Theor Appl Genet (1993) 87:568–576

THEORETICAL AND APPLIED GENETICS

Multilocus structure in *Pinus contorta* Dougl.

R.-C. Yang, F. C. Yeh

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

Received: 25 October 1992 / Accepted: 17 May 1993

Abstract. We studied isozyme variation at 21 loci in 66 populations from three subspecies of *Pinus contorta* Dougl.; 35 in spp. *latifolia*, 20 in spp *contorta* and 11 in spp. *murrayana*. The objectives were to assess gametic disequilibria and multilocus structure. There was considerable differentiation of allele frequencies at 19 polymorphic loci across the 66 populations and within the subspecies. Allele frequencies at many loci correlated with geographic variables. Genetic variability varied considerably among populations within subspecies but the subspecies means were similar. The mean number of polymorphic loci and the mean heterozygosity over 19 polymorphic loci were, respectively, 13 and 0.194 in *latifolia*, 12 and 0.196 in *murrayana*, and 12 and 0.180 in *contorta*. The mean heterozygosity correlated with longitude and altitude across the 66 populations and with latitude in *latifolia*. Gametic disequilibria were evident in 40 populations; 29 in *latifolia*, eight in *murrayana* and three in *contorta*. Gametic disequilibria correlated with latitude across the 66 populations and with longitude in *latifolia*. The single-locus F_{ST} averaged 0.0339 in *latifolia*, 0.0567 in *murrayana*, and 0.0764 in *contorta*. The multilocus F_{STM} was 0.1227 in *latifolia*, 0.2926 in *murrayana*, and 0.3328 in *contorta*. Multilocus Wahlund and founder effects, migration patterns, and natural selection, probably played significant roles in generating and maintaining the multilocus genetic structure in *P. contorta* in general and the subspecies *latifolia* in particular.

Key words: *Pinus contorta* Dougl. – Isozymes – Gametic disequilibria – Multilocus structure

Communicated by P. M. A. Tigerstedt
Correspondence to: F. C. Yeh

Introduction

Pinus contorta Dougl. is a widely distributed and variable North American conifer with four regional forms meriting subspecies rank; *P. contorta* spp. *bolanderi*, *P. contorta* spp. *contorta*, *P. contorta* spp. *latifolia* and *P. contorta* spp. *murrayana* (Critchfield 1980). Of these, *P. contorta* spp. *latifolia* has been the subject of many isozyme surveys due to its commercial importance. Allele frequencies in *P. contorta* spp. *latifolia* from British Columbia, Alberta and the Yukon correlated strongly with the geographic variability (Yeh and Layton 1979; Dancik and Yeh 1983; Yeh et al. 1985). However, the contribution of multilocus structuring to the population subdivision was not assessed in these studies.

Multilocus structure is important when there is gametic disequilibrium or nonrandom association of nonalleles between the independent as well as the linked loci (Hedrick et al. 1978; Barton and Clark 1990). Empirical studies showed that gametic disequilibrium was the result of many evolutionary mechanisms, including epistatic selection (e.g., Clegg et al. 1972; Brown et al. 1977), chance from population subdivision and founder effect (e.g., Smouse and Neel 1977; Muona and Szmidt 1985; Waller and Knight 1989), and genetic hitchhiking (Laurie-Ahlberg and Weir 1979). When natural populations are in gametic equilibrium, *F*-statistics (Wright 1951; Nei 1973) and its extension to multiple alleles and loci (Weir and Cockerham 1984; Long 1986) are adequate descriptions of the population differentiation. However, the presence of gametic disequilibrium in natural populations would require the specification of additional parameters for the new sources of genetic variation, such as the multilocus Wahlund effects (Feldman and Christiansen 1975; Sinnock 1975).

The prediction for outcrossing plants such as conifers has been that any initial gametic disequilibria will rapidly decay without strong epistatic selection. Nevertheless, this is a prediction of long-term behavior. In the establishment of a forest after a fire, for example, there could initially be gametic disequilibria that slowly decay for closely-linked genes particularly if there was nonrandom mating. This consideration is fitting for *P. contorta* since most of its forests are thought to be of fire origin (Critchfield 1980). We collected allozyme data from 66 populations in three subspecies of *P. contorta*; 35 in *latifolia*, 20 in *contorta*, and 11 in *murrayana*. Our objectives were to investigate the intensity of gametic disequilibria and to explain the importance of multilocus associations to the population subdivision in this conifer.

Materials and methods

Seed collection

The 66 populations were bulk-stand and single-tree collections of *P. contorta* spp. *contorta P. contorta* spp. *murrayana* and *P. contorta* spp. *latifolia*, made within their natural range in western North America (see Table 1). Each population contained several hundred trees, from which 15 to 35 cone-bearing trees a minimum of 46 metres apart were randomly sampled. Each population represented a locality that was uniform with respect to climate, landform, soil, and vegetation. Cones were stored in sacs in a locked lath-house for less than 6 months until air dried and then opened by placing in a kiln (52.2 °C) for 12 h. Seed extraction was completed in a vibratory cone-shaker, and the seeds cleaned in a hand-operated "Clipper" and regulated air column.

Electrophoretic procedures

Protein extraction and horizontal starch-gel electrophoresis were as described (Yeh and Layton 1979; Yeh and O'Malley 1980). We analyzed 75 megagametophytes (female gametophytes) per population for variation in 14 enzymes; acid phosphatase (*APH*, EC 3.1.3.2), aconitase (*Aco*, EC 4.2.1.3), alcohol dehydrogenase (*Adh*, EC 1.1.1.1), aspartate aminotransferase (*Aat*, EC 2.6.1.1), diaphorase (*Dia*, EC 1.6.4.3), glucose-6-phosphate dehydrogenase (*G6p*, EC 1.1.1.49), glutamate dehydrogenase (*Gdh*, EC 1.4.1.3), isocitrate dehydrogenase (*Idh*, EC 1.1.1.42), malate dehydrogenase (*Mdh*, EC 1.1.1.37), malic enzyme (*Me*, EC 1.1.1.40), peptidase (*Pep*, EC 3.4.13.1), phosphoglucose isomerase (*Pgi*, EC 5.3.1.9), phosphoglucomutase (*Pgm*, 2.7.5.1), phosphogluconate dehydrogenase (*Pgd*, EC 1.1.1.44). Gels were prepared with Electrostarch, lot 307 (Electrostarch Co., Madison, Wis.). Interpretation of banding patterns followed Yeh and Layton (1979) for *P. contorta* spp. *latifolia*. The enzyme was identified by its abbreviation and a hyphenated numeral in decreasing order of anodal mobility if controlled by multiple loci. The most common allozyme for the species at each locus was designated *100*. Additional allozymes were given numerical values according to migration relative to the *100* allozyme. The inheritance of the isozyme was confirmed from the segregation of the progeny of heterozygous maternal trees (Yeh, unpublished) and by comparison with the same enzymes in *P. contorta* spp. *latifolia* (Yeh and Layton 1979) and other conifers (Yeh and El-Kassaby 1980; Yeh and O'Malley 1980).

Measuring multilocus associations

We extended the measure of multilocus association in humans (Smouse and Neel 1977) to haplotype data in *P. contorta*. We followed the assumptions of Smouse and Neel (1977) but haploid data need not consider the departure from panmixia. Consider a diallelic locus, A, with a frequency array $p_1A_1 + (1-p_1)A_2$ and define the random variable X_1 that takes the values 1 and 0 for haplotypes A_1 and A_2, respectively. Similarly, a second diallelic locus, B, with the frequency array $p_2B_1 + (1-p_2)B_2$ has the random variable X_2 with values 1 and 0 for haplotypes B_1 and B_2, respectively. Then, the joint gametic distribution between locus A and locus B is:

Gamete	X_1	X_2	Frequency
A_1B_1	1	1	g_{11}
A_1B_2	1	0	g_{12}
A_2B_1	0	1	g_{21}
A_2B_2	0	0	g_{22}

with $g_{11} + g_{12} + g_{21} + g_{22} = 1$. From this distribution, we have the means (μ_1, μ_2), variances (σ_1^2, σ_2^2), covariance (σ_{12}), and correlation (ρ_{12}) as follows:

$$\begin{aligned} &\mu_1 = E(X_1) = p_1 && \mu_2 = E(X_2) = p_2 \\ &\sigma_1^2 = \mathrm{Var}(X_1) = p_1(1-p_1) && \sigma_2^2 = \mathrm{Var}(X_2) = p_2(1-p_2) \\ &\sigma_{12} = \mathrm{Cov}(X_1, X_2) \\ &\quad = g_{11} - p_1p_2 = D_{12} && \rho_{12} = \frac{D_{12}}{\sqrt{p_1(1-p_1)p_2(1-p_2)}} \end{aligned} \quad (1)$$

The covariance between the variables X_1 and $X_2(\sigma_{12})$ is the gametic disequilibrium (D_{12}) between locus A and locus B. If there are m such random variables defined for the m loci, then the m variances and $m(m-1)$ covariances are the diagonal and the off-diagonal elements of an $m \times m$ covariance matrix (Σ). i.e.,

$$\Sigma = \begin{pmatrix} p_1(1-p_1) & D_{21} & \cdots & D_{m1} \\ D_{12} & p_2(1-p_2) & \cdots & D_{m2} \\ \cdot & \cdot & \cdot & \cdot \\ D_{1m} & D_{2m} & \cdots & p_m(1-p_m) \end{pmatrix}$$

$$= \begin{pmatrix} \sigma_1^2 & \sigma_{21} & \cdots & \sigma_{m1} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{m2} \\ \cdot & \cdot & \cdot & \cdot \\ \sigma_{1m} & \sigma_{2m} & \cdots & \sigma_m^2 \end{pmatrix} \quad (2)$$

Note that $D_{jl} = D_{lj}$ and $\sigma_{jl} = \sigma_{lj}$ for all j and l. The corresponding correlation matrix $(\boldsymbol{P})$ is

$$\boldsymbol{P} = \mathrm{diag}\{\sigma_l^{-1}\}\,\Sigma\,\mathrm{diag}\{\sigma_l^{-1}\} = \begin{pmatrix} 1 & \rho_{21} & \cdots & \rho_{m1} \\ \rho_{12} & 1 & \cdots & \rho_{m2} \\ \cdot & \cdot & \cdot & \cdot \\ \rho_{1m} & \rho_{2m} & \cdots & 1 \end{pmatrix} \quad (3)$$

where ρ_{jl} (for $j \neq l, j, l = 1, 2, \ldots, m$) is given in equation (1).

We can extend this procedure to the case of multiple alleles. Suppose there are r alleles at locus A and s alleles at locus B. We define $(r-1)$ variables for the r alleles at locus A and $(s-1)$ random variables for the s alleles at locus B. The correlation matrix, $\boldsymbol{P}$, which consists of four partitioned matrices, is of the form:

$$\boldsymbol{P} = \begin{pmatrix} \boldsymbol{P}_{11} & \boldsymbol{P}_{12} \\ \boldsymbol{P}_{21} & \boldsymbol{P}_{22} \end{pmatrix} \quad (4)$$

here $\boldsymbol{P}_{11}$ contains $(r-1)(r-2)$ intralocus correlations for locus A, $\boldsymbol{P}_{22}$ contains $(s-1)(s-2)$ intralocus correlations for locus B, and $\boldsymbol{P}_{12}$ or its reciprocal $\boldsymbol{P}_{21}$ contains $(r-1)$ $(s-1)$ interlocus correlations. With multiple loci, the random variables for the successive loci are simply appended and this generates a large correlation matrix.

To simplify interpretation of the above correlation structure, we defined an "effective correlation" (ρ_e) as follows. For any particular correlation matrix $\boldsymbol{P}$ of rank k, there exists an equicorrelation matrix $\boldsymbol{P}_e$ (Mardia et al. 1979) such that the determinants of these two matrices are equal, i.e., $|\boldsymbol{P}| = |\boldsymbol{P}_e|$, where

$$\boldsymbol{P}_e = \begin{pmatrix} 1 & \rho_e & \cdots & \rho_e \\ \rho_e & 1 & \cdots & \rho_e \\ \cdot & \cdot & \cdot & \cdot \\ \rho_e & \rho_e & \cdots & 1 \end{pmatrix} = (1-\rho_e)\boldsymbol{I}_k + \rho_e \boldsymbol{J}_k \tag{5}$$

The $\boldsymbol{I}_k$ is an identity matrix of order k and $\boldsymbol{J}_k$ is a $k \times k$ matrix of 1s. Since a monomorphic locus has zero column rank, it follows that k is the number of polymorphic loci and satisfies the condition $k \leqq m$.

The relationship between effective correlation (ρ_e) and the determinant of matrix $\boldsymbol{P}$ is set by the following polynomial (Mardia et al. 1979):

$$|\boldsymbol{P}| = |\boldsymbol{P}_e| = (1-\rho_e)^{k-1}[1+(k-1)\rho_e] \tag{6}$$

Clearly, when $\rho_e = 0, |\boldsymbol{P}| = 1$ and when $\rho_e = 1, |\boldsymbol{P}| = 0$. Numerical calculations of the intermediate values (not presented here) revealed that an increase in the effective correlation and the number of polymorphic loci induced a decrease in $|\boldsymbol{P}|$. For example, with two polymorphic loci ($k = 2$), $|\boldsymbol{P}|$ decreased from 0.99 to 0.19 when ρ_e increased from 0.1 to 0.9. With $\rho_e = 0.2, |P|$ decreased from 0.96 to 0.00032 when k increased from 2 to 100. In general, there is a single solution on this interval since for a given k the polynormial in ρ_e is a monotonically decreasing function on the $[0, 1]$ interval. Therefore, all values of ρ_e reported would necessarily be positive although the acceptable values of ρ_e would lie on the $[-1/(k-1), 1]$ interval.

The estimation of $\boldsymbol{P}$ was obtained by computing the sample correlation matrix $\boldsymbol{R}$ for a set of variables $\{X_1, X_2, \ldots, X_m\}$ from a sample of size n. An asymptotic χ^2 test criterion (Kendall and Stuart 1966) inspected the complete independence of these variables as follows:

$$-\left(n - \frac{2m+11}{6}\right)\log|\boldsymbol{R}| \sim \chi^2_{df = m(m-1)/2} \tag{7}$$

and with the assumption that multi-nominally distrubuted variates approximate a multivariate normal distribution (Smouse and Neel 1977).

Population structure

Analysis of variance of allele frequencies. Following Cockerham (1969) and Weir (1990), we considered a diallelic locus A with alleles A and a, and defined an indicator variable x to be 1 for allele A and 0 for allele a. Long (1986) extended the definition to the multiallelic case. For a mixture of I populations, the total variance in the populations could be apportioned into the within- and among-population components by considering the following linear model:

$$x_{ij} = p + a_i + w_{ij} \tag{8}$$

where x_{ij} is the observed value of the j^{th} haplotype in the i^{th} subpopulation and p is the frequency of allele A. The effect a_i, is for the populations, and the effect w_{ij} is for the haplotypes within the populations. From an analysis of variance, the mean square among the populations (MSA) and the mean square within the populations (MSW) have the following expectations:

$$E(MSA) = p(1-p)(1-F_{ST}) + np(1-p)F_{ST} \tag{9}$$

$$E(MSW) = p(1-p)(1-F_{ST})$$

where n is the number of haplotypes in each population (n is the weighted number if populations have uneven numbers of haplotypes) and F_{ST} is a measure of population differentiation. Thus, the F_{ST} can be estimated as:

$$\hat{F}_{ST} = \frac{MSA - MSW}{MSA + (n-1)MSW} \tag{10}$$

as shown in Weir (1990, pp 145–152). The ratio of MSA/MSW in the analysis of variance has an approximate F-distribution (Cockerham 1969; Long 1986) and is appropriate for testing the null hypothesis $H_0: F_{ST} = 0$.

Multilocus population structure. We conducted a multivariate analysis of variance to partition the total mean squares and mean cross-products matrix for a set of 19 loci in a mixed pool of I populations into the among- ($\boldsymbol{MSA}$) and the within-population ($\boldsymbol{MSW}$) components. Here I is 35 in *latifolia*, 11 in *murrayana*, 20 in *contorta* and 66 in the whole species. Testing for the null hypothesis that the frequencies of the multilocus haplotypes are the same in all the populations makes use of the test statistic, $|\boldsymbol{MSW}|/|\boldsymbol{MSA} + \boldsymbol{MSW}|$, which is distributed approximately as Wilk's Λ. Therefore, Rao's (1965) F-approximation to Wilk's Λ distribution was used for the hypothesis testing. From the relationships equivalent to those given in equation (9), the matrix of multilocus F_{ST} statistics was estimated as:

$$\boldsymbol{F}_{ST} = [\boldsymbol{MSA} + (n-1)\boldsymbol{MSW}]^{-1/2}(\boldsymbol{MSA} - \boldsymbol{MSW})][\boldsymbol{MSA} + (n-1)\boldsymbol{MSW}]^{-1/2} \tag{11}$$

where $\boldsymbol{MSA}$ and $\boldsymbol{MSW}$ are the mean square and the mean cross-product matrices among and within populations, respectively.

There is no universal agreement on how best to define a single measurable statistic that adequately describes the multilocus population differentiation using the matrix $\boldsymbol{F}_{ST}$ in equation (11). For a symmetric square matrix, such as the matrix $\boldsymbol{F}_{ST}$, three common functions of the matrix can be defined: (1) the arithmetic mean of all eigenvalues or $(1/m)$tr $\boldsymbol{F}_{ST}$ (tr refers the trace of the matrix); (2) the geometric mean of all eigenvalues or $|\boldsymbol{F}_{ST}|^{1/m}$; and (3) the largest eigenvalue (λ_1). If both $\boldsymbol{MSA}$ and $\boldsymbol{MSW}$ are diagonal (i.e., no gametic disequilibrium), then $\boldsymbol{F}_{ST}$ is the diagonal and the diagonal elements of $\boldsymbol{F}_{ST}$ are the single-locus $\boldsymbol{F}_{ST}$-statistics in equation (10). Long (1986) did not consider the effect of gametic disequilibrium so that his multilocus $\boldsymbol{F}_{ST}$ $[(1/m)\text{tr } \boldsymbol{F}_{ST}]$ was an average of the $\boldsymbol{F}_{ST}$-statistics over individual loci. However, assuming interlocus independence among natural populations was probably unwarranted. For example, population subdivision due to a multilocus Wahlund effect could generate a substantial amount of gametic disequilibrium (Nei and Li 1973; Feldman and Christiansen 1975; Sinnock 1975). The geometric mean of all eigenvalues of matrix $\boldsymbol{F}_{ST}$ ($|\boldsymbol{F}_{ST}|^{1/m}$) is likely to be a better measure of multilocus differentiation. However, its estimation critically depends on the among-population covariance matrix being positive definite. In this study, the among-population covariance matrix was estimated from substraction of the mean squares and the mean corss-product matrices among and within populations, and was not positive definite in each subspecies. The number of negative eigenvalues out of 19 was nine in *murrayana*, seven in *contorta*, four in *latifolia* and one in the 66 populations. Clearly, we could not use all the eigenvalues to construct the multilocus population differentiation as the estimates of matrix $\boldsymbol{F}_{ST}$ were not positive definite

in the subspecies. Thus, the best measure of multilocus population differentiation (F_{STM}) in this study was the largest eigenvalue of matrix $\boldsymbol{F}_{ST}$ (λ_1):

$$\hat{F}_{STM} = \hat{\lambda}_1 = e_1' \hat{\boldsymbol{F}}_{ST} e_1 \quad (12)$$

subject to the condition that $e_1' e_1 = 1$ and where e_1 is the eigenvector corresponding to the largest eigenvalue. Further, if $F_{ST1}, F_{ST2}, \ldots, F_{STm}$ are the single-locus $\boldsymbol{F}_{ST}$-statistics for the m loci, then it can be shown that $F_{STM} \geqq \max(F_{ST1}, F_{ST2}, \ldots, F_{ST})$, with equality if, and only if, there is complete among-population interlocus independence (i.e., $\boldsymbol{F}_{ST}$ is a diagonal matrix). F_{STM} is an adequate descriptor of the multilocus population differentiation since the leading eigenvalue extracts the maximum proportion of information from the matrix $\boldsymbol{F}_{ST}$ in each subspecies.

Results

Allele frequency and genetic variability

Two (*Pep-1* and *Pep-2*) of twenty-one loci assayed were monomorphic. The remaining 19 (*Aat-1, Aat-2, Aco, Adh, Aph, Dia-2, Dia-3, Gdh, G6p, Idh, Mdh-1, Mdh-2, Mdh-3, Mdh-4, Me, Pgi, Pgm, 6Pg-1, 6Pg-2*) were polymorphic in at least some populations. The *100* allele predominated at each locus ($\geqq 50\%$) in each population, and one or more of the other alleles occurred in low frequency. The numbers of alleles at the polymorphic loci over the 66 populations were 5, 8, 4, 6, 6, 7, 3, 3, 3, 3, 5, 8, 8, 7, 4, 4, 4, and 5, respectively. The allele frequencies at the polymorphic loci are not presented due to the large size of the data.

Allele frequencies differed among populations in *P. contorta*. The contingency G^2 test for homogeneity across the 66 populations suggested differentiation at 16 of 19 polymorphic loci; *Mdh-1*, *Mdh-2* and *6Pg-2* were homogeneous across the populations. Allele frequencies were also heterogeneous across populations within the subspecies. Loci that were homogeneous within subspecies numbered four (*Gdh; Mdh-1; Mdh-2; 6Pg-2*) in *latifolia*, five (*Dia-3; Gdh; Mdh-1; Mdh-2; 6Pg-2*) in *murrayana*, and one (*Mdh-1*) in *contorta*.

The allele frequencies and geographic variables showed clinal variation patterns. When all the 66 populations were analyzed, 7 (*Aat-1; Aat-2; Aph; Dia-2; Dia-3; Mdh-4; Pgi*) of 19 loci displayed significant linear correlations with latitude, longitude or altitude. The allele frequencies within the subspecies also correlated with geographic variables at eight loci, of which four (*Aco, Adh, Mdh-2, 6Pg-1*) were effects specific to the subspecies. The loci *Aat-1, Aat-2, Dia-2, Dia-3*, and *Mdh-2* correlated with latitude, longitude and altitude in *latifolia*. The loci *Aat-1* and *Aco* in *murrayana* exhibited clinal variation patterns with latitude, longitude and altitude. In *contorta*, the loci *Idh* and *6Pg-1* correlated with latitude, longitude and altitude.

Genetic variability varied considerably among populations within the species, but the average of subspecies was similar (Table 1). The mean number of polymorphic loci and mean heterozygosity over the 19 polymorphic loci were, respectively, 13 land 0.194 in

Table 1. Geographic location, genetic diversity and multilocus association in *Pinus contorta*

Population	Latitude (°N)	Longitude (°W)	Altitude (m)	Polymorphic loci (%)	Average heterozygosity	Multilocus association*	
						$\lvert \boldsymbol{R} \rvert$	r_e
latifolia							
1	51.06	121.40	1814	63.16	0.19	0.263	0.204[a]
2	52.00	121.12	991	78.95	0.19	0.150	0.208[b]
3	51.59	123.45	1059	73.68	0.17	0.180	0.208[a]
4	54.39	127.03	518	59.90	0.17	0.204	0.247[c]
5	53.01	123.14	983	73.68	0.20	0.037	0.315[c]
6	56.02	122.05	725	63.16	0.16	0.315	0.187
7	57.00	122.24	1113	68.42	0.19	0.164	0.229[c]
8	58.40	124.10	762	73.68	0.21	0.110	0.244[c]
9	59.59	128.33	640	68.42	0.16	0.148	0.238[c]
10	61.10	129.20	884	73.68	0.19	0.057	0.288[c]
11	62.14	136.18	671	63.16	0.19	0.151	0.254[c]
12	63.18	136.28	876	59.90	0.17	0.329	0.198[a]
13	60.41	136.11	747	68.42	0.17	0.126	0.250[c]
14	59.48	133.47	789	63.16	0.16	0.248	0.210[c]
15	57.29	130.13	815	78.95	0.21	0.052	0.277[c]
16	50.02	118.34	1137	78.95	0.19	0.077	0.253[c]
17	50.49	116.26	1173	73.68	0.18	0.096	0.254[c]
18	49.34	116.04	1661	68.42	0.20	0.173	0.225[c]
19	54.38	127.26	1005	73.68	0.18	0.093	0.255[c]
20	49.54	118.12	579	73.68	0.16	0.224	0.190

Table 1. *(Continued)*

Population	Latitude (°N)	Longitude (°W)	Altitude (m)	Polymorphic loci (%)	Average heterozygosity	Multilocus association*	
						$\|R\|$	r_e
21	55.33	122.33	732	53.63	0.16	0.264	0.241[c]
22	53.52	121.48	838	63.16	0.16	0.249	0.209[c]
23	58.39	124.46	1137	68.42	0.17	0.129	0.249[c]
24	53.16	117.09	1204	73.68	0.18	0.140	0.226[c]
25	51.01	115.02	1501	68.42	0.19	0.178	0.223[c]
26	49.26	114.25	1379	78.95	0.19	0.123	0.222[c]
27	50.43	119.27	1524	63.16	0.16	0.228	0.218[c]
28	50.42	119.11	777	78.95	0.17	0.061	0.268[c]
29	49.04	120.46	1128	53.63	0.17	0.344	0.211[c]
30	49.04	120.55	1539	63.16	0.18	0.165	0.246[c]
31	47.47	120.56	762	59.90	0.15	0.465	0.157
32	46.04	121.27	1219	78.95	0.20	0.235	0.175
33	45.30	110.45	2134	68.42	0.17	0.198	0.214[b]
34	46.45	110.30	2134	73.68	0.17	0.291	0.168
35	48.45	116.30	1219	73.68	0.18	0.402	0.139
Mean				68.95	0.18	0.191	0.226
murrayana							
36	45.23	121.50	549	63.16	0.17	0.440	0.151
37	45.18	121.45	1280	73.68	0.17	0.299	0.166
38	44.08	121.38	1707	84.21	0.19	0.128	0.206[b]
39	43.19	121.39	1676	78.95	0.21	0.217	0.181
40	41.16	121.55	1219	63.16	0.17	0.391	0.164
41	41.13	122.30	2134	52.63	0.15	0.433	0.182[b]
42	41.11	120.10	2179	42.11	0.14	0.657	0.150
43	39.53	121.07	1646	57.90	0.17	0.512	0.145
44	37.10	119.12	2256	57.90	0.18	0.476	0.154
45	36.06	118.32	2164	63.16	0.19	0.325	0.183
46	34.18	116.59	2316	47.37	0.15	0.457	0.193[a]
Mean				62.11	0.17	0.394	0.171
contorta							
47	53.39	132.04	23	84.21	0.17	0.306	0.144
48	53.49	132.08	23	73.68	0.17	0.152	0.220[c]
49	52.03	126.43	152	47.37	0.13	0.601	0.149
50	50.31	126.36	15	57.90	0.16	0.367	0.186[b]
51	54.27	128.35	198	57.90	0.15	0.177	0.260[c]
52	49.09	123.06	6	78.95	0.17	0.218	0.180
53	48.34	123.34	350	78.95	0.17	0.141	0.212[a]
54	54.04	128.41	76	36.84	0.16	0.820	0.112
55	47.38	124.18	30	52.63	0.15	0.475	0.169
56	46.26	124.03	15	59.90	0.18	0.507	0.146
57	45.43	123.56	15	42.11	0.14	0.676	0.144
58	45.13	123.57	15	63.16	0.15	0.297	0.192[b]
59	44.34	124.04	15	59.90	0.15	0.308	0.205[a]
60	43.50	124.09	15	73.68	0.20	0.208	0.196
61	43.30	124.14	15	78.95	0.20	0.150	0.208[b]
62	42.46	124.31	15	68.42	0.16	0.326	0.170
63	42.15	124.24	15	63.16	0.19	0.379	0.167
64	41.50	123.53	1097	57.90	0.18	0.332	0.197[a]
65	40.47	124.13	15	52.63	0.17	0.550	0.148
66	47.25	122.40	76	78.95	0.18	0.212	0.183
Mean				63.16	0.17	0.360	0.179
Grand mean				65.79	0.17	0.276	0.203

* $|R|$ is the determinant of the sample correlation matrix R, which is the unbiased estimate of the population correlation matrix, P; r_e is the effective correlation as defined in equation (6)

[a,b,c] Significant at $P \leq 0.1$, 0.05 and 0.01, respectively

latifolia, 12 and 0.196 in *murrayana*, and 12 and 0.180 in *contorta*. The distribution of mean heterozygosity correlated with longitude and altitude across the 66 populations and with latitude in *contorta*.

Multilocus associations

With up to eight alleles per locus in the same or in different populations, the number of intralocus and interlocus correlations was exceedingly large. Therefore, it was necessary to reduce the data to manageable proportions. Following Clegg et al. (1972), we created two allelic classes per locus per population from the *100* allele and a "synthetic" allele consisting of all the other alleles combined. With only two alleles per locus, the correlation matrix consisted of *only* the interlocus correlations. This simplified our interpretation of the multilocus structure *P. contorta* from the analysis of the correlation matrix.

In the last two columns of Table 1 we present the determinants of the estimated correlation matrices, $|\boldsymbol{R}|$, and their corresponding effective correlations (r_e). An apparent and expected relationship between $|\boldsymbol{R}|$ and r_e revealed that those high interlocus correlations corresponded to a low determinant of the estimated correlation matrix, and vice versa. A significant linear correlation between $|\boldsymbol{R}|$ and r_e was found within subspecies ($r = -0.859$, $P < 0.001$ in *latifolia*; $r = -0.706$, $P < 0.016$ in *murrayana*; $r = -0.829$, $P < 0.001$ in *contorta*) and across the 66 populations ($r = -0.846$, $P < 0.001$). Therefore, r_e is a reliable descriptor of gametic disequilibrium in *P. contorta*.

The $|\boldsymbol{R}|$ values were significant in 40 populations, suggesting the presence of gametic disequilibria (Table 1). Of these, 29 were in *latifolia*, three in *murrayana* and eight in *contorta*. The $|\boldsymbol{R}|$ values averaged 0.191 in *latifolia*, 0.394 in *contorta* and 0.276 in *murrayana*. Judging from the intensity and the number of significant $|\boldsymbol{R}|$ and the level of significance, there were greater gametic disequilibria in *latifolia* than in *murrayana* and *contorta*. Gametic disequilibria correlated with latitude and longitude only in *latifolia*. This is not surprising since a large proporation of the populations in *murrayana* and *contorta* were in gametic equilibria. Gametic disequilibria also correlated with altitude across the 66 populations.

Population structure

Single-locus population differentiation ($\boldsymbol{F}_{ST}$) at the 19 loci (Table 2) differed from zero ($P < 0.01$) in *latifolia*, varying from 0.0146 (*Gdh*) to 0.0749 (*Aco*) and averag-

Table 2. Single-locus (F_{ST}) and multilocus population differentiation ($\boldsymbol{F}_{STM}$) and multilocus contributions (loadings) of individual loci to population differentiation in *Pinus contorta*

Locus	Subspecies						Total	
	latifolia		*murrayana*		*contorta*		F_{ST}	Loading
	F_{ST}	Loading	F_{ST}	Loading	F_{ST}	Loading		
Aat-1	0.021[a]	−0.150	0.030[a]	−0.046	0.078[a]	0.029	0.052[a]	−0.155
Aat-2	0.059[a]	0.615	0.022[a]	0.078[a]	0.064[a]	−0.225	0.061[a]	−0.149
Aco	0.075[a]	−0.239	0.066[a]	0.304	0.111[a]	−0.245	0.084[a]	0.336
Adh	0.023[a]	0.174	0.113[a]	−0.323	0.075[a]	0.276	0.055[a]	0.274
Aph	0.033[a]	0.150	0.037[a]	−0.207	0.098[a]	0.028	0.072[a]	0.231
Dia-2	0.021[a]	0.010	0.043[a]	0.205	0.047[a]	0.172	0.033[a]	−0.035
Dia-3	0.028[a]	0.190	0.018[a]	0.038	0.105[a]	0.144	0.050[a]	0.013
Gdh	0.015[a]	0.126	0.054[a]	0.043	0.083[a]	−0.051	0.043[a]	−0.053
G6p	0.071[a]	−0.335	0.093[a]	−0.179	0·077[a]	0.303	0.075[a]	0.305
Idh	0.034[a]	0.096	0.001	0.113	0.155[a]	−0.539	0.067[a]	−0.357
Mdh-1	0.016[a]	−0.115	0.000	−0.020	0.000	−0.007	0.013[a]	0.050
Mdh-2	0.024[a]	−0.153	0.030[a]	0.143	0.053[a]	0.119	0.037[a]	0.057
Mdh-3	0.031[a]	0.276	0.134[a]	0.364	0.124[a]	0.340	0.085[a]	0.217
Mdh-4	0.044[a]	−0.267	0.130[a]	0.035	0.051[a]	0.125	0.060[a]	−0.070
Me	0.016[a]	−0.069	0.057[a]	−0.021	0.067[a]	0.070	0.037[a]	−0.081
Pgi	0.025[a]	0.192	0.159[a]	−0.670	0.062[a]	0.139	0.107[a]	0.460
Pgm	0.046[a]	−0.233	0.023[a]	−0.120	0.029[a]	0.074	0.037[a]	0.020
6Pg-1	0.044[a]	−0.040	0.038[a]	0.152	0.148[a]	−0.444	0.093[a]	−0.458
6Pg-2	0.019[a]	−0.197	0.030[a]	0.143	0.028[a]	−0.066	0.022[a]	−0.047
Mean	0.034		0.057		0.076		0.057	
F_{STM}		0.123		0.239		0.333		0.195

[a] Significant at $P \leqq 0.01$

ing 0.0339. In *murrayana*, the F_{ST} for *Idh* and *Mdh-1* did not differ from zero; the 17 significant estimates ranged from 0.0183 (*Dia-3*) to 0.1587 (*Pgi*), with a mean of 0.0567. The F_{ST} for *Mdh-1* in *contorta* was not different from zero. The 18 significant F_{ST} s ranged from 0.0278 (*6pg-2*) to 0.1545 (*Idh*), averaging 0.0764. The F_{ST} estimates across the 66 populations were significantly different from zero ($P < 0.01$) at all loci, varying from 0.0130 (*Mdh-1*) to 0.1069 (*Pgi*) and averaging 0.0570.

The estimate of multilocus population differentiation (F_{STM}) across the 19 loci was 0.1227 in *latifolia*, 0.2926 in *murrayana*, 0.3328 in *contorta* and 0.1954 across the 66 populations (Table 2). These estimates were more than three times the estimates of the F_{ST} averages (Table 2). However, the order of F_{STM} (*latifolia* < *Murrayana* < *contorta*) conformed to the order of the F_{ST} averages.

The relative contribution of the loci to multilocus structuring was assessed by an inspection of their eigenvectors (loadings), and there was considerable variation among the loci (Table 2). Although most loci contributed disparately to the loading, *Aco*, *Adh*, *G6p*, *Mdh-3* and *Pgi* consistently scored high in all the subspecies. The highest loading to the F_{STM} was *Aat-2* (0.615) in *latifolia*, *Pgi* (−0.670) in *murrayana*, and *Idh* (−0.539) in *contorta*. As noted before, the allele frequencies at these loci correlated with geographic variables.

Discussion

There was considerable differentiation of allele frequencies among populations within the subspecies and across the 66 populations. The allele frequencies at many loci correlated with the geographic variables. Correlations between altitude and *Aat-1* and between latitude and *6Pg-1* were described for *P. contorta* spp. *latifolia* (Yeh and Layton 1979). In fact, clinal patterns of the differentiation associated with geographical variables often are the rule, rather than the exception, for most conifers (Stern and Roche 1974).

The level of isozyme variability in this study is comparable to the range-wide survey (Wheeler and Guries 1982), confirming the existence of a high level of isozyme variability in *P. contorta* relative to most other conifers. High levels of isozyme variability in conifers are well documented (Loveless and Hamrick 1984). Occurrence over broad environmental spectra, large population size, longevity, high fecundities, predominance of outcrossing, and extensive gene flow via pollen and seed dispersal, are some plausible biotic factors responsible for the high levels of isozyme variability in conifers (Loveless and Hamrick 1984).

The mean heterozygosity increased eastward and with altitudes across the 66 populations and decreased to the north in *latifolia*. Neutral alleles with migration and hybridization of the genotypes could produce these isozyme patterns. However, for many loci chosen at random, such correlations should not predominate. The 19 loci could not be a random sample of the genome, but certainly there was a random element in how they were chosen. Whether these nonrandom variation patterns were the result of drift or selection on the isozymes, or the coadapted gene complexes that they mark, is unknown. From the patterns so far found, predictions about what genotypes should be found in other areas of the species distribution can probably be made. This is important in certifying the seed sources for reforestation and in designing gene conservation strategies.

The single-locus F_{ST} averages suggest the following ranking of within-subspecies population differentiation: *latifolia* < *murrayana* < *contorta* (Table 2). Yeh and Layton (1979) gave Nei's (1973) G_{ST} at 4.11% for *latifolia* from British Columbia and the Yukon. Our G_{ST} (data not shown) averaged over 19 loci for *latifolia* was comparable, at 4.55%. The F_{ST} (3.39%) was less than G_{ST} because G_{ST} was estimated using sample size (n) while F_{ST} was estimated using degrees of freedom ($n-1$). Nevertheless, our F_{ST} is larger than the 1.8% reported for *latifolia* from Alberta (Dancik and Yeh 1983). In general, our results corroborate the previous finding that over 90% of the total gene diversity in *P. contorta* resided within local populations. This high level of within-population isozyme variation has been found in many conifers (see Yeh 1989 for review). Near complete outcrossing (Yeh and Layton 1979; Perry and Dancik 1986; Epperson and Allard 1987) and extensive gene flow (Epperson and Allard 1989) might in part be responsible for the relatively low differentiation in *P. contorta*. Gametic disequilibria were detected in 40 of 66 populations. The level of gametic disequilibria in the northern populations was greater than in the southern populations. Yeh et al. (1985) suggested that southern populations occurred in large and dense stands with the potential for extensive gene flow, whereas northern populations occurred in small and isolated stands and probably adapted to narrow ecological niches. Therefore, further north in the range of *P. contorta*, the importance of nonadaptive forces, such as genetic drift and founder effect, and adaptive forces, such as environmental changes in temperature, precipitation, and photoperiod on the multilocus structure, have probably increased.

Gametic disequilibria contributed greatly to the multilocus structuring in *P. contorta*. The F_{STM} was much larger than its expected value, $\max(F_{ST1}, \ldots, F_{ST2}, \ldots, F_{STm})$, under complete interlocus independence in all the subspecies (Table 2). For example, the estimate of F_{STM} for *latifolia* (0.123) was almost twice its expected value (i.e., F_{ST} for *G6p* at 0.071). This increase in the level of population differentiation is not unexpected since a multilocus approach to analyzing

isozyme data takes into account the interaction among the loci (Yeh et al. 1985).

Epperson and Allard (1987) detected significant pairwise disequilibria mainly between four tightly-linked loci (*Aat-1/Per-1/Per-2/Pgi-2*) in *latifolia* from Washington. If gametic disequilibria in our sampled populations were mainly between the linked instead of the independent loci, we would expect most two-locus haplotype correlations to develop between tightly-linked loci and to be common to all the subspecies (data not shown due to the size of matrix of pairwise correlations). The proportion of significant two-locus haplotype correlations among the populations were 30/171 in *latifolia*, 34/153 in *contorta*, and 14/136 in *murrayana*. A large proportion of the significant two-locus haplotype correlations were unique to the subspecies: 21/30 in *latifolia*, 22/34 in *contorta*, and 9/14 in *murrayana*. Only two (*Dia-2/Dia-3*; *Gdh/G6p*) were common to the subspecies.

Murrayana and *latifolia* had the same six two-locus correlations (*Dia-2/Dia-3*, *Gdh/G6p*, *Aat-2/Mdh-3*, *Dia-3/Mdh-4*, *Mdh-4/Pgm* and *Gdh/6Pg-2*). There were seven shared two-locus correlations (*Aat-1/Aat-2*; *Aat-2/G6p*; *Aat-2/Dia-2*; *Dia-2/Dia-3*; *Dia-3/G6p*; *Gdh/G6h*; *Adh/6Pg-1*) between *murrayana* and *contorta*. The number of significant two-locus correlations common to *latifolia* and *contorta* was four (*Aco/Dia-2*; *Aat-1/Dia-3*; *Dia-2/Dia-3*; *Gdh/G6p*). Of these shared two-locus gametic disequilibria, the confirmed linkage groups are *Aat-1/Aat-2*, *Aat-1/Dia-3* and *Gdh/G6p* with recombination frequencies 0.262, 0.170 and 0.234, respectively (Yeh, unpublished data). Based on the linkage map for *P. contorta* (Yeh, unpublished data), most ($\geqq 80\%$) two-locus haplotype correlations were not between the linked loci. Therefore, the gametic disequilibria in this sample of *P. contorta* were mostly unique and developed between independent loci.

Gametic disequlibria have not been well-investigated in outcrossing forest trees. However, the prevalence of gametic disequilibria in this survey and other trees examined to-date (e.g., Boyle 1985; Muona and Szmidt 1985; Roberds and Brotschol 1985; Yeh and Morgan 1987) is contrary to the notion that gametic disequilibria are insignificant in outcrossing plants. One cause of gametic disequilibria in conifers is restricted effective population size (founder effect) due to uneven seed production, partial selfing, and uneven numbers of male and female parents contributing to the pollen pool (Muona and Szmidt 1985). Much of the *latifolia* forests are thought to be of fire origin and, jointly with the serotinous nature of their cones (Critchfield 1980), extreme bottlenecks could follow. Since gametic disequilibria generated by severe bottlenecks would remain in the population for a long time even after an increase in population size (Avery and Hill 1979), restricted effective population size (founder effect) is a plausible cause of gametic disequilibria in *P. contorta* in general and the subspecies *latifolia* in particular.

The nonrandom associations of alleles at different loci in *P. contorta* could also be caused in part by the biased representation of the parental genotypes in sampled populations. Allele frequencies at different loci varied markedly from one sampled population to the other. Thus, a substantial amount of added gametic disequilibria would follow upon the mixing of such populations. This is the result of a multilocus Wahlund effect (Nei and Li 1973; Feldman and Christiansen 1975; Sinnock 1975) and can be an important determinant for generating and maintaining the gametic disequilibria in this conifer. While a high level of heterozygosity in *P. contorta* and other conifers requires the sampling of more than one megagametophyte per tree to maximize the number of sampled gametes from segregation, the likelihood of sampling identical gametes within trees could contribute to the association of gene frequencies at different loci. With given technical and financial constraints, there must exist an optimal allocation for sampling gametes within and among trees to minimize sampling effects on generating gametic disequilibria. This issue remains to be investigated.

Whatever the causes of gametic disequilibira in sampled *P. contorta* populations, our findings have significance to its genetic improvement through selective breeding. When gametic disequilibria are as prevalent and differ among the populations, the pooling of select trees from populations in gametic disequilibria into one breeding populations would disrupt the disequilibria (Muona 1982; Yeh 1989). This reduces the breeding value of the progenies and increases the rate of decay of additive genetic variance in the breeding population (Karlin 1978). The pooling of select trees into one breeding population from sub-populations each of which differs in allele frequencies is also problematic when estimating the genetic covariances among the relatives for construction of the selection indices, which requires the assumption of gametic equilibria (Baker 1986). Thus, even if gametic equilibria exist in natural populations of *P. contorta*, the conventional selection indices are not applicable for progenies from crosses among parent trees from sub-populations differing in allele frequencies.

Knowledge of multilocus structuring in *P. contorta* also aids with the planning of its genetic conservation. The optimal sampling strategy for *in situ* conservation of the multilocus structure may differ from that of the individual genes as the sample size for conserving single genes (Brown and Moran 1981) may not be large enough for conserving the gene-combinations. Thus, a sampling theory comparable to the single-locus model is needed to address this issue. Our analysis of the multilocus structure in *P. contorta* strengthens the thesis that evolutionary forces such as the multilocus

Wahlund and founder effects, migration patterns, and natural selection, have probably played significant roles in generating and maintaining the multilocus genetic structure in this conifer.

Acknowledgements. We thank A. H. D. Brown, J. H. Roberds and R. Wright for helpful comments on an earlier draft of this paper, K. Illingworth for sampling the populations and P. Kagawa and D. C. Trust for laboratory assistance. This research was supported in part by grants from the Natural Sciences and Engineering Research Council of Canada, A2282 and IRC 8607, British Columbia Ministry of Forests, and Petawawa National Forestry Institute to F.C.Y.

References

Avery PJ, Hill WG (1979) Distribution of linkage disequilibrium with selection and finite population size. Genet Res 33:29–48

Baker RJ (1986) Selection indices in plant breeding. CRC Press, Boca Raton, Florida, USA

Barton N, Clark A (1990) Population structure and processes in evolution. In: Wöhrmann K, Jain SK (eds) Population biology. Springer-Verlag, Berlin, pp 115–173

Boyle TJB (1985) The mating system and population structure of black spruce in central New Brunswick and its influence on tree improvement strategy. PhD dissertation, University of New Burnswick, Canada

Brown AHD, Moran GF (1981) Isozymes and the genetic resources of forest trees. In: Conkle MT (ed) Proc Symp on Isozymes of North American Forest Trees and Insects. US For Serv Tech Rep PSW-48, pp 1–10

Brown AHD, Nevo E, Zohary D (1977) Association of alleles at esterase loci in wild barley *Hordeum spontaneum.* Nature 268:430–431

Clegg MT, Allard RW, Kahler AL (1972) Is the gene the unit of selection? Evidence from two experimental populations. Proc Natl Acad Sci USA 69:2474–2478

Cockerham CC (1969) Variance of gene frequencies. Evolution 23:72–84

Critchfield WB (1980) Genetics of lodgepole pine. USDA For Serv Res Pap WO-37

Dancik BP, Yeh FC (1983) Alllozyme variability and evolution of lodgepole pine (*Pinus contorta* var *latifolia*) and jack pine (*P. banksiana*) in Alberta. Can J Genet Cytol 25:57–64

Epperson BK, Allard RW (1987) Linkage disequilibrium between allozymes in natural populations of lodgepole pine. Genetics 115:341–352

Epperson BK, Allard RW (1989) Spatial autocorrelation analysis of the distribution of genotypes within populations of lodgepole pine. Genetics 121:369–377

Feldman MW, Christiansen FB (1975) The effect of population subdivision on two loci without selection. Genet Res 24:151–162

Hedrick P. Jain S, Holden L (1978) Multilocus systems in evolution. Evol Biol 11:101–184

Karlin S (1978) Mathematical population genetics. Academic Press, New York

Kendall MG, Stuart A (1966) The advanced theory of statistics, vol 3. Design and analysis, and time series. Charles Griffin, London

Laurie-Ahlberg CC, Weir BS (1979) Allozymic variation and linkage disequilibrium in some laboratory populations of *Drosophila melanogaster.* Genetics 92:1295–1314

Long JC (1986) The allelic correlation structure of Gainj- and Kalam-speaking people. 1. The simulation and interpretation of Wright's F statistics. Genetics 112:629–647

Loveless MD, Hamrick JL (1984) Ecological determinants of genetic structure in plant populations. Annu Rev Ecol Syst 15:65–95

Mardia KV, Kent JT, Bibby JM (1979) Multivariate analysis. Academic Press, London

Muona O (1982) Potential causes for multilocus structure in predominantly outcrossing populations. Silva Fenn 16: 107–114

Muona O, Szmidt AE (1985) A multilocus study of natural populations of *Pinus sylvestris.* In: Gregorius HR (ed) Population genetics in forestry. Springer-Verlag, Berlin, pp 241–252

Nei M (1973) Analysis of gene diversity in subdivided populations. Proc Natl Acad Sci USA 70:3321–3323

Nei M, Li WH (1973) Linkage disequilibrium in subdivided populations. Genetics 75:213–219

Perry DJ, Dancik BP (1986) Mating system dynamics of lodgepole pine in Alberta, Canada, Silvae Genet 35:190–195

Rao CR (1965) Linear statistical inference and its applications. John Wiley and Sons, New York

Roberds JH, Brotschol JV (1985) Linkage disequilibrium among allozyme loci in natural populations of *Liriodendron tulipifera* L. Silvae Genet 34:137–141

Sinnock P (1975) The Wahlund effect for the two locus model. Am Nat 109:565–570

Smouse PE, Neel JV (1977) Multivariate analysis of gametic disequilibrium in the Yanomama. Genetics 85:733–752

Stern K, Roche L (1974) Genetics of forest ecosystems. Springer-Verlag, New York

Waller DM, Knight SE (1989) Genetic consequences of outcrossing in the Cleistogamous annual, *Impatiens capensis.* III. Interlocus associations. Heredity 63:1–9

Weir BS (1990) Genetic data analysis. Sinauer Associates, Sunderland, Massachusetts, USA

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

Wheeler NC, Guries RP (1982) Population structure, gene diversity, and morphologivcal variation in *Pinus contorta* Dougl. Can J For Res 12:595–606

Wright S (1951) The genetical structure of populations. Ann Eugen 15:323–354

Yeh FC (1989) Isozyme analysis for revealing population structure for use in breeding strategies. In: Gibson GL, Griffin AR, Matheson AC (eds) Breeding topical trees: population structure and genetic improvement strategies in clonal and seedling forestry. Oxford Forestry Institute, Oxford, UK, and Winrock international Arlington Virginia, USA, pp 119–131

Yeh FC, El-Kassaby YA (1980) Enzyme variation in natural populations of Sitka spruce (*Picea sitchensis*). I. Genetic variation patterns among trees from 10 IUFRO provenances. Can J For Res 10:415–422

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

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

Yeh FC, Chiliak WM, Dancik BP, Illingworth K, Trust DC, Pryhitka BA (1985) Population differentiation in lodgepole pine, *Pinus contorta* spp. *Latifolia*: a discriminant analysis of allozyme variation. Can J Genet Cytol 27:210–218