Isozymes and the Genetic Resources of Forest Trees[1]

A. H. D. Brown[2] and G. F. Moran[3]

Abstract: Genetic data are an essential prerequisite for analysing the genetic structure of tree populations. The isozyme technique is the best currently available method for obtaining such data. Despite several shortcomings, isozyme data directly evaluate the genetic resources of forest trees, and can thus be used to monitor and manipulate these resources. For example, preliminary isozyme data indicate that domestication of cultivated species has generally reduced variation within populations. Geographic differences among natural populations are less evident in trees than in herbaceous plants, but need to be considered in sampling and conservation strategies. In dealing with remnant population size (N_r) and distribution, management programs should recognize that the half-life of heterozygosity is about 1.4 N_r generations, whereas the number of trees required to retain half the current alleles after a bottleneck is about the square root of the original population size.

The growing awareness of two major theses provide the occasion of this opening contribution to the symposium. The first is that the genetic resources of economic plant species, and in particular of forest trees species, are finite. Yet, these resources have been managed without concern for their conservation or renewal. It is now clear that our remaining forest resources must be used prudently to guarantee a long future for managed species. The second thesis is that the isozyme technique is well suited for monitoring and assisting the manipulation of our genetic heritage, at least in part (Brown 1978). The last decade has witnessed the growth and impact of this technique in population biology. The ensuing contributions to this symposium will amply demonstrate the ways in which electrophoretic data can contribute to our genetic knowledge of populations of forest trees.

This paper first considers the properties of isozyme data and then discusses how the data can be used to monitor and conserve the genetic resources of forest trees.

ISOZYMES: BONANZA OR BLIND ALLEY IN FOREST GENETICS?

Phenotypic as Opposed to Genetic Evaluation

Two radically different approaches can be taken when measuring differences among individuals, populations or species. The first we define as the "phenotypic" approach. It is motivated by the argument that it deals directly with characters of economic or biological importance. The ideal of this approach is to know the phenotypic expression of yield, quality, and pest and disease resistance in all possible environments. Clearly, this ideal is unattainable, because the need for decisions limits, in space and time, the number of environments that may be tested. Quantitative genetics attempts to overcome these limitations by using a predictive model. The model assumes that unspecified polygenes can simulate the genetic and developmental complexities that underly phenotypic measurements and their variation. Various classes of effects on phenotypic variation—additive, dominance, epistasis, genotype x environment interaction—are defined (Libby and others 1969). The predictive power, however, is severely limited to the genotypes tested, the environments experienced, and the mode of measurement. Of course phenotypic (agronomic or silvicultural) assessment and phenotypic selection are essential in any plant breeding program. It is less certain whether such data provide valid measures of genetic variation and genetic similarity among populations.

The second approach to evaluation, the "genetic" approach, is entirely different in emphasis. Its aim is to detect genetic differences as close as possible to the DNA (deoxyribonucleic acid) level. The ideal of this approach is to know all the DNA sequences. Such basic data would form direct measures of how genetically variable is one population, and how similar it is to another, entirely free of environmental effects. Data from different species would be directly comparable. Again, this ideal is far from being a reality. Perhaps it may be attainable for very restricted classes of DNA (for example, organellar DNA, highly repeated DNA) or in very simple organisms (Smith 1979). But the need for "genetic," as distinct from "phenotypic," evaluation is apparent.

Because we cannot presently measure variation directly at the DNA level, we must lay down criteria for the choice of techniques that are currently feasible. *Table 1* summarizes four criteria first suggested by Hubby and Lewontin (1966; see also Lewontin 1974). These criteria require that the effect of an allelic substitution at a locus be detectable (C), and distinguishable from the substitution of any other allele at the same locus (A), or other loci (B).

---

[1] Presented at the Symposium on Isozymes of North American Forest Trees and Forest Insects, July 27, 1979, Berkeley, Calif.

[2] Principal Research Scientist, Commonwealth Scientific and Industrial Research Organization (CSIRO), Division of Plant Industry, Canberra, Australia.

[3] Experimental Officer, CSIRO, Division of Forest Research, Canberra, Australia.

Table 1—*The Hubby-Lewontin criteria for genetic markers, and the properties and pitfalls of the isozyme technique with respect to each criterion*

Criteria to be met	Properties of isozymes	Problems occurring
A. Alleles are distinct in individuals	Codominance; free of epistatic or environmental effects	Genetic or environmental mobility; mobility modification
B. Effects are locus-specific	Enzyme specificity	Nonspecific assays; duplication or polyploid
C. All substitutions are detectable	Charge-state separation, irrespective of function or variation	Detectable class of base substitutions is restricted and biased in variation
D. Loci are sampled randomly	Assay availability is independent of variation	Loci assayed are a limited class and differ in variation

Further, the sample of loci studied should be random, irrespective of function, or likely level of polymorphism (D).

The Isozyme Method

As perceived by Hubby and Lewontin (1966), the isozyme technique meets these criteria more closely than any preexisting method. Its general properties relevant to the criteria are also listed *(table 1)*. In short, one can study discrete Mendelian genes. This applies even in long-lived organisms such as trees, because individual seeds or seedlings can be assayed as members of progeny arrays.

An additional and important advantage of the method is the facility with which an array of enzymatic loci can be assayed using material in minute quantities, and with minimal preparation. The isozyme technique, therefore, represents the best currently available method for measuring genetic variation, close to the DNA level, relatively free of environmental effects, and on a manageable number of samples.

Pitfalls and Shortcomings of Isozyme Data

Some of the problems of isozymes associated with each of the Hubby-Lewontin criteria are listed in *table 1*. First mention is of some known cases where simple Mendelian codominant inheritance does not apply. Post-translational modification, either genetic (Law 1967) or environmental (Cullis 1977) in origin, is a possibility of which isozyme geneticists are becoming increasingly aware. The best way to avoid this pitfall is to verify by a Mendelian analysis that a set of isozyme variants are *allozymes*, that is, they are coded by alleles at a locus. This is especially the case for enzymes which do not form heteromultimers, or which are assayed nonspecifically (for example, esterases), or which are known to be affected by environment (for example, peroxidases).

Enzyme assays vary in their specificity. Diploid barley has at least 10 genetically defined esterase loci (Hvid and Nielsen 1977). With such nonspecific enzyme assays, a Mendelian analysis is essential. Duplication of genetic material, as in polyploidy, can often lead to the inability to ascribe a particular variant to a single locus. Special chromosomal stocks, such as nullisomics (Hart and Langston 1977) or haploids, if available, provide a definitive solution to this problem. This problem is rare in forest species, however, as diploidy is the common condition (Libby and others 1969).

With respect to the third criterion, two problems arise. The first is that electrophoresis does not detect all single base substitutions. About 23 percent of substitutions are synonymous and do not alter the amino acid because of redundancy in the code. About 67 percent of amino acid replacements do not alter net charge (Marshall and Brown 1975b). Some fraction of these replacements may be detectable by more refined electrophoretic techniques, or by altered heat sensitivity of the enzyme. Also, apparently long sequences of DNA inserted within the coding sequence of many eukaryotic genes may or may not be removed in the production of mRNA (messenger ribonucleic acid) (Crick 1979). These sequences are not translated, and are thus beyond the reach of electrophoretic surveys. Their function is not known, but preliminary evidence suggests that they can be quite variable. Crick (1979) describes the discovery of inserted sequences and split genes as "a mini-revolution in molecular genetics".

The other problem with respect to the third criterion is that, given access to a restricted class of base substitutions at a locus, can we assume that this class is unbiased with respect to the level of overall variability? Unfortunately, mounting evidence indicates that we cannot. For example, careful hybridization studies of variation in the DNA of sea urchins reveal a remarkably high level—4 percent—of base pair mismatching among individuals (Britten and others 1978). This mismatching is computed to be an order of magnitude greater than that predicted from electrophoretic heterozygosity, even after the latter has been corrected for synonymy and electrophoretically silent amino acid replacements. Isozymes are not so "close to the gene level" after all.

The fourth criterion raises the same two questions, but with respect to the loci sampled. Electrophoresis is used to study variation in the structural region of nuclear genes for certain restricted classes of proteins. Several studies of variation at these loci, either within functional classes of enzymes (Johnson 1974), or in relation to their physical parameters (Koehn and Eanes 1978), indicate that such loci differ in their tendency to vary genetically. This point must be taken into account in comparative studies of data on different species from different laboratories that may not have covered comparable sets of loci. Leigh Brown and Langley (1979) have suggested that the enzymes typically studied may overestimate genetic variation. They found

markedly less variation (4 percent heterozygosity in about 54 of the abundant proteins of adult *Drosophila*) than that predicted from allozyme studies (14 percent). Also, the inability of comparative electrophoretic divergence to measure evolutionary morphological divergence has been stressed by Wilson and others (1977). The lack of information on eukaryotic regulation systems and developmental sequences, and of data on sequence variation in regulatory loci or variation in their chromosomal location are severe shortcomings. Fortunately, the future application of other techniques of molecular biology to population studies does offer the hope of complementing the isozyme picture. These techniques include restriction nuclease analysis of organellar DNA's (Scowcroft[4]), DNA hybridization (Hall and others 1976), and ultimately cloning of specific sequences and DNA sequencing, and characterization of highly repeated classes of DNA for their base sequences and chromosomal location.

A major shortcoming—perhaps the greatest of all—is that the adaptive or selective significance of the bulk of allozyme variation remains obscure (Lewontin 1974). Is this variation adaptive or is it "evolutionary noise"? A less esoteric way of framing this question is to ask whether the genes studied by isozyme techniques have any relation at all to the characters of interest to the plant breeder. Is the study of allozyme variation a blind alley for forest tree breeding? Preliminary studies have shown it is possible to examine this question with a combination of population, physiological, and biochemical approaches. For example, alcohol dehydrogenase polymorphism in certain plant species shows population differentiation, selection responses, adaptive differences in tolerance to waterlogging, and in germination rates, and parallel variation in enzyme activity (Brown 1979). Much more evidence must be assembled, however, before a general answer is possible.

FOREST TREES AND THEIR GENETIC RESOURCES

The genetic resources of forest tree species consist of both primary and secondary gene pools (Frankel 1977). The primary reservoirs are those of undisturbed natural forests; the secondary sources include the remaining trees after ecosystem disturbance or harvesting, and the productive populations planted from local or exotic seed. Many authors have reviewed the genetic research on forest trees from the standpoint either of utilization or of genetic conservation (Libby and others 1969, Richardson 1970, Stern and Roche 1974). We now outline some issues in the status, handling, and conservation of genetic resources, which arise from the distinctive features of forest trees, in contrast to crop plants.

Size and Individuality

As large organisms, forest trees can readily be the target of individual phenotypic selection for growth of timber quality characteristics. This facility raises the question of the role of individual selection when collecting reproductive material. Should seed be taken only from a limited number of outstanding or "plus" trees? To answer this question, we first consider natural stands and ask whether the sampling is for conservation or breeding. Is the goal to obtain the largest amount of potentially useful genetic variation from that population; or is it to select a limited number of trees as parents in a seed orchard or breeding program? If the goal is *conservation,* Marshall and Brown[5] (1975a) have argued against sampling biased by phenotype. The sampling pattern at a site should be random with respect to phenotype, but representative with respect to ecological variation within the site. *Where* a plant grows is, perhaps, as reliable a guide as *what* it looks like when we consider its likely contribution to the genetic resources of the species. If the sampling aim is restricted to *breeding,* then the collection of seed from "plus" trees is debatable. It is difficult, however, to avoid unwanted biases from, for example, ephemeral selection criteria, diverse age structure of the population, and physical or biotic microenvironmental effects. If the "plus" trees are to be propagated clonally, one may have more justification for selection. In plantations, with a uniform age structure, and more random distribution of genotypes is microenvironments, it is usually easier to designate "plus" trees. Still, they could amount to a random sample. Most important, however, is the *number* of trees sampled. This number should not be restricted solely on the grounds that few of the available trees are outstanding. Genetic data that show the effect of biased as opposed to random sampling in tree populations are obviously needed.

A second practical consequence of tree size contrasts with the situation in most crop and forage species. The culture of a single plant uses considerable resources, and severely limits the population numbers in seed orchards, and *ex situ* conserved stands (Bouvarel 1970).

Longevity

As long-lived plants, the phenotypes of mature forest trees sum up a long history of environmental variation in space and time. This fact raises the issue of genic variation and fitness. Is genic variation maintained by heterozygote advantage in most environments? This may or may not mean that heterozygotes show less phenotypic plasticity for certain morphological traits (Bradshaw 1965). Or, are the individual alleles at polymorphic loci adapted to cope with different environments? Testing for isozyme-environ-

[4]Scowcroft, W. R. Nucleotide polymorphism in chloroplast DNA of *Nicotiana debneyi.* Theor. Appl. Genet. (In press.)

[5]Marshall, D. R. and A. H. D. Brown. Theory of forage plant collection. *In* Genetic Resources in Forage Plants. R.A. Bray and J.G. McIver, eds. CSIRO, Melbourne (In press).

ment associations must take account of this complex of possibilities and of conflicting predictions from different modes of selection.

Longevity and the delay in maturity in forest species, compared with annual crops, raise a practical issue in conservation by seed storage (Wang 1975). The viability of stored seed can approach the duration of the life cycle. If so, it must be regenerated as often as it takes to grow one crop of seed. Stored seed has an important role in the distribution of genetic resources without repeated collecting. What, then, is its role in genetic conservation of trees? Stern and Roche (1974) view *in situ* conservation as an interim measure until conservation is possible by "other means." In contrast, it can be argued that seed storage is essentially a provisional measure until *ex situ* stands are established, or a temporary insurance against the accidental destruction of *in situ* stands or plantations. The role of seed storage is severely limited in the case of tropical species with "recalcitrant" seeds.

Fecundity

The number of seeds produced by a single tree, even in one crop, may be enormous. This high fecundity enables tree species to tolerate high genetic loads and therefore, high levels of heterozygosity. From the standpoint of seed collection, this is one reason why the number of seeds sampled per mature individual can and should be much greater in trees than in crop or pasture species. Another reason is that the subsequent multiplication in plantations of any seed sample from trees is a slow and costly process. More specific recommendations on sample size are given below.

Outcrossing

Almost all the forest trees in which the breeding system has been studied are outbreeders. From this we can infer that heterozygosity is beneficial. A corollary is that inbreeding depression is likely to be evident in progeny if selfing rates are high. Further, the evidence suggests that outbreeders have less scope for the evolution of coadapted complexes of genes (at least in the sense of genes in linkage disequilibrium) than inbreeders (Brown 1979). Because of maternal heterozygosity and outcrossing, the seed crop on an individual tree is genetically heterogeneous (in contrast to inbred crops).

Harvesting

The link between economic utilization and the threat to genetic resources in forest species is more direct and final than in most other economic species. Logging destroys the individual as a genetic resource, and renewal requires reforestation based on sound genecological principles.

Stage of Domestication

Forest trees are wild plants at the earliest stages of domestication (Libby 1973). The time scale of the domestication process will be long. Unless the use of vegetative propagation becomes widespread, breeding and agronomic advances in themselves will not be the major causes of genetic erosion, as they are in agricultural crop species. The chief threat to agricultural genetic resources is the replacement of the older, more variable strains with modern advanced cultivars. The principal task in their conservation is the collection and preservation in gene resource centers, of the dwindling stocks of land races (Frankel 1974). The major threats to the integrity or, indeed, the survival of the genetic resources of forest trees are (a) ecosystem destruction and changing patterns of land use, (b) widespread planting with a few fast-growing species, (c) selective harvest of trees, and (d) plantation or reforestation practices based on limited or maladapted germplasm. Several case histories of tropical pines, teak, eucalypts, and others were reviewed at the Third World Consultation on Forest Tree Breeding, Canberra, 1977. The desirability and need for *in situ* conservation was frequently stressed.

With these general features of forest trees as a background, we now return to considering the importance of population genetic research based on the isozyme technique.

ISOZYME STUDIES AND GENE RESOURCE MANAGEMENT IN FOREST TREES

What are the roles that genetic evaluation based upon the isozyme technique might play in the management of the genetic resources of forest species (Feret and Bergmann 1976)? We would like to stress two interrelated functions—monitoring and decisionmaking.

Isozymes are ideal for monitoring:

(a) the current level of genetic variability in populations of economic species. This would furnish a base-line for evaluating the genetic consequences of any practice, as well as furnishing data for comparing levels of variability in different species (Hamrick and others 1979);
(b) increased genetic vulnerability in plantations as indicated by a decline in variability because of selection;
(c) loss of variation resulting from ecosystem destruction;
(d) the extent of divergence and variation in heterozygosity between natural populations of a species;
(e) the differentiation within a species between natural and planted populations, especially in other countries where exotic species (for example *Pinus radiata* D. Don, *Eucalyptus grandis* W. Hill ex Maiden) may have undergone adaptive "naturalization" (Frankel 1977);
(f) the extent of effective cross-pollination in seed orchards;
(g) pollen contamination of seed orchards;

(h) pollen contamination of *in situ* reserves by plantations;
(i) the inbreeding load in severely bottlenecked populations, such as those that regenerate from decimated stands following clear felling;
(j) seed and clone identification.

Several of these topics will be discussed in this symposium. As an example, we quote some of our own and other results to illustrate the impact of domestication on levels of genetic diversity. *Table 2* gives estimates of genetic diversity (H) (Nei 1975), and alleles per locus (A) in four cases. The statistic H is the mean expected panmictic heterozygosity($H = 1-\sum p_i^2$ Where pi is the frequency of the i^{th} allele). In barley (Hordeum), the sequence wild progenitor → primitive varieties or land races → advanced cultivars shows a fall in genetic variation both within lines and in the total sample. In contrast, domestication in *Phlox* is accompanied by increased total diversity. This increase probably arose from the deliberate collection and continual use of interesting variants. Darwin described an analogous consequence of artificial selection in the domestication of pigeons. In forest trees, the preliminary data indicate important changes in diversity levels with management, especially in *P. radiata.*

Recently, Hamrick and others (1979) have concluded from an analysis of published data that trees are genetically more variable on the average than herbaceous plants, because they are widespread, long-lived, large, fecund, and outcrossing species. This conclusion differs from that of animal data (Powell 1975, Nevo 1978): many vertebrates have less than one-half the heterozygosity of invertebrates. Is longevity or size of an organism (the attribute that Frankel and Soulé[6] use as an index of conservation strategies) a conflicting indicator of genetic variability in plants as compared with animals? Perhaps the plant literature is limited and biased by studies of inbreeding, colonizing, exotic annuals on the one hand, and the genecological studies of forest trees on the other. Several of these latter studies, including ours on mating systems, deliberately focused *only* on variable loci.

Table 3 summarizes some estimates of population genic diversity per variable locus ($\overline{H}$) and population differentiation ($D_{ST}/\overline{H}$, where $D_{ST}=HT-\overline{H}$ [Nei 1975]) in tree species. These estimates are compared with the same statistics in inbreeding and outbreeding herbaceous plants. As a group, the wind-pollinated conifers show the least degree of population differentiation. Trees pollinated by insects or birds show about three times as much, and are comparable with herbaceous outbreeders. *Eucalyptus caesia* Benth., a bird-pollinated, and geographically restricted species shows the highest differentiation of these trees, approaching the level for inbreeding herbs. Clearly, a wide range of population structures is found in trees.

DECISIONS

We will now outline some issues in which management decisions could be made with more precision if extensive isozyme data were available.

Sampling Strategies

How large should samples be and how should they be disposed among trees and among populations? Answers to this question depend on the purpose of sampling. This may be (a) to assemble the genetic resources in a target area for direct use, breeding or conservation; (b) to measure population genetic structure at individual loci, or obtain a sample that faithfully reflects gene frequencies; (c) to

[6]Frankel, O. H. and M. Soulé. Conservation and evolution. (Manuscript in preparation.)

Table 2—*Comparative genetic diversity (H) and alleles per locus (A) of domesticated lines and related natural populations*

Species	Lines	Allozyme loci[1]	Within lines		Total		Source[2]
			A	H	A	H_T	
Hordeum vulgare	32 Scandinavian varieties	11[1]	1.02	0.007	1.55	0.127	a
	Composite cross XXI, F_{17}	17[1]	—	—	1.53	.094	b
	11 Iran land races	17	1.21	.072	1.94	.180	c
H. spontaneum	28 Israel populations	17[1]	1.56	.109	4.12	.216	b
Phlox drummondii	16 Cultivars	17	1.12	.041	1.41	.133	d
	10 Populations	17	1.25	.063	1.29	.089	d
Pinus sylvestris	2 Seed orchards	3	3.5	.36	3.7	.37	e
	3 Sweden populations	3[1]	4.3	.33	5.3	.38	d
Pinus radiata	1 Seed orchard (30 clones)	18[1]	—	—	1.67	.118	f
	20 Breeding clones	19	—	—	1.74	.145	f
	5 California populations	19	1.84	.139	2.63	.161	f

[1]*Loci:* More loci are in the original source, but only comparable loci are included here

[2]*Sources:* (a) Almgard and Landegren (1974), (b) Nevo, E., D. Zohary, A. H. D. Brown, and M. Haber, 1979. Genetic diversity and environmental associations of wild barley, *Hordeum spontaneum* in Israel. Evolution. (In press.) (c) Brown (unpublished). (d) See Brown (1979) for references, (e) Rudin and Lindgren (1977), (t) Moran (unpublished).

Table 3—*Population genic diversity per variable locus ($\overline{H}$) and differentiation ($D_{ST}/\overline{H}$) in trees*

Species	Variable loci	Populations	$\overline{H}$	DST/ $\overline{H}$ Percent	Source[1]
Abies lasiocarpa	1	3	0.40	3.0	(a)
Picea abies	4	8	.42	4.5	(a)
P. abies	6	10	.40	4.7	(a)
P. abies	4	8	.35	2.4	(a)
P. abies	10	4	.41	2.9	(b)
P. engelmannii	1	3	.47	12	(a)
Pinus longaeva	11	5	.47	3.6	(c)
P. nigra clusiana	3	7	.48	13	(d)
laricio	4	12	.30	6.9	(d)
nigricans	3	11	.52	12	(d)
pallasiana	4	10	.30	7.0	(d)
P. radiata	19	5	.14	15	(e)
P. rigida	15	4	.17	1.1	(b)
P. sylvestris	3	3	.33	14	(a)
P. virginiana	2	4	.29	2.6	(g)
Pseudotsuga menziesii	2	6	.55	3.0	(h)
P. menziesii	4	9	.59	2.9	(i)
Eucalyptus caesia	8	10	0.11	78	(j)
E. cloeziana	5	17	.25	17	(k)
E. delagatensis	4	8	.34	24	(a)
E. obliqua	3	7	.39	21	(a)
E. pauciflora	7	3	.28	2.2	(a)
Ficus carica	2	4	.53	2.3	(a)

	Summary		Average		Weighted[2] Average	
			$\overline{H}$	DST/ $\overline{H}$ [3]	$\overline{H}$	DST/ $\overline{H}$
Outbreeders	Wind-pollinated		0.372	7.4	0.298	6.8
	Animal-pollinated		.317	16	.265	20
	Herbaceous	(b)	.325	22	.282	18
Inbreeders		(b)	.126	118	.133	102

[1]*Source:* (a) see Brown (1979) for references, (b) Lundkvist (1979), (c) Hiebert, see Hamrick and others (1979), (d) Bonnet-Masimbert and Bikay-Bikay (1978), (e) Moran (unpublished) (f) Guries and Ledig (1977), (g) Witter and Feret (1978), (h) Mejnartowicz (1976), (i) Yang and others (1977), (j) Moran and Hopper (unpublished), (k) Turnbull (unpublished).

[2] Weighted by the number of variable loci in each study.

[3]Computed by dividing the sum of D_{ST} estimates by the sum of $\overline{H}$ estimates.

measure multilocus associations; or (d) to estimate parameters of the mating system.

Sampling for the first purpose has been discussed extensively by Marshall and Brown[5] (1975a). The key to this sampling is to maximize the number of *locally common* alleles in a collection whose total size is strictly limited. Locally common alleles are defined here as alleles that assume a frequency greater than 0.10 in only one locality. Alleles with two such occurrences are termed sporadically common. The existence of locally common alleles in plant populations is discussed elsewhere (Brown 1978). The limited data from forest species (*table 4*) indicate that of all alleles detected, locally common alleles make up a lesser (but significant) fraction, compared with herbaceous plants. Our rule of thumb has been to sample about 50 plants at random and visit as many contrasting sites as possible. If this is difficult or time consuming, then samples of fewer plants from more sites should be taken (Marshall and Brown[5]).

In forest trees, this rule applies directly to the sampling of vegetative material. For seed samples, however, it requires variation because of the high fecundity, the outbreeding system, and the difficulty of sampling each additional tree. Samples from fewer trees (n), but more seed per tree (k), such that $nk > 100$ are in order. In practice, however, much larger samples than this are taken in forestry to allow for seed requests, and to avoid early resampling or regeneration.

Samples for research, or for the representation of site allele frequencies, must be based on a reasonable number of trees. A minimum value for n follows from the limiting formula for the variance of allele frequencies in the seed generation (Brown and others 1975) with large k and assuming $p = q = 0.5$

$$\lim_{k\to\infty} o^2_p = (1 + 3F)^2 /32n(1 + F)$$

Table 4—*The number of alleles with various kinds of distribution*

Species	Populations	Loci	Kinds of enzymes	Common			Rare alleles	
				Widespread	Sporadic	Localized	Widespread	Localized
Picea abies	11	4	4	10	1	1	1	1
P. abies	10	6	4	12	—	2	3	1
P. abies	4	11	6	19	3	3	16	13
Pinus longaeva	5	14	7	27	3	—	3	—
P. nigra clusiana	7	4	1	9	3	1	1	—
laricio	12	4	1	8	1	—	1	1
nigricans	11	4	1	10	—	—	2	1
pallasiana	10	4	1	7	—	1	1	3
P. radiata	5	19	13	26	1	3	10	10
P. rigida	4	15	10	19	2	1	8	—
Pseudotsuga menziesii	6	2	1	7	—	—	4	1
P. menziesii	9	4	3	12	—	—	—	—
Eucalyptus caesia	10	8	6	12	1	6	—	2
E. cloeziana	17[1]	5	5	9	3	1	1	—
E. delagatensis	8	4	4	8	1	—	1	—
E. obliqua	7	3	3	6	—	2	4	2
Summary				Percentage of variants				
Wind-pollinated trees		91	—	41	8	7	27	17
Animal-pollinated trees		20	—	39	13	23	75	10
Wild herbaceous (8 entries[2])		14	—	31	18	28	7	16
Cultivars (4 entries[2])		25	—	45	14	23	4	14

[1] These populations were grouped into seven geographic regions.
[2] From Brown (1978). See *table 3* for sources of remaining data.

where F is the inbreeding coefficient.

The minimum number of trees required to ensure that the difference (d), between the sample allele frequency and its true population frequency, is exceeded on less than 5 percent of occasions is

$$n = (1 + 3F)^2/8(1 + F)d^2$$

$$\simeq 0.2\ d^{-2}, \text{ assuming } F < 0.10$$

Ford = 0.1, we would require large samples from at least n = 20 trees. A sample of 10 trees from each of two populations can detect differences in gene frequency between them of 0.2 at the 5 percent level of significance.

Much larger samples are needed to detect weak multilocus associations in the adult generation (Brown 1975). Finally, sampling for the study of mating systems unfortunately is ideally served by one contrary principle: the sampling of trees from within the same microenvironment and in close proximity (Brown and others 1975). Otherwise, microgeographic differentiation can lead to inflated estimates of self-fertilization. In general, more information on allelic frequency patterns could be used to assess optimal sampling strategies within and among sites.

Harvesting in Naturally Regenerating Systems

What and where to log and what to leave? Selective harvesting followed by natural or assisted reforestation is a common practice. There are important genetical issues for which we have little genetic data to inform us. If size or growth rate is related to heterozygosity, will selective logging amount to selection against heterozygotes and, therefore, lead to destabilized genetic equilibria (Charlesworth 1972) and the loss of certain polymorphic alleles? Dysgenic selection could also arise in an additive model, simply by removing the preferred alleles. Perhaps it would be a better strategy, genetically, to clear fell most of the area and leave certain areas untouched as a seed source. A major variable here would be the pattern and scale of microdifferentiation. For example, the study of Mitton and others (1977) of *Pinus ponderosa* Dougl. ex Laws., indicates that reserved areas should be located on both north- and south-facing slopes, and that of Bergmann (1978) of *Picea abies* K indicates that elevational gradients are important.

Critical Population Sizes for Conservation

How large a population should be set aside and managed as a long-term source of genetic diversity (conservation *in situ*)?

As the need for conservation becomes more pressing, ecologists and geneticists are being asked to recommend the minimum size of a viable population. As yet there is no single answer to this question. Frankel and Soulé[6] have suggested 50 for the basic irreducible number for

conservation genetics. They regard this number as the absolute minimum below which inbreeding ($\Delta F = 1 - (1/2N)$) would be unacceptably greater than 1 percent per generation. Of course, we refer to *effective* population number—actual numbers being somewhat larger. This reasoning, however, does not apply to inbreeding or to apomictic plants. Furthermore, populations of this size have little scope for adaptive evolution, a prerequisite for long-term survival. Franklin (see Frankel and Soulé[6]) has proposed the population size of 500, so that the effects of drift would be negligible when compared with those of selection of measurable intensity.

Latter and Frankel (in Frankel 1970) have suggested using the equilibrium formula of Crow and Kimura (1970) which is

$$H_e = 4Nu/(1 + 4Nu) = \Theta/(1 + \Theta)$$

in which

H_e = equilibrium heterozygosity

N = effective population number

u = mutation rate

$\Theta = 4Nu$

If we assume $H_e = 0.3$ and $u = 2 \times 10^{-6}$ (Mukai and Cockerham 1977), both figures for electrophoretically detectable mutation, we obtain the number, $N \simeq 50{,}000$. This is the effective number required to prevent neutral heterozygosity declining at all from its present level. Thus, we have a range of answers (from 50 to 50,000) based on a range of rationales. The answer will vary for different species, so that no rationale is universally applicable.

Another way to look at this issue is to introduce the time element into the calculations, and ask what is *the half-life of heterozygosity* in generations, given a particular effective size? The loss of one-half the heterozygosity is biologically equivalent to one round of complete selfing.

Let t_j denote the time in generations required for the current heterozygosity (H_o) to fall to a value $j\,H_o (o < j < 1)$ in a restricted population of effective size $N_r u$ and neutral mutation rate u. If $H_e = 4N_rU/(1 + 4N_ru)$, the new equilibrium heterozygosity, and assuming $[j\,H_o > H_e]$, the number of generations is

$$t_j = \frac{\log_j \{[jH_o - H_e]/[H_o - H_e]\}}{\log_j \{1 - 2u - 1/(2N_r)\}}$$

In particular, the half-life of heterozygosity is approximately

$$t_{0.5} \simeq 1.4\, N_r$$

If we assume that the "time scale of concern" (Frankel 1974) for forest species is of the order of 100 generations or 1000 years (Frankel and Soulé[6]), and that we wish to preserve at least one-half the heterozygosity by that date, the effective number required is $N_r = 70$.

The loss of heterozygosity is but one dimension to the problem of population size. A second and potentially more important problem is the numbers of alleles lost in a bottleneck. We may again ask an analogous question.

If the original population size is N_o, what size of relic population, N_r, is required to retain half the alleles currently present? We can use the sampling theory for neutral alleles (Ewens 1972) to attempt an answer to this question. We will assume two values for $\Theta = 4N_o$, namely $\Theta = 0.5$ or 1.0. In this section, u denotes the *total* mutation rate.

In the population, the actual number of alleles is

$$n_a = \theta \int_{1/2N_o}^{1} (1-x)^{\theta - 1} x^{-1} dx$$

For $\Theta = 1$, $\quad n_a|_{\Theta=1} = \log_e(2N_o)$

and for $\Theta = 0.5$, $\quad n_a|_{\Theta=0.5} \simeq [\log_e(8N_o)]/2$

From Ewens (1972), the expected number of alleles retained in the sample (the relic population) is

$$n_r = \sum_{i=1}^{2N_r} \frac{\theta}{\theta + i - 1}$$

Using an integral approximation, this summation can be evaluated as

$$n_r \simeq \Theta \left\{ \log_e [(\Theta + 2N_r - 1)/\Theta] \right\} + 0.5$$

We wish to find the values of N_r which will yield $n_r/n_a = 0.5$, to retain half the alleles.

For $\Theta = 1$; $N_r = 0.43(N_o)^{1/2}$, and for $\Theta = 0.5$;

$$N_r = 0.26(N_o)^{1/2}$$

Unfortunately, rather large standard errors are associated with the number of alleles per locus retained in a relic population. Applying the same approximation procedure to Ewens' (1972) variance formula, we obtain for the variance,

$$\text{var}(n_r) \simeq \Theta \left\{ \log_e [(\Theta + 2N_r - 1)/\Theta] - 1 \right\}$$

However, if one considers that a large number of independent loci ($v > 100$) are polymorphic in the population, it is possible to ask what is the number required to ensure survival of one-half of all the alleles with 95 percent certainty. For $\Theta = 1$, the effective population size (N_r)required for this is

$$N_r \simeq 0.43 N_o^{1/2} \times \exp\left[2\left\{\log_e(0.1N_o)\right\}/v\right]^{1/2}$$

Original population size (N_o)	Number of polymorphic loci (v)		
	100	1000	∞
10,000	62	48	43
100,000	209	158	136

These figures indicate that, as a rough guide, the number to retain to ensure one-half the alleles survive with 95 percent certainty is about the square root of the original population size. More precise calculations would be possible when more estimates of Θ, u and v are available.

Population Distribution

Should only one large population be set aside, or is it sounder genetically to divide the requirements into several partially isolated subpopulations?

If s is the number of subpopulations, each of effective size N_S, and m is the migration rate between them in the island model (Nei 1975), the total heterozygosity maintained in the whole ensemble (H_T) is

$$H_T = 1 - [1+u \{ s/m+4sN_s \}]^{-1}$$

From this formula, it can be seen that H_T will increase under subdivision as $s/ m + 4sN_s$ increases. The island model of migration is the most conservative from this viewpoint. The quantity sN_s nominally equals N_r, the total number retained from the original population. This quantity, however, is not likely to remain constant with increasing s; rather, it will decline because of the need for buffer zones around each subpopulation. More and more subpopulations of smaller size may render each one steadily more vulnerable, so that the total (N_r) declines as s increases. The requirement that the quantity (s/m) increase with increasing s means that, as the number of subpopulations and hence their proximity increases, a less than commensurate proportional increase in migration will occur between them. The optimum value of s to maximize H_T could be defined if we had information on buffer-zone requirements and on gene flow with increasing s.

CONCLUSIONS

The isozyme technique is a quantum jump forward in research on the genetics of forest trees. It does more than just measure another set of characteristics, because it enables a direct genetic evaluation of the status of genetic resources. Despite the real limitations of isozyme data, which may in the near future be partly met by the application of newer molecular techniques in population studies, it is possible to make decisions on the manipulation and conservation of these resources, aided by such data. An isozyme laboratory may be considered demanding of personnel and financial resources. Yet these costs should be compared with those of establishing and maintaining large-scale field trials at several sites for several years. It is our view that genetic evaluation is as essential a component as phenotypic evaluation in research programs aimed at supporting decisions on the management of the genetic resources of forest trees.

Preliminary isozyme studies in several cultivated plants (for example, barley, tomatoes, *Phlox,* radiata pine), have quantified the impact of domestication on genetic variation. The erosion of genetic resources and, hence, the greater genetic vulnerability of planted populations, is evident. Comparative studies of variation patterns in trees and herbaceous plants emphasize the diversity of population genetic structures in plants. The degree of geographic differentiation among populations is more limited in trees than in herbs, but nonetheless significantly affects sampling and conservation strategies. Estimates of present day heterozygosity, mutation rates, gene flow, and differentiation permit an objective approach to questions of remnant population size and distribution. Several recommendations of minimum populations size may be compared within practical and species-specific constraints. Effective population sizes of about 70 allow one-half the current heterozygosity to remain after 100 generations, whereas the number of trees required to ensure with 95 percent certainty that one-half the current alleles survive the bottleneck is approximately the square root of the current population size. As further isozyme data accrue, more precise estimates and strategies will be possible.

Acknowledgments

We are grateful to our colleagues A.G. Brown, K.G. Eldridge, O.H. Frankel, A.C. Matheson, R.N. Oram and J. W. Turnbull for their critical reading of the manuscript.

LITERATURE CITED

Almgard, G., and U. Landegren.
1974. **Isoenzymatic variation used for the identification of barley eultivars.** Zeitschrift für Pflanzenzüchtung 72:63-73.

Bergmann, F.
1978. **The allelic distribution at an acid phosphatase locus in Norway Spruce (*Picea abies*) along similar climatic gradients.** Theor. Appl. **Genet. 52:51-64.**

Bonnett-Masimbert, M., and V. Bikay-Bikay.
1978. **Variabilité intraspécifique des isozymes de la glutamate-oxalo-acetate transaminase chez *Pinus nigra* Arnold intérêt pour la taxonomie des sous espèces.** Silvae Genetica 27:71-79.

Bouvarel, P.
1970. **The conservation of gene resources of forest trees.** *In* Genetic resources in plants. O. H. Frankel and E. Bennett, eds. p. 523-530. Blackwell, Oxford.

Bradshaw, A. D.
1965. **Evolutionary significance of phenotypic plasticity in plants.** Adv. Genet. 13:115-155.

Britten, R. J., A. Cetta, and E. H. Davidson.
1978. **The single-copy DNA sequence polymorphism of the sea urchin *Strongylocentrotus purpuratus.*** Cell 15:1175-1186.

Brown, A. H. D.
1975. **Sample sizes required to detect linkage disequilibrium between two and three loci.** Theor. Popul. Biol. 8:184-201.

Brown, A. H. D.
1978. **Isozymes, plant population genetic structure and genetic conservation.** Theor. Appl. Genet. 52:145-157.
Brown, A. H. D.
1979. **Enzyme polymorphism in plant populations.** Theor. Popul. Biol. 15:1-42.
Brown, A. H. D., A. C. Matheson, and K. G. Eldridge.
1975. **Estimation of the mating system of *Eucalyptus obliqua* L'Hérit. using allozyme polymorphisms.** Aust. J. Biol. Sci. 23:931-949.
Charlesworth, Brian.
1972. **Selection in populations with overlapping generations. III. Conditions for genetic equilibrium**. Theor. Popul. Biol. 3:377-395.
Crick, Francis.
1979. **Split genes and RNA splicing.** Science 204:264-271.
Crow, James F., and Motoo Kimura.
1970. **An introduction to population genetics theory.** p. 322. Harper and Row, New York.
Cullis. C. A.
1977. **Molecular aspects of the environmental induction of heritable changes in flax.** Heredity 38:129-154.
Ewens, W. J.
1972. **The sampling theory of selectively neutral alleles.** Theor. Popul. Biol. 3:87-112.
Feret, P. P., and F. Bergmann.
1976. **Gel electrophoresis of proteins and enzymes.** *In* Modern methods in forest genetics. J. P. Miksche, ed. p. 49-78. Springer-Verlag, Berlin.
Frankel, O. H.
1970. **Variation—the essence of life.** Proceedings of Linnean Society of N.S.W. 95:158-169.
Frankel, O. H.
1974. **Genetic conservation: our evolutionary responsibility.** Genetics 78:53-65.
Frankel, O. H.
1977. **Philosophy and strategy of genetic conservation in plants.** 3d World Consult. For. Tree Breeding. Canberra 1:1-11.
Guries, R. P., and F. T. Ledig.
1977. **Analysis of population structure from allozyme frequencies.** Proc. South. For. Tree Improv. Conf. 14:246-253.
Hall, R. B., J. P. Miksche, and K. M. Hansen.
1976. **Nucleic acid extraction, purification, reannealing and hybridization methods.** *In* Modern methods in forest genetics. J. P. Miksche, ed. p. 19-48. Springer-Verlag. Berlin.
Hamrick, J. L., Y. B. Linhart, and J. B. Mitton.
1979. **Relationships between life history characteristics and electrophoretically-detectable genetic variation in plants.** Ann. Rev. Ecol. Syst. 10:173-200.
Hart, G. E., and P. J. Langston.
1977. **Chromosomal location and evolution of isozyme structural genes in hexaploid wheat.** Heredity 39:263-277.
Hubby, J. L., and R. C. Lewontin.
1966. **A molecular approach to the study of genetic heterozygosity in natural populations. I. The number of alleles at different loci in *Drosophila pseudoobscura*.** Genetics 54:577-594.
Hvid, Søren, and Gunnar Nielsen.
1977. **Esterase isoenzyme variants in barley.** Hereditas 87:155-162.
Johnson, George B.
1974. **Enzyme polymorphism and metabolism.** Science 184:28-37.
Koehn, Richard K., and Walter F. Eanes.
1978. **Molecular structure and protein variation within and among populations.** Evol. Biol. 11:39-100.
Law, G. R. J.
1967. **Alkaline phosphatase and leucine aminopeptidase association in plasma of the chicken.** Science 156:1106-1107.
Leigh Brown, A. J., and C. H. Langley.
1979. **Reevaluation of level of genic heterozygosity in natural population of *Drosophila melanogaster* by two dimensional electrophoresis.** Proc. Natl. Acad. Sci. 76:2381-2384.
Lewontin, R. C.
1974. **The genetic basis of evolutionary change.** 346 p. Columbia Univ. Press, New York.
Libby, W. J.
1973. **Domestication strategies for forest trees.** Can. J. Forest Res. 3:265-276.
Libby, W. J., R. F. Stettler, and F. W. Seitz.
1969. **Forest genetics and forest tree breeding.** Ann. Rev. Genet. 3:469-494.
Lundkvist, Kenneth.
1979. **Allozyme frequency distributions in four Swedish populations of Norway spruce (*Picea abies* K.). I. Estimates of genetic variation within and among populations, genetic linkage and among mating system parameters.** Hereditas 90:127-143.
Marshall, D. R., and A. H. D. Brown.
1975a. **Optimum sampling strategies in genetic conservation.** *In* Genetic resources for today and tomorrow. O. H. Frankel and J. G. Hawkes, eds., p. 53-80. Cambridge Univ. Press, Cambridge.
Marshall, D. R., and A. H. D. Brown.
1975b. **The charge state model of protein polymorphisms in natural populations.** J. Mot. Evol. 6:149-63.
Mejnartowicz, L.
1976. **Genetic investigations on Douglas-fir (*Pseudotsuga menziesii* [Mirb.] Franco) populations.** Arboretum Kornickie 21:126-187.
Mitton, J. B., Y. B. l.inhart, J. L. Hamrick, and J. S. Beckman.
1977. **Observations on the genetic structure and mating system of ponderosa pine in the Colorado front range.** Theor. Appl. Genet. 51:5-13.
Muskai, Terumi, and C. Clark Cockerham.
1977. **Spontaneous mutation rates at enzyme loci in *Drosophila melanogaster.*** Proc. Natl. Acad. Sci. 74:2514-2517.
Nei, Masatoshi.
1975. **Molecular population genetics and evolution.** 288 p. Elsevier, New York.
Nevo, Eviatar.
1978. **Genetic variation in natural populations: patterns and theory.** Theor. Popul. Biol. 13:121-177.
Powell, Jeffry R.
1975. **Protein variation in natural populations of animals.** Evol. Biol. 8:79-119.
Richardson, S. D.
1970. **Gene pools in forestry.** *In* Genetic resources in plants. O. H. Frankel and E. Bennett, eds. p. 353-365. Blackwell, Oxford.
Rudin, Dag, and Dag Lindgren.
1977. **Isozyme studies in seed orchards.** Studia Forestalia Suecica 139:1-23.
Smith, M.
1979. **The first complete nucleotide sequencing of an organism's DNA.** Amer. Sci. 67:57-67.
Stern, Klaus, and Laurence Roche.
1974. **Genetics of forest ecosystems.** 330 p. Springer-Verlag, Berlin.
Wang, B. S. P.
1975. **Tree seed and pollen storage for genetic conservation: possibilities and limitations.** *In* Methodology of conservation of forest genetic resources. p. 93-103. FAO, Rome.
Wilson, Allan C., Steven C. Carlson, and Thomas J. White.
1977. **Biochemical evolution.** Ann. Rev. Biochem. 46:573-639.
Witter, M. S., and P. P. Feret.
1978. **Inheritance of glutamate oxaloacetate transaminase isozymes in Virginia pine megagametophytes.** Silvae Genetica 27:128-134.
Yang, J.-Ch., T. M. Ching, and K. K. Ching.
1977. **Isoenzyme variation of coastal Douglas-fir. 1. A study of geographic variation in three enzyme systems.** Silvae Genetica 26:10-18.