PATTERNS OF GENE FLOW AND GEOGRAPHIC STRUCTURE IN *PINUS CONTORTA* DOUGL.

Rong-Cai Yang & Francis C. Yeh

Department of Renewable Resources, University of Alberta, Edmonton, Alberta, Canada T6G 2H1

Received May 22, 1995; accepted August 25, 1995

ABSTRACT

We analyzed a range-wide allozyme variation in 66 populations from three subspecies of *Pinus contorta* Dougl.: 20 in *P. contorta* spp. *contorta*, 35 in *P. contorta* spp. *latifolia* and 11 in *P. contorta* spp. *murrayana*. Our objectives were (i) to assess geographic variation in effective population size (N) and rare alleles (≤ 0.05) and (ii) to characterize the extent and patterns of gene flow among populations in each subspecies. There was substantial variation in the estimate of total allele numbers and N among subspecies and among populations within subspecies. The number of private alleles (occurred only in one population) in each population ranged between 0 and 4, averaging 0.6, 0.7 and 1.0 for subspecies *latifolia*, *contorta* and *murrayana*, respectively. The number of rare alleles shared with other populations [a(s)] and per-locus average heterozygosity ($\bar{h}$) also varied considerably across populations. Linear correlation estimates of N and a(s) with latitude of populations were significant in subspecies *latifolia* and *murrayana*, not in spp. *contorta*. In contrast, $\bar{h}$ correlated significantly with latitude only in spp. *contorta*. The number of migrants exchanged between populations was > 1 in each subspecies, large enough to obscure the among-population genetic differentiation from theoretical expectation. Significant pattern of "isolation by distance" was evident only in ssp. *contorta*. Biogeographical analysis suggested the existence of a 'centre of diversity' for spp. *latifolia* in north-central British Columbia.

Key words: *Pinus contorta* Dougl., allozymes, effective population size, rare alleles, isolation by distance, geographic variation

INTRODUCTION

Pinus contorta Dougl. is one of the most widely distributed and variable species in North America. As an aggressive pioneer in a variety of edaphic and climatic conditions, it has evolved as an extremely variable species with respect to morphological and biochemical traits (*e.g.*, CRITCHFIELD 1957, 1980; FORREST 1980; WHEELER & GURIES 1982a; YING *et al.* 1985; YANG & YEH 1993). Single-locus (WHEELER & GURIES 1982a) and multilocus diversity studies (YANG & YEH 1993) have shown that the majority of genetic variation resides within populations and little among-population differentiation at allozyme loci. This observed population structure has been thought to be due to extensive gene flow and large effective population size (EPPERSON & ALLARD 1989) since *P. contorta* Dougl. is capable of producing abundant pollen and small, light-weighted seeds, thereby allowing for rapid and long-distance dispersals (CRITCHFIELD 1980).

WRIGHT (1931) showed that, for selectively neutral alleles, genetic variation within populations is proportional to $4N\mu$ and genetic variation among populations is proportional to 4Nm, where N is the effective population size, μ is the mutation rate and m is the migration rate. He further showed that selection of strength, s, can overpower the effects of gene flow and random genetic drift if $4Ns > 4Nm$ and $4Ns > 1$, respectively (WRIGHT 1931). Despite the importance of gene flow, random genetic drift, selection, effective population size and mutation, empirical study to assess their influence on genetic variation and population differentiation is lacking for *Pinus contorta* Dougl.. While most conifers have shown to exhibit extensive gene flow (GOVINDARAJU 1988), little information is available on the patterns of gene flow among populations. In a continuous distributed species such as *P.contorta* Dougl., for example, it remains unclear whether the extent of gene flow between populations depends on their geographic distances, *i.e.*, "isolation by distance", as originally proposed by WRIGHT (1943).

Recently, SCHOEN and BROWN (1991) estimated N of many plants, including three conifers. These authors found a substantial amount of variation in N among populations. For *P. contorta* Dougl. which is widespread, it is important to know whether the among-population variation in N is associated with geographic variation since N effective sizes of marginal populations

in spp. *latifolia* were hypothesized to be smaller than those of central populations due to effects of random drift, increased inbreeding and/or selection (YEH & LAYTON 1979; YEH *et. al.* 1985).

We conducted a range-wide allozyme survey of 66 populations in three subspecies of *P. contorta* Dougl. and the multilocus genetic structure has been reported (YANG & YEH 1993). This study assesses the patterns of gene flow and geographic structure in effective population size for this conifer. Our objectives were (1) to determine whether the among-population variations in N and rare alleles were associated with the geography of the population in each subspecies and (2) to investigate the relationship between gene flow and geographic distance for pairs of populations in each subspecies, thereby identifying possible patterns of isolation by distance.

MATERIALS AND METHODS

Allozyme data

The 66 populations from three subspecies represent geographically and ecologically distinct regions of *Pinus contorta* Dougl.'s range (CRITCHFIELD 1980; WHEELER & CRITCHFIELD 1985): 20 in *P. contorta* spp. *contorta*, 35 in *P. contorta* spp. *latifolia* and 11 in *P. contorta* spp. *murrayana*, were sampled from the species' natural range in western North America (Fig. 1 and Table 1). Each population was represented by a random sample of 15 to 35 cone-bearing trees with a minimum of 46 m apart. A bulked sample of 75 megagametophytes (haploid maternal tissues) per population were electrophoretically assayed for variation in 14 enzymes. Tissue preparation, electrophoretic conditions, and enzymes scored were previously described (YANG & YEH 1993). Of 21 loci scored, two (*Pep–1* and *Pep–2*) were monomorphic in all population. The remaining 19 loci (*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 and were used in subsequent analyses.

Estimating rare alleles and effective population size

We recorded the number of alleles (k_{ij}) observed in a sample of n_{ij} genes at the ith locus from the jth population (i = 1, 2, ..., I; j = 1, 2, ..., J). Thus, the total number of alleles observed in a population (a_T) was $a_T = \Sigma_{i=1}^{I} k_{ij}$. Note that the index (j) for the population was omitted from a_T for simpler notation. Using estimated allele frequencies, we divided the total observed number of alleles into those which were common (≥ 0.05) and those which were rare (< 0.05). For our purpose, we further divided rare alleles into those which occurred only in one sampled population ['private' alleles, a(1)] and those which were shared only by subsets of populations [shared alleles, a(s)]. Populations sharing rare alleles are more likely to have a recent coancestry either because of mutations in recent history, or because more distantly related populations lost such alleles following divergence and isolation (WHEELER & GURIES 1982a). On the other hand, average frequency of private alleles across populations is a sensitive predictor of gene flow (SLATKIN 1985; BARTON & SLATKIN 1986). However, when identifying patterns of gene flow (see below), we employed a measure of gene flow based on WRIGHT's (1943) F_{ST} because F_{ST} uses all of the gene-frequency data and thus is less prone to errors in misidentifying alleles than the private alleles method (SLATKIN & BARTON 1989).

The effective population size (N) was estimated following CHAKRABORTY and NEEL (1989). This method extended EWENS' (1972) sampling theory for neutral alleles into multiple loci and populations. With I loci and J populations, I mutation rates (μ_1, ..., μ_I) and J effective population sizes (N_1, ..., N_J) need to be estimated. If each population is at steady state under mutation-drift balance (neutral hypothesis), EWENS (1972) has shown that k_{ij} is a sufficient statistic for estimating the above parameters. We carried out the Ewens-Watterson test for neutrality at allozyme loci (MANLY 1985) and results indicated that all 19 polymorphic loci were selectively neutral. A different test based on the comparison of observed variance of heterozygosity with its expectation under neutrality (YEH & LAYTON 1979) also suggested neutrality at 23 polymorphic loci in spp. *latifolia*. Using the total number of alleles (a_T) of polymorphic loci in at least one of the J populations, we obtained the maximum likelihood estimates (MLEs) of N_j:

$$\hat{N}_j = \frac{a_T - I}{\sum_{i=1}^{I} 4\mu_i(1-P_i)[\psi(n_{ij}+\theta_{ij}) - \psi(1+\theta_{ij})]} \quad [1]$$

where $P_i = Q_i / (1 - Q_i)$ with Q_i being the probability that all J populations were monomorphic at the ith locus, $\theta_{ij} = 4N_j\mu_i$ and $\psi(.)$ was the di-Gamma function with $\psi(\alpha) = [\ln \Gamma(\alpha)]' = \Gamma'(\alpha)/\Gamma(\alpha)$ and $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$. Because μ and N appeared only in product form ($\theta_{ij} = 4N_j\mu_i$), it was necessary to assume an average mutation rate ($\bar{\mu}$) in order to obtain MLEs of N_j.

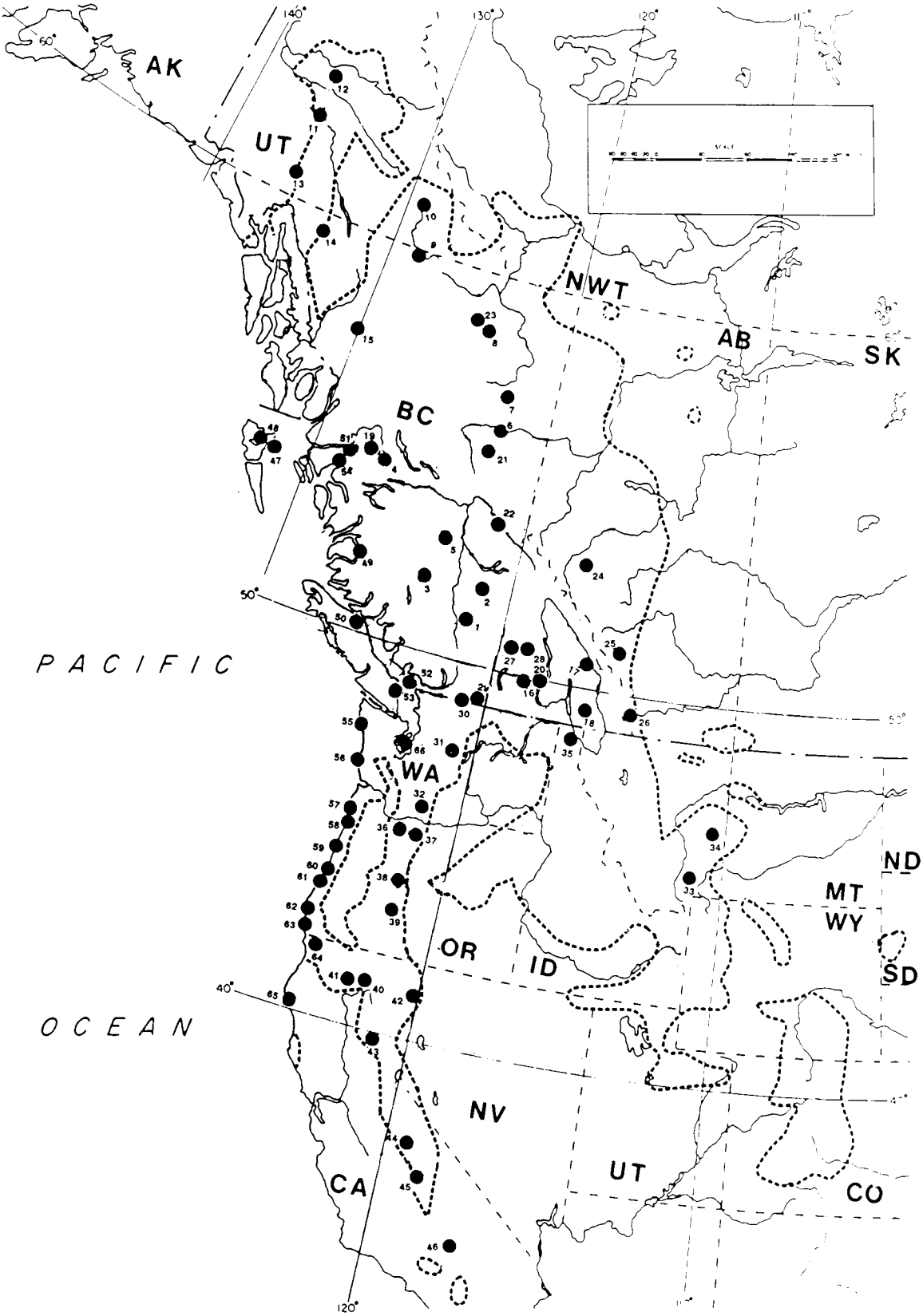

Figure 1 Natural range for *Pinus contorta* Dougl. Delimited by dash lines (– – – –) and location of 66 sampled populations (●)

Table 1 Latitudes (°N), longitudes (°W) and altitudes (m) and estimates of total allele number (a_T), private allele [a(l)] and shared rare alleles [a(s)], mean heterozygosity ($\bar{h}$), effective population size (N) (± standard error SE) for 66 populations of *Pinus contorta* Dougl.

Population	Prov./State	Latitude	Longitude	Altitude	a_T	a(l)	a(s)	$\bar{h}$	N ± SE
latifolia									
1 Lime	BC	51.06	121.40	1814	41	0	8	0.207	6391 ± 1399
2 Oie	BC	52.00	121.12	991	48	2	13	0.205	8419 ± 1850
3 Chilco	BC	51.59	123.45	1059	43	0	10	0.176	6886 ± 1647
4 Telkwa	BC	54.39	127.03	518	36	0	4	0.192	4911 ± 1117
5 Udy	BC	53.01	123.14	983	43	1	12	0.212	6987 ± 1509
6 Hudson	BC	56.02	122.05	725	37	0	5	0.171	5156 ± 1249
7 Pink	BC	57.00	122.24	1113	40	0	5	0.198	6079 ± 1363
8 Tetsa	BC	58.40	124.10	762	41	1	5	0.227	6446 ± 1338
9 Lower	BC	59.59	128.33	640	39	2	8	0.183	5755 ± 1346
10 Frances	UT	61.10	129.20	884	41	1	8	0.207	6393 ± 1397
11 Carmacks	UT	62.14	136.18	671	39	0	6	0.199	5793 ± 1293
12 Ethel	UT	63.18	136.28	876	37	0	6	0.187	5189 ± 1198
13 Takhini	UT	60.41	136.11	747	41	1	8	0.183	6330 ± 1481
14 Atlin	BC	59.48	133.47	789	37	1	5	0.172	5156 ± 1249
15 Kinaskan	BC	57.29	130.13	815	44	3	6	0.230	7336 ± 1510
16 Bisson	BC	50.02	118.34	1137	41	0	9	0.201	6376 ± 1418
17 Cartwright	BC	50.49	116.26	1173	41	0	7	0.194	6359 ± 1440
18 Sawdust	BC	49.34	116.04	1661	43	1	9	0.207	6975 ± 1522
19 Telkwa High	BC	54.38	127.26	1005	41	1	10	0.188	6343 ± 1462
20 Inonosklin	BC	49.54	118.12	579	42	0	8	0.188	6633 ± 1526
21 Mt Iemoray	BC	55.33	122.33	732	35	0	2	0.177	4593 ± 1095
22 Purden	BC	53.52	121.48	838	39	2	5	0.179	5747 ± 1359
23 Stone	BC	58.39	124.46	1137	37	0	5	0.184	5182 ± 1208
24 Hinton	AB	53.16	117.09	1204	39	0	5	0.202	5799 ± 1286
25 Kananaskis	AB	51.01	115.02	1501	42	0	7	0.208	6687 ± 1456
26 Carbondale	AB	49.26	114.25	1379	42	0	7	0.200	6665 ± 1483
27 Flyhills	BC	50.43	119.27	1524	38	1	5	0.184	5469 ± 1276
28 Larch Hills	BC	50.42	119.11	777	41	1	8	0.189	6346 ± 1458
29 Manning L.	BC	49.04	120.46	1128	39	0	8	0.187	5764 ± 1333
30 Manning H.	BC	49.04	120.55	1539	40	0	5	0.197	6076 ± 1367
31 Stevens Pa.	WA	47.47	120.56	762	38	1	7	0.166	5431 ± 1338
32 Trout	WA	46.04	121.27	1219	45	1	9	0.224	7607 ± 1593
33 Yellowstone	MT	45.30	110.45	2134	41	0	11	0.191	6350 ± 1452
34 Lewis	MT	46.45	110.30	2134	43	1	12	0.184	6908 ± 1613
36 Bonners F.	ID	48.45	116.30	1219	41	0	9	0.206	6390 ± 1401
Mean					**40.4**	**0.6**	**7.3**	**0.194**	**6198 ± 803**
murrayana									
36 Zig Zag	OR	45.23	121.50	549	43	1	8	0.187	6916 ± 1600
37 Mt. Hood	OR	45.18	121.45	1280	43	2	10	0.192	6931 ± 1579
38 Broken Top	OR	44.08	121.38	1707	44	2	8	0.221	7307 ± 1539
39 Chemule	OR	43.19	121.39	1676	44	2	7	0.234	7347 ± 1499
40 Mcleod	CA	41.16	121.55	1219	38	1	2	0.193	5489 ± 1248
41 Cumboot	CA	41.13	122.30	2134	38	0	8	0.181	5463 ± 1286
42 Patterson	CA	41.11	120.10	2179	33	0	2	0.168	4006 ± 980
43 Bucks	CA	39.53	121.07	1646	37	2	4	0.185	5183 ± 1206
44 Huntington	CA	37.10	119.12	2256	37	1	4	0.199	5214 ± 1164
45 Camp	CA	36.06	118.32	2164	41	0	5	0.216	6416 ± 1371
46 Big Bear	CA	34.13	116.59	2316	32	0	3	0.177	3732 ± 890
Mean					**39.1**	**1.0**	**5.5**	**0.196**	**5819 ± 1265**

Table 1 (continued)

Population Prov./State		Latitude	Longitude	Altitude	a_T	a(l)	a(s)	$\bar{h}$	N ± SE
contorta									
47 Mayer L.	BC	53.39	132.04	23	43	1	12	0.187	6918 ± 1598
48 Masset Rd.	BC	53.49	132.08	23	41	0	9	0.177	6315 ± 1506
49 Bentinck	BC	52.03	126.43	152	33	0	6	0.145	3970 ± 1051
50 Cracroft	BC	50.31	126.36	15	39	0	6	0.162	5707 ± 1427
51 Terrace	BC	54.27	128.35	198	35	0	6	0.162	4566 ± 1141
52 Lulu	BC	49.09	123.06	6	43	1	12	0.191	6930 ± 1581
53 Malahat	BC	48.34	123.34	350	44	1	12	0.190	7215 ± 1651
54 Kitimat	BC	54.04	128.41	76	28	0	0	0.170	2577 ± 627
55 Queets	WA	47.38	124.18	30	36	1	6	0.170	4867 ± 1184
56 LongBeach	WA	46.26	124.03	15	36	0	4	0.188	4902 ± 1130
57 Manzanita	OR	45.43	123.56	15	36	2	6	0.168	4862 ± 1192
58 Pacific C.	OR	45.13	123.57	15	36	0	6	0.169	4865 ± 1188
59 Newport	OR	44.34	124.04	15	39	0	6	0.175	5738 ± 1373
60 Carter	OR	43.50	124.09	15	42	0	8	0.211	6695 ± 1447
61 Rauser D.	OR	43.30	124.14	15	43	0	11	0.213	6991 ± 1504
62 Port Oxford	OR	42.46	124.31	15	42	2	11	0.171	6588 ± 1596
63 Pistol	OR	42.15	124.24	15	35	0	3	0.201	4638 ± 1031
64 Coon Mts.	CA	41.50	123.53	1097	41	4	6	0.196	6364 ± 1433
65 Samoa	CA	40.47	124.13	15	31	0	3	0.176	3443 ± 824
66 Port Orch.	WA	47.25	122.40	76	42	1	11	0.187	6630 ± 1531
Mean					**38.3**	**0.7**	**7.2**	**0.181**	**5540 ± 1316**

Following SCHOEN and BROWN (1991), we used $\bar{\mu} = 10^{-5}$. Further, we estimated the expected heterozygosity under mutation-drift balance [$\theta_j = h_j/(1 - h_j)$, h_j being the average per-locus heterozygosity in the *j*th population] and this estimate of θ_j was employed to give initial values of μ_i and N_j (CHAKRABORTY & NEEL 1989). In our study, n_{ij} was the number of haploid megagametophytes assayed for the *i*th locus and the *j*th population (n_{ij} = 75 for all populations), J = 20 in spp. *contorta*, J = 35 in spp. *latifolia* and J = 11 in spp. *murrayana*, and I = 19 in all 66 populations.

Identifying patterns of gene flow

We estimated SLATKIN (1993)'s index of 'genetic similarity' ($\hat{M}$) from the relationship, $F_{ST} = 1/(4M + 1)$, to identify the pattern of gene dispersal among populations in each subspecies. Then, we plotted $\hat{M}$ against geographic distance in kilometres (d) on a log-log graph. If there was a pattern of isolation by distance, SLATKIN (1993) predicted a linear relationship between $\log_{10}(\hat{M})$ and $\log_{10}(d)$:

$$\log_{10}(\hat{M}) = a + b \log_{10}(d) \qquad [2]$$

Equation [2] is approximately true under a variety of population structure models simulated by SLATKIN (1993). In estimating $\hat{M}$, F_{ST} was a measure of genetic divergence between a pair of sampled populations (WEIR 1990, p.167). We chose to employ WEIR and COCKERHAM'S (1984) $\hat{\theta}$ because it is an unbiased estimator of F_{ST} and has a relatively low bias when used as an estimator of M in Wright's island model (COCKERHAM & WEIR 1993).

Our use of $\hat{M}$ in characterizing patterns of gene flow in *P. contorta* Dougl. is different from the approach based on the relationship between NEI's (1972) genetic distance (D) and geographic distance (*e.g.*, LI & ADAMS 1989; XIE *et al.* 1992). In contrast to D that is dependent on mutation rates, $\hat{M}$ is a direct measure of gene exchange among populations and is independent of mutation rates (SLATKIN 1993). Consequently, SLATKIN (1993) has suggested that D is more appropriate for estimating divergence times between different species whereas $\hat{M}$ is more appropriate for characterizing the extent and pattern of gene dispersal between populations within a single species.

$\hat{M}$ and d were computed for all possible pairs of populations in each subspecies; 595 (35*34/2) pairs in spp. *latifolia*, 55 (11*10/2) in spp. *murrayana* and 190 (20*19/2) in spp. *contorta*. Consequently, we estimated

Table 2 Correlations between geographic and genetic variables in three subspecies of *P. contorta* Dougl

Geographic variable	a(s)	$\bar{h}$	N
latifolia (n = 35)			
Latitude (°N)	−0.421*	−0.049	−0.372*
Longitude (°W)	−0.298	−0.073	−0.287
Altitude (m)	0.384*	0.162	0.306
murrayana (n = 11)			
Latitude (°N)	0.652*	0.181	0.673*
Longitude (°W)	0.512	0.125	0.541
Altitude (m)	−0.408	−0.065	−0.564
contorta (n = 20)			
Latitude (°N)	0.068	−0.468*	−0.130
Longitude (°W)	−0.010	−0.271	−0.104
Altitude (m)	−0.018	0.140	0.152

* siginificant at $P \leq 0.05$

the regression coefficient (b) using equation [2]. A significant regression ($b < 0$) would indicate the presence of isolation by distance. However, there was no parametric test for significance of the b because values of $\hat{M}$ from different population pairs were not independent. For this reason, we obtained an empirical distribution of b by bootstrapping (EFRON 1982; WEIR 1990). Each bootstrap sample was drawn from the data by sampling (with replacement) n times from the n pairs of $\log_{10}(\hat{M})$ and $\log_{10}(d)$, where n = 595 for spp. *latifolia*, n = 55 for spp. *murrayana* and n = 190 for spp. *contorta*. The estimate of b was computed directly from the data and from each of 1000 bootstrap samples. A 95% confidence interval for b was constructed as the interval between the 26th to the 975th of 1000 ordered bootstrap estimates of b. We applied the same resampling procedure to obtain a 95% confidence interval for a, the intercept in equation [2].

RESULTS

Allele distribution and effective population size

The total number of alleles varied considerably among populations, from $a_T = 28$ for population 54 to $a_T = 48$ for population 2 (Table 1). The average per locus ranged between 1.5 and 2.5 alleles over 19 polymorphic loci. On average, *spp. murrayana* had 0.8 more alleles per population than spp. *contorta* while spp. *latifolia* had 1.3 and 2.1 more alleles per population than spp. *murrayana* and *contorta*, respectively. Estimates of mean heterozygosity ($\bar{h}$) ranged from 0.162 (populations 50 and 51) to 0.230 (population 15). While spp. *latifolia* had more alleles than spp. *murrayana*, they had similar mean heterozygosities. However, both subspecies had higher mean heterozygosity than spp. *contorta*. The effective number of alleles per locus was computed as $(1 - \bar{h})^{-1}$ and ranged between 1.19 (Populations 50 and 51) and 1.3 (Population 15). They were less than the observed number of alleles (1.50 – 2.50), suggesting the occurrence of rare alleles.

The number of private alleles per population ranged between zero and four across 66 populations (Table 1). The portion of populations having zero, one, two, three and four private alleles were 19/35, 12/35, 3/35, 1/35 and 0/35 in spp. *latifolia*; 4/11, 3/11 and 4/11, 0/11 and 0/11 in spp. *murrayana*; and 12/20, 5/20 and 2/20, 0/20 and 1/20 in spp. *contorta*. Each population shared varying number of rare alleles with one or more other populations, from two (population 21) to 13 (population 13) in spp. *latifolia*, two (populations 40 and 42) to 10 (population 37) in spp. *murrayana* and zero (population 54) to 12 (populations 47, 52 and 53) in spp. *contorta*. Population 54 had no private alleles nor shared rare alleles with any other population because its allele frequencies were greater than 0.05 across all 19 polymorphic loci.

Table 3 Mean and range of pairwise geographic distances (d) and indices of genetic similarity ($\hat{M}$) in three subspecies of *P. contorta* Dougl.

	Mean	Range
latifilia (n = 595)		
d	826.4	7 – 2565
$\hat{M}$	6.2	1.7 – 36.0
murrayana (n = 55)		
d	523.8	7 – 1303
$\hat{M}$	4.0	0.9 – 19.7
contorta (n = 190)		
d	638.6	11 – 1566
$\hat{M}$	3.7	0.5 – 17.4

Table 4 Estimates of intercept (a) and slope (b) from the regression of genetic similarity (M) on geographic distance (d) on a log-log scale in three subspecies of *Pinus contorta* Dougl.

Subspecies	a	L_a	U_a	b	L_b	U_b
latifolia	0.833	0.716	0.954	–0.032	–0.074	0.010
contorta	0.924	0.674	1.203	–0.169	–0.274	–0.077
murrayana	1.129	0.409	1.588	–0.238	–0.409	0.028

There was a substantialt of variation in the estimates of N across populations, from 2577 ± 627 for population 54 to 8419 ± 1850 for population 2 (Table 1). The mean values of N for the three subspecies showed the following ranking: *latifolia* (N = 6198) > *murrayana* (N = 5819) > *contorta* (N = 5539). The ranges of N for *latifolia*, *murrayana* and *contorta* were 4593 to 8419, 3732 to 7347 and 2577 to 7215, respectively. The coefficients of variation (CV) in N ranked as follows: *latifolia* (CV = 13%) < *murrayana* (CV = 22%) < *contorta* (CV = 24%).

Estimates of correlation between geographic (latitude, longtude and altitude) and genetic [N, a(s) and $\bar{h}$] variables are in Table 2. Longitude did not correlate significantly with genetic variables while altitude correlated significantly only with a(s), in spp. *latifolia*. However, latitude exhibited different patterns of association with genetic variables in the three subspecies. Estimates for N and a(s) with latitude were significantly negative in *latifolia*, significantly positive in *murrayana* but insignificant in *contorta*. On the other hand, correlation between $\bar{h}$ and latitude was significant only in *contorta*.

Patterns of gene dispersal among populations

Figures 2a, 2b and 2c are plots of all pairwise $\hat{M}$ values against their geographic distance (d) in a logarithmic scale for subspecies, *latifolia*, *murrayana* and *contorta*, respectively. Genetic similarity ($\hat{M}$) and geographic distance (d) varied considerably among population pairs within each subspecies (Table 3). The dashed line ($\hat{M}$ = 1) in each figure represents Wright's (1943) criterion to indicate the relative strengths of gene flow and random drift. According to WRIGHT (1943), random drift would result in substantial local differentiation if $\hat{M} < 1$, but not if $\hat{M} > 1$. There was no population pair in spp. *latifolia* and only one (populations 42 and 46, $\hat{M}$ = 0.894, d = 835 km) in spp. *murrayana* for which $\hat{M} <$ 1. However, eight population pairs in spp. *contorta* showed $\hat{M} < 1$. Of these, seven involved population 49 with populations 51 ($\hat{M}$ = 0.834, d = 280), 54 ($\hat{M}$ = 0.727, d = 260), 55 ($\hat{M}$ = 0.917, d = 542), 56 ($\hat{M}$ = 0.736, d = 665), 57 ($\hat{M}$ = 0.536, d = 763), 60 ($\hat{M}$ = 0.839 This corresponded to the ran-king of mean values

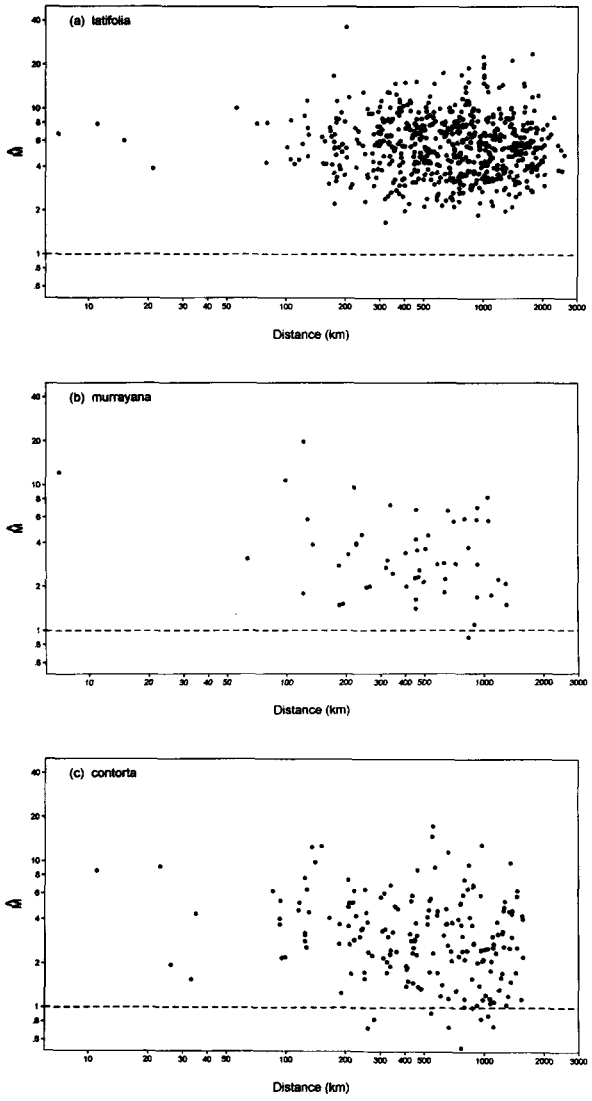

Figure 2 Log $_{10}(\hat{M})$ plotted against log $_{10}$ (distance in km) for three subspecies of *Pinus contorta* Dougl.: (a) *latifolia*, (b) *murrayana*, (c) *contorta*. A slight but significant pattern of isolation by distance was detected in ssp. *contorta*.

remaining pair was between populations 54 and 58 ($\hat{M}$ = 0.881, d = 1050). This corresponded to the ranking of mean values of $\hat{M}$: *latifolia* ($\hat{M}$ = 6.24) > *murrayana* ($\hat{M}$ = 4.04) > *contorta* ($\hat{M}$ = 3.69).

Estimates of regression of $\log_{10}(\hat{M})$ on $\log_{10}$(d) [b in equation (2)] for the three subspecies were all negative and small (Table 4). Judging from the confidence intervals for b generated by bootstrapping, the value of b was significantly different from zero in spp. *contorta*, not in *latifolia* and *murrayana*. Thus, there was a significant pattern of isolation by distance only in spp. *contorta*. To dismiss the possibility that such pattern in spp. *contorta* was due primarily to population 49 as it had low rates of gene flow with many other populations, we excluded population 49 from the analysis. The recomputed regression coefficient (b = –0.152) and its 95% confidence interval (–0.254, –0.041) again suggested the existence of isolation by distance in spp. *contorta*.

DISCUSSIONS

Our range-wide electrophoretic survey of *P. contorta* Dougl. attempted to sample as many populations as possible to cover each subspecies' geographic range delimited by CRITCHFIELD (1957) and others. We sampled 66 populations, 15 – 35 trees per population and 75 megagametophytes per bulk seed collection. Thus, considering that gametes were random in sampled populations, the "effective" sample size per population (MORRIS & SPIETH 1978; LI & ADAMS 1989) for estimating allele frequencies and other genetic parameters ranged from 22 [= 1 / (75^{-1} + 30^{-1})] to 37 [= 1 / (75^{-1} + 70^{-1})]. This range is comparable with those given in other range-wide allozyme surveys [*e.g.*, range 14–39 for *Pseudotsuga menziesii* (LI & ADAMS 1989)].

Since there is a nearly perfect relationship between the total allele number ($a_T = \Sigma_{i=1}^{I} k_{ij}$) and effective population size (N), a direct consequence of our choice of constant mutation rate $\bar{\mu} = 10^{-5}$ (cf. equation [1]), geographic variation in N also reflects that in a_T. It may be argued that estimates of the absolute values of N are dependent on the choice of $\bar{\mu}$. However, our primary interest was in revealing patterns of geographic variation in N. SCHOEN & BROWN (1991) used the same procedure to analyze variation in N in inbreeding versus outbreeding plants.

Estimates of effective population size (N) were over 2,000 in all 66 *P. contorta* Dougl. populations (Table 1). By estimating the neighbourhood area (A) based on CRAWFORD'S (1984) method, EPPERSON and ALLARD (1989) inferred that N might have been at least 1000 trees historically and perhaps several times larger in the two spp. *latifolia* populations from Washington. Large estimates of N in *P. contorta* are probably expected because this conifer is capable of producing a large amount of sound seeds (0.2 – 3.2 million /acre) in open or closed (serotinous) cones (CRITCHFIELD 1980). SCHOEN and BROWN (1991, Fig. 2) used CHAKRABOTY and NEEL's (1989) procedure to estimate effective population sizes for three other conifers: *Pseudotsuga menziesii* (N = 2,000 – 5,000), *Picea abies* (N = 9,000 – 16,000) and *Pinus sylvestris* (N = 8,000 – 10,000). Thus, many conifers, including *P. contorta* Dougl., have large effective population sizes.

Also apparent from Table 1 is a substantial variation in N among and within subspecies. The among-

subspecies variation may be in part explained by the subspecies' life-history and ecological attributes because the subspecies also represent geographically and ecologically distinct regions (CRITCHFIELD 1980; WHEELER & CRITCHFIELD 1985). Much of spp. *latifolia* forests is thought to be of fire origin. Serotinous cones are present in many stands and they contain a large seed reserve released following fire (CRITCHFIELD 1980; WHEELER & CRITCHFIELD 1985). This leads to the rapid establishment of large and even-aged stands with dense stocking. On the other hand, cone serotiny is absent or infrequent in subspecies *murrayana* and *contorta*. Thus, their seeds in open (non-serotinous) cones can be released at any time after maturity. Seeds released in this way may often meet with unfavorable site and climatic conditions for seed germination and seedling establishment. For example, in the Sierra Nevada, the timing and extent of seed germination and seedling establishment of spp. *murrayana* depend critically on meadow variation in topography, water availability, and microclimates (HELMS & RATLIFF 1987). Consequently, spp. *murrayana* and *contorta* stands are probably smaller and less even-aged than spp. *latifolia* stands. Furthermore, since spp. *contorta* is linearly and uniformly distributed along the Pacific Coast (CRITCHFIELD 1957, 1980), the effective population size is expected to be even smaller and the opportunity for local non-adaptive differentiation is greater (WRIGHT 1943). The ranking of averages of N for three subspecies [*latifolia* (N = 6,198) > *murrayana* (N = 5,819) > *contorta* (N = 5,539)] appears to support these expectations.

Geographic patterns in effective population size differed from one subspecies to another (Table 2). There was a trend towards reduced N in northern populations in spp. *latifolia* ($r = -0.372$, $P \leq 0.05$) and the opposite trend was true in spp. *murrayana* ($r = 0.673$, $P \leq 0.05$). However, such latitudinal affinity was not apparent in spp. *contorta*. This is in accordance with CRITCHFIELD's (1957) original demarcation of these subspecies. Being separated from around the Columbia River in Oregon, spp. *latifolia* extended northwards up to the interior Yukon and spp. *murrayana* extended southwards up to the high mountains of southern California and northern Baja California (CRITCHFIELD 1957). In marginal habitats, population density is likely lower and levels of inbreeding may be higher because there often is lack of suitable micro-sites and competitive exclusion by other, more adapted species (YEH & LAYTON 1979). Lack of latitudinal affinity in spp. *contorta* may reflect that all populations in this Coastal subspecies have experienced the extreme edaphic and climatic conditions (CRITCHFIELD 1980) so that among-population fluctuation in N is not associated with the geographic location of the populations.

The among-population variations in $\bar{h}$ and a_T (Table 1) deserve special mentioning. Value of $\bar{h}$ is dependent on allele numbers and frequencies which represent allele "richness" and "evenness", the two complementary components of genetic diversity (BROWN and WEIR 1983). It is of interest to inquire how much of the variation in $\bar{h}$ is attributable to the variation in a_T. We computed coefficients of determination (r^2) for three subspecies: $r^2 = 0.348$ for spp. *latifolia*, 0.550 for spp. *murrayana* and 0.292 for spp. *contorta*.Thus, about 30% to 55% of the variation in $\bar{h}$ can be explained by that in a_T. Accordingly, reduction in the number of alleles leads to decreased heterozygosity. For example, the estimates of $\bar{h}$ in this study were higher than those in YANG and YEH (1993) because the estimates in YANG and YEH (1993) were based on two alleles, the most frequent and a "synthetic" allele consisting of all other observed alleles combined.

There was generally extensive gene flow ($\hat{M} > 1$) among populations in each subspecies of *P. contorta* Dougl. (Fig. 2). The number of private alleles per population was very low, averaging 0.6, 1.0 and 0.7 for subspecies *latifolia*, *murrayana* and *contorta*, respectively (Table 1). In fact, more than half of the populations did not have private alleles, suggesting extensive gene flow among populations (SLATKIN 1985). Furthermore, gene exchange between pairs of populations appeared to be independent of their geographic distances. Judging from confidence intervals for the regression of genetic similarity on geographic distance by bootstrapping, this is certainly true for subspecies *latifolia* and *murrayana* (Table 4). EPPERSON and ALLARD (1989) also noted that two population of spp. *latifolia* separated by only 11 kilometers differed at some allozyme loci in allele frequencies for as much as populations at different ends of the vast geographic range of the subspecies (YEH & LAYTON 1979; WHEELER & GURIES 1982a). Thus, WRIGHT's (1943) island model may be invoked to explain the observed patterns of gene flow in subspecies *latifolia* and *murrayana*. Extensive gene flow coupled with large effective population size leads to the scenario that "very little nonadaptive differentiation is to be expected and only rather strong differences in the action of selection avoid swamping" (WRIGHT 1943, p. 135).

Despite extensive gene flow, a slight but significant pattern of isolation by distance was found in spp. *contorta* (Table 4). This observed pattern in spp. *contorta* may be best explained by invoking a linear continuum model (WRIGHT 1943) or one dimensional stepping stone model (KIMURA & WEISS 1964) of population structure as this subspecies is linearly distributed along the Pacific Coast (Fig. 1) and the

linear distribution predicts the formation of isolation by distance (WRIGHT 1943). The smaller effective population sizes were also observed for the populations in spp. *contorta* (Table 1). Nevertheless, spp. *contorta* occupies a variety of extreme habitats along the coast, ranging from bogs and muskeg in the far north to cliffs, sand dunes and rocky sites in the central and southern portion of its distribution (CRITCHFIELD 1980). Conceivably, the dynamics of migration, divergent selection and random drift may often reach a steady state in this subspecies (WRIGHT 1931; SLATKIN 1985). Should this occur, it would be difficult for subspecies *contorta* to develop a strong pattern of isolation by distance.

Our observed patterns of gene dispersal among populations across the natural range of *P. contorta* Dougl. also give insight into the species' biogeography and natural history. Since the present range of spp. *murrayana* is largely in unglaciation areas and might not have been influenced by the Wisconsin glaciation (WHEELER & GURIES 1982b), our discussion will focus on the two northerly subspecies, *latifolia* and *contorta*.

Geographically restricted gene dispersal among populations detected in spp. *contorta* suggests that the adjacent populations are genetically more similar than distant populations. Biogeographically, this pattern is consistent with the paleobotanists' view that spp. *contorta* survived the Pleistocene glaciation south of the ice and migrated north to its current range following glacial retreat (*e.g.*, HOPKINS *et al.* 1981). WHEELER and GURIES (1982b) argued that should this model be correct, the genetic constitution of northern populations could be predicted by (i) a close genetic affinity; (ii) a reduced frequency and distribution of rare alleles and (iii) a reduced level of genetic diversity. Since none of these conditions were observed in their study, these authors rejected the above model and proposed that their data were better explained by a multiple north-cost refugia model. In contrast, we observed a significant negative correlation between latitude and average heterozygosity ($r = -0.466$, $P \leq 0.05$; Table 2), suggesting that northern populations were significantly less variable than southern populations. However, correlation of latitude with the number of shared rare alleles was not significant and the pattern of isolation by distance was rather weak. Therefore, our indirect genetic evidence favoring a single, southern refugium was suggestive, but inconclusive. In order to obtain the more conclusive evidence, it is necessary to have more extensive sampling of trees, particularly within populations, to allow for accurate assessment of spatial distribution of genetic diversity (EPPERSON & ALLARD 1989).

Lack of isolation by distance in spp. *latifolia* is expected if the populations originate from multiple refugia following Pleistocene glacial retreat. Such biogeographical model has been suggested earlier (WHEELER and GURIES 1982b; WHEELER and CRITCHFIELD 1985). We again used WHEELER and GURIES' (1982b) three predictive conditions to verify this model. This time, both genetic similarity and mean heterozygosity showed no clinal (latitudinal) pattern but the number of shared rare alleles did exhibit a latitudinal cline ($r = -0.421$, $P \leq 0.05$; Table 2). Since our data are consistent with most of WHEELER and GURIES' (1982b) criteria, we favor the multiple refugia model.

Three putative migratory pathways following Pleistocene glacial retreat have been proposed (WHEELER & CRITCHFIELD 1985, Fig. 9) for spp. *latifolia*. They included one from the Yukon refugium and two from the southern refugia, all pointing to north-central British Columbia. If this proposal is correct, then spp. *latifolia* populations in north-central British Columbia may constitute a 'centre of diversity'. Should such a centre exist, it would be of great value to improvement, management and conservation of genetic resources in spp. *latifolia* because it presumably would have assembled genetic variants originated from different parts of the natural range. Accordingly, this centre of diversity would be the target place for designing *in situ* sites or sampling for *ex situ* conservation (MARSHALL & BROWN 1975; YANG & YEH 1992). We have calculated $\bar{h}$ for each of three groups, Yukon (populations 10, 11, 12, 13, 14), central (populations 7, 8, 9, 15, 23) and southern (populations 16, 17, 18, 20, 27, 28, 29, 30, 31, 35). Their ranking, central group (0.204) > southern group (0.192) > Yukon group (0.189), seems to support the proposal advanced by WHEELER and CRITCHFIELD (1985). However, additional sampling of populations, particularly around Kinaskan (Population 15) where $\bar{h}$ is highest (Table 1), would be needed to confirm the hypothesis of a centre of diversity in north-central British Columbia.

ACKNOWLEDGMENTS

We thank N. C. Wheeler, J. L. Hamrick and an anonymous reviewer for many valuable comments, K. Ritland and M. Slatkin for helpful discussion during the course of this research and M. Slatkin for providing a copy of his program for estimating $\hat{M}$. This research was supported in part by grants from the Natural Sciences and Engineering Research Council of Canada (A2282) and Petawawa National Forestry Institute to F. C. Y.

REFERENCES

BARTON, N. H. & SLATKIN, M. 1986: A quasi-equilibrium theory of the distribution of rare alleles in a subdivided population. *Heredity* **56**:409–415.

BROWN, A. H. D. & WEIR, B. S. 1983: Measuring genetic variability in plant populations. *In:* Isozymes in plant genetics and breeding, Part A. (eds. S. D. Tanksley and T. J. Orton). pp. 219–239. Elsevier Science Publishers B. V., Amsterdam.

CHAKRABORTY, R. & NEEL, J. V. 1989: Description and validation of a method for simultaneous estimation of effective population size and mutation rate from human. *Proc. Natl. Acad. Sci. USA* **86**:9407–9411.

COCKERHAM, C. C. & WEIR, B. S. 1993: Estimation of gene flow from *F*-Statistics. *Evolution* **47**:855–863.

CRAWFORD, T. J. 1984: The estimation of neighbourhood parameters for plant populations. *Heredity* **52**:273–283.

CRITCHFIELD, W. B. 1957: Geographic variation in *Pinus contorta*. Maria Moors Cabot Foundation Publication No. 3, Harvard University, Cambridge, MA, 118 pp.

CRITCHFIELD, W. B. 1980: Genetics of lodgepole pine. U. S. For. Serv. Res. Pap. WO–37, 57 pp.

EFRON, B. 1982: The jackknife, the bootstrap and other resampling plans. S. I. A. M., PA.

EPPERSON, B. K. & ALLARD, R. W, 1989: Spatial autocorrelation analysis of the distribution of genotypes within populations of lodgepole pine. *Genetics* **121**:369–377.

EWENS, W. J. 1972: The sampling theory of selectively neutral alleles. *Theor. Pop. Biol.* **3**:87–112.

FORREST, G. I. 1980: Geographic variation in the monoterpenes of *Pinus contorta* oleoresin. *Bioch. Syst. Ecol.* **8**:343–359.

GOVINDARAJU, D. R. 1988: Relationship between dispersal ability and levels of gene flow in plants. *Oikos* **52**:31–35.

HAMRICK, J. L., GODT, M. J. W. & SHERMAN-BROYLES, S. L. 1992: Factors influencing levels of genetic diversity in woody plant species. *New Forests* **6**:95–124.

HELMS, J. A. & RATLIFF, R. D. 1987: Germination and establishment of *Pinus contorta* var. *murrayana* (Pinaceae) in mountain meadows of Yosemite National Park, California. *Madroño* **34**:77–90.

HOPKINS, D. M., SMITH, P. A. & MATTHEWS JR., J. V. 1981: Dated wood from Alaska and the Yukon: implications for forest refugia in Beringia. *Quat. Res.* (N. Y.) **15**: 217–249.

KIMURA, M. & WEISS, G. H. 1964: The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics* **49**:561–576.

LI, P. & ADAMS, W. T. 1989: Range-wide patterns of allozyme variation in Douglas-fir (*Pseudotsuga menziesii*). *Can. J. For. Res.* **19**:149–161.

MANLY, B. F. J. 1985: The Statistics of Natural Selection. Chapman and Hall, London, 484 pp.

MARSHALL, D. R. & BROWN, A. H. D. 1975: Optimum sampling strategies in genetic conservation. *In*: Crop Genetic Resources for Today and Tomorrow. (eds. O. H. Frankel and J. G. Hawkes). pp. 53–80. Cambridge Univ. Press, London.

MORRIS, R. W. & SPIETH, P. T. 1978: Sampling strategies for using female gametophytes to estimate heterozygosity in conifers. *Theor. Appl. Genet.* **51**:217–222.

NEI, M. 1972: Genetic distance between populations. *Am. Nat.* **106**:283–292.

NEI, M. 1973: Analysis of gene diversity in subdivided populations. *Proc. Natl. Acad. Sci. USA* **70**:3321–3323.

SCHOEN, D. J. & BROWN, A. H. D. 1991: Intraspecific variation in population gene diversity and effective population size correlates with the mating system in plants. *Proc. Natl. Acad. Sci. USA* **88**:4494–4497.

SLATKIN, M. 1985: Rare alleles as indicators of gene flow. *Evolution* **39**:53–65.

SLATKIN, M. 1993: Isolation by distance in equilibrium and non-equilibrium populations. *Evolution* **47**:264–279.

SLATKIN, M. & BARTON, N. H. 1989: A comparison of three indirect methods for estimating average levels of gene flow. *Evolution* **43**:1349–1368.

WEIR, B. S. 1990: Genetic data analysis. Sinauer Associates, Sunderland, MA, 377 pp.

WEIR, B. S. & COCKERHAM, C. C. 1984: Estimating *F*-statistics for the analysis of population structure. *Evolution* **38**:1358–1370.

WHEELER, N. C. & CRITCHFIELD, W. B. 1985: The distribution and botanical characteristics of lodgepole pine: Biogeographical and management implications. *In:* Lodgepole pine: The species and its management. (eds. D. M. Baumgartner, R. G. Krebill, J. T. Arnott, and G. F. Weetmen). pp. 1–13. Washington State Univ. Cooperative Extension Service, Pullman, WA.

WHEELER, N. C. & GURIES, R. P. 1982A: Population structure, genetic diversity, and morphological variation in *Pinus contorta* Dougl. *Can. J. For. Res.* **12**:595–606.

WHEELER, N. C. & GURIES, R. P. 1982B: Biogeography of lodgepole pine. *Can. J. Bot.* **60**:1805–1814.

WRIGHT, S. 1931: Evolution in Mendelian populations. *Genetics* **16**:97–159.

WRIGHT, S. 1943: Isolation by distance. *Genetics* **28**:114 –138.

XIE, C. Y., DANCIK, B. P. & YEH, F. C. 1992: Genetic structure of *Thuja orientalis*. *Bioch. Syst. Ecol.* **20**:433–441.

YANG, R. -C. & YEH, F. C. 1992: Genetic consequences of *in situ* and *ex situ* conservation of forest trees. *For. Chron.* **68**:720–729.

YANG, R. -C. & YEH, F. C. 1993: Multilocus structure in *Pinus contorta* Dougl. *Theor. Appl. Genet.* **87**:568–576.

YEH, F. C. & 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, F. C., CHELIAK, W. M., DANCIK, B. P., ILLINGWORTH. K., TRUST, D. C. & PRYHITKA, B. A. 1985: Population differentiation in lodgepole pine, *Pinus contorta* spp. *latifolia*: a discriminant analysis of allozyme variation. *Can. J. Genet. Cytol.* **27**:210–218.

YING, C. C., ILLINGWORTH, K. & CARLSON, M. 1985: Geographic variation in lodgepole pine and its implications for tree improvement in British Columbia. *In*: Lodgepole pine: The species and its management. (eds. D. M. Baumgartner, R. G. Krebill, J. T. Arnott, and G. F. Weetmen). pp. 45–53. Washington State Univ. Cooperative Extension Service, Pullman, WA.