

# RAPDs and allozymes exhibit similar levels of diversity and differentiation among populations and races of Douglas-fir

NOTICE:  
This Material  
may be protected  
by copyright  
law.

JAN E. AAGAARD, KONSTANTIN V. KRUTOVSKII\*\* & STEVEN H. STRAUSS\*

Department of Forest Science, Oregon State University, Corvallis, OR 97331-7501, U.S.A.

Thirty-six nuclear-encoded RAPD loci and 20 allozyme loci were studied to compare levels of diversity and differentiation among populations and races of the widespread North American conifer, Douglas-fir [*Pseudotsuga menziesii* (Mirb.) Franco]. RAPD assays used diploid seed embryo DNA from 22 to 36 trees in each of six populations that sampled the three major races (two populations per race). A comparable allozyme data set for nearby populations was constructed from a published study. RAPDs of organelle origin were excluded by hybridization of blotted RAPD gels with chloroplast and mitochondrial DNA-enriched probes. RAPD and allozyme markers had similar levels of diversity within populations ( $H_s = 0.22 \pm 0.03$  and  $0.16 \pm 0.03$ , respectively) and differentiation among populations ( $G_{ST} = 0.34 \pm 0.07$  and  $0.29 \pm 0.07$ , respectively). When the allozyme data set was transformed into dominant, biallelic markers to study how RAPDs may bias diversity estimates, resampling studies showed that simulated  $H_s$  and  $H_T$  were reduced by half regardless of sample size. Because observed diversity for RAPDs was equivalent to, or higher than, that of allozymes, our simulations suggest that RAPD markers may contain substantially higher levels of inherent, but hidden, diversity. In contrast, the simulations showed that estimates of  $G_{ST}$  using RAPDs should not be significantly biased at the population sizes we employed.

**Keywords:** allozymes, DNA markers, dominance simulation, Douglas-fir, genetic diversity, RAPDs.

## Introduction

Allozyme markers have been used extensively to describe the genetic structure of populations. This has led to several useful generalizations about the levels and partitioning of diversity in relation to factors such as plant mating system, mode of pollination and geographical distribution (Hamrick & Godt, 1989). However, it is also clear that allozymes do not represent a random sample of genomes, and thus may bias some population genetic inferences. Allozymes reflect variability in protein coding sequences and may thus be selectively constrained in ways that noncoding regions are not (reviewed by Clegg, 1989); for example, it appears that balancing selection can act on allozymes (Altukhov, 1991; Karl & Avise, 1992; Raybould *et al.*, 1996), resulting in

overestimation of allelic similarity among populations compared to that at neutral loci (reviewed by Altukhov, 1990, 1991). Random DNA markers, however, should be virtually free of these influences.

RAPD (randomly amplified polymorphic DNA) markers (Williams *et al.*, 1990) have been increasingly employed for population studies in recent years. They are technically simpler than other DNA-based markers (e.g. RFLPs), facilitate studies of large numbers of loci, and are expected to provide a far more random sample of genomic DNA than do allozymes. Although there have been only a few studies comparing the markers directly, most qualitative comparisons have suggested that RAPDs and allozymes reveal similar patterns of genetic diversity, but that RAPDs tend to provide more diagnostic population-, race- or species-specific markers (e.g. Liu & Furnier, 1993; Peakall *et al.*, 1995). These characteristics can be important for population studies and for informing decisions about population conservation.

\*\*Permanent address: Laboratory of Population Genetics, N. I. Vavilov Institute of General Genetics, Russian Academy of Sciences, 117809 GSP-1, Moscow B-333, Russia.

\*Correspondence. E-mail: strauss@fsl.orst.edu

However, RAPDs also have significant limitations compared to allozymes. Most important is that the large majority of RAPD loci provide biallelic, dominant markers. These factors limit inferences from RAPDs, especially when studied in diploid tissues, and may bias some population genetic parameters when compared with codominant, multiallelic markers such as allozymes (Lynch & Milligan, 1994); for example, recent empirical studies of conifers (Isabel *et al.*, 1995; Szmidt *et al.*, 1996) have suggested that inferring RAPD allele frequencies from diploid tissues may result in overestimates of population differentiation when compared with allozyme markers.

To understand better the limitations and advantages of RAPDs compared to allozyme markers, and to give further insights into the structure of genetic variation in Douglas-fir [*Pseudotsuga menziesii* (Mirb.) Franco], RAPD and allozyme diversity were compared in a sample of populations and races. Douglas-fir is a long-lived, monoecious species and is the dominant forest tree throughout most of its range in western North America. Although they can readily hybridize (Rehfeldt, 1977), the coastal variety (var. *menziesii*) and the interior variety (var. *glauca*) are distinct, based on needle and cone morphology, growth characteristics (Rehfeldt, 1977) and allozymes (Li & Adams, 1989). Allozymes also distinguish two major races within the interior variety (north interior and south interior; Li & Adams, 1989). The two varieties have a long evolutionary history, appearing to be distinct from at least the Miocene (13 Myr ago; reviewed by Critchfield, 1984). Similar to other studies, we find that RAPD phenotypes tend to be more strongly differentiated at both the population and racial levels than are allozymes. In addition, despite the tendency for estimates of intrapopulation diversity to be strongly deflated by dominance and biallelism, we find that RAPD loci have similar or higher levels of polymorphism than do allozymes.

## Materials and methods

### Douglas-fir populations

Seeds representing two populations from each of the three races of Douglas-fir (coastal, north interior, and south interior) were selected from existing seed collections for RAPD analysis. Coastal and north interior populations (Fig. 1) were selected from single-tree seed collections stored at the USDA Forest Service Pacific Northwest Research Station (Corvallis, OR). Coastal seed collections were

originally made from natural populations in western Washington (Smith Creek; 34 maternal trees) and western Oregon (Lacomb; 22 maternal trees). Maternal trees were separated by ~50 to 100 m. Each of two north interior seed collections was originally made from three contiguous natural stands

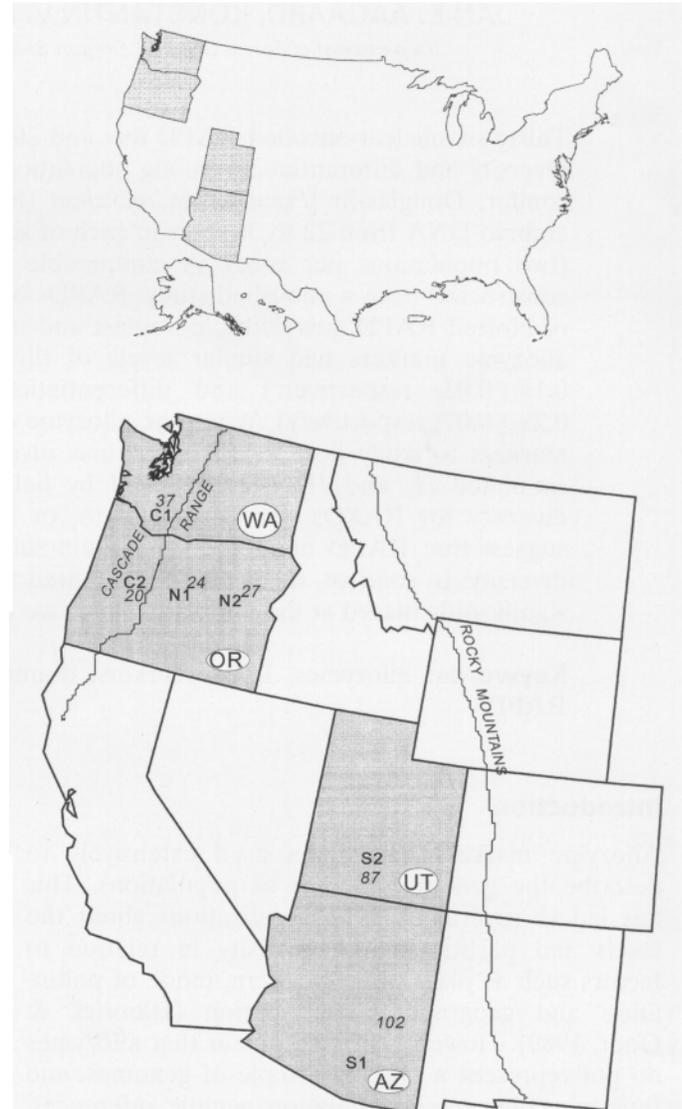


Fig. 1 Locations of Douglas-fir populations used for the RAPD and allozyme studies. The six populations of Douglas-fir used in the RAPD study from the coastal race (C1, Smith Creek; C2, Lacomb), north interior race (N1, Wildcat Mountain; N2, Baldy Mountain) and south interior race (S1, Coronado; S2, Dixie) were geographically matched with six populations from Li & Adams's (1989) allozyme study (C1, 37; C2, 20; N1, 24; N2, 27; S1, 102; S2, 87). All populations were located within the states of Arizona (AZ), Oregon (OR), Utah (UT) and Washington (WA).

(five to 13 maternal trees per stand) separated by 300 m of elevation in central (Wildcat Mountain; 36 maternal trees in total) and east-central Oregon (Baldy Mountain; 30 maternal trees in total). Trees sampled within stands were separated by ~50 to 100 m.

South interior populations were selected from bulked seed collections stored at the USDA Forest Service Lucky Peak Nursery (Boise, ID). The southern Arizona bulked seed collection (Coronado) was made from a single stand of 30 maternal trees, and the southern Utah bulked seed collection (Dixie) was made from two contiguous stands of 22 maternal trees each; 30 and 29 seeds were analysed from the Coronado and Dixie collections, respectively.

#### *DNA extraction, RAPD amplification, and scoring of RAPD markers*

Total genomic DNA was extracted from diploid seed embryos using a modified CTAB protocol described previously (Aagaard *et al.*, 1995). A total of nine RAPD primers identified in a related study (Aagaard, 1997: UBC 234, UBC 328, UBC 330, UBC 336, UBC 411, UBC 428, UBC 460, UBC 504 and UBC 570) were used to amplify RAPD fragments using a standard RAPD protocol (Aagaard *et al.*, 1995). Blots of RAPD gels were hybridized with enriched mitochondrial and chloroplast DNAs to allow identification of RAPD fragments of organellar origin (Aagaard, 1997); fragments of mitochondrial origin are frequent among RAPD profiles of Douglas-fir (Aagaard *et al.*, 1995).

Only those RAPD fragments which were distinct and highly reproducible over multiple (at least four) amplifications were scored. All RAPD fragments of putative organellar origin were excluded from analyses. Allele frequencies were calculated from null homozygote frequencies assuming panmixia and corrected for dominance according to Lynch & Milligan (1994, equation 2a). Like most conifers, Douglas-fir exhibits high outcrossing rates and genotype frequencies of seed embryos that closely approximate Hardy-Weinberg proportions (Neale & Adams, 1985); even substantial deviations from Hardy-Weinberg frequencies (e.g.  $F_{IS} = 0.20$ ) would have very little effect on estimates of diversity parameters for our RAPD data set (Aagaard, 1997).

#### *Data analysis*

Nei's (1973) gene diversity statistics were calculated for RAPDs and were corrected for small sample size

(Nei, 1978) and small number of populations (Nei, 1986) using the computer program GENESTAT-PC 3.3 (Lewis, 1994). A hierarchical analysis employed only the small sample size correction in order to allow for the decomposition of gene diversity into components attributable to population and race differentiation (Nei, 1986). Calculations of population diversity ( $H_S$  and  $H_T$ ) were calculated both with and without completely monomorphic loci included (99% criterion). Standard errors for  $G_{ST}$  values were calculated by jackknifing over loci (Weir & Cockerham, 1984). Nei's (1972) genetic distance was calculated between pairs of populations.

We also analysed phenotypic differentiation of RAPD fragments with the Analysis of Molecular Variance (AMOVA) approach of Excoffier *et al.* (1992). Simple dissimilarity [ $1 - (\text{the fraction of matches})$ ] was employed as a measure of genetic distance between individuals (e.g. Apostol *et al.*, 1993) using RAPDPLOT 2.4 (Black, 1996) to calculate the fraction of matches ( $M$ ):

$$M = NAB/NT,$$

where  $NAB$  is the total number of matches in individuals A and B (i.e. both fragments absent or present) and  $NT$  is the total number of loci scored in the overall study. The computer program winamova (Excoffier, 1993) was used for all AMOVA calculations. WINAMOVA was also used for permutation tests of differentiation among nearby stands within populations of the north interior race.

In order to compare RAPD diversity and differentiation with allozymes, data from Li & Adams's (1989) allozyme study of Douglas-fir were reanalysed. The six populations sampled with RAPDs in this study were geographically matched to the closest populations from Li & Adams's study (Fig. 1); these paired populations are referred to below as if they were the same populations sampled with both RAPDs and allozymes. Statistical comparisons of mean within-population diversity ( $H_S$ ) and total diversity ( $H_T$ ) calculated from RAPD and allozyme data were made using a nonparametric rank sum test (Mann-Whitney test) applied to single-locus estimates of each parameter. This nonparametric test was used in order to avoid assumptions of equal variance for RAPDs and allozymes. In addition, a G-test for goodness of fit for multiple classes (Sokal & Rohlf, 1981) was used to compare the distribution of RAPD and allozyme diversity ( $H_S$  and  $H_T$ ). For the G-test, diversities at single loci were rounded and pooled into one of six classes (0.0, 0.1, ..., 0.5). Because diversity for RAPDs cannot exceed 0.50, allozyme loci with diversities above 0.50 were pooled

into the 0.5 class for the G-test. Comparisons of population differentiation ( $G_{ST}$ ) for RAPDs and allozymes were also made with the rank sum test using single-locus estimates of  $G_{ST}$ , and the jackknifed variance was used to compare  $G_{ST}$ s using a two-sample t-test.

#### *Assessment of sample size*

In order to assess the adequacy of the sample sizes (loci and individuals) in the RAPD study, a resampling approach was used (bootstrapping; Efron & Tibshirani, 1991) similar to that used elsewhere to study the adequacy of genotype sampling in germplasm and population collections (e.g. Thormann et al., 1994; Bonnin et al., 1996). We studied the effects of bootstrapping (with replacement) over loci and over individuals separately to provide estimates of the variance in population genetic parameters ( $H_S$ ,  $H_T$  and  $G_{ST}$ ) associated with different sample sizes of each. Sample size varied between five and 50 in increments of five in order to bracket the sample sizes used in the RAPD study (mean = 30 individuals per population and 31 loci). The variance at each increment was calculated from the variance among 250 bootstrap populations.

#### *Simulating dominance and biallelism of markers*

To help understand the effects of RAPD dominance and biallelism on the estimates, and to facilitate comparisons between RAPDs and allozymes, dominance and biallelism were simulated in the allozyme dataset using a computer model. From allozyme allele frequencies in the six Douglas-fir populations, six populations of 1000 individuals were generated each with genotypes that maintained the original allele frequencies within populations. A total of 250 subpopulations of  $n$  individuals were drawn, with replacement from each of the six populations, and population genetic parameters ( $H_S$ ,  $H_T$  and  $G_{ST}$ ) were calculated for each set of 250 subpopulations in three ways. First, for a codominant dataset, calculations were made from all alleles present in all the subpopulations. Secondly, the same subpopulations were used to construct a dominant, biallelic dataset intended to imitate the methods used in scoring and calculation of RAPD allele frequencies. In this case, one allele was randomly selected as dominant and the rest were considered to be recessive; the synthetic null allele frequency was then calculated from null homozygote frequencies assuming Hardy-Weinberg equilibrium, and diversity and differentiation parameters calculated as for the original RAPD

data. Finally, null allele frequencies were corrected for dominance using Lynch & Milligan's data (1994, equation 2a) as described above. The number of individuals within subsamples ( $n$ ) was varied from 10 to 200 to bracket the range of sample sizes that might reasonably be employed in population studies.

#### *Results*

##### *Pooling of stands and adequacy of RAPD sample size*

No evidence of RAPD differentiation was found among the nearby stands sampled in either of the north interior populations based on aMova permutation analysis ( $P = 1.00$ ); stands were thus pooled for further analysis.

Variance in population genetic parameters ( $H_S$ ,  $H_T$  and  $G_{ST}$ ) did not decrease substantially when sample sizes larger than 31 loci and 30 individuals (the actual sample sizes used in the RAPD study reported here) were generated from bootstrapping (Fig. 2). Variance decreased sharply when sample sizes increased from five to 15 loci or five to 20 individuals, and above these levels variance decreased only slowly with increasing sample size. Variances from bootstrapping over loci were approximately fivefold higher than variances from bootstrapping over individuals throughout the range of sample sizes studied.

##### *RAPD diversity and differentiation*

Of the 36 RAPD fragments scored, five fragments (14%) were completely monomorphic. Based on the remaining 31 RAPD markers, two of which were race-specific, average within-population gene diversity ( $H_S$ ) was 0.22 and total diversity ( $H_T$ ) was 0.32 (Table 1). When monomorphic loci were included,  $H_S$  and  $H_T$  decreased to 0.19 and 0.27, respectively. Diversities in coastal and north interior populations were similar, and were consistently higher than in south interior populations. Differentiation of populations within races showed an inverse trend to that for intrapopulation diversity.  $G_{ST}$  and genetic distances were lower among both coastal and north interior populations than among south interior populations (Table 2). The majority of differentiation was caused by variation among races ( $G_{ST}$  for races = 0.25), which was approximately fivefold higher than among populations within races ( $G_{ST} = 0.05$ ). Results from AMOVA gave a nearly identical pattern of population and racial differentiation; 8% ( $P = 0.002$ ) of the RAPD variation was

distributed among populations within races and 29% ( $P = 0.004$ ) was distributed among races.

#### Allozyme diversity and differentiation

Of 20 allozyme loci, two (10%) were completely monomorphic. There were no diagnostic allozyme markers, although several loci had low frequency alleles restricted to a single population. Based on the 18 polymorphic allozyme loci, average diversity within populations ( $H_s$ ) was 0.16 and total diversity ( $H_T$ ) was 0.22 (Table 1). As with RAPDs, diversity for allozymes was highest in coastal and north interior populations and lowest in south interior populations. Differentiation among populations for allozymes was also similar to that found for RAPDs (Table 2). Differentiation was considerably higher among south interior populations than among the north interior or coastal populations, and population differentiation was principally caused by variation

among races ( $G_{ST} = 0.22$ ), rather than variation among populations within races ( $G_{ST} = 0.05$ ).

#### Comparison of diversity and differentiation between RAPDs and allozymes

In most cases, diversity and differentiation for RAPDs were higher than for allozymes (Tables 1 and 2). However, based on the rank sum test, there was only weak statistical support for higher total RAPD diversity ( $H_T$ :  $P = 0.06$ ), and the higher average within-population RAPD diversity was not statistically supported ( $H_s$ :  $P = 0.21$ ). Similarly, based on single-locus estimates of overall  $G_{ST}$  (differentiation among both populations and races) used in the rank sum test, differentiation for RAPDs was not significantly different from allozyme differentiation ( $P = 0.24$ ), nor were  $G_{ST}$  s different using jackknifed variances ( $P > 0.30$ ).

The G-test did not show significant differences in the distribution of within-population diversity among RAPD and allozyme loci ( $P = 0.45$ ; Fig. 3). Allozymes, however, tended to have a larger proportion of their loci in the low-diversity  $H_s$  classes (0.1 and 0.2) than RAPDs (45% and 25% vs. 28% and 14%, respectively). Similarly, allozymes had a larger proportion of loci in the 0.1 and 0.2 classes for total diversity ( $H_T$ ) than did RAPDs (45% and 20% vs. 19% and 8%, respectively), but the overall distributions were not significantly different ( $P = 0.32$ ).

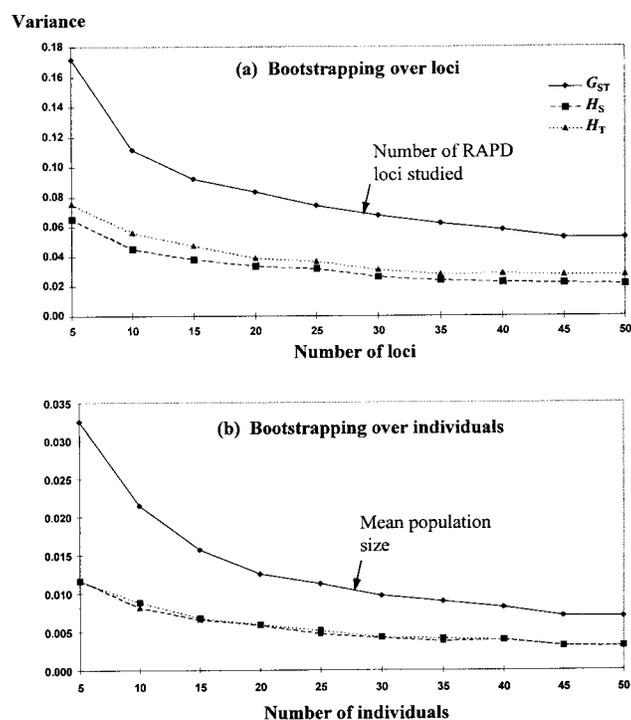


Fig. 2 Bootstrapping over loci and individuals to determine adequacy of sample sizes. The variance in population genetic parameters (within-population gene diversity,  $H_s$ ; total gene diversity,  $H_T$ ; and population differentiation,  $G_{ST}$ ) changes depending on the number of (a) loci and (b) individuals sampled. Variances were calculated from 250 bootstrap samples, selected with replacement from a dataset of six populations (mean of 30 individuals per population) and 31 loci. Monomorphic loci were excluded.

#### Dominance simulations

When dominance and biallelism were simulated based on the allozyme dataset, diversity within populations ( $H_s$ ) and total diversity ( $H_T$ ) were found to be approximately half the value found for codominant allozymes over the full range of sample sizes investigated (Fig. 4).  $H_s$  for the codominant allozymes was nearly constant, but  $H_s$  for the simulated dominant markers (corrected according to Lynch & Milligan, 1994) increased initially with increasing sample size until ~30 individuals were sampled within each population.  $H_T$  changed little over the sample sizes studied for either codominant allozymes or simulated dominant markers.

Despite lower diversity for the simulated dominant markers, differentiation among populations ( $G_{ST}$ ) was greater for dominant markers than codominant allozymes when sample sizes were smaller than 30 individuals per population (Fig. 4). When 30 individuals were sampled,  $G_{ST}$  was nearly the same for both dominant and codominant markers ( $G_{ST} = 0.31$ ), although the variance for dominant

**Table 1** Expected within-population ( $H_c$  and  $H_s$ ) and total ( $H_T$ ) gene diversities and their standard errors (SE) calculated among six populations from three races of Douglas-fir for RAPD and allozyme markers based on polymorphic loci (99% criterion; 31 loci and 18 loci, respectively), and on all loci including monomorphic ones (36 loci and 20 loci, respectively)

Population (sample size)	RAPDs		Allozymes	
	Polymorphic loci	All loci	Polymorphic loci	All loci
$H_c \pm SE$				
Coastal race				
Smith Creek (34)	0.27 ± 0.03	0.23 ± 0.03	0.20 ± 0.05	0.18 ± 0.04
Lacomb (22)	0.25 ± 0.04	0.21 ± 0.04	0.20 ± 0.05	0.19 ± 0.05
North interior race				
Wildcat Mountain (36)	0.27 ± 0.04	0.23 ± 0.04	0.18 ± 0.05	0.16 ± 0.05
Baldy Mountain (30)	0.24 ± 0.04	0.20 ± 0.04	0.22 ± 0.05	0.20 ± 0.05
South interior race				
Coronado (30)	0.15 ± 0.04	0.13 ± 0.03	0.08 ± 0.03	0.07 ± 0.03
Dixie (29)	0.15 ± 0.03	0.13 ± 0.03	0.10 ± 0.04	0.09 ± 0.04
Means				
$H_s$	0.22 ± 0.03	0.19 ± 0.03	0.16 ± 0.03	0.15 ± 0.03
$H_T$	0.32 ± 0.03	0.27 ± 0.03	0.22 ± 0.05	0.20 ± 0.05

**Table 2** Genetic differentiation (Nei's  $G_{ST}$  and standard genetic distance,  $D$ ) in Douglas-fir based on RAPDs (31 loci) and allozymes (18 loci)

	RAPDs		Allozymes	
	$G_{ST} \pm SE$	$D$	$G_{ST} \pm SE$	$D$
Genetic differentiation among populations within races†				
Coastal	0.05 ± 0.02	0.02	0.10 ± 0.06	0.02
North interior	0.07 ± 0.03	0.03	0.06 ± 0.04	0.01
South interior	0.25 ± 0.08	0.06	0.17 ± 0.09	0.02
Without regard to races	0.34 ± 0.07	0.16	0.29 ± 0.07	0.08
Proportion of total genetic differentiation‡ attributable to variation				
Among populations within races	0.05 ± 0.01		0.05 ± 0.02	
Among races	0.25 ± 0.07		0.22 ± 0.06	

†Calculations of  $G_{ST}$  employed corrections for both small sample size (Nei, 1978) and small number of populations (Nei, 1986).

‡Hierarchical analyses of the proportion of differentiation resulting from differences among races and populations employed only the correction for small sample size, as specified in Nei (1986).

markers was nearly threefold greater ( $SD = 0.033$  and  $0.012$ , respectively).  $G_{ST}$  values for dominant markers above sample sizes of 30 were similar, although slightly lower, than for codominant allozymes. Lynch and Milligan's correction had very little effect on  $H_S$ ,  $H_T$  or  $G_{ST}$  values compared with uncorrected dominant markers (data not shown); for example, at a sample size of 30, uncorrected values were only 1% different from the corrected values for either  $H_S$  or  $H_T$ , and 2% different for  $G_{ST}$ .

## Discussion

### RAPDs vs. allozymes

Differences between RAPDs and allozymes were found to be modest and not statistically significant for most population genetic parameters. They exhibited comparable levels of intrapopulation diversity and population differentiation. These results are concordant with most aspects of earlier studies. Liu & Furnier (1993) found that diversities for allozymes

and RAPDs were similar within populations of trembling aspen (*Populus tremuloides*;  $H_e = 0.32 \pm 0.06$  and  $0.30 \pm 0.11$ , respectively), and slightly higher for RAPDs than for allozymes within populations of bigtooth aspen (*Populus grandidentata*;  $H_e = 0.35 \pm 0.08$  and  $0.28 \pm 0.11$ , respectively). Peakall *et al.* (1995) compared RAPD and allozyme diversity in populations of buffalograss (*Buchloe dactyloides*) using an AMOVA approach to quantify the distribution of variation in both markers. They

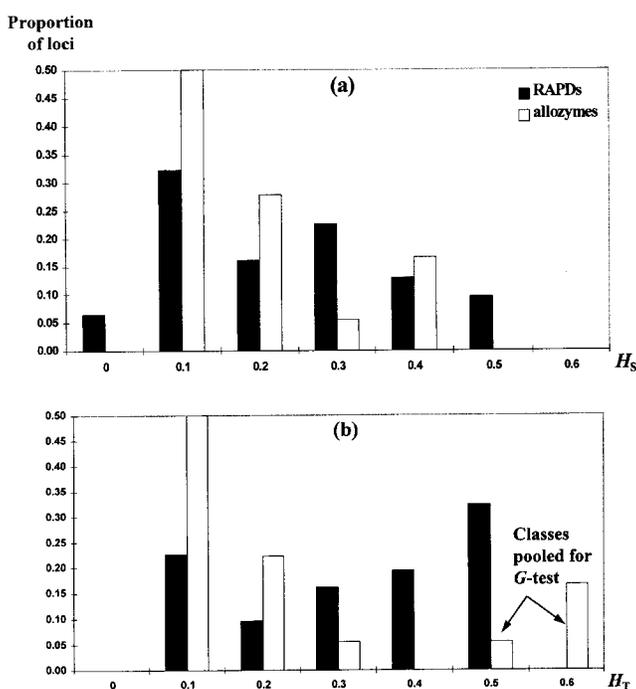


Fig. 3 Gene diversity based on RAPDs and allozymes in Douglas-fir. (a) Distribution of average, single-locus, within-population gene diversities ( $H_S$ ), and (b) total gene diversities ( $H_T$ ) for RAPDs (31 loci) and allozymes (18 loci). Gene diversities were calculated among all six populations from the three races of Douglas-fir. For the G-test comparing the distribution of  $H_T$  for RAPDs and allozymes (see text), allozyme loci in the 0.5 class were pooled together with the 0.6 class.

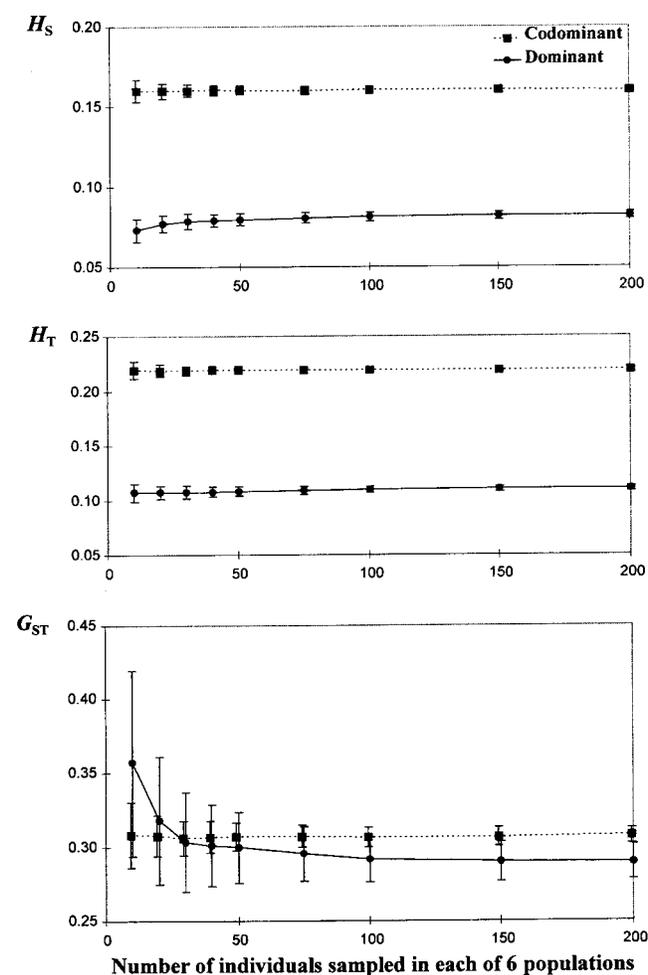


Fig. 4 Levels of diversity and differentiation for codominant, multiallelic allozymes vs. biallelic, dominant markers simulated from allozyme data, vary as a function of sample size. Values of within-population gene diversity ( $H_S$ ), total gene diversity ( $H_T$ ), and population differentiation ( $G_{ST}$ ) were plotted over a range of population sizes varying between five and 200, and are corrected for dominance using Lynch & Milligan's data (1994, equation 2a). Standard deviations (error bars) were calculated among 250 samples at each level of sampling and represent the variance resulting from sampling of individuals from each master population of 1000 individuals.

found that RAPD differentiation was slightly higher than allozyme differentiation among populations ( $D_{ST} = 0.61$  and  $0.55$ , respectively), and a smaller proportion of the total RAPD variation was located within populations of buffalograss than for allozymes (32% vs. 45%, respectively).

Several studies have found that comparisons between RAPDs and allozymes can be influenced by dominance of RAPDs when analysed in diploid tissue. Isabel et al. (1995) found nearly identical heterozygosity for RAPDs and allozymes within populations of black spruce (*Picea mariana*) using RAPD allele frequencies that were inferred from haploid megagametophyte tissues of tree seeds ( $H_o = 0.35$  and  $0.34$ , respectively). RAPD differentiation was very low in general, but more than threefold higher than for allozymes ( $G_{ST} = 0.03$  and  $0.01$ , respectively). However, Isabel et al. (1995) estimated that computing RAPD allele frequencies indirectly from RAPD phenotypes observed in diploid tissues of the same maternal trees could have caused as much as a twofold downward bias in  $H_S$  and a 1.8-fold downward bias in  $H_T$ , whereas  $G_{ST}$  could have been upwardly biased severalfold.

Szmidt et al. (1996) also used haploid megagametophyte tissues to study two populations of Scots pine (*Pinus sylvestris*). Based on RAPD allele frequencies estimated from analysis of haploid megagametophytes, one population had nearly identical RAPD and allozyme diversity ( $H_e = 0.30 \pm 0.04$  and  $0.28 \pm 0.05$ , respectively), but the second population had greater RAPD diversity ( $H_e = 0.36 \pm 0.03$  and  $0.26 \pm 0.05$ ) for RAPDs and allozymes, respectively. Szmidt et al. (1996) estimated that indirect calculations of RAPD allele frequencies from RAPD phenotypes in diploid tissues would have shown downwards biased RAPD diversity within both populations ( $H_S$  from diploid tissues was estimated to be  $0.19$  and  $0.22$ , respectively). Similarly, although  $G_{ST}$  values were nearly identical for RAPDs and allozymes when RAPD allele frequencies were inferred directly ( $0.02$  for both markers), calculation of RAPD allele frequencies from inferred diploid phenotypes of trees caused a modest overestimate of RAPD differentiation ( $G_{ST} = 0.06$ ). These studies show that the dominance of RAPD markers can influence comparisons between RAPDs and allozymes when studying diploid tissues.

#### *Effects of RAPD dominance*

Although allele frequencies among populations of Douglas-fir were calculated indirectly using diploid

tissues and Hardy-Weinberg assumptions, the large discrepancy between RAPDs and allozymes predicted by Isabel et al. (1995) and Szmidt et al. (1996) was not found. Assuming diversity and differentiation for RAPDs and allozymes are indeed similar in Douglas-fir, two reasons may explain the lack of expected bias in this study. First, 50% larger sample sizes were used than employed in the Szmidt et al. (1996) study, increasing the likelihood of detecting null homozygotes and thus reducing the expected bias in  $G_{ST}$  (Fig. 2). The importance of sample size for estimating allele frequencies with dominant markers is widely known and has been discussed elsewhere (Lynch & Milligan, 1994).

Secondly, because of the lower proportion of loci with low frequency null alleles in the present study compared with that in Szmidt et al. (1996), the diploid RAPD data reported here should be less subject to bias from dominance. This bias results predominantly from low frequency null alleles (Lynch & Milligan, 1994); when amplifying RAPDs from diploid tissues, at least one null homozygote must be included in a population sample in order to detect polymorphism at that locus. The potential bias can be described by summing binomial distributions (Weir, 1996); for example, with a null allele frequency of  $0.22$  (the minimal detectable null allele frequency in Szmidt et al.'s (1996) study using diploid tissues), the probability of including at least one null homozygote in a sample of size 20 from a randomly mating population is only  $0.63$ . In contrast, when the null allele frequency is  $0.67$  (i.e. the null allele is twice as common as the dominant allele), the probability of including at least one null homozygote in a sample size of 20 is  $0.99$ . For most RAPD markers (80%) in Szmidt et al.'s (1996) study, the dominant allele was the most common one, whereas only about half (56%) of the RAPD loci in the present study had the dominant allele as the most common one.

#### *Simulations of dominance and biallelism*

In addition to dominance, comparisons between RAPD and allozyme data are complicated by the biallelic nature of RAPDs. The RAPD null allele is likely to be a synthetic allelic class, potentially resulting from many types of mutations (alleles) which all confer the null RAPD phenotype (i.e. lack of amplification of a fragment of specified size). In contrast, the dominant phenotype is likely to result from a single or much smaller number of alleles. When the combined effects of RAPD dominance and biallelism were simulated using the Douglas-fir

allozyme data, it was found that diversity within populations ( $H_s$ ) and total gene diversity ( $H_T$ ) were half the value of multiallelic, codominant allozymes regardless of sample size (Fig. 4). Thus, although large sample sizes (or haploid tissues) alleviate the problem of determining null homozygote frequencies with dominant markers, it is likely that the synthetic null allele can continue to mask a large amount of underlying polymorphism. Based on this simulation model, it would be expected that allozyme diversity would be twice as high as diversity calculated from RAPD phenotypes, assuming that RAPDs had an identical level of inherent polymorphism to that observed at allozyme loci. Doubling the estimates of Douglas-fir RAPD diversity to enable a more reasonable comparison to allozymes gives values of 0.44 and 0.64 for  $H_s$  and  $H_T$ , respectively. These are well above the 95% confidence intervals of the allozyme data (upper limits of 0.22 and 0.32, respectively).

The high diversity that was found for RAPDs, despite their inability to identify as many alleles per locus as allozymes, might result from an inherently higher rate of detectable mutations, and weaker degree of selective constraint, at RAPD compared to allozyme loci. RAPDs are presumed to result primarily from amplification of the noncoding DNA that comprises the bulk of eukaryotic genomes. Gymnosperms, including Douglas-fir, have particularly large genomes (O'Brien *et al.*, 1996). Nucleotide substitution rates for noncoding sequences are generally more than twice the rate for coding sequences, and are similar to synonymous substitution rates (reviewed by Nei, 1987; Moriyama & Powell, 1996). Similarly, noncoding regions are likely to contain higher proportions of repetitive DNA, which host various forms of replication-associated mutation and non-Mendelian recombination (e.g. unequal crossing-over: reviewed by Charlesworth *et al.*, 1994), which may cause much greater polymorphism within populations compared to coding loci (e.g. Scribner *et al.*, 1994). RAPD markers are sometimes amplified from regions of repetitive DNA (e.g. Williams *et al.*, 1990), and thus could reflect these high rates of mutation. Finally, protein encoding loci are known to be under strong selective constraint, as evidenced by the large differences typically observed between synonymous and nonsynonymous substitution rates (e.g. Moriyama & Powell, 1996). RAPDs should be virtually free of this strong constraint. The net result of these factors would be higher average levels of polymorphism at RAPD loci compared to allozyme loci.

## Acknowledgements

We wish to thank Jen Sceppa for writing the resampling programs, Joe Felsenstein for reviewing the concept and results of our resampling studies, Mike Lynch for reviewing resampling results and commenting on the effects of RAPD dominance, Svetlana Erofeeva for timely work on the simulation program, Frank Sorensen and Dick Jeffers of the USDA Forest Service for providing access to seed collections, and Sheila Vollmer for providing laboratory guidance. J.E.A. was supported in part by NSF grants DEB 9300083 and BSR 895702 to S.H.S.

## References

- AAGAARD, J. E. 1997. *Genetic Diversity and Differentiation in Douglas-fir from RAPD Markers of Nuclear and Mitochondrial Origin*. M.Sc. Thesis, Oregon State University, Corvallis, OR.
- AAGAARD, J. E., VOLLMER, S. S., SORENSEN, F. C. AND STRAUSS, S. H. 1995. Mitochondrial DNA products among RAPD profiles are frequent and strongly differentiated between races of Douglas-fir. *Mol. Ecol.*, **4**, 441-447.
- ALTUKHOV, Y. P. 1990. *Population Genetics: Diversity and Stability*. Harwood Academic Publishers, London.
- ALTUKHOV, Y. P. 1991. The role of balancing selection and overdominance in maintaining allozyme polymorphism. *Genetica*, **85**, 79-90.
- APOSTOL, B. L., BLACK, W. C., IV, MILLER, B. R., REITER, P. AND BEATY, J. B. 1993. Estimation of family numbers at an oviposition site using RAPD-PCR markers: applications to the mosquito *Aedes aegypti*. *Theor. Appl. Genet.*, **86**, 991-1000.
- BLACK, W. C., IV. 1996. *RAPD PLOT 2.4*. Department of Microbiology, Colorado State University, Fort Collins, CO.
- BONNIN, L., HUGUET, T., GHERARDI, M., PROSPERI, J. AND OLIVIERI, L. 1996. High level of polymorphism and spatial structure in a selfing plant species, *Medicago truncatula* (Leguminosae), shown using RAPD markers. *Am. J. Bot.*, **83**, 843-855.
- CHARLESWORTH, B., SNTÉGOWSKI, P. AND STEPHAN, W. 1994. The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature*, **371**, 215-220.
- CLEGG, M. T. 1989. Molecular diversity in plant populations. In: Brown, A. H. D., Clegg, M. T., Kahler, A. L. and Weir, B. S. (eds) *Plant Population Genetics, Breeding, and Genetic Resources*, pp. 98-115. Sinauer Associates, Sunderland, MA.
- CRITCHFIELD, W. B. 1984. Impact of the Pleistocene on the genetic structure of North American conifers. In: Lanner, R. M. (ed.) *Proceedings of the 8th North American Forest Biology Workshop*, pp. 70-118. Utah State University, Logan, UT.
- EFRON, B. AND TIBSHIRANI, R. 1991. Statistical data analysis in the computer age. *Science*, **253**, 390-395.

- EXCOFFIER, L. 1993. WINamova. Genetics and Biometry Laboratory, University of Geneva, Carouge, Switzerland.
- EXCOFFIER, L., SMOUSE, P. E. AND QUATTRO, J. M. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, *131*, 479-491.
- HAMRICK, J. L. AND GODT, M. J. 1989. Allozyme diversity in plant species. In: Brown, A. H. D., Clegg, M. T., Kahler, A. L. and Weir, B. S. (eds) *Plant Population Genetics, Breeding, and Genetic Resources*, pp. 43-63. Sinauer Associates, Sunderland, MA.
- ISABEL, N., BEAULIEU, J. AND BOUSUET, J. 1995. Complete congruence between gene diversity estimates derived from genotypic data at enzyme and random amplified polymorphic DNA loci in black spruce. *Proc. Natl. Acad. Sci. U.S.A.*, *92*, 6369-6373.
- KARL, S. A. AND AVISE, J. C. 1992. Balancing selection at allozyme loci in oysters: implications from nuclear RFLPs. *Science*, *256*, 100-102.
- LEWIS, P. o. 1994. *GENESTAT-PC, a 3.3*. Department of Statistics, North Carolina State University, Raleigh, NC.
- LI, P. AND ADAMS, w. T. 1989. Range-wide patterns of allozyme variation in Douglas-fir (*Pseudotsuga menziesii*). *Can. J. Forest Res.*, *19*, 149-161.
- LIU, z. AND FuRNIER, G. R. 1993. Comparison of allozyme, RFLP, and RAPD markers for revealing genetic variation within and between trembling aspen and bigtooth aspen. *Theor. Appl. Genet.*, *87*, 97-105.
- LYNCH, M. AND MILLIGAN, B. G. 1994. Analysis of population genetic structure with RAPD markers. *Mol. Ecol.*, *3*, 91-99.
- MORIYAMA, E. N. AND POWELL, J. R. 1996. IntraSpecific nuclear DNA variation in *Drosophila*. *Mol. Biol. Evol.*, *13*, 261-277.
- NEALE, D. B. AND ADAMS, w. T. 1985. The mating system in natural and shelterwood stands of Douglas-fir. *Theor. Appl. Genet.*, *71*, 201-207.
- NEI, M. 1972. Genetic distance between populations. *Am. Nat.*, *106*, 283-292.
- NEI, M. 1973. Analysis of gene diversity in subdivided populations. *Proc. Natl. Acad. Sci. U.S.A.*, *70*, 3321-3323.
- NEI, M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, *89*, S83-S90.
- NEI, M. 1986. Definition and estimation of fixation indices. *Evolution*, *40*, 643-645.
- NEI, M. 1987. *Molecular Evolutionary Genetics*. Columbia University Press, New York.
- O'BRIEN, I. E. W., SMITH, D. R., GARDNER, R. C. AND MURRAY, B. G. 1996. Flow cytometric determination of genome size in *Pinus*. *Plant Sci.*, *115*, 91-99.
- PEAKALL, R., SMOUSE, P. E. AND HUFF, D. R. 1995. Evolutionary implications of allozyme and RAPD variation in diploid populations of dioecious buffalograss *Buchloe dactyloides*. *Mol. Ecol.*, *4*, 135-147.
- RAYBOULD, A. F., MOGG, R. J. AND CLARKE, R. T. 1996. The genetic structure of *Beta vulgaris ssp. maritima* (sea beet) populations: RFLPs and isozymes show different patterns of gene flow. *Heredity*, *77*, 245-250.
- REHFELDT, G. E. 1977. Growth and cold hardiness of intervarietal hybrids of Douglas-fir. *Theor. Appl. Genet.*, *50*, 3-15.
- SCRIBNER, K. T., ARNTZEN, J. W. AND BURKE, T. 1994. Comparative analysis of intra- and interpopulation genetic diversity in *Bufo bufo*, using allozyme, singlelocus microsatellite, minisatellite, and multilocus minisatellite data. *Biol. Evol.*, *11*, 737-748.
- SOKAL, R. R. AND ROHLF, F. J. 1981. *Biometry*, 2nd edn. W. H. Freeman and Co., New York.
- SZMIDT, A. E., WANG, x. AND Lu, M. 1996. Empirical assessment of allozyme and RAPD variation in *Pinus sylvestris* L. using haploid tissue analysis. *Heredity*, *76*, 412-420.
- THORMANN, C. E., FERREIRA, M. E., CAMARGO, L. E., TIVANG, J. G. AND OSBORN, T. c. 1994. Comparison of RFLP and RAPD markers to estimating genetic relationships within and among cruciferous species. *Theor. Appl. Genet.*, *88*, 973-980.
- WEIR, B. s. 1996. *Genetic Data Analysis 11*. Sinauer Associates, Sunderland, MA.
- WEIR, B. s. AND COCKERHAM, c. c. 1984. Estimating F-statistics for the analysis of population structure. *Evolution*, *38*, 1358-1370.
- WILLIAMS, J. G., KUBELIK, A. R., LIVAK, K. J., RAFALSKI, J. A. AND TINGEY, s. v. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucl. Acids Res.*, *18*, 6531-6535.