

Genetic consequences of *in situ* and *ex situ* conservation of forest trees

by Rong-Cai Yang and Francis C. Yeh

To counteract loss of genetic diversity crucial for current and future tree improvement, tree breeders have conserved forest genetic resources *in situ* in their natural ecosystems in protected areas or *ex situ* in plantations, seed orchards, and breeding arboreta. This article reviews the genetic consequences of these two conservation methods in terms of single-locus and multilocus population structure from electrophoretic studies of natural forests and breeding populations. Although natural forest populations have maintained high level of genetic diversity and exhibited low level of population differentiation, loss of genetic diversity would occur during the entire conservation process, from population establishment to management of breeding and production populations. Since forest trees are still at their earliest stage of domestication in Canada, loss of genetic diversity comes primarily from the initial sampling process during population establishment. We discuss the optimal sampling strategy during population establishment to conserve common and widespread alleles, common and localized alleles, rare and widespread alleles, and rare and localized alleles. We also discuss three methods for studying the multilocus structure of forest trees and show how such information would be useful for conserving co-adapted gene complexes. We conclude that being small and maintained in controlled environments, *ex situ* conserved populations would retain less genetic diversity than *in situ* conserved forest populations. While *ex situ* conservation is operationally convenient for the short-term gains in tree improvement, we believe *in situ* conservation is essential for renewing the genetic diversity to meet the changing environments of an uncertain future.

Afin de contrebalancer la perte de la diversité génétique essentielle à l'amélioration génétique des arbres en cours et pour l'avenir, les généticiens ont conservé les ressources génétiques forestières *in situ* au sein de leurs écosystèmes naturels dans des zones protégées ou *ex situ* dans des plantations, des vergers à graine et des arboretums destinés aux croisements génétiques. Cet article fait état des conséquences génétiques de ces deux méthodes de conservation en terme de structure de population issue d'un seul endroit et répartie en plusieurs endroits à partir des études électrophorétiques des forêts naturelles et des populations destinées aux croisements. Même si les populations des forêts naturelles ont maintenu un haut niveau de diversité génétique et ont démontré un faible degré de différenciation des populations, une perte de la diversité génétique pourrait survenir au cours de l'ensemble du processus de conservation, soit entre la mise en place des populations et à la production. Compte tenu que les espèces forestières sont encore au tout début du processus de domestication au Canada, la diminution de la diversité génétique surgit principalement suite à l'échantillonnage initial effectué lors de l'établissement de la population. Nous discutons de la stratégie optimale d'échantillonnage à utiliser lors de l'établissement de la population afin de conserver les allèles communs et largement répandus, les allèles communs et localisés, les allèles rares et largement répandus, ainsi que les allèles rares et localisés. Nous abordons également trois méthodes pour étudier la structure des arbres répartis en plusieurs endroits et nous démontrons également qu'une telle information pourrait être utile pour conserver des complexes génétiques susceptibles d'être utilisés. Nous concluons que les petites populations conservées *ex situ* et maintenues dans un environnement contrôlé contiendraient moins de diversité génétique que les populations forestières conservées *in situ*. Même si la conservation *ex situ* est opérationnellement utile pour obtenir des gains à court terme en amélioration génétique, nous croyons que la conservation *in situ* est essentielle au renouvellement de la diversité génétique afin de faire face à des environnements en transformation dans un avenir incertain.

Introduction

The forest is Canada's most important and valuable natural resource, covering about 45% of the total land base (Brand 1990). It provides nationwide economic, environmental, and recreational benefits. However, the provision of such benefits is being threatened because the level of biodiversity in forest ecosystems is declining at an accelerating rate. The responsible factors for loss of biodiversity range from global environmental effects (e.g., urban and agriculture development, water and air pollution, and global climatic change) to breeders' intensified efforts on tree improvement. Therefore, conservation of biodiversity has been the subject of strong debate among Canadian scientists.

Ledig (1986) gave three distinct reasons why tree breeders should be concerned with the maintenance of biodiversity: (1) genetic uniformity increases vulnerability to pests and other environmental stresses (2) genetic diversity permits construction of new strains which have breeding values in future and (3) the loss of species diversity may reduce the stability of entire

ecosystems. Thus, declining genetic diversity in both natural and breeding populations will restrict the potential for genetic adjustment to the ever changing natural, economic, and social components of the environment (Frankel 1977; Namkoong 1986).

Forest genetic resources consist of both primary and secondary gene pools (Frankel 1977). The primary gene pools are the undisturbed natural forests and the secondary gene pools are the remaining forests after ecosystem disturbance (including breeding populations). Many believe that genetic diversity in the primary gene pools can be conserved without the need to maintain it. However, as natural areas become progressively modified by human and non-human interferences, maintaining genetic diversity of natural forests and their ecosystems will increasingly depend on application of specific technologies (OTA 1987).

In situ and *ex situ* conservation are the two principal methods to maintain genetic diversity (Frankel and Soulé 1981). *In situ* conservation maintains the natural forests in their native environments in protected areas. Thus, the protected forests evolve continuously and dynamically with their living and non-living environments (Frankel and Soulé 1981). On the other hand, breeding populations or collections

Department of Forest Science, University of Alberta, Edmonton, Alberta, Canada T6G 2H1.

from disturbed ecosystems are preferably grown in controlled (*ex situ*) environments either for easier and better genetic manipulation or for safety reasons. While tree breeders improve the genetic structure of breeding populations to maximize economic gain, natural selection and other evolutionary forces shape the genetic structure of *in situ* conserved natural populations. Therefore, during the conservation process, from the establishment of populations to the managing process, changes in genetic diversity through *in situ* and *ex situ* conservation will be very different. In this article, we first review the electrophoretic studies on population structure of forest trees, with special reference to the optimal population structure for gene conservation and/or tree improvement. We then examine and predict the loss of genetic diversity for *in situ* and *ex situ* conservation from sub-optimal sampling of population structure during population establishment and during subsequent managing processes. Finally, we elucidate the contribution of multilocus associations to the development of population structure and its implication for genetic conservation.

Characterization of Population Genetic Structure

Level of genetic diversity in a population is affected by an array of ecological, life history and genetic characteristics that collectively define the population genetic structure. The population genetic structure can be characterized by two sets of measurable parameters (Brown 1989). The first set is the geographic parameters which describe and quantify the geographic variation patterns of populations at the single-locus level. The second set is the genomic parameters which describe and quantify the non-random associations of genes or gametic disequilibria of populations. The genomic variation patterns of forest populations are far less investigated than the geographic variation patterns. This is due to their more severe sampling problems than the single-locus patterns and the associated statistical and analytical complexity when studying associations of multiple alleles among many loci (Brown 1984). Both sets of parameters describe the outcomes of selective and non-selective events in natural and domesticated populations.

Geographic Variation Patterns

Electrophoretic studies of natural populations of plants have been reviewed by several workers (Brown 1979; Loveless and Hamrick 1984; Hamrick and Godt 1990). The most common measures of geographic variation are percentage of polymorphic loci (P), average number of alleles per locus (A), gene diversity index (H) and proportion of total diversity among populations (G_{st}). A significant proportion of isozyme variation of plant species can be explained by the variation in their life history and ecological characteristics (Loveless and Hamrick 1984). Consequently, many studies (Moran and Hopper 1987) have focused on whether or not a species' geographic range is a reliable predictor of the distribution of its genetic diversity.

Theoretically, loss of genetic diversity to random drift, founder effect and directional selection is more likely for geographically restricted species with few individuals than for widespread species with many individuals (Frankel and Soulé 1981). Many electrophoretic studies have confirmed

this theory (Hamrick and Godt 1990). However, contrary to expectations, some widespread species exhibited higher levels of population differentiation than some geographically narrow species (Loveless and Hamrick 1984). Thus, geographic range alone is not a reliable predictor of the population genetic structure. Other characteristics such as mating system and population size can overwhelm the effect of geographic range. Based on an extended allozyme data, Hamrick and Godt (1990) suggested the combined effect of geographic range and mating system was a better predictor of a species' population structure.

As large and long-lived plants with widespread distribution, high fecundity, predominantly outcrossing mode and wind pollination, natural populations of forest trees maintained high level of genetic variation and exhibited low level of population differentiation (Yeh 1989). The estimates of gene diversity and proportion of population differentiation for eleven North American tree species were summarized (Table 1), each with relatively large sample of loci and populations to minimize the sampling bias. Levels of isozyme variation as measured by gene diversity were moderate to high for most species in comparison to the 0.126 average for thirteen inbred plant species (Brown 1979). The observed low levels of isozyme variation in *Thuja plicata* and *Pinus resinosa* were due to restricted population size and severe bottlenecks (Fowler and Morris 1977; Yeh 1988). Levels of differentiation among forest populations were low but statistically significant (Table 1), ranging from 1% for *Picea mariana* (Boyle 1985) to 16.8% for *Pinus radiata* (Moran *et al.* 1988). These estimates of differentiation among forest populations differed sharply from the near 100% average population differentiation in 13 inbred plant species (Brown 1979). Thus, diversifying selection and the outcrossing mode might be important determinants of the level and distribution of genetic diversity in forest trees. Nevertheless, non-selective historical events, such as past migration patterns and changes in population size and colonization, with associated founder effects, could also be important determinants of level and distribution of genetic diversity in forest trees (Yeh 1989).

The geographic variation patterns within forest tree species can be summarized as follows: (1) on average, predominantly outcrossing forest trees exhibit higher level of genetic diversity than inbred species (2) most of the genetic diversity is maintained within populations (3) clinal patterns, particularly associations of isozyme variation with the location of population, are often observed; thus the small G_{st} estimate should not be ignored and (4) departure of the observed heterozygosity from the Hardy-Weinberg expectation in either direction is the rule rather than the exception and this may suggest that selection and population structuring are common features of forest populations.

Population Structure and Sampling Strategies

The electrophoretic studies on amount and geographic distributions of isozyme variation in natural populations are conducive to implementing efficient and practical genetic conservation programs. Both *in situ* and *ex situ* conservation programs aim at capturing the maximum number of alleles per locus in a sample of given size. However, it is difficult to capture rare alleles unless the sample size is large and/or when many populations are sampled. What is the optimal

Table 1. Average heterozygosity per locus (H) and differentiation (G_{st}) in selected tree species

Species	H	G_{st}	Reference
<i>Pinus contorta</i>	0.160	0.041	Yeh and Layton (1979)
	0.119		Wheeler and Guries (1982)
<i>P. banksiana</i>	0.115	0.024	Dancik and Yeh (1983)
<i>P. resinosa</i>	0.000	0.000	Fowler and Morris (1977)
<i>P. rigida</i>	0.146	0.023	Guries and Ledig (1982)
<i>P. monticola</i>	0.177		Steinhoff <i>et al.</i> (1983)
<i>P. radiata</i>	0.089	0.162	Moran <i>et al.</i> (1988)
<i>Picea mariana</i>	0.229	0.010	Boyle (1985)
<i>P. glauca</i>	0.183		Cheliak <i>et al.</i> (1984)
<i>P. sitchensis</i>	0.147	0.079	Yeh and El-Kassaby (1981)
<i>Pseudotsuga menziesii</i>			
Coastal	0.155	0.026	Yeh and O'Malley (1980)
Interior	0.175	0.043	Yeh (1981)
<i>Thuja plicata</i>	0.039	0.033	Yeh (1988)
13 Inbred species	0.126	1.000	Brown (1979)

sampling strategy that captures the maximum number of alleles is the question of major interest and importance to forest geneticists. The answer to this question depends largely on the knowledge of allele distributions within and among forest populations.

In situ conservation of forest trees sets aside representative populations of sufficient size in a unique and easily managed forest ecosystem. One frequently asked question is: "Should we set aside one large population or several partially isolated subpopulations?" Based on data from a study of population structure in coastal Douglas-fir, *Pseudotsuga menziesii* (Mirb.) Franco at two locations, Adams (1981) argued that populations within each region are similarly adapted because the mean number of alleles is similar among populations within a region and because no particular population has appreciably fewer alleles than the whole region. Thus, the optimal strategy for *in situ* conservation of coastal Douglas-fir is to set aside a single, large population with the maximum number of alleles.

Although a considerable amount of genetic diversity is potentially conserved *in situ* in a large population, this does not necessarily imply that the best sampling strategy for *ex situ* conservation is to sample from just a large population. The ability to sample an allele depends on its presence and frequency in the population (Marshall and Brown 1975). Allele distribution can be described by two variables. First, alleles can be divided into those which are common (≥ 0.05) and those which are rare (< 0.05). Potentially a population can maintain many alleles. However, the number of common alleles per population is usually less than four and the remaining alleles are rare. Second, the alleles can be further categorized as to whether they are widespread over populations or localized to few populations. Adams (1981) defined any allele occurring in $\geq 25\%$ of populations as a widespread allele; otherwise it is a localized allele. This kind of classification results in four conceptual classes of alleles: (1) common, widespread; (2) common, localized; (3) rare, widespread; and (4) rare, localized (Table 2).

The first class of allele (common, widespread) is captured regardless of the sampling strategy employed. Table 3 presents the sample size required to be 95% certain in capturing at least one copy of each common allele (≥ 0.05) for five allelic profiles. These allelic profiles simulate different allelic distributions, ranging from the more even distributions

Table 2. The Marshall-Brown (1975) two-way classification of allele distribution

Allele occurrence	Allele frequency ¹	
	Common (C) ≥ 0.05	Rare (R) < 0.05
Widespread (W) $\geq 25\%$	(C,W)	(R,W)
Localized (L) $< 25\%$	(C,L)	(R,L)

¹Allele frequency refers to the frequency of an allele in a population while allele occurrence refers to the percentage of populations that have the allele in question.

Table 3. Size (N_r) of *ex situ* conserved population required to be 95% certain of obtaining at least one copy of each common allele (> 0.05) for five allele profiles (Modified from Marshall and Brown, 1975)

Allele frequency	Heterotic model	Neutral model			Mutation-selection model
		$\theta^1 = 0.5$	$\theta = 1.0$	$\theta = 2.0$	
p_1	0.25	0.76	0.63	0.49	0.80
p_2	0.25	0.20	0.23	0.22	0.05
p_3	0.25	0.00	0.09	0.12	0.05
p_4	0.25	0.00	0.00	0.07	0.05
p_s^2	0.00	0.04	0.05	0.10	0.05
N_r	19	15	37	43	80

¹ $\theta (= 4N_e\mu)$, N_e = effective population size and μ = mutation rate) measures the level of polymorphism of neutral loci.

² $p_s = 1 - (p_1 + p_2 + p_3 + p_4)$ = sum of frequencies of rare alleles.

of heterotic models at one extreme to the highly skewed distributions of mutation-selection balance model at the other extreme (Marshall and Brown 1975). The three neutral models are between the two extremes and they differ from each other only in their polymorphic levels (i.e., number of common alleles and allelic frequencies). A random sample of 15 gametes permits inclusion of two common alleles in the neutral model of $\theta = 0.5$. Under the skewed distribution of mutation-selection balance model, a random sample of 80 gametes is sufficient to obtain the required alleles with 95% certainty for each allelic profile. In the limiting case of 20 alleles each with 0.05 frequency, a random sample of 120 gametes will accomplish the same objective. The probability of including the fourth class of allele (rare, localized) in a sample of limited size is very small. Thus, obsession with conserving such alleles will lead to the impractical sampling strategies of saving everything (Zobel 1977).

The remaining two classes of alleles are relevant to formulating optimal sampling strategies. The second class of allele (common, localized) is of great importance in framing the sampling strategies. Since such alleles are locally common, the way to increase the probability of including them in a sample of limited size is to sample more populations at the cost of fewer individuals per population. Because population sampling is more expensive and time-consuming than individual-tree collection within population, the choice of this sampling strategy depends primarily on the proportion of such locally common alleles in the population. Table 4 presents the proportion of alleles in each of the four classes among populations of coastal Douglas-fir in British Columbia, Washington, and Oregon. Those locally common

Table 4. Percentage of alleles with four types of distributions (common widespread, C,W; common localized, C,L; rare widespread, R,W; and rare localized, R,L) among populations of Douglas-fir (*Pseudotsuga menziesii*)

Source	Allozyme loci			
	(C,W)	(C,L)	(R,W)	(R,L)
11 British Columbia populations ¹	18	65	11	13
4 Longview, Washington populations ²	10	70	13	10
4 Springfield Oregon populations ²	10	63	17	13

¹From Yeh and O'Malley (1981).

²From Adams (1981).

alleles in class 2 make up an appreciable proportion (11–17%) of the coastal Douglas-fir alleles. Therefore, the optimal sampling strategy for *ex situ* collections of coastal Douglas-fir would be to sample fewer trees from each of many populations to cover a wide spectrum of environments (Adams 1981). In sampling, populations exhibiting significant genetic differentiation would be obvious targets.

The locally common alleles in class 2 are also important to tree breeders. Common alleles presumably represent adaptive variants maintained in populations by some form of balancing selection (Dobzhansky 1970). Thus, they represent adaptations to particular local environments and provide tree breeders with restricted (localized) source of disease and pest resistance, and stress tolerances.

Inclusion of the rare and widespread alleles in class 3 requires the following consideration: What is the expected number of alleles included in a sample of given size? Marshall and Brown (1975) argued that the number of populations to be sampled is irrelevant to the formulation of sampling strategies in this case. There is the optimal sample size beyond which there is little or slight increase in the probability of capturing additional alleles even with a marked increase in the sample size. The optimal sample size can be estimated for the neutral model (Brown 1989). Despite the complex "sampling" structure involved in the natural populations, approximation of this model to actual data is very good for rare alleles (Chapman 1984). For neutral models, the expected number of alleles (n_a) in an equilibrium population of size N depends on N and the level of polymorphism (θ), and is

$$n_a = \int_{\frac{1}{2N}}^1 \phi(x) dx = \theta \int_{\frac{1}{2N}}^1 (1-x)^{\theta-1} x^{-1} dx \quad (1)$$

The probability density $\phi(x)$ is approximately proportional to the reciprocal of allelic frequency ($1/x$) when the allele is rare ($x < 0.05$). Thus, formula (1) can be rewritten as

$$\begin{aligned} n_a &\approx \theta \int_{\frac{1}{2N}}^{0.05} \frac{1}{x} dx + \theta \int_{0.05}^1 \frac{(1-x)^{\theta-1}}{x} dx \\ &= \theta [\log_e(2N) - 3] + \theta \int_{0.05}^1 \frac{(1-x)^{\theta-1}}{x} dx \end{aligned} \quad (2)$$

The first part in the right hand side of formula (2) is the expected number of rare alleles and the second part is the expected number of common alleles in the population. For

example, when $\theta = 1$, there are approximate three common alleles and $[\log_e(2N) - 3]$ rare alleles in a population of N individuals. For a sample of limited size (N_r), not all of these alleles can be included. The rare alleles are more likely to be lost than the common ones in the sampling. Hence, we are seeking the answer to the question: What is the expected number of alleles present in the population but not included in the sample? Following Ewen's (1972) sampling theory for selectively neutral alleles, this number is

$$n_a - n_r = \theta \log_e \left(\frac{N}{N_r} \right) - \theta(\theta-1) \left(\frac{1}{2N_r} - \frac{1}{2N} \right)$$

where n_r is the expected number of alleles included in the sample.

Brown (1989) computed an approximate 95% confidence interval for the number of alleles n_r in the sample, assuming approximation to the normal distribution. He then plotted the lower limit of n_r as a fraction of all alleles in population (n_a) against the sample size (N_r), expressed as the percent of the original population size (N). From this plot, Brown concluded that a random sample of 10% of the original population size should retain at least 70% of the alleles in the population with 95% certainty. If Douglas-fir's results (Table 4) are typical of what might be found in other conifers, then Brown's conclusion suggests that a substantial amount of rare widespread alleles should be captured in a sample consisting of 10% of the original population. Thus, 10% of the original population should be the manageable size for *ex situ* collecting and conservation of forest genetic diversity.

Decay of Genetic Diversity in Conserved Populations

Population genetic theory predicts that presence of small number of individuals sustained over many generations in a partially or completely isolated population will lead to depletion of genetic diversity. The reduction of heterozygosity can be very rapid if population size (N) is small (Fig. 1). The expected heterozygosity under random mating declines by a factor of $[1 - (2N)^{-1}]$ per generation. For example, with $N = 4$, after 10 generations of random drift, only 26% of the heterozygous genotypes at a locus is left in the population. Restricted population size also increases the probability of loss of neutral alleles that are negatively correlated with selected alleles at other loci (Namkoong and Roberds 1982).

Inbreeding accelerates the reduction in heterozygosity (Fig. 2) because it further reduces the effective population size (N_e). Under the mixed random mating and selfing model, N_e is reduced to $\frac{1}{2}(1+t)N$, where t is outcrossing rate (Strobeck 1983; Pollak 1987). Thus, the effective population size is always smaller than the actual size unless the outcrossing rate is unity. For example, with $N = 4$, and $t = 0.8$, $N_e = 3.6$. Most tree species have an outcrossing rate greater than 0.8 (Yeh 1989; Muona 1990). Therefore, inbreeding has relatively less impact on the decay of genetic diversity than the impact of small population size in forest trees.

Fitness-related traits such as survival and growth correlated with the level of heterozygosity of populations (Ledig 1986). In addition, a loss of genetic diversity in forest trees will

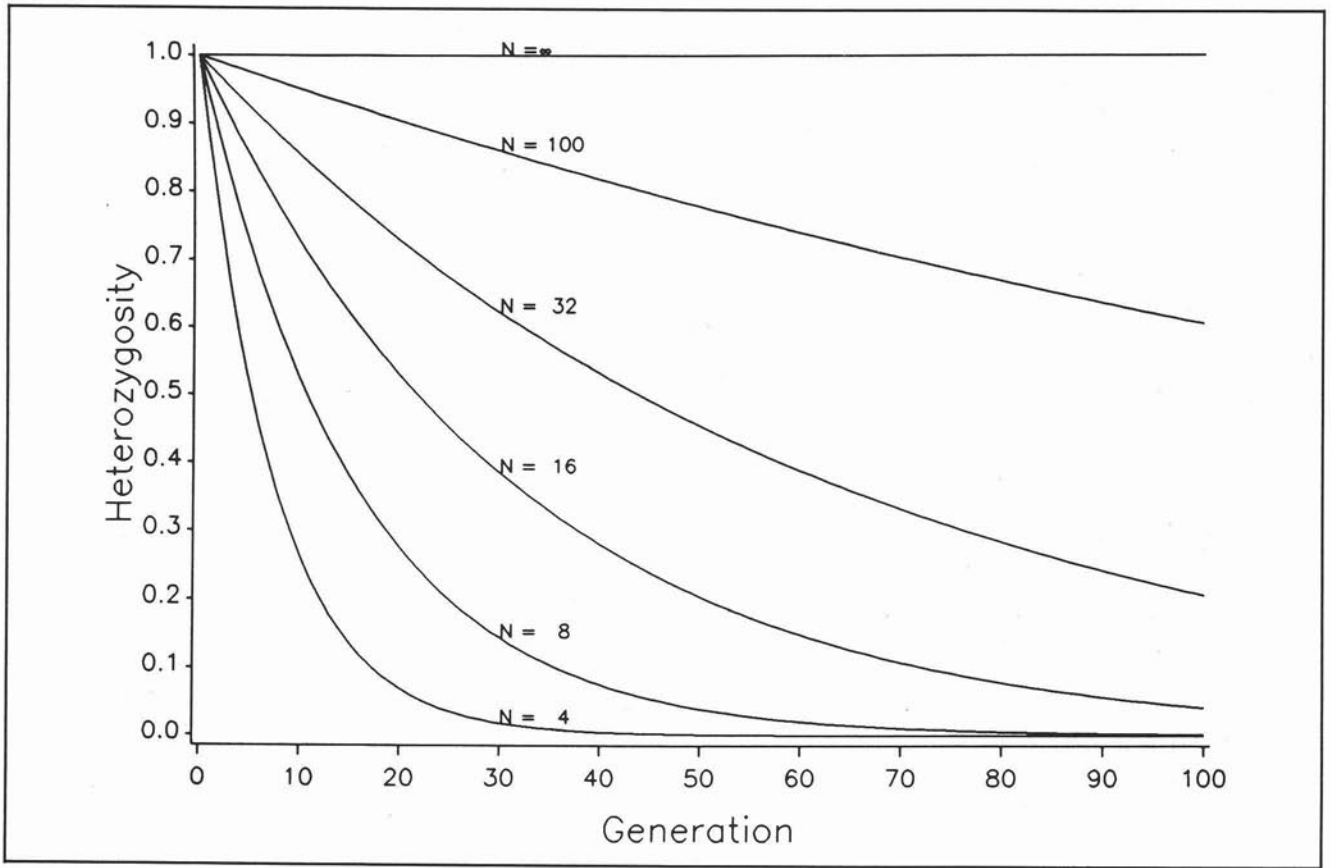


Figure 1. Decay of genetic diversity in finite populations of size (N).

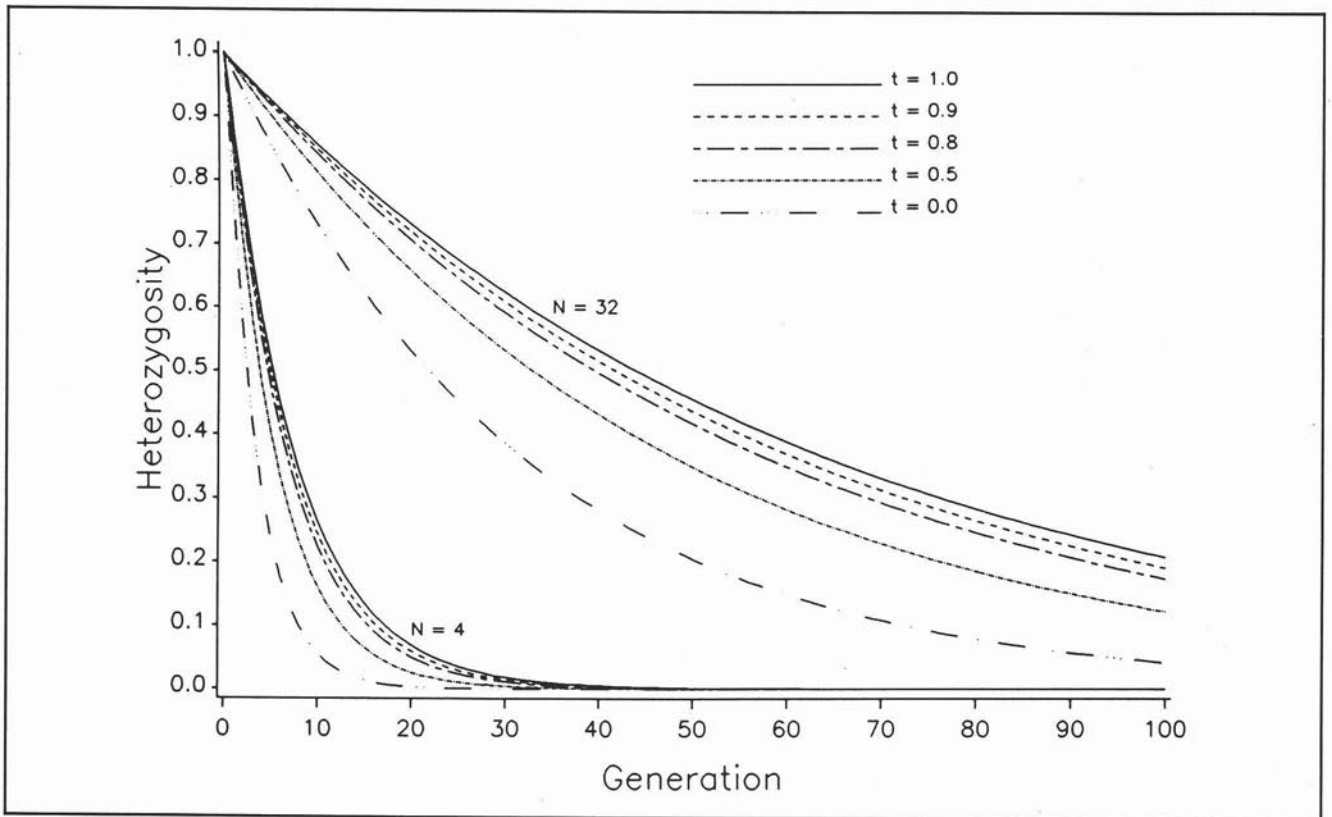


Figure 2. Effect of inbreeding on decay of genetic diversity in populations of size $N = 4$ and $N = 32$, respectively. The extent of inbreeding ranges from random mating ($t = 1$) to completely selfing ($t = 0$).

result in a loss of ability to respond to the changing requirements and changing environments (Namkoong 1986). Thus, it is important to maintain high levels of heterozygosity in the conserved forest populations.

Breeding for multiple populations rather than simply increasing population size is a suitable cure when dealing with the problem of restricted breeding population size in forest trees (Namkoong *et al.* 1980). Because of sampling variation and differences in selection criteria or uncontrolled error, the genetic gain will not be the same among the subdivided populations. However, average gain of some sub-populations is expected to be greater than the average of single large population due to variation among the sub-populations. Furthermore, different sub-populations will use and fix different alleles. Therefore, selection among sub-populations and controlled crossing among them in future breeding cycles will give greater gains and regenerate any lost but useful variation in the recombined populations.

The above prediction of long-term behaviour of small populations may prove inaccurate for the following reasons. First, random genetic drift is a stochastic process and is affected by a number of factors other than small population size and inbreeding. Second, the sampling effect was formulated under the neutral model. Heterozygosity advantages or inbreeding depression may result in the maintenance of greater variation than that expectation under the neutral model. Third, the formulation is often the approximation with respect to the assumptions and the underlying conditions. Thus, direct monitoring of genetic diversity in the managing populations is needed for forest trees.

Ex situ conserved populations and breeding populations consist of individual trees selected from natural stands. It is desirable to compare the level of genetic diversity in natural populations with that of the *ex situ* conserved populations and breeding populations. The level of genetic diversity in the *ex situ* conserved populations and breeding populations is expected to be lower than the genetic diversity in the natural populations. This is because two stages of sampling effect occurred during the establishment and maintenance of breeding populations. The first stage of sampling is a statistical effect and results in a limited number of alleles being included in the sample. The second stage of sampling is the genetic effect and it causes the decay of genetic diversity due to small population size. Genetic improvement of forest trees in Canada is still at the first generation. The loss of genetic diversity in one generation should be very small. Thus, reduced genetic diversity in the breeding populations, if any, should be primarily due to the first stage of sampling effect. Another difficulty in comparing breeding and natural populations is that the breeding population is often the result of sampling from a number of geographically differentiated natural populations. This is the sampling strategy for the locally common alleles in class 2 (see previous section). The resulting population may have even higher genetic diversity than a particular natural population.

Muona and Harju (1989) estimated heterozygosity (H), number of alleles per locus (A), multilocus outcrossing rate (t_m) and the inbreeding effective population size ($N_{e(f)}$) in two seed orchards and three natural stands of scots pine, *Pinus sylvestris* (Table 5). While the average heterozygosity in seed orchards and natural stands was similar, there was a slight decrease in the number of alleles per locus in the

Table 5. Estimation of inbreeding effective population size ($N_{e(f)}$), heterozygosity (H), number of alleles per locus (A) and multilocus outcrossing rate (t_m) in seed orchards and natural stands of *Pinus sylvestris* (From Muona and Harju (1989))

Population	$N_{e(f)}$	H	A	t_m
<i>Seed orchards</i>				
Viitaselka (25 clones)	23.2	0.276	2.2	0.976
Vilhelminmaki (28 clones)	18.5	0.259	2.4	0.978
<i>Natural stands</i>				
Karvia	—	0.271	2.5	0.986
Padajoki	—	0.293	2.7	0.878
Kangasniemi	—	0.286	2.8	0.919

seed orchards. These results corroborate the finding in *Pinus radiata* (Brown and Moran 1981) and *Pseudotsuga menziesii* (Adams 1981). The outcrossing estimate was higher in the seed orchards than in the two natural stands (Table 5). Moran *et al.* (1989) presented similar results for *Pinus radiata* and *Eucalyptus regnans*. The higher outcrossing in seed orchards was attributed to the effective reduction of family structuring in seed orchards as compared to the level of family structuring in natural stands (Muona 1990). Muona and Harju (1989) did not estimate $N_{e(f)}$ for the natural stands which they assumed 'large' due to extensive pollen flow and the lack of differentiation among populations. However, the estimate of effective population size in the two seed orchards was low. This suggests that decay of genetic diversity will likely occur in later generations.

Multilocus Population Structure

Electrophoretic studies on geographic variation patterns assume random recombinations of genes at different loci. In such cases, single-locus estimates of the genetic parameters is adequate for describing geographic variation patterns. However, many selective and non-selective evolutionary forces cause the non-random associations of alleles among loci. Therefore, it is necessary to study the joint effects of genes at different loci on population structure of forest trees (multilocus structure).

A general conclusion from electrophoretic studies on geographic variation is that population difference in the frequency of allozyme is small. This contrasts with the greater morphological differentiation among populations in provenance-family tests (Guries and Ledig 1982; Wheeler and Guries 1982). Since one important prerequisite of implementing conservation programs is to identify and characterize the differentiated populations, the concern among tree breeders is that isozymes cannot accurately identify and characterize the populations due to their limited differentiation among populations. Such concerns have led us to examine the multilocus structure and its impact on forest population differentiation.

Another reason to study multilocus structure is the need to resolve the contentious issue — just how important are co-adapted gene complexes in the adaptation of natural population to their environments and in domesticated plants for productivity? Brock (1971, 1977) argued that the notion of co-adaptation, not being supported by adequate data, is conjectural. He suggested that new variants created by mutagenesis could be an attractive and potentially more efficient alternative to genetic conservation. Mutagenesis and techniques of molecular biology have the potential to increase

the genetic diversity of populations. At this time, however, it is far from clear how co-adapted gene complexes, if they do exist, can be manipulated by mutation and molecular techniques. Therefore, it is vital for tree breeders or conservation geneticists to establish the factual evidences of the existence of co-adaptation in both natural and breeding populations.

Co-adaptation also implies that the interlocus combination of alleles found in different ecogeographical regions will be nonrandom and correlated with local habitat (Allard 1977). This leads to a different conservation strategy than the one for the random combination of alleles. Later we will discuss the implication of multilocus structure on tree improvement.

The key to analyzing multilocus structure is to study gametic disequilibrium. We use the term "gametic disequilibrium" rather than "linkage disequilibrium" to refer to non-random associations between independents as well as linked loci. Consider a pair of loci *A* and *B*, gametic disequilibrium between allele *A_i* at locus *A* and allele *B_j* at locus *B* is measured by

$$D_{ij} = g_{ij} - p_i q_j \quad (3)$$

where g_{ij} = observed frequency of gamete $A_i B_j$

p_i = frequency of allele A_i

q_j = frequency of allele B_j

and

$$0 \leq D_{ij} \leq \text{Min}[p_i(1-q_j), (1-p_i)q_j]$$

$$-\text{Min}[p_i q_j, (1-p_i)(1-q_j)] \leq D_{ij} \leq 0$$

This definition of gametic disequilibrium can be extended to the case of multiple loci (Bennett 1954). We will describe three different approaches to analyze multilocus structure in natural populations of forest trees: (a) correlation between zygotic loci (Smouse and Neel 1977), (b) multilocus heterozygosity (Brown *et al.* 1980; Brown and Feldman 1981), and (c) discrete multivariate log-linear analysis (Feinberg 1980).

(a) Discriminant Analysis of Population Differentiation in Lodgepole Pine

Yeh *et al.* (1985) used the correlation between zygotic loci (Smouse and Neel 1977) to study differentiation in seventeen populations of lodgepole pine, *Pinus contorta* spp. *latifolia* from western North America. The first step was to code the allozyme profiles of parent trees. For example, genotypes *AA*, *Aa* and *aa* were represented by the vector $Y' = (1, 1/2, 0)$. The next step was the construction of the matrix of correlations between pairs of loci (*R*) for each population:

$$R = \begin{pmatrix} 1 & r_{12} & \dots & r_{1k} \\ r_{21} & 1 & \dots & r_{2k} \\ \cdot & \cdot & \cdot & \cdot \\ r_{k1} & r_{k2} & \dots & 1 \end{pmatrix}$$

The genetic expectation of the pairwise correlation between the *i*th and *j*th loci [$E(r_{ij})$] is well known (Hill and Robertson 1968, Wright 1969) and is:

$$E(r_{ij}) = \frac{D_{ij}}{\sqrt{p_i(1-p_i)q_j(1-q_j)}}$$

Finally, canonical discriminant functions were established on the basis of the correlation matrices.

The first two discriminant functions in lodgepole pine accounted for 38% of the total variation in the 20 polymorphic loci (Yeh *et al.* 1985). Scatter plots of the 17 populations on the axes of these two functions showed that while southern populations were grouped into a cluster, northern populations were disjointly distributed over the whole domain. Thus, there was greater differentiation among the northern populations. This is most consistent with the population structure of lodgepole pine. The northern populations often occurred in small, more isolated stands and must adapt to narrow ecological niches. The southern populations occurred in large, dense stands with potential for extensive gene flow and must adapt to the relatively stable environments. The fact that single-locus analysis of population structure in two earlier studies (Yeh and Layton 1979; Dancik and Yeh 1983) revealed only slight population differentiation of the same species demonstrates the importance of studying gametic disequilibria to determine population structure and argue for the multilocus approach to differentiate among populations of lodgepole pine.

The more differentiated northern populations of lodgepole pine would require a different sampling strategy than the less differentiated southern populations. In the north, alleles in populations are common and localized and the optimal conservation strategy is to sample as many populations as possible. The alleles in the southern population are common or rare but widespread. Thus, a random sample of 50 trees per site optimize the latitudinal and altitudinal contrasts was the recommended sampling strategy for genetic conservation of southern populations in lodgepole pine (Yeh *et al.* 1985).

(b) Multilocus Structure in Three Sub-species of *Pinus contorta*

Studies of gametic disequilibrium in natural populations typically require the computation of individual pair-wise disequilibrium coefficients (Brown *et al.* 1980). However, with an increase in number of loci and alleles per locus, the number of pair-wise disequilibrium coefficients quickly multiplies beyond comprehension. For all practical purposes, it is desirable to have a set of summary statistics that adequately describes the multilocus structure within and among subdivided populations. One such set of statistics was based on the distribution $f(K)$ of the number of heterozygous loci in two randomly chosen gametes (Brown *et al.* 1980). Brown and Feldman (1981) further demonstrated that the variance of $f(K)$ could be partitioned into the single-locus and the two-locus effects of population structure.

We have studied the influence of multilocus associations on the population structure in three sub-species of lodgepole pine using a modification of Brown *et al.* (1980) and Brown and Feldman (1981). We detected significant gametic disequilibria in all 66 populations from a set of modified summary statistics that assent to individual configuration of gametic disequilibrium phases. Two-locus gametic disequilibria accounted for over sixty percent of the average variance and about a third of the total

Table 6. Components of total and average variances in number of heterozygous loci in three sub-species of *Pinus contorta*.

Component	<i>latifolia</i>	<i>murrayana</i>	<i>contorta</i>
<i>Single-locus effect</i>			
Mean diversity	2.128	2.037	1.962
Variance of diversity	0.103	0.147	0.188
Wahlund effect	0.032	0.045	0.068
Total	2.263	2.229	2.218
<i>Two-locus effect</i>			
Mean disequilibria (MD)	0.605	0.658	0.291
Wahlund effect (WC)	0.310	0.788	0.931
AI (MD × WC)	0.002	0.005	0.003
Variance of disequilibria	0.276	0.157	0.142
Covariation of interaction	3.426	2.274	2.549
Average variance	6.437	5.131	4.947
% two-locus effect	66.9	60.3	60.3
Total variance	3.180	3.680	3.443
% two-locus effect	28.8	39.4	35.6

variance within subspecies (Table 6). Moderate two-locus Wahlund effect and high covariation in the interaction of disequilibria and the two-locus Wahlund effect suggested both founder effect and population subdivision as the major causes for the observed gametic disequilibria in natural populations of lodgepole pine.

These findings have important implications for breeding of lodgepole pine. First, the pooling of trees from populations in gametic disequilibria into a single breeding population will disrupt the apparent disequilibria (Yeh 1989). This increases the rate of decay of additive genetic variance in the breeding population. Under such conditions, there is selective advantage in subdividing the breeding population, with repeated cycles of within and between sub-population selection to exploit "multiple peak epistasis" (Yeh 1989). Second, pooling of selected trees into a single breeding population from sub-populations each differing in allele frequencies is problematic when estimating genetic covariances among relatives for the construction of selection indices, which requires the assumption of gametic equilibrium (Baker 1986). Thus, even if gametic equilibrium exists in natural populations of lodgepole pine, conventional selection indices cannot be constructed for progeny from crosses among trees from sub-populations each having different allele frequencies.

(c) Log-linear Analysis of Multilocus Structure

The collapsing of multilocus data to generate summary statistics to describe forest populations inevitably results in loss of information on the multilocus structure. Log-linear analysis (Fienberg 1980) is an alternative in studying the structure of the entire multilocus array in natural and domesticated populations in agricultural species (Jana and Pietrzak 1988; Zhang *et al.* 1990; Yang *et al.* 1991). It involves the following steps: (1) construct an n-way contingency table for n loci (2) fit different models of multilocus associations to the observed data (3) test for goodness-of-fit of the different models, and (4) select the best-fitting model by the step-wise procedures and/or residual analysis. Using this procedure, Yeh and Morgan (1987) detected five of forty-five significant two-locus associations in pollen gamete pool of a coastal Douglas-fir natural population. The residual analysis indicated the apparent associations were attributable to the relative excess of the rare paternal gametes. Since the ten

polymorphic loci studied were physically unlinked, the apparent gametic disequilibria were caused by biased representation of parental genotypes in the sampled pollen pool (Yeh and Morgan 1987).

Conclusions

Compared with agricultural crops, Canadian forest trees are wild plants at their earliest stages of domestication. Thus, the majority of Canada's forest genetic resources is being maintained *in situ* in their native and stable ecosystems at the pre-breeding age. However, as tree improvement activities intensify and when ecosystems are being altered by human and non-human interventions at an alarming rate, *ex situ* conservation of genetic resources will become a common practice in forest management. We analyzed the genetic consequences of the entire *in situ* and *ex situ* conservation process and reached the following conclusions:

(1) Loss of forest genetic diversity in Canada will mainly occur during the initial population establishment and during the managing process. There is no clear evidence that selective breeding has reduced the level of genetic diversity in the breeding populations because tree breeding is in the first generation in Canada. Thus, sampling during population establishment is the only contributing factor that has caused any potential loss of forest genetic diversity.

(2) *Ex situ* conserved or breeding populations are typically small, isolated, and maintained in controlled environments. While it is possible to maximize the genetic gain in *ex situ* conserved or breeding populations in the short-term, generations of selective breeding of such genetically restricted populations will lead to reduced genetic diversity. This restricts the potential for genetic adjustment of populations to meet future demands.

(3) Genetic diversity of *in situ* conserved populations is dynamically maintained and it evolves with the changing environments. Therefore, *in situ* conservation of genetic diversity is essential for the long-term success and survival of tree improvement.

(4) While single-locus population structure provides a preliminary guide for prescribing a genetic conservation strategy, an enhanced understanding of multilocus structure of forest populations will optimize the conservation strategy and minimize the loss of genetic diversity.

Acknowledgements

This work was supported in part by grants from the Natural Sciences and Engineering Research Council of Canada, A2282 and IRC 8607, and from the Alberta Forest Products Association and the Alberta Forest Service to F.C.Y.

References

- Adams, W.T. 1981. Population genetics and gene conservation in Pacific Northwest conifers. In: G.G.E. Scudder and J.L. Reveal (Eds.). *Evolution Today, Proceedings of the Second International Congress of Systematic and Evolutionary Biology*. pp. 401-415. Hunt Institute for Botanical Documentation, Carnegie-Mellon University, Pittsburgh, PA, USA.
- Allard, R.W. 1977. Coadaptation in plant populations. In: A. Muhammed, R. Aksel and R.C. von Borstel (Eds.). *Genetic Diversity in Plants*. pp. 223-231. Plenum Press, New York.

- Baker, R.J. 1986.** Selection Indices in Plant Breeding. CRC Press, Boca Raton, FL. USA.
- Bennett, J.H. 1954.** On the theory of random mating. *Ann. Eugen.* 18: 311-317.
- Boyle, T.J.B. 1985.** The mating system and population structure of black spruce in central New Brunswick and its influence on tree improvement strategy. Ph.D. dissertation, University of New Brunswick.
- Brand, D.G. 1990.** Advances in Canadian forest research: an introduction. *Can. J. For. Res.* 20: 373-374.
- Brock, R.B. 1971.** The role of induced mutation in plant improvement. *Radiat. Bot.* 11: 181-196.
- Brock, R.B. 1977.** Prospects and perspectives in mutation breeding. *In: A. Muhammed, R. Aksel and R.C. von Borstel (Eds.). Genetic Diversity in Plants.* pp. 117-132. Plenum Press, New York.
- Brown, A.H.D. 1979.** Enzyme polymorphism in plant populations. *Theor. Pop. Biol.* 15: 1-42.
- Brown, A.H.D. 1989.** The case of core collections. *In: A.D.H. Brown, O.H. Frankel, D.R. Marshall and J.T. Williams (Eds.). The Use of Plant Genetic Resources.* pp. 136-156. Cambridge University Press, Cambridge, U.K.
- Brown, A.H.D. and M.W. Feldman. 1981.** Population structure of multilocus associations. *Proc. Natl. Acad. Sci. USA* 78: 5913-5916.
- Brown, A.H.D. and G.F. Moran. 1981.** Isozymes and the genetic resources of forest trees. *In: M.T. Conkle (Ed.). Proceedings of Symposium on Isozymes of North American Forest Trees and Insects.* pp.1-10. U.S. For. Serv. Tech. Rep. PSW-48.
- Brown, A.H.D., M.W. Feldman and E. Nevo. 1980.** Multilocus structure of natural populations of *Hordeum spontaneum*. *Genetics* 96: 523-536.
- Chapman, C.G.D. 1984.** On the size of a genebank and the genetic variation it contains. *In: J.H.W. Holden and J.T. Williams (Eds.). Crop Genetic Resources: Conservation and Evaluation.* pp. 102-119, George Allen & Unwin, London.
- Cheliak, W.M., J.A. Pitel and G. Murray. 1984.** Population structure and the mating system of white spruce. *Can. J. For. Res.* 15: 301-308.
- Dancik, B.P. and F.C. Yeh. 1983.** Allozyme variability and evolution of lodgepole pine (*Pinus contorta* var. *latifolia*) and jack pine (*P. banksiana*) in Alberta. *Can. J. Genet. Cytol.* 25: 57-64.
- Dobzhansky, Th. 1970.** Genetics of the Evolutionary Process. Columbia University Press, New York.
- Ewen, W.J. 1972.** The sampling theory of selectively neutral alleles. *Theor. Pop. Biol.* 3: 87-112.
- Fienberg, S.E. 1980.** The Analysis of Cross-classified Categorical Data. MIT Press, Cambridge, MA, USA.
- Fowler, D.P. and R.W. Morris. 1977.** Genetic diversity in red pine: evidence for low gene heterozygosity. *Can. J. For. Res.* 7: 343-347.
- Frankel, O.H. 1977.** Philosophy and strategy of genetic conservation in plants. 3rd World Consult. For. Tree Breeding. Canberra 1: 1-11.
- Frankel, O.H. and M.E. Soulé. 1981.** Conservation and Evolution. Cambridge University Press, Cambridge, London.
- Guries, R.P. and F.T. Ledig. 1982.** Genetic diversity and population structure in pitch pine (*Pinus rigida* Mill.). *Evolution* 36: 387-402.
- Hamrick, J.L. and M.J.W. Godt. 1990.** Allozyme diversity in plant species. *In: A.H.D. Brown, M.T. Clegg, A.L. Kahler, and B.S. Weir (Eds.). Plant Population Genetics, Breeding, and Genetic Resources.* pp. 43-63. Sinauer Associates Inc., Sunderland, MA, USA.
- Hill, W.G. and A. Robertson. 1968.** Linkage disequilibrium in finite populations. *Theor. Appl. Genet.* 38: 226-231.
- Jana, S. and L.N. Pietrzak. 1988.** Comparative assessment of genetic diversity in wild and primitive cultivated barley in a centre of diversity. *Genetics* 119: 981-990.
- Ledig, F.T. 1986.** Conservation strategies for forest gene resources. *For. Ecol. Manage.* 14: 77-90.
- Loveless, M.D. and J.L. Hamrick. 1984.** Ecological determinants of genetic structure in plant populations. *Ann. Rev. Ecol. Syst.* 15: 65-95.
- Marshall, D.R. and A.H.D. Brown. 1975.** Optimum sampling strategies in genetic conservation. *In: O.H. Frankel and J.G. Hawkes (Eds.). Crop Genetic Resources for Today and Tomorrow.* pp. 53-80. Cambridge University Press, Cambridge, London.
- Moran, G.F. and S.D. Hopper. 1987.** Conservation of genetic resources of rare and widespread eucalypts in remnant vegetation. *In: D.A. Saunders, G.W. Arnold, A.A. Burbidge, and A.J.M. Hopinks (Eds.). Nature Conservation: the Role of Remnants of native vegetation.* pp. 151-162. Surrey Beatty, Sydney.
- Moran, G.F., J.C. Bell and K.G. Eldridge. 1988.** The genetic structure and the conservation of the five natural populations of *Pinus radiata*. *Can. J. For. Res.* 18: 506-514.
- Moran, G.F., J.C. Bell and A.R. Griffin. 1989.** Reduction in levels of inbreeding in a seed orchard of *Eucalyptus regnans* compared with natural populations. *Silv. Genet.* 38: 32-36.
- Muona, O. 1990.** Population genetics in forest tree improvement. *In: A.H.D. Brown, M.T. Clegg, A.L. Kahler and B.S. Weir (Eds.). Plant Population Genetics, Breeding, and Genetic Resources.* pp. 282-298. Sinauer Associates Inc., Sunderland, MA, USA.
- Muona, O. and A. Harju. 1989.** Effective population size, genetic variability, and mating system in natural stands and seed orchards of *Pinus sylvestris*. *Silvae Genet.* 38: 221-228.
- Namkoong, G. 1986.** Genetics and the forests of the future. *Unasylva* 38: 2-18.
- Namkoong, G. and J.H. Roberds. 1982.** Short-term loss of neutral alleles in small-population breeding. *Silvae Genet.* 31: 1-6.
- Namkoong, G., R.D. Barnes and J. Burley. 1980.** A Philosophy of Breeding Strategy for Tropical Forest Trees. Commonwealth Forestry Institute, University of Oxford, Oxford.
- Office of Technology Assessment (OTA). 1987.** Technologies to maintain biological diversity. U.S. Government Printing Office, Washington, D.C.
- Pollak, E. 1987.** On the theory of partially inbreeding finite populations I. Partial selfing. *Genetics* 117: 353-360.
- Smouse, P.E. and J.V. Neel. 1977.** Multivariate analysis of gametic disequilibrium in the Yanomama. *Genetics* 85: 733-752.
- Strobeck, C. 1983.** The use of coefficients of identity to study random drift. *Lecture Notes on Biomathematics.* 52: 1-13.
- Steinhoff, R.J., D.G. Joyce and L. Fins. 1983.** Isozyme variation in *Pinus monticola*. *Can. J. For. Res.* 13: 1122-1132.

Wheeler, N.C. and R.P. Guries. 1982. Population structure, genic diversity and morphological variation in *Pinus contorta* Dougl. Can. J. For. Res. 12: 595-606.

Wright, S. 1969. Evolution and the Genetics of Populations. Vol. 2. The Theory of Gene Frequencies. University of Chicago Press, Chicago.

Yang, R.-C., S. Jana and J.M. Clarke. 1991. Phenotypic diversity and associations of some potential drought-response characters in durum wheat. Crop Sci. 31: 1484-1491.

Yeh, F.C. 1981. Analysis of gene diversity in some species of conifers. In M.T. Conkle (Ed.). Proceedings of Symposium on Isozymes of North American Forest Trees and Insects. pp. 48-52. U.S. For. Serv. Tech. Rep. PSW-48.

Yeh, F.C. 1988. Isozyme variation of *Thuja plicata* (Cupressaceae) in British Columbia. Biochem. Syst. Ecol. 16: 373-377.

Yeh, F.C. 1989. Isozyme analysis for revealing population structure for use in breeding strategies. In G.L. Gibson, A.R. Griffin, and A.C. Matheson (Eds.). Breeding tropical trees: Population structure and genetic improvement strategies in clonal and seedling forestry. pp. 119-131. Oxford Forestry Institute, Oxford, UK and Winrock International, Arlington, VA, USA.

Yeh, F.C. and Y.A. El-Kassaby. 1981. Enzyme variation in natural populations of Sitka spruce (*Picea sitchensis*). Genetic variation patterns among trees from 10 IUFRO provenances. Can. J. For. Res. 10: 415-422.

Yeh, F.C. and C. Layton. 1979. The organization of genetic variability in central and marginal populations of lodgepole pine *Pinus contorta* spp. *latifolia*. Can. J. Genet. Cytol. 21: 487-503.

Yeh, F.C. and K. Morgan. 1987. Mating system and multilocus associations in a natural population of *Pseudotsuga menziesii* (Mirb.) Franco. Theor. Appl. Genet. 73: 799-808.

Yeh, F.C. and D.M. O'Malley. 1980. Enzyme variation in natural populations of Douglas-fir (*Pseudotsuga menziesii* (Mirb.) Franco) from British Columbia. 1. Genetic variation patterns in coastal populations. Silvae Genet. 29: 83-92.

Yeh, F.C., W.M. Cheliak, B.P. Dancik, K. Illingworth, D.C. Trust and B.A. Pryhitka. 1985. Population differentiation in lodgepole pine, *Pinus contorta* spp. *latifolia*: a discriminant analysis of allozyme variation. Can. J. Genet. Cytol. 27: 210-218.

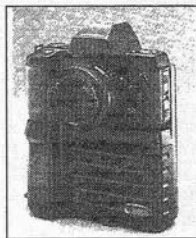
Zhang, Q., M.A. Saghai Maroof and R.W. Allard. 1990. Worldwide pattern of multilocus structure in barley determined by discrete log-linear multivariate analyses. Theor. Appl. Genet. 80: 121-128.

Zobel, B. 1977. Gene conservation — as viewed by a forest tree breeder. For. Ecol. Manage. 1: 339-344.

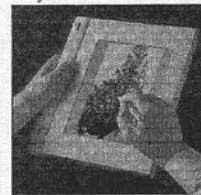
PenARBOR™ For tree taper analysis

PenARBOR™ is a new computerized system, based on image processing, for semi-automatic and interactive measurement of stem properties. It works with images taken with a special high resolution (1500x1200 pixels) digital camera which can store images in it or transfer them rapidly to a computer through a SCSI bus (the same bus used for connecting hard disks to computers).

Images taken by the camera can be analysed in the field with a graphical pen-based computer based on a 386 or 486 processor or in-house with a conventional desktop computer (Macintosh or for IBM compatible running Microsoft Windows 3.1). A version for GO Corp Penpoint operating system is also under study.



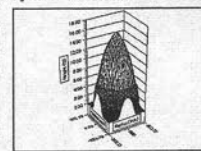
The camera is lightweight, compact and uses standard 35 mm lenses so that field of view and viewing distances from the tree can be easily changed by the operator. Because it produces images of excellent quality, the camera can be used for other applications like tele-detection or any field or laboratory measurement (like leaf area index or wood cells).



Tree properties like:

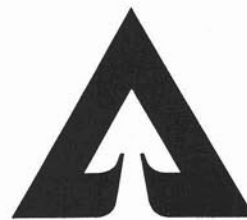
- Tree height & volume
- Stem widths & curvature
- Crown shape and size

can be done by simple pen or mouse gestures over an image. A utility software is also provided for visualization of data produced by PenARBOR™.



Régent Instruments Inc.

P.O. Box 75, Lac-St-Charles, Québec, CANADA, G0A 2H0
Tel & Fax: 418-849-0244 (Régent Guay)
in collaboration with:
Forestry Canada, Québec Region (Dr. C.H. Ung)



Weyerhaeuser Canada

A CIF/IFC Corporate Sustaining Member