crushed in liquid nitrogen with a mortar and pestle, and proteins were extracted using a phosphate polyvinyl pyrolidone buffer (Mitton et al. 1979). Samples were run on five buffer systems (Soltis et al. 1983) to resolve 19 loci from 13 enzyme systems using 10% starch gels: alcohol dehydrogenase (*Adh*), adenylate kinase (*Ak2*), diaphorase (*Dia2*), fluorescent esterase (*Fe1*, *Fe2*), glutamate dehydrogenase (*Gdh*), glutamate oxaloacetate transaminase (*Got1*, *Got2*), isocitrate dehydrogenase (*Idh*), malate dehydrogenase (*Mdh1*, *Mdh2*), phoshoglucomutase (*Pgm1*, *Pgm2*), phoshoglucoisomerase (*Pgi2*), shikimate dehydrogenase (*Skdh*), triose-phosphate isomerase (*Tpi1*, *Tpi2*), 6-phosphogluconic dehydrogenase (*6Pgd1*, *6Pgd2*). The genetic basis of allozyme banding patterns was inferred from segregation patterns. For enzymes with

Table 1 A Description of each subpopulation in the multiple-population breeding population (MPBS) option

Criteria	Multiple population breeding strategy (MBPS)					
	ELA	ELB	HD	NCS	SCS	GGC
# Generations since initial selections	3	3	2	1	1	2
Origin	Natural	Natural	Plantation	Natural	Natural	Natural & plantation
Selection traits	Growth	Growth	Growth, sweep, quality	Growth	Growth	Growth
Selection method/age	Offspring, age 3	Offspring, age 3	Offspring, age 12	Parental, age 8	Parental, age 8	Offspring, age 8
Census size ($N = 132$)	25	30	33	21	14	9

Table 1 B Description of each subpopulation in the enriched elite breeding (HOPE) option

Criteria	Elite enrichment strategy (HOPE)	
	Elite	Main
# Generations since initial selections	3	1
Origin	Natural stand	Natural stand
Selection traits	Growth	Growth/straightness
Selection method/ selection	Offspring selections, age 3	Parental selections, age 8
Census size($N = 90$)	50	40

more than one locus, isozymes were numbered sequentially. The most anodal was designated as the first locus and alleles were numbered in a similar manner. Both monomorphic and polymorphic data are reported but only for those loci which exhibited clearly interpretable banding patterns. These 19 loci were not a random sample of all possible loci; they were selected for analysis because they had proved to be polymorphic in a rangewide survey of genetic diversity in loblolly pine (Hamrick, unpublished data). As a result, the absolute measures of genetic diversity in this study will be higher than reported in previous studies since monomorphic loci have been excluded.

Genetic diversity estimators

1. Average heterozygosity and number of alleles: Genetic diversity for natural stands, MBPS and HOPE (among subpopulations and within each subpopulation) was measured across all 19 loci using five standard population genetic parameters: $\%P$ = percentage of polymorphic loci, He = expected average heterozygosity among subpopulations, A = average number of alleles per locus, Apoly = average number of alleles per polymorphic locus and the total of all alleles across all loci.

2. Degree of population subdivisions: F-statistic. Wright's F-statistics are the most widely used set of descriptors for the degree of population subdivision although there are different approaches to estimation if there are small samples, multiple loci and multiple alleles (Nei 1973; Nei and Chesser 1983; Weir and Cockerham 1984; Nei 1986). Wright's fixation indices or F statistics, F_{it}, F_{is} and F_{st} (Wright 1965), measure the correlation between uniting gametes under various sampling schemes. The calculation of F-statistics is achieved by calculating the degree of genetic diversity (Ho, He, Hs) at different levels in the sampling scheme (within-individual, within-subpopulation and among subpopulations) and is analogous to an analysis of variance.

Under a model including drift, but excluding mutation and other perturbing forces such as assortative mating, F-statistics can be interpreted as descent measures. F_{it} and F_{is} measure the probability that the two genes within an individual are identical-by-descent, F_{it} taking into account variation among subpopulations and F_{is} ignoring variation among subpopulations. F_{st} measures the probability that genes from different individuals within the same subpopulation are identical by descent. F_{st} can be used as a measure of differentiation among subpopulations since subpopulations become more genetically divergent as individuals within subpopulations become more related.

Weir and Cockerham (1984) chose to use the symbols F, f, and θ, for Wright's parameters F_{it}, F_{is}, and F_{st} because of the considerable confusion in the literature surrounding the parameters and the statistics used to estimate them. For example, one common statistic used to estimate F_{it} is as follows:

$$F_{it} = \frac{Ho}{p(1-p)}$$

where Ho is the observed heterozygosity and p the average gene frequency. This was the formula given by Wright, which was intended to be used not for estimation but as a description of the level of inbreeding expected in an infinitely large but subdivided population. When such parameters are estimated from a sample, statistical sampling error is introduced and leads to biased estimates if not explicitly accounted for. Similar problems with bias occur with the traditional formula for F_{st}:

$$F_{st} = \frac{s^2}{p(1-p)}$$

where s is the standard deviation of subpopulation gene frequencies and p is the average gene frequency. This formula provides an accurate estimate only when the number of subpopulations sampled is large and large samples are taken within each subpopulation. Nei's coefficient of population differentiation, G_{st}, has been often used as an estimate of Wright's F_{st}, but has also been shown to depend heavily on sample size within subpopulations and the number of subpopulations sampled (Cockerham and Weir 1993).

We present Nei's estimates alongside estimates obtained using the formulas presented by Weir and Cockerham (1984) which take statistical sampling into account. While the sampling distribution for the Weir and Cockerham estimators is unknown, tests of specific hypotheses can be carried out by bootstrapping across loci. Each independent locus provides, in effect, a replication of the genetic sampling process that creates relatedness among individuals within subpopulations and thus differentiation among subpopulations, so the empirical distribution obtained by resampling loci with replacement provides a natural context within which to test hypotheses about the distribution of such estimators (Weir 1990 p. 140). We were particularly interested in testing the null hypothesis of $\theta = 0$. A total of

999 bootstrap data sets were generated by choosing loci at random (and with replacement) from the original collection of loci and θ was estimated for each of these data sets. A 95% confidence region was constructed using the 1 000 estimates. The original estimate of θ was declared significant if the confidence region did not include zero.

The χ^2-test of heterogeneity in allelic frequencies across subpopulations is very sensitive to small, expected values (Weir 1990 pp. 76–77) and the experiment-wise error rate is quite high for multiple loci (Weir 1990 p. 109). Multiple chi-square tests carry a high probability of accepting a chi-square value as significant when it is not truly different from zero. For 19 loci, two subpopulations and a threshold of 1%, the experiment-wise error rate carries a 34% probability for spurious rejection of the null hypothesis.

Among subpopulations, chi-square tests of heterogeneity in allele frequencies detect subpopulation differentiation. Positive values for F_{it} and F_{is} represent deviations from Hardy-Weinberg genotypic proportions among and within subpopulations. Positive values indicate an excess of homozygotes, and negative values indicate an excess of heterozygotes. A χ^2-statistic, used to detect heterogeneity in allele frequencies among populations (Workman and Niswander 1970) and the appropriate degrees of freedom (df) are given as:

$$\chi^2 = F_{st}\, 2n(a-1) \text{ and } df = (a-1)(n-1)/2$$

where a is the total number of alleles at the locus and n is the number of subpopulations. Significant χ^2 values indicate that allele frequencies are truly different among subpopulations at that locus and that the breeding population as a whole is not a panmictic population.

3. Genetic distance and co-ancestry distance. Genetic distance (D) and identity (I) (Nei 1972) were calculated for all pairwise combinations of populations to measure accumulated electrophoretic differences between populations. These were obtained using GENESTAT 3.31 (Lewis and Whitkus 1989) to facilitate direct comparisons with previous studies of natural populations. The estimator θ also provides a measure of genetic distance when estimated in a pairwise fashion for two subpopulations only. When genetic drift is the only force causing changes in gene frequencies among subpopulations, the distance measure is obtained as $d = -\ln(1-\theta)$ which is proportional to the time since divergence of the two subpopulations (Weir 1990, p. 167). This measure of distance is preferable to Nei's genetic distance when divergence is low and not likely to have been greatly affected by mutation.

4. Variance-effective population size. Variance-effective population size (Ne_v) is defined as the effective number that renders different populations comparable with respect to changes in gene frequency due to genetic sampling and is related to the number in the progeny generation (Crow and Kimura 1970, pp. 352–357).

For a population of monoecious diploids, N_{t-1} is the census, k represents the number of successful gametes per individual, V_k is the variance around the mean number of gametes, $\bar{k}$ (Crow and Kimura 1970, pp. 352–357) and the variance-effective population size is:

$$Ne_v = \frac{(N_{t-1}\bar{k}) - 1}{(V_k/\bar{k}) + \bar{k} - 1}$$

Results

The multiple-population (MPBS) aggregate had genetic diversity comparable to that of natural stands (Table 2) even though its effective population sizes were lower ($Ne_v = 5$–22) than that of the natural stands (Table 3). The average expected heterozygosity was lower for MPBS ($He = 0.180$) than for natural stands ($He = 0.213$) but these values were not detectably different. The average alleles per locus (A) were equivalent; for polymorphic loci (Apoly), MPBS had a comparable number of alleles per locus relative to the natural population. The aggregate MPBS population captured almost as many total alleles (50) as sampling in three natural stands (52) (Table 2); the number of polymorphic loci changed from 93 to 90% relative to natural stands. The MPBS option exhibited the highest level of genetic diversity for a given effective population size.

Table 2 Comparative estimates of genetic diversity for three breeding strategies and natural stands of loblolly pine. MPBS refers to a multiple population breeding strategy, HOPE refers to a pyramidal breeding strategy and ELITE is an elite, third-generation breeding population managed for short-term gain. Expected average heterozygosity (He), the percentage of polymorphic loci (%P), the number of alleles per locus (A), the number of alleles per polymorphic loci (A poly) are reported

Estimates of genetic diversity	Natural stands	MPBS aggregate	HOPE aggregate	ELITE
He	0.213	0.180	0.181	0.171
%P	100	90	85	75
A	2.63	2.50	2.40	2.05
A poly	2.63	2.67	2.65	2.40
Total alleles	52	50	47	41
Total N	144	132	90	50

The HOPE option was comprised of fewer selections ($N = 90$) than MPBS but had higher effective population sizes ($Ne_v = 40$–41) (Table 3). There were fewer polymorphic loci (85%) and the average number of alleles per locus was lower than either the MPBS or natural populations (Table 2). Most of the genetic diversity in natural stands was still found within HOPE. However, fewer total alleles (47) were represented in the HOPE aggregate.

Third-generation elite selections exhibited a decrease in genetic diversity relative to natural stands (Table 2). The elite subpopulations (ELITE, ELA, ELB) had less genetic diversity than either the MPBS or the HOPE aggregate (Tables 2, 3) and they exhibited greater allele loss (Fig. 1 A, B). The elite subpopulation alone had a total of 41 alleles as opposed to 50 for MPBS, 47 for HOPE and 52 for the natural population. Rare allele distribution shifted upward in the selected third-generation population; allele loss seemed highest in the 0.01 range (Fig. 1 B).

Genetic sampling across subpopulations was more important than the number of breeding cycles. There were no trends associated with generational level. For example, NCS and SCS subpopulations were both composed of first-generation selections yet SCS had only 68% polymorphic loci compared with 84% polymorphic loci for NCS (Table 4). Values for SCS and NCS overlapped with ELA and ELB, replicates drawn from a third-generation population (Table 4).

For MBPS and HOPE, there was no more differentiation than what has been previously reported for natural conifer populations. For MPBS, allelic frequencies diverged among subpopulations. Allele frequencies at 6 out of 18 polymorphic loci differed at the 1% signifi-

Table 3 Variance-effective population sizes (Ne_v) for MPBS and HOPE subpopulations. The original selections were made in natural stands and these selections were bred for two cycles. G3 refers to third-generation elite selections

Sub-populations	Initial stands generations: (1)	Ne_v across		Census number (N) for G3	Ne_v/N ratio for G3
		G2	G3		
HOPE					
ELITE	34	29	41	50	0.82
MAIN	40			40	1.00
MPBS					
ELA	21	22	17	25	0.68
ELB	25	17	21	30	0.70
HD	13	5	–	33	0.15
NCS	21	–	–	21	1.00
SCS	14	–	–	14	1.00
GGC	9	–	–	9	1.00

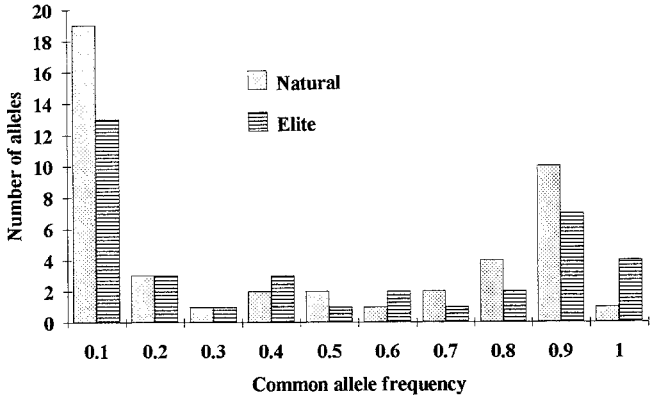

Fig. 1 A Distribution of common alleles for natural loblolly pine and third-generation elite selections

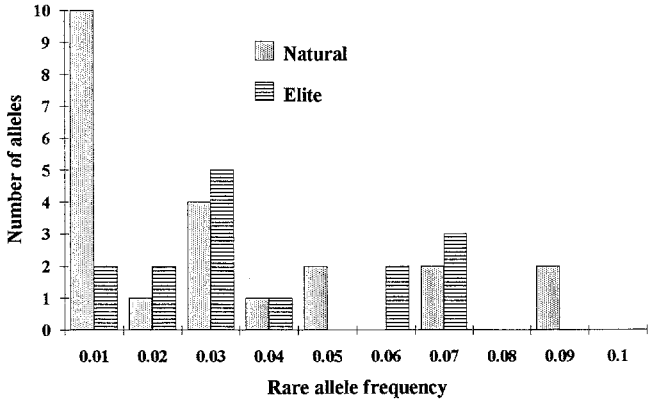

Fig. 1 B Distribution of rare alleles for natural loblolly pine and third-generation elite selections

cance level among the six subpopulations (Table 5). There was a statistically significant θ value (0.038 or 3.8%) and a higher G_{st} value (0.050 or 5%) for MPBS which indicated genetic differentiation among subpopulations (Table 5). Genetic differentiation was not greater for the more highly selected (ELA, ELB, ELITE) subpopulations nor for the exotic (HD) subpopulation of southern Africa origin (Table 4).

The HOPE option showed no divergence among allelic frequencies (Table 2) and the value for θ (0.0106) and for G_{st} (0.0114 or 1.1%) was not different from zero (Table 6). Both MPBS and HOPE subpopulations were no more divergent (1 to 5%) than natural stands (1.4%) within the Atlantic Seaboard population (Tables 3 and 5).

Estimates of G_{st} were slightly biased upward with respect to the among-population component of the total variance. The bias arises from both the number of individuals sampled per subpopulation and the number of subpopulation. We reported G_{st} because it is useful to compare among studies and it has been used widely in the forest genetics literature. For unbalanced sampling schemes such as ours, the estimator θ was actually more appropriate because it was unbiased with respect to sample size imbalance. This is important for a direct comparison among the natural population (three stands), MPBS (six subpopulations) and HOPE (two subpopulations) because both subpopulation numbers and sizes are unbalanced.

For MPBS, most loci were in Hardy-Weinberg genotypic proportions and appeared to be neutral with respect to selection (Table 5). The chi-square analysis of Wright's fixation index for each locus in each population showed significant deviations from zero in only 5 out of 96 cases (one negative, four positive). The chi-square analysis for HOPE showed similar results (Table 6). Of 36 tests, only 3 showed significant deviations (two negative, one positive) at the 1% level or higher. For MPBS the F_{is} value was 0.02 and for HOPE the F_{is} value was 0.05 (nearly zero); also indicating that for most loci the genotypes are in Hardy-Weinberg proportions.

Measures of co-ancestry and genetic distances supported the similarity between selected and natural populations (Tables 7, 8). We reported Nei's genetic identities because they have been widely reported in the forestry literature but pairwise θ values may be more appropriate for estimates of genetic distance for plant-breeding programs (Table 8). The pairwise θ appropriately reflects short-term drift whereas Nei's genetic distances reflect both mutation and drift which are more appropriate for changing gene frequencies over a longer time-frame.

Table 4 Estimates of genetic diversity for the multiple population option (MPBS) and for the HOPE option for *P. taeda* L. Expected average heterozygosity (He), the percentage of polymorphic loci (%P), the number of alleles per locus (*A*) and the number of alleles per polymorphic loci (*A* poly) are reported. Gen refers to the generational level of the subpopulations; loblolly pine breeding programs historically have not used overlapping generations

Option/sub	Gen	*Ho*	*He*	*%P*	*A*	Apoly
NAT-VA	0	0.199 (0.010)	0.210 (0.034)	89	2.21	2.35
NAT-NC	0	0.186 (0.010)	0.213 (0.034)	95	2.32	2.39
NAT-SC	0	0.185 (0.010)	0.210 (0.034)	95	2.32	2.39
MPBS/ELA	3	0.168 (0.015)	0.183 (0.043)	79	2.05	2.33
MPBS/ELB	3	0.176 (0.017)	0.185 (0.045)	68	1.89	2.31
MPBS/HD	2	0.152 (0.014)	0.174 (0.045)	63	1.89	2.42
MPBS/NCS	1	0.152 (0.018)	0.171 (0.042)	84	2.05	2.25
MPBS/SCS	1	0.182 (0.023)	0.198 (0.051)	68	1.95	2.38
MPBS/GGC	2	0.200 (0.030)	0.204 (0.036)	79	1.84	2.07
HOPE/elite	3	0.171 (0.012)	0.185 (0.045)	75	2.05	2.40
HOPE/main	1	0.164 (0.013)	0.195 (0.045)	89	2.21	2.35

Table 5 Number of alleles at each polymorphic locus, heterogeneity chi-square values (with degrees of freedom) and genetic diversity statistics for MPBS. Two types of population structure statistics are reported. F_{it}, F_{is} and G_{st} are shown next to F, f and θ for each locus

Locus	Observed alleles	χ^2 (*df*)[a]	F_{it}	F	F_{is}	f	G_{st}	θ
Adh	3	15.68 (10)	−0.039	−0.033	−0.070	−0.047	0.029	0.013
Pgm1	2	7.41 (5)	−0.069	−0.063	−0.100	−0.077	0.028	0.013
Pgm2	3	***21.18 (10)***	−0.022	−0.011	−0.073	−0.049	0.048	0.037
Tpil	2	4.68 (5)	−0.019	−0.016	−0.038	−0.014	0.018	−0.001
Tpi2	–	–	–	–	–	–	–	–
Dia2	3	15.45 (10)	0.494	0.497	0.481	0.499	0.026	−0.004
Pgi2	5	***34.34 (20)***	0.059	0.066	0.025	0.048	0.035	0.019
6Pgd1	5	***53.00 (20)***	0.020	0.035	−0.053	−0.030	0.069	0.063
6Pgd2	3	12.88 (10)	−0.030	−0.024	−0.056	−0.033	0.025	0.008
Idh	2	13.72 (5)	−0.004	0.008	−0.059	−.036	0.052	0.042
Skdh	3	***26.80 (10)***	0.035	0.052	−0.045	−0.021	0.077	0.072
Fe1	3	11.84 (10)	0.103	0.109	0.075	0.099	0.031	0.011
Fe2	3	***39.68 (5)***	0.161	0.189	0.013	0.036	0.150	0.159
Gdh	2	7.27 (5)	0.374	0.378	0.356	0.377	0.028	0.002
Got1	2	4.99 (5)	−0.023	−0.019	−0.043	−0.020	0.019	0.000
Got2	2	6.99 (5)	−0.059	−0.053	−0.088	−0.065	0.026	0.011
Ak2	2	5.15 (5)	−0.027	−0.023	−0.048	−0.024	0.019	0.001
Mdh1	2	***18.86 (5)***	0.663	0.668	0.637	0.651	0.072	0.050
Mdh2	2	5.98 (5)	0.083	0.089	0.062	0.085	0.023	0.003
Mean			0.069	0.080	0.020	0.043	0.050	0.038

[a] Bold italicized values are statistically significant at the 1% level

Genetic identities between selected and natural populations were lower ($I = 0.914$–0.955) than genetic identities within either of these ($I = 0.930$–0.999 and $I = 0.995$–0.998, respectively in Table 7). In studies of natural populations of *P. taeda*, genetic identity (*I*) values were rarely lower than 0.95. Genetic identities for selected and natural populations indicated drift after one or two generations of selection (Table 7). This was due to the combined effect of rare allele loss and to changes in gene frequencies. The latter had little effect on the genetic diversity measures. For example, if allele frequencies shifted from $p = 0.6$ and $q = 0.4$ to $p = 0.4$ and $q = 0.6$ there would not be a change in average heterozygosity (*He*) but the genetic identity value (*I*) will decrease. Drift appears to have changed the genetic composition of selected populations and had altered genetic diversity through loss of rare alleles.

Genetic identity estimates also supported the similarity among MPBS subpopulations. Genetic identity was not greater for the more highly selected subpopulations (ELA, ELB, ELITE) or for an exotic (HD) subpopulation originating from southern Africa. There was less difference between *HD* and the other subpopulations than there was between the two elite replicate subpopulations (ELA and ELB) (Table 7).

There were also some low-frequency alleles ($q < 15\%$) in the selected populations which were not present in the natural stands. In MPBS, there were nine alleles present which were not found in the natural populations. Of these, eight were found in selections originating from natural stands along the Atlantic Seaboard. This is not unusual since the natural stands were local samples rather than progenitors of actual selections. If more stands had been sampled from

Table 6 Number of alleles at each polymorphic locus, heterogeneity chi-square values (with degrees of freedom) and genetic diversity statistics for HOPE. F_{it}, F_{is} and G_{st} are shown next to F, f and θ for each locus

Locus	Observed alleles	χ^2 (df)[a]	F_{it}	F	F_{is}	f	G_{st}	θ
Adh	3	2.66 (2)	−0.049	−0.045	−0.053	−0.042	0.004	−0.003
Pgm1	2	0.16 (1)	−0.046	−0.046	−0.048	−0.036	0.001	−0.009
Pgm2	3	4.58 (2)	−0.018	−0.005	−0.032	−0.020	0.013	0.015
Tpi1	2	0.05 (1)	−0.023	−0.023	−0.023	−0.012	0.001	−0.011
Tpi2	–	–	–	–	–	–	–	–
Dia2	3	4.10 (2)	0.492	0.494	0.488	0.497	0.006	−0.005
Pgi2	4	3.38 (3)	0.101	0.103	0.100	0.111	0.002	−0.009
6Pgd1	4	9.21 (4)	−0.018	−0.012	−0.024	−0.013	0.006	0.001
6Pgd2	3	2.50 (2)	−0.041	−0.037	−0.044	−0.033	0.004	−0.003
Idh	1	–	–	–	–	–	–	–
Skdh	3	2.53 (2)	−0.032	−0.024	−0.040	−0.029	0.008	0.005
Fe1	3	3.12 (2)	0.131	0.138	0.124	0.136	0.008	0.003
Fe2	2	4.32 (1)	0.141	0.161	0.120	0.131	0.024	0.035
Gdh	2	0.24 (1)	0.517	0.517	0.516	0.524	0.001	−0.015
Got1	2	5.11 (1)	−0.023	+0.006	−0.053	−0.041	0.028	0.046
Got2	2	5.29 (1)	−0.090	−0.058	−0.123	−0.112	0.030	0.048
Ak2	2	0.61 (1)	−0.017	−0.014	−0.020	−0.009	0.003	−0.004
Mdh1	2	1.26 (1)	−0.006	0.001	−0.013	−0.001	0.007	0.003
Mdh2	2	1.54 (1)	0.037	0.046	0.029	0.040	0.009	0.005
Mean			0.062	0.076	0.053	0.066	0.009	0.0110

[a] No values were statistically significant at the 1% level

Table 7 Nei's genetic identities (above) and genetic distances (below) among populations in natural stands and subpopulation within MPBS and HOPE breeding-conservation options

	Natural stands			Multiple population breeding system					HOPE		
No.	48	48	48	25	30	33	21	14	10	50	40
Code	VA	NC	SC	ELA	ELB	HD	NCS	SCS	GGC	Elite	Main
VA	–	0.998	0.996	0.949	0.922	0.935	0.951	0.946	0.932	0.940	0.953
NC	0.002	–	0.995	0.948	0.914	0.932	0.949	0.941	0.931	0.936	0.950
SC	0.004	0.005	–	0.953	0.925	0.934	0.955	0.944	0.936	0.944	0.955
ELA	0.053	0.054	0.049	–	0.984	0.989	0.999	0.992	0.982	–	–
ELB	0.081	0.089	0.078	0.016	–	0.985	0.986	0.987	0.966	–	–
HD	0.068	0.070	0.068	0.011	0.015	–	0.992	0.996	0.979	–	–
NCS	0.051	0.052	0.046	0.001	0.014	0.008	–	0.930	0.924	–	–
SCS	0.055	0.061	0.057	0.008	0.013	0.004	0.073	–	0.967	–	–
GGC	0.070	0.071	0.066	0.018	0.034	0.021	0.079	0.031	–	–	–
Elite	0.062	0.066	0.057							0.003	0.997
Main	0.048	0.052	0.046								

Table 8 Pairwise values for θ genetic identities within the MPBS option

	Multiple population breeding system					
Code	A-1	A-2	B	C	D	E
A-1						
A-2	0.0651					
B	0.0505	0.0629				
C	0.0318	0.0576	0.0010			
D	0.0253	0.0688	0.0200	0.0210		
E	−0.0024	0.0297	0.0175	−0.0162	0.0206	

this region then more rare alleles would have been detected.

The ninth allele was the only contribution unique to the exotic germplasm from southern Africa. The HD subpopulation originating from southern Africa had one genotype with a novel allele at the PG1–2 locus. This allele, observed in one selection, has not been observed in any previous range-wide loblolly-pine populations (J. Hamrick, unpublished data). This unique allele could have arisen from either ancestral migration from mixed loblolly-slash pine (*Pinus elliottii*) plantations in southern Africa where phenotypic selections were made or from a spontaneous mutation.

Discussion

Most of the genetic diversity found in natural stands is still present in the selected populations. This was especially true for the aggregate MPBS and HOPE options. The third-generation elite subpopulation ex-

hibited only a slight decrease in genetic diversity relative to natural stands. Allele loss was confined to low-frequency alleles ($q < 15\%$). Slight genetic differentiation was observed among either MPBS or HOPE subpopulations; most genetic diversity still exists within each subpopulation rather than among subpopulations.

Genetic diversity for MPBS and HOPE versus natural stands

Aggregate breeding populations such as MPBS and HOPE conserved more genetic diversity than the elite population (ELITE) alone. Genetic diversity for MPBS was high although the MPBS breeding population size had lower effective numbers (ranging from 15 to 100% of census number) for subpopulations. Still, there was some question as to whether genetic diversity should be higher for the aggregate MPBS population and whether additional subpopulations should be added.

Over 94% of the variability for selectively neutral isozyme loci was still within a subpopulation rather than among subpopulations. There was sufficient genetic diversity with six (or fewer) subpopulations. Additional subpopulations drawn from the periphery of the species' range may not clearly increase genetic diversity of this MPBS population at this time. Product uncertainty, long-term adaptability and reforestation demand are the appropriate criteria to justify additional subpopulations because genetic diversity levels are sufficiently high in the breeding population at present.

The assertion that these MPBS subpopulations have sufficient genetic diversity must be qualified with two assumptions. First, it is assumed that rare alleles ($q < 10\%$) are not important to long-term genetic gain or adaptability. The crux of the forest conservation question lies in the importance of rare alleles and here there is little agreement. The most cogent argument is that rare alleles contribute little to overall fitness value, arise largely as unfavorable mutations and may be evolutionary relics (Lindgren and Gregorius 1976; Brown 1989). Conversely, loss of rare alleles may not compromise long-term flexibility to future pests, climate changes, and domestication goals but these losses could indicate similar non-directional loss in alleles at loci which control quantitative traits.

Second, genetic diversity estimates apply to the breeding population. The genetic diversity in breeding populations translates into sufficient genetic diversity in the reforestation effort (i.e., seed orchards or forest plantations) only if selections from each subpopulation are equally represented in seed orchards and in plantations.

The drop in average heterozygosity observed in selected germplasm resulted from the loss of low-frequency alleles and from random shifts in gene frequencies. The selected MPBS population included eight indigenous alleles absent in the natural stands. The most likely explanation is that the level of genetic diversity within the three natural stands could be under-represented relative to the selected population. If more natural stands had been sampled within the region, more low-frequency alleles would have been revealed. This may well be the case for the original first-generation selections which were sampled from individual stands, i.e., 34 first-generation selections represent 34 different stands. A less likely explanation is that the genetic composition of the natural stands has been altered since loblolly-pine domestication programs were established four decades ago. Single seed-tree colonization, fire and other disturbances can result in a series of bottlenecks in contemporary natural stands.

Genetic diversity for elite, third-generation selections

Third-generation elite selections had similar genetic diversity levels as other selected populations, regardless of generational differences. The elite selections still surpassed the range-wide levels reported for natural populations of radiata pine (*Pinus radiata*) (Moran et al. 1988).

Genetic diversity was higher than anticipated for ELITE, ELA and ELB subpopulations for two reasons. First, the variance-effective population size was quite high relative to census number. The resulting 2% change in average heterozygosity was too small to be detected with 19 loci (Nei and Roychoudhury 1974; Nei et al. 1977). Also, multiple alleles mitigated against a detectable decrease in average heterozygosity because there are more heterozygous combinations with multiple alleles.

Second, there have been no bottlenecks in the brief history of *P. taeda* as a domesticated population (see ELITE, ELA, ELB in Table 3). The ELITE subpopulation has had similar variance-effective numbers throughout domestication (29 and 41 in the first and second breeding cycles, respectively). The slight dip in the first cycle of breeding was due to the use of a four-male tester mating design in some breeding programs; the second cycle of breeding is higher because selections were used as equally as male and female parents. ELA and ELB represented smaller replicates of ELITE but the trend was the same for variance-effective numbers: they varied only slightly through each breeding cycle (Table 3).

The elite selections did have fewer alleles than natural populations (41 versus 52) which reflects a loss of alleles in the 1% frequency range (Fig. 1 B). Allele number tended to be a more sensitive indicator of population sizes than average heterozygosity and it is not clear how valuable these rare alleles are to the long-term success of the species.

Genetic differentiation among subpopulations within MPBS and HOPE

Genetic divergence, as measured by neutral loci, has not yet occurred among MPBS or HOPE subpopulations.

There was little genetic differentiation in any of the selected populations, including the exotic selections from southern Africa. One explanation is that these selections were drawn mostly from the center of the species' natural range. In future breeding cycles, the level of genetic diversity for MPBS is expected to increase. Rare alleles will be fixed through drift in some subpopulations but lost in others, resulting in an aggregate breeding population with a higher level of genetic diversity than a single, panmictic population of comparable size. It may be possible to hasten this through wide crossing or recurrent backcrossing to exotic relatives to capture low-frequency novel alleles.

The estimates of genetic divergence based on neutral loci are conservative for MPBS but possibly overestimated for HOPE. If both neutral loci and loci under selection were available, MPBS would be expected to have higher levels of genetic diversity than either HOPE, the elite population alone, or natural populations. In MPBS, the geographic origins and selection goals differ among subpopulations. Along the same lines, HOPE would have lower genetic diversity because all selections in both the elite and main subpopulations have a common selection goal and a common geographic origin.

We have assumed that the alleles at the 19 loci are effectively neutral and that selection applied to these populations should not affect allele frequencies. Any changes in genetic diversity observed between the selected populations were considered to be the result of genetic drift (i.e., fluctuations in population sizes). Our results appear to be consistent with this observation. For selected loblolly pine, only low-frequency alleles were lost, average heterozygosity declined within a range predicted from neutral allele theory, and genotypic proportions were generally within Hardy-Weinberg expectations. This study does not attempt to address the question of how much genetic diversity in adaptive traits has been lost due to directional selection in domestication programs. Future studies must await markers which are tightly linked to quantitative trait loci (QTLs) (Williams and Neale 1992; Bachmann 1994; Groover et al. 1994) which are consistent enough to be used to track loci under selection in subdivided breeding populations.

References

Bachmann K (1994) Molecular markers in plant ecology. New Phytol 126:403–418

Barnes RD (1984) A multiple population breeding strategy for Zimbabwe. In: Barnes RD, Gibson GL (eds) Proceedings IUFRO provenance and genetic improvement strategies in tropical forest trees. Commonwealth Forestry Institute, Oxford UK

Bramel-Cox PJ, Cox TS (1988) Use of wild germplasm in sorghum improvement. 43rd Annual Corn and Sorghum Conference 43: 13–26

Brown AHD (1989) Core collections: a practical approach to genetic resources management. Genome 31:818–824

Burdon RD, Namkoong G (1983) Short note: multiple populations and sublines. Silvae Genet 32:221–222

Cockerham CC, Weir BS (1993) Estimation of gene flow from F-statistics. Evolution 47:855–863

Cotterill PP (1984) A plan for breeding radiata pine. Silvae Genet 33: 84–90

Crow JF, Kimura M (1970) An introduction to population genetics theory. Harper and Row, New York

Eriksson G, Namkoong G, Roberds JH (1994) Dynamic gene conservation for uncertain futures. For Ecol Management 62: 15–37

Groover AT, Devey M, Fiddler T, Lee J, Megraw R, Mitchell-Olds T, Sherman B, Vujcic S, Williams CG, Neale DB (1994) Identification of quantitative trait loci influencing wood specific gravity in an outbred pedigree of loblolly pine. Genetics 138(4): 1293–1300

Hamrick JL, Godt MJW, Sherman-Broyles SL (1992) Factors affecting levels of genetic diversity in woody plant species. New Forests 6:95–124

Kannenberg LW (1984) Utilization of genetic diversity in crop breeding. In: Yeatman CW, Kafton D, Wilkes G (eds) Plant genetic resources: a conservation imperative, AAAS Selected Symposium 87th edn. Westview Press, Boulder, Colorado

Kidd G (1993) Analyzing the U.S. corn-genetics business. Biotechnology 11(9):980

Lewis PO, Whitkus R (1989) GENESTAT for microcomputers. Am Soc Plant Taxon Newslett 2:15–16

Lindgren D, Gregorius H (1976) Inbreeding and coancestry. In: Proceedings, IUFRO Joint Meeting on Advanced Generation Breeding, Bordeaux France, pp 49–72

Lowe WJ, van Buijtenen JP (1986) The development of a sublining system in an operational tree improvement program. In: Proceedings IUFRO Conference on Breeding theory, Progeny testing and Seed Orchards, Williamsburg, Virginia, pp 98–106

Mahalovich MF (1989) Modelling positive assortative mating and elite populations in recurrent selection programs for general combining ability. PhD dissertation, NC State University, Raleigh, North Carolina

McKeand SE, Bridgwater FE (1992) Third-generation breeding strategy for the North Carolina State University-Industry cooperative tree improvement program. In: Proceedings, IUFRO Conference on Breeding Tropical Trees, pp 234–240

McKeand SE, Li B, Hatcher A, Weir RJ (1988) Stability parameter estimates for sstem volume for loblolly pine families growing in regions in southeastern U.S. For Sci 38(1):10–17

Mitton JB, Linhart YB, Sturgeon KB, Hamrick JL (1979) Allozyme polymorphism detected in mature needle tissue of ponderosa pine. J Hered 70:86–89

Moran GF, Bell JC, Eldridge KG (1988) The genetic structure and genetic conservation of the five natural populations of *Pinus radiata*. Can J For Res 18:506–514

Namkoong G (1984) A control concept of gene conservation. Silvae Genet 33:160–163

Nei M (1972) Genetic distance between populations. Am Nat 106: 282–292

Nei M (1973) Analysis of gene diversity in subdivided populations. Proc Natl 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 (1986) Definition and estimation of fixation indices. Evolution 40:643–645

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

Nei M, Roychoudhury AK (1974) Sampling variances of heterozygosity and genetic distance. Genetics 76:379–390

Nei M, Maruyama T, Chakraborty R (1975) The bottleneck effect and genetic variability in populations. Evolution 29:1–10

Soltis DE, Haufler CH, Darrow DC, Gastony GC (1983) Starch gel electrophoresis of ferns: a compilation of grinding buffers, gel and electrode buffers and staining schedules. Am Fern Jour 73:9–27

Weir BS (1990) Genetic data analyses. Sinauer, Sunderland, Massachusetts

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

Williams CG, Lambeth CC (1993) Experimental elite population using *Pinus taeda* L. In: Proceedings, IUFRO Conference on Breeding Tropical Trees, Cali Colombia, pp 223–233
Williams CG, Neale DB (1992) Conifer wood quality and marker-assisted selection: a case study. Can J For Res 22:1009–1017
Workman PL, Niswander JD (1970) Population studies on southwestern Indian tribes. II. Local genetic differentiation in the Papago. Am J Hum Genet 22:24–49
Wright S (1931) Evolution in Mendelian populations. Genetics 16: 97–159
Wright S (1943) Isolation by distance. Genetics 28:114–138
Wright S (1965) The interpretation of population structure by F-statistics with special regard to systems of mating. Evolution 19:355–420
Wright S (1977) Evolution and the genetics of populations, vol 3. Experimental results and evolutionary deductions. University of Chicago Press, Chicago