Theor Appl Genet (1994) 88:465–471

F. C. Yeh · J. Shi · R. Yang · J. Hong · Z. Ye

Genetic diversity and multilocus associations in *Cunninghamia lanceolata* (Lamb.) Hook from The People's Republic of China

Received: 20 June 1993 / Accepted: 9 September 1993

Abstract Open-pollinated seeds were assayed for allozyme polymorphisms at 24 loci to assess genetic diversity and multilocous associations in 16 populations of *Cunninghamia lanceolata* (Lamb.) Hook in the People's Republic of China. On average, the percentage of polymorphic loci was 88.0, the number of alleles per locus was 3.0, and the expected heterozygosity was 0.394. The distribution of genetic diversity was not correlated with the geographic and climatic variables of the populations. However, allele frequencies correlated linearly with the mean annual temperature of the populations at *Mdh-1*, *Mdh-2*, *Mnr-2*, *Pgi-1*, and *Skdh-1* and with the altitude of the populations at *Aph-4* and *6Pg-2*. Of the total gene diversity 6% was attributed to among-population differentiation; 94% resided within populations. Two-locus gametic disequilibria were found in 15 of the 16 populations, and higher-order gametic disquilibria were significant in most populations. The gametic disequilibria did not correlate with geographic and climatic variables. The results suggest that population subdivision, founder effect, occurrence across diverse environments, a mating system dominated by inbreeding, and historical events from 2000 years of cultivation are contributing factors in the generation and maintenance of the multilocus genetic structure in this conifer.

Key words Genetic structure · Multilocus associations · Allozyme polymorphism · *Cunninghamia lanceolata*

Communicated by P. M. A. Tigerstedt

F. C. Yeh (✉) · J. Shi[1] · R. Yang · J. H. Hong[2] · Z. Ye[1]
Department of Forest Science, University of Alberta, Edmonton, AB, Canada, T6G 2H1

Present addresses:
[1] Forest Genetics Laboratory, Nanjing Forestry University, Nanjing 210037, PRC
[2] Chinese Academy of Forestry, Beijing, PRC

Introduction

Naturally occuring conifers with wide ranges characteristically display larger amounts of variation and little population differentiation at isozyme loci (Hamrick et al. 1981). Longevity, high fecundity, outcrossing mode, extensive gene flow, and large effective population size have been specified as determinants of genetic structure in plant populations (Loveless and Hamrick 1984). In conifers, the genetic structure is also influenced by disjunct distribution and environmental gradients (Xie et al. 1992), and multilocus Wahlund and founder effects (Yang and Yeh 1993). However, little is known about the genetic structure of wide-ranging conifers that have been submitted to an extensive period of cultivation. In such species, the relative importance of Wahlund and founder effects, epistatic selection, the chance of population subdivision, and other evolutionary mechanisms on the genetic structure may be strikingly different from that reported for the naturally occurring conifers.

Cunninghamia lanceolata (Lamb.) Hook is a fast-growing conifer with a wide range and great afforestation value in the People's Republic of China. Centred along the Nanling Mountains and spanning about ten degrees of latitude, fifteen degrees of longitude, and, 1,800 meters of elevation, the species has been under cultivation for more than 2000 years (Chen and Shi 1983). The diversity of the sites and climatic conditions in which *C. lanceolata* grows and the preference of the local people for different characters and for propagation both by rooted cuttings and seedlings have led to the evolution of an extremely variable species with respect to cold and disease resistance, form, growth, wood, and physiology (Chen and Shi 1987). A survey of two populations at 10 isozyme loci showed high levels of genetic diversity and inbreeding in *C. lanceolata* (Müller-Starck and Liu 1989a,b). In the study reported here, we investigated 24 isozyme loci in each of 16 populations sampled from the natural

range of *C. lanceolata*. Our objective was to determine whether 2000 years of cultivation has had any apparent effect on the genetic diversity and the multilocus genetic structure in this conifer.

Materials and methods

Seed collection

The 16 population studies were bulk population collections made within the present range of *C. lanceolata* in southeastern China. The geographic and climatic information of the sampled populations is presented in Table 1. Each population had at least several hundred trees, from which 50 or more cone-bearing trees situated a minimum of 50 m apart were randomly sampled. As far as could be ascertained, each sampled population represented a locality that was uniform with respect to climate, landform, soil, and vegetation.

Electrophoretic procedures

Seeds were germinated in a growth chamber under 24 h of light, 25 °C, and 85% relative humidity for 5–7 days until the length of radicle emergence was about 1 cm. For each germinated seed, the megagametophyte (female gametophyte) was separated from its corresponding embryo for electrophoretic analysis. The extraction of proteins from individual megagametophyte and horizontal strach gel electrophoresis were as described previously (Yeh and Layton 1979; Yeh and O'Malley 1980). Per population 60–100 megagametophytes were analyzed for variation in nine enzymes: aspartate aminotransferase (AAT, EC 2.6.11), aconitase (ACO, EC 4.2.1.3), acid phosphatase (APH, EC 3.1.3.2), isocitrate dehydrogenase (LDH, EC 2.6.1.1), malate dehydrogenase (MDH, EC 1.1.1.37), manadione reductase (MNR, EC 1.6.99.2), phosphoglucose isomerase (PGI, EC 5.3.1.9), shikimate dehyrogenase (SKDH, EC 1.1.1.25), and 6-phosphogluconate dehydrogenase (6-PGD, EC 1.1.1.44). Gels (12%) were prepared with Cannaught strach (Cannaught, Toronto).

When interpreting the electrophoretic banding patterns we followed the principles that have been outlined for *Pinus contorta* spp. latifolia (Yeh and Layton 1979). Each 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 designed 100, with additional allozymes being given numerical values according to their migration relative to the 100 allozyme. The inheritance of electrophoretic bands was interpreted as allozymes at genetic loci from segregation patterns f progeny of heterozygous maternal trees (unpublished data) and by comparison with the same enzymes in other conifers (Yeh and Layton 1979; Yeh and O'Malley 1980; Yeh and El-Kassaby 1980).

Analysis of multilocus structure using moments of empirical distribution

We constructed the empirical distribution of the number of heterozygous loci in a sample of 60–100 gametes assayed for each population. There were $1770\{60(60-1)/2\}$ and $4950\{100(100-1)/2\}$ possible comparisons for the 60 and 100 gametes, respectively. From the resulting distribution, we estimated the observed mean and the observed second to fourth moments about the mean. To validate our estimation, we obtained the second moment about the mean (i.e., variance) using Eq. 15 of Brown et al. (1980). The two methods of computation yielded identical results.

When the distribution of alleles at different loci are independent, the expectations of sample moments are functions of the gene diversity (Brown et al. 1980), and the approximate sampling variances of the sample moments can be computed using Eq. 9.9 of Kendall (1947). The estimation of sampling variance of the second moment followed Brown et al. (1980). We extended this for estimating the sampling variances of the third and fourth moments (Appendix). Using these sampling variances and assuming that sampling distributions of the sample moments approximated normality, we constructed the upper and lower 95% confidence limits of the sample moments as described in the Appendix to test for the presence of gametic disequilibria (associations among loci).

The total and average variance in the number of heterozygous loci (second moment) found in random pairs of gametes in the mixture of 16 populations was partitioned into the single- and two-locus effects (Brown and Feldman 1981). The total variance had three single-locus and three two-locus effects. The three single-locus effects were mean gene diversity (MH), variance in gene diversity (VH), and single-locus Wahlund effect (WH). The three two-locus effects were mean disequilibrium (MD), two-locus Wahlund's covariance (WC), and MD × WC (AI). The average variance had one single-locus effect in mean gene diversity (MH) and four two-locus effects in MD, AI, variance of disequilibrium (VD), and covariation in the interaction between MD and WC (CI). We generalized Brown and Feldman (1981) by allowing for unequal population size. The relative size of the t^{th} population is measured by w_t with $\Sigma w_t = 1$.

Table 1 Geographic and environmental variables of 16 populations of *Cunninghamia lanceolata* (Lamb.) Hook

Population	Latitude (°N)	Longitude (°E)	Altitude (m)	Precipitation (mm)	Mean temperature (°C)
1 Lechan, Guangdong	25.08	113.20	200	1800	19.0
2 Rongshui, Guangxi	25.05	109.14	340	1852	19.3
3 Pubei, Guangxi	22.18	109.18	250	1904	21.4
4 Liping, Giuzhou	26.14	109.18	500	1322	15.7
5 Jianghua, Hunan	25.00	114.47	310	1512	19.1
6 Hiutong, Hunan	26.50	109.44	380	1295	16.5
7 Jianou, Fujian	26.40	118.10	320	1813	18.8
8 Pingbian, Yunnan	22.56	103.41	1450	1659	16.5
9 Xishou, Yunnan	23.30	104.50	1962	1296	15.8
10 Quannan, Jiangxi	24.45	114.31	500	1675	18.5
11 Lean, Jiangxi	27.24	115.48	185	1662	17.0
12 Tonggu, Jiangxi	28.33	114.15	650	1735	16.3
13 Xiushui, Jiangxi	29.20	114.10	117	1580	16.5
14 Qijiang, Sichuan	29.00	106.50	900	1200	18.3
15 Xinxie, Henan	31.35	116.50	280	1120	15.3
16 Shanchen, Henan	31.47	115.20	430	1300	12.6

Results

Allelic frequency and genetic diversity

A total of 94 alleles from 24 loci was studied. Two loci (*Aph-1* and *Aph-2*) were monomorphic in all 16 populations while the remaining 22 loci (*Aat-1*, *Aat-2*, *Aat-3*, *Aco*, *Aph-3*, *Aph-4*, *Aph-5*, *Aph-6*, *Idh-1*, *Idh-2*, *Mdh-1*, *Mdh-2*, *Mdh-3*, *Mdh-4*, *Mnr-1*, *Mnr-2*, *Pgi-1*, *Pgi-2*, *Skdh-1*, *Skdh-2*, *6Pgd-1*, *6Pgd-2*) were polymorphic in at least some of the populations. We present only the frequency of the most common allele at each polymorphic loci for each population due to the large size of the data set (Table 2).

Single populations were polymorphic (P), on average, at 88% of their loci (Table 2), and most such loci segregated for 3 alleles. The number of alleles per locus (n_a) ranged from 2.13 to 3.17, averaging 3.0 (Table 2). The expected heterozygosity ranged from 0.236 to 0.452 with an average (h) of 0.394 (Table 2). The distribution of genetic variability among the 16 populations did not correlate with the geographic and climatic variables of the populations.

The likelihood ratio chi-square tests indicated the homogeneity of allele frequencies only at *Mdh-4*. Allele frequencies correlated linearly with the mean annual temperature of populations at *Mdh-1*, *Mdh-2*, *Mnr-2*, *Pgi-1*, and *Skdh-1* and with the altitude of populations at *Aph-4* and *6Pgd-2*.

Nei's (1973) G_{st} showed that, on average, about 6% of the total gene diversity was attributed to the among-population differentiation, 94% resided within populations (Table 3). However, the estimates of G_{st} varied considerably among the isozyme loci from 0.65% at *Mdh-4* to 21.63% at *Aph-6* (Table 3).

Multilocus structure

Of the 16 variances observed 15 exceeded their upper 95% confidence limits (Table 4), indicating the significance of two-locus gametic disequilibria in *C. lanceolata*. The observed third and fourth moments exceeded their upper or lower 95% confidence limits in all populations except for the third moment in Populations 4, 7, and 8 (Table 4). This signifies the predominance of three- and four-locus gametic disequilibria, since the observed third and fourth moments are functions of the three- and four-locus gametic disequilibria respectively, as well as to the lower-order gametic disequilibria (Brown et al. 1980).

The components of varaince in number of heterozygous loci in pairs of random gametes in the mixed pool of 16 populations (Table 5) indicated that 53% of the

Table 2 Common allele frequency ($\times$ 0.001) and genetic variability[a] in 16 populations of *Cunninghamia lanceolata* (Lamb.) Hook (n_a number of alleles per locus, P percentage of polymorphic loci at 0.99 criterion, h heterozygosity)

Locus	Population															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Aat-1	996	841	810	829	1000	835	933	957	943	914	914	950	900	949	929	856
Aat-2	509	587	690	1000	533	709	667	671	729	514	657	650	686	696	586	756
Aat-3	661	619	810	986	567	658	733	557	671	629	629	617	614	646	743	899
Aco	592	767	633	800	940	757	640	406	357	450	357	350	414	494	400	629
Aph-3	796	714	661	617	567	729	610	–	632	790	667	632	833	704	700	816
Aph-4	875	743	814	817	667	855	814	–	611	895	750	789	759	929	684	623
Aph-5	767	818	707	886	632	640	790	647	433	567	508	552	493	500	455	654
Aph-6	793	731	661	857	825	747	754	821	588	750	612	625	507	636	612	714
Idh-1	814	800	714	986	860	841	880	900	729	700	814	650	650	810	757	729
Idh-2	373	471	414	343	383	329	433	429	551	313	441	509	448	367	443	489
Mdh-1	797	800	814	1000	850	875	933	914	942	899	943	877	900	949	826	1000
Mdh-2	814	800	814	1000	850	850	933	929	943	871	943	900	900	949	814	1000
Mdh-3	949	971	949	1000	933	863	983	971	986	943	929	983	857	975	929	1000
Mdh-4	949	957	948	1000	933	888	983	971	986	957	929	983	868	974	941	1000
Mnr-1	839	706	593	523	933	722	933	929	955	913	851	833	768	816	776	639
Mnr-2	964	829	831	1000	900	950	833	971	838	928	871	950	943	823	942	989
Pgi-1	661	614	576	989	517	638	633	929	943	771	971	933	757	810	870	778
Pgi-2	593	643	441	486	667	823	767	529	514	657	629	467	629	658	609	678
Skdh-1	700	509	837	529	840	586	780	485	400	483	443	267	567	506	586	557
Skdh-2	420	462	592	800	560	514	580	443	414	433	457	400	600	468	471	457
6Pgd-1	458	457	509	1000	450	500	600	486	500	386	507	450	429	430	443	633
6Pgd-2	864	929	966	986	964	888	867	800	809	871	864	850	900	886	886	944
n_a($\times$ 0.001)	275	288	279	213	254	308	283	263	283	317	304	308	304	308	317	267
P($\times$ 0.10%)	917	917	917	625	875	917	917	833	917	917	917	917	917	917	917	750
h($\times$ 0.001)	401	440	430	236	386	435	363	330	389	425	393	412	452	402	423	389

	n_a($\times$ 0.001)	P($\times$ 0.10%)	h($\times$ 0.001)
Grand average	300	880	394

[a] Genetic variability was based on 22 polymorphic and 2 monomorphic loci

Table 3 Gene diversity of 16 populations of *Cunninghamia lanceolata* (Lamb.) Hook

Locus	H_T	H_S	G_{ST}
Aat-1	0.1694	0.1651	0.0251
Aat-2	0.4325	0.4199	0.0290
Aat-3	0.4363	0.4123	0.0552
Aco	0.5932	0.5340	0.0998
Aph-3	0.4592	0.4474	0.0256
Aph-4	0.3585	0.3463	0.0341
Aph-5	0.5562	0.4872	0.1242
Aph-6	0.5389	0.4223	0.2163
Idh-1	0.3340	0.3244	0.0289
Idh-2	0.7208	0.6973	0.0325
Mdh-1	0.1918	0.1870	0.0248
Mdh-2	0.1927	0.1880	0.0246
Mdh-3	0.0927	0.0911	0.0176
Mdh-4	0.0871	0.0865	0.0065
Mnr-1	0.3380	0.3016	0.1077
Mnr-2	0.1662	0.1620	0.0254
Pgi-1	0.3506	0.3070	0.1242
Pgi-2	0.5746	0.5535	0.0385
Skdh-1	0.6218	0.5886	0.0535
Skdh-2	0.6540	0.5921	0.0947
6Pgd-1	0.6258	0.5974	0.0455
6Pgd-2	0.1989	0.1957	0.0162
Mean	0.3952	0.3685	0.0568

Table 5 Components of total and average variances in number of heterozygous loci in *Cunninghamia lanceolata*

Component[a]	
Single-locus effect	
Mean gene diversity (MH)	4.086
Variance in gene diversity (VH)	0.307
Wahlund effect (WH)	−0.113
Total	4.280
Two-locus effect	
Mean linkage disequilibrium (MD)	1.092
Wahlund's covariance (WC)	3.720
Interaction between MD and WC (AI)	0.055
Variance of disequilibria	1.116
Covariance of interaction (CI)	−0.072
Average variance (MH + MD + AI + VD + CI)	6.277
Total variance (MH + VH + WH + MD + WC + AI)	9.147

[a] See Brown and Feldman (1981) for a full description of these terms

total variance was due to the two-locus effects; 47% was due to the single-locus effects. Of the two-locus effects, the covariance of allele frequencies or Wahlund's covariance (WC) was the most prominent, accounting for 64%. The variance of disequilibrium (VD) and mean disequilibrium (MD) were 19% and 17% of the two-locus effect, respectively. The interaction between MD and WC and the covariation in the interaction between MD and WC were insignificant.

Discussion

The level of genetic diversity in *C. lanceolata* found in this study ($n_a = 3.00$; $P = 88.0$; $h = 0.394$) and in another survey of two populations at ten isozyme loci ($n_a = 4.15$; $P = 100.0$; $h = 0.299$) by Müller-Starck and Liu (1989a) is greater than the average ($n_a = 2.29$; $P = 67.7$; $h = 0.207$) that has been reported for other conifers (Loveless and Hamrick 1984). The biotic factors cited by Hamrick et al. (1981) and Loveless and Hamrick (1984), including the occurrence of the species over broad environmental spectra, large population size, long generation times, high fecundities, and the potential for long-distance gene flow via pollen and seed dispersal, likely contributed to the high level of genetic diversity observed in *C. lanceolata*.

The genetic differentiation in terms of G_{st} that has been reported for conifers ranges from a low of 0.000 in *Pinus resinosa* (Flower and Morris 1977) to a high of 0.241 in *Pseudotsuga menziesii* (Li and Adams 1988). Relative to G_{st} estimates obtained from studied involving wide-ranging populations and large samples of loci

Table 4 Estimates of multilocus associations in 16 populations of *Cunninghamia lanceolata* ($\sigma^2_K(H_0)$ is the expected variance under the null hypothesis (H_0) of complete interlocus independence; S^2_K is observed variance and U is its upper 95% confidence limit; M_3 is observed third moment, and L_3 and U_3 are, respectively, its lower and upper 95% confidence limit; M_4 is observed fourth moment, and U_4 is its upper 95% limit)

Population	$\sigma^2_K(H_0)$	S^2_K	U	M_3	L_3	U_3	M_4	U_4
1	4.35	6.55	5.89	−9.92	−4.66	6.16	181.97	99.73
2	4.65	6.27	6.16	−12.32	−4.79	6.09	202.28	109.00
3	4.63	8.21	6.28	−10.69	−5.25	6.56	274.36	112.54
4	2.49	5.09	3.29	1.92	−1.72	2.47	63.01	30.73
5	4.06	8.89	5.48	−6.01	−4.10	5.59	261.74	86.57
6	4.64	10.79	6.04	22.91	−4.19	6.00	677.26	106.22
7	4.15	7.44	5.61	−1.77	−4.08	5.97	197.91	90.74
8	3.38	4.45	4.48	−2.35	−3.06	3.77	83.62	58.24
9	4.02	7.02	5.32	−4.77	−4.10	4.66	189.38	81.52
10	4.20	9.46	5.58	−6.64	−4.17	5.35	291.59	90.43
11	4.19	6.40	5.56	−6.46	−4.23	5.20	167.60	89.65
12	4.09	7.85	5.52	−10.24	−4.61	5.13	228.25	87.57
13	4.58	8.91	6.07	−10.68	−4.81	5.90	294.78	106.45
14	4.07	6.61	5.32	−8.00	−3.78	4.69	165.93	82.20
15	4.51	7.55	5.98	−6.55	−4.79	5.67	219.91	103.03
16	3.79	6.44	4.87	−8.40	−3.13	3.82	152.66	68.01

(i.e., Yeh and El-Kassaby 1980; Li and Adams 1988; Lagercrantz and Ryman 1990; Xie et al. 1992), our G_{st} estimate of 0.057 for *C. lanceolata* falls around the average of the differentiation spectrum. The extent of population differentiation in *C. lanceolata* would probably be higher if we were to sample many remote or marginal populations.

Correlations of allele frequency with geographic and environmental variables are well-documented in conifers (e.g., Yeh and Layton 1979; Xie et al. 1992). Whether the loci in this study that correlated with the mean annual temperature (*Mdh-1*, *Mdh-2*, *Mnr-2*, *Pgi-1*, and *Skdh-1*) and altitude (*Aph-4* and *6Pgd-2*) of the populations were markers of adaptive gene complexes and (or) are locally adapted to the environment and responded to selection is unknown. This is because neutral alleles with migration and hybridization of genotypes along the altitudinal and temperature gradients across geographic regions could produce the observed patterns in this study. Nevertheless, from the patterns so far found at *Aph-4*, *6Pgd-2*, *Mdh-1*, *Mdh-2*, *Mnr-2*, *Pgi-1*, and *Skdh-1* predictions as to what their allele frequencies should be in other parts of the species distribution can be made in *C. lanceolata*. This is of practical importance in the certification of seed sources of reforestation and genetic conservation in this conifer.

The comparison of the single-locus Wahlund effect (WH) and the two-locus Wahlund effect (WC) indicated neglible WH and an appreciable amount of WC in *C. lanceolata* (Table 5). There is thus a greater level of population differentiation when many loci are considered jointly, which supports the notion that multilocus analysis of gene frequently data accounts for interactions among loci that are not observable using the averaging single-locus effects (Yeh et al. 1985).

The single-locus components accounted for only 47% of the total observed variance in the number of heterozygous loci in pairs of random gametes in the mixed pool of 16 populations (Table 5). Therefore, there is more than a doubling of this variance due to gametic disequilibria within and among populations. The prevalence of gametic disequilibria in this survey and other conifers examined to date (e.g., Boyle 1985; Mouna and Szmidt 1985; Roberds and Brotschol 1985; Yeh amd Morgan 1987; Yang and Yeh 1993) is contrary to the notion that gametic disequilibria are not important in outcrossing plants. The high two-locus Wahlund effect and the low interaction between mean disequilibrium and the two-locus Wahlund's covariance in *C. lanceolata* (Table 5) are characteristic of multilocus associations that are associated with a strong effect of population subdivision (Brown and Feldman 1981). The multilocus Wahlund effect has also been hypothesized to be a major cause of gametic disequilibrium in 40 populations of *Pinus contorta* due to the pooling of populations that differed markedly in allele frequencies at many loci (Yang and Yeh 1993). The component due to variation in disequilibrium accounted for 19% of the two-locus effect. This indicates that the most frequent gametic types usually differed from one population to another. Such a result would arise under the founder and/or the diversifying selection hypothesis. It has been suggested that a restricted effective population size (founder effect) was a possible cause of the gametic disequilibria in three populations of *Pinus sylvestris* (Muona Szmidt 1985). These authors concluded that the decrease in the effective population size might be due to uneven seed production, partial selfing, and uneven numbers of male and female parents contributing to the pollen pool. Founder effect has also been cited to be a cause of gametic disequilibria in *Pinus contorta* as the result of extreme bottlenecks due to the fire origin of much of the forests and the serotinous nature of the cones (Yang and Yeh 1993).

The species *C. lanceolata* has been cultivated for more than 2000 years in China (Chen and Shi 1983) where non-local sources were introduced liberally and incessantly to reforest local areas in large numbers prior to the 1950. The movement of *C. lanceolata* among the regions could be extensive and over long distances due to the vast waterways in China. Such pooling of the *C. lanceolata* genes into local populations from initially differentiated populations across the regions (Wahlund's covariance) probably contributed to the high level of genetic diversity and generated the extensive gametic disequilibria that have been found in our sampled populations. The gametic disequilibria could be further generated over time by means of a breeding system that is dominated by inbreeding (Müller-Starck and Liu 1989b) and/or as a result of the preferences of the local people for different tree characters and for vegetative propagation by rooted cutting. Thus, long-term cultivation and mating system have played major roles in influencing the genetic diversity and structure and the direction of evolution of *C. lanceolata*.

Our findings have important implications for selective breeding of *C. lanceolata*. When gametic disequilibria are widespread and when population subdivision is a major cause of gametic disequilibria in *C. lanceolata*, the pooling of select trees from populations in gametic disequilibria into a single breeding population will disrupt the disequilibrium (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). Under such conditions, there is a selective advantage in sub-dividing the breeding population, with repeated cycles of within- and between-sub-population selection (Katz and Enfield 1977) to exploit "multiple peak epistasis". The pooling of select trees into a single breeding population from sub-populations each differing in allele frequencies is also problematic when estimating genetic covariances among relatives for the construction of selection indices, which requires the assumption of gametic disequilibrium (Baker 1986). Thus, even if natural populations of *C. lanceolata* are in gametic equilibrium, conventional selection indices are not applicable for progeny from crosses among trees from sub-populations each having different allele frequencies.

A knowledge of the multilocus structure of *C. lanceolata* is also conductive to the efficient design and implementation of a practical strategy for its genetic conservation. The prominent multilocus Wahlund effect indicates that the optimal streategy for ex situ collections of *C. lanceolata* would be to sample fewer trees from each of many populations to cover a wide spectrum of the environment. Populations that exhibit significant genetic differentiation would be the obvious targets for sampling. Such a strategy is similar to the sampling of locally common genes in forest populations (Yang and Yeh 1992) and depends primarily on the proportion of locally common gametes in the populations. We anticipate that multilocus analysis of population structure as presented in this study will have a major impact on the management, improvement, and conservation of forest trees.

Appendix

Construction of 95% confidence intervals for observed moments in the number of heterozygous loci of two random chosen gametes under complete interlocus independence.

Consider m independent loci in an infinite population of haploid gametes and define random variable K as the number of heterozygous loci of two random chosen gametes. Then, K is distributed as sum of m independent Bernoulli distributions (Brown et al. 1980):

$$K = \sum_{j=1}^{m} X_j, \tag{A1}$$

where X_j takes either 0 or 1, depending on whether the j^{th} locus is homozygous or heterozygous when two random chosen gametes are compared. Probability that this locus is homozygous is $\sum_i p_{ji}^2$,where p_{ij} is the population frequency of i^{th} allele at the j^{th} locus. Probability that it is heterozygous is $1-\sum_i p_{ji}^2$, which equals the gene diversity at the j^{th} locus, h_j. With this outset, the moment-generating function of distribution of K is:

$$\phi(\theta) = \prod_{j=1}^{m} (h_j e^{\theta} + 1 - h_j), \tag{A2}$$

from which we have the related comulative function

$$\begin{aligned}\psi(\theta) &= \log\phi(\theta) = \Sigma\log(h_j e^{\theta} + 1 - h_j) \\ &= \Sigma\log\left[1 + h_j\theta + h_j\frac{\theta^2}{2!} + h_j\frac{\theta^3}{3!} + \cdots\right] \\ &= \kappa_1\theta + \kappa_2\frac{\theta^2}{2!} + \kappa_3\frac{\theta^3}{3!} + \cdots.\end{aligned} \tag{A3}$$

The i^{th} comulant is obtained from the i^{th} derivative of $\psi(\theta)$ with respect to θ, evaluated for $\theta = 0$,

$$\kappa_i = \left.\frac{d^i\psi(\theta)}{d\theta^i}\right|_{\theta=0}, \tag{A4}$$

for $i = 1, 2, 3, \ldots$. We now write out the first eight cumulants

$$\kappa_1 = \Sigma h_j$$
$$\kappa_2 = \Sigma h_j - \Sigma h_j^2$$
$$\kappa_3 = \Sigma h_j - 3\Sigma h_j^2 + 2\Sigma h_j^3$$
$$\kappa_4 = \Sigma h_j - 7\Sigma h_j^2 + 12\Sigma h_j^3 - 6\Sigma h_j^4$$
$$\kappa_5 = \Sigma h_j - 15\Sigma h_j^2 + 50\Sigma h_j^3 - 60\Sigma h_j^4 + 24\Sigma h_j^5$$
$$\kappa_6 = \Sigma h_j - 31\Sigma h_j^2 + 180\Sigma h_j^3 - 390\Sigma h_j^4 + 360\Sigma h_j^5 - 120\Sigma h_j^6$$
$$\kappa_7 = \Sigma h_j - 63\Sigma h_j^2 + 602\Sigma h_j^3 - 2100\Sigma h_j^4 + 3360\Sigma h_j^5 - 2520\Sigma h_j^6 + 720\Sigma h_j^7$$
$$\kappa_8 = \Sigma h_j - 127\Sigma h_j^2 + 1932\Sigma h_j^3 - 10206\Sigma h_j^4 + 25200\Sigma h_j^5 - 31920\Sigma h_j^6 + 20160\Sigma h_j^7 - 5040\Sigma h_j^8. \tag{A5}$$

The second to eight central moments can be obtained using the κ values in Eq. A5, and Eq. 3.30 of Kendall (1947):

$$\mu_2 = \kappa_2$$
$$\mu_3 = \kappa_3$$
$$\mu_4 = \kappa_4 + 3\kappa_2^2$$
$$\mu_5 = \kappa_5 + 10\kappa_3\kappa_2$$
$$\mu_6 = \kappa_6 + 15\kappa_4\kappa_2 + 10\kappa_3^2 + 15\kappa_2^3$$
$$\mu_7 = \kappa_7 + 21\kappa_5\kappa_2 + 35\kappa_4\kappa_3 + 105\kappa_3\kappa_2^2$$
$$\mu_8 = \kappa_8 + 28\kappa_6\kappa_2 + 56\kappa_5\kappa_3 + 35\kappa_4^2 + 210\kappa_4\kappa_2^2 + 280\kappa_3^2\kappa_2 + 105\kappa_2^4 \tag{A6}$$

Inserting κ_2–κ_4 in Eq. A5 into μ_2–μ_4 in (A6) will yield Eqs. 3–5 of Brown et al. (1980).

Under complete inter-locus independence, observed second to fourth moments (m_2–m_4) will be estimates of their respective expected values (μ_2–μ_4). The null hypotheses under this condition are:

H_0': $\mu_2 = \Sigma h_j - \Sigma h_j^2$

H_0'': $\mu_3 = \Sigma h_j - 3\Sigma h_j^2 + 2\Sigma h_j^3$

H_0''': $\mu_4 = \Sigma h_j - 7\Sigma h_j^2 + 12\Sigma h_j^3 - 6\Sigma h_j^4 + 3[\Sigma h_j - \Sigma h_j^2]^2$.

Under the null hypotheses, sampling variances of observed second to fourth moments are approximately given by using Eq. 9.9 of Kendall (1947):

$$\mathrm{Var}(m_2|H_0') = \frac{1}{n}(\mu_4 - \mu_2^2)$$

$$\mathrm{Var}(m_3|H_0'') = \frac{1}{n}(\mu_6 - \mu_3^2 + 9\mu_2^3 - 6\mu_2\mu_4)$$

$$\mathrm{Var}(m_4|H_0''') = \frac{1}{n}(\mu_8 - \mu_4^2 + 16\mu_3^2\mu_2 - 8\mu_5\mu_3). \tag{A7}$$

From Eq. A5 and A6 above, these sampling variances can be estimated by replacing the parametric values of the $\{h_j\}$ by their estimates from samples of n gametes. Using estimates of the sampling variances and assuming sampling distributions of m_2–m_4 approximate normality, the 95% confidence intervals (CI) for m_2–m_4 are:

$$CI_2 \simeq \mu_2 \pm 1.96\sqrt{\mathrm{Var}[m_2|H_0']}$$

$$CI_3 \simeq \mu_3 \pm 1.96\sqrt{\mathrm{Var}[m_3|H_0'']}$$

$$CI_4 \simeq \mu_4 \pm 1.96\sqrt{\mathrm{Var}[m_4|H_0''']}, \tag{A8}$$

where $\pm$ refer to upper (+) and lower (−)95% confidence limits, respectively.

Acknowledgements We thank two anonymous reviewers for their suggestions. This research was supported in part by grants from the Natural Sciences and Engineering Research Council of Canada, A2282 and IRC 8607, and Petawawa National Forestry Institute to F.C.Y.

References

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

Boyle TJB (1985) The mating system and population structure of black spruce in central New Brunswick and its influence on tree improvement strategy. PhD thesis, University of New Brunswik, Fredricton, N.B.

Brown AHD, Feldman MW (1981) Population structure of multilocus associations. Proc Natl Acad Sci USA 78:5913–5916

Brown AHD, Feldman MW, Nevo E (1980) Multilocus structure of natural populations of *Hordeum sponteneum*. Genetics 96:523–536

Chen YW, Shi JS (1983) Some fundamental problems in the genetic improvement of Chinese fir. J Nanjing Inst For 4:1–21

Chen YW, Shi JS (1987) Genetic variation and improvement program of Chinese fir. J Subtrop For Inst 1:1–19

Fowler DP, Morris RW (1977) Genetic diversity in red pine: evidence for low heterozygosity. Can J For Res 7:341–347

Hamrick JL, Mitton JB, Linhart YB (1981) Levels of genetic variation in trees: influences of life history characteristics. In: Conkle MT (ed) Proc Symp Isozymes North Am For Trees For Insects. USDA For Serv Genet Tech Rep PSW-48, pp 35–41. Pacific SW Forest and Range Exp Stn, Berkeley, Calif

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

Kartz AJ, Enfield FD (1977) Response to selection for increased pupal weight in *Tribolium castaneum* as related population structure. Genet Res 30:237–246

Kendall MG (1947) The advanced theory of statistics, vol 1, 3rd edn. Charles Griffin, London

Lagercrantz U, Rayman N (1990) Genetic structure of Norway spruce (*Picea abies*): concordance of morphological and allozyme variation. Evolution 44:38–53

Li P, Adams WT (1988) Range-wide patterns of allozyme variation in Douglas-fir (*Pseudotsuga menziesil*). Can J For Res 19:149–161

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

Müller-Starck G, Liu YQ (1989a) Genetics of *Cuninghamia lanceolata* Hook. 2. Genetic variation within and between two provenance samples. Silvae Genet 38:172–177

Müller-Starck G, Liu YQ (1989b) Inference on the reproductive system of *Cunninghamia lanceolata*. For Ecol Manage 29:187–198

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

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

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

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

Xie CY, Dancik BP, Yeh FC (1992) Genetic structure of *Thuja orientalis*. Biochem Syst Ecol 20:433–441

Yang R-C, Yeh FC (1992) Genetic consequences of in situ and ex situ conservation of forest trees. For Chron 68:720–729

Yang R-C, Yeh FC (1993) Multilocus structure in *Pinus contorta* Dougl. Theor Appl Genet 87:568–576

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

Yeh FC, El-Kassaby YA (1980) Enzyme variation in natural populations of Sitka spruce (*Picea stichensis*). 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, Morgan KM (1987) Mating system and multilocus associations in a natural population of *Pseudotsuga menziesii* (Mirb.) Franco. Theor Appl Genet 73:799–808

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, Cheliak 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