

# Genetic structure of black spruce on two contrasting sites

Timothy Boyle,\*  
Chaiyasit Liengsiri† and  
Chana Piewluang†

\*Petawawa National Forestry Institute, Chalk River,  
Ontario, Canada KOJ 1J0.

†ASEAN-Canada Forest Tree Seed Centre,  
Muak Lek, Saraburi 18180, Thailand.

Several hundred dominant and co-dominant individuals in two populations of black spruce, located within 2 km of each other and representing upland and lowland site types, were sampled and mapped. Multilocus isozyme genotypes were determined and the genetic organization of the two populations examined by means of nearest-neighbour analysis and spatial autocorrelation of alleles. On the upland site type, population density was much higher and variation in age considerably less than on the lowland site type, which was interpreted as evidence for fire origin in the former. On the upland site, the mature trees were more inbred, possibly due to the conditions under which regeneration occurred and because of weak selective pressure. On the lowland site, clustering of similar genotypes was detectable on a local scale but the mature trees were not inbred. The two populations were very similar in terms of allozyme frequencies.

## INTRODUCTION

The interaction of genetic forces, in particular migration and selection, will affect the distribution of genetic variation within populations, and this can influence the efficacy of genetic sampling (Ledig, 1974). Although genetic organization within populations has received far less attention than among-population variation, there is some evidence to suggest that genes and genotypes are not distributed at random within populations of conifers. For example, Linhart *et al.* (1981) detected significant differences in allele frequencies among sub-populations of ponderosa pine (*Pinus ponderosa* Laws.) separated by less than 100 metres in a spatially heterogeneous environment. The absence of spatial heterogeneity in pollen allele frequencies in six populations of black spruce (*Picea mariana* [Mill.] B.S.P.) in central New Brunswick, together with other evidence, led Boyle and Morgenstern (1985a) to conclude that family clustering was weakly developed.

There has been some debate over genetic differences between so-called ‘upland’ and ‘lowland’ black spruce populations. The former are characteristically located on clay soils and regenerate after fire, and are therefore typically even-aged. In contrast, lowland populations are found on organic soils in moist or wet conditions. Fire is not

so frequent and regeneration often occurs without disturbance. This results in multi-aged populations in which overall population density is high, but the density of dominant and co-dominant individuals is frequently lower than on upland sites. Despite the terminology used, topographical distinction between the two site types is often subtle.

In this study, two populations from northern Ontario were sampled and multilocus allozyme genotypes determined for several hundred trees. Although separated by only 2 km, one population was on a typical lowland site, whereas the other was a characteristic upland population. The distribution of genetic variation in the two populations was studied by means of nearest-neighbour analysis and spatial autocorrelation.

## MATERIALS AND METHODS

### Sampling

Two geographically proximate, but ecologically contrasting, populations of black spruce were identified in Kennedy Township, near Cochrane, Ontario (Lat. 49°07'N; Long. 80°42'E). Site 1 was located in a low-lying, wet area, adjacent to a swampy creek, and the soil consisted of a thick

layer of organic matter. The arboreal flora consisted almost purely of black spruce with a wide range of sizes. Site 2, which was 2 km to the south, was located on a drier site with a much thinner layer of organic matter overlying a clay soil. The composition of the tree cover was similar to site 1, but with very little variation in tree size.

On both sites, an area was mapped and dominant and co-dominant trees marked prior to a commercial logging operation. On site 1, 206 trees were marked in an area of 0.971 ha. On site 2, 367 trees were marked in 0.284 ha. Cones and/or foliage was collected by felling. Some trees were lost during the felling operation, resulting in the collection of material from 175 trees on site 1 and 364 on site 2. For about 10 per cent of the sampled trees on each site (12 on site 1, 30 on site 2), a basal disc was cut and the age of the tree determined by ring counts.

#### Laboratory procedures

Using procedures described by Boyle and Morgestern (1985b), maternal genotypes were established from haploid megametophytic tissue of at least 10 seeds per tree for the 13 allozyme loci listed in table 1. Vegetative tissue was used for those trees from which no cones were collected, following the methods described by Liengsiri *et al.* (1990). In order to confirm the interpretation of zymograms from the haploid and diploid material, both methods were used for a small number of trees from which both seed and vegetative tissue had been collected.

#### Statistical analyses

The genetic diversity of the two populations was expressed as the mean expected heterozygosity over all loci, where the expected heterozygosity at

any one locus is defined as:

$$h_e = 1 - \sum_i p_i^2$$

$p_i$  being the frequency of the  $i$ th allele. One minus the ratio of observed to expected heterozygosity provides an estimate of the fixation index of the population (Wright, 1965).

The degree of differentiation among the two populations was expressed in terms of Nei's  $G_{st}$  statistic (Nei, 1973) and genetic distance (Nei, 1972). The standard error of genetic distance was calculated according to the method of Nei and Roychoudhury (1974).

The spatial arrangement of trees, as well as individual genotypes, was examined by means of nearest-neighbour analysis (Clark and Evans, 1954) and spatial autocorrelation (Cliff and Ord, 1981). In nearest-neighbour analysis, the distance to the nearest individual having the same attribute is calculated and compared with the expectation under a null hypothesis of a completely random distribution. In an infinitely large population, this expected value  $r_E$  is given as:

$$r_E = 1/2\sqrt{\rho}$$

where  $\rho$  is the average density of individuals possessing some attribute, and the standard error of this expectation is:

$$\sigma(r_E) = 0.26136/\sqrt{(N\rho)}$$

where  $N$  is the number of distance measurements in the sample. The ratio of mean observed distance to nearest neighbour ( $r_A$ ) to the expected value is then an indication of departure from randomness. Values less than one may be taken to indicate some form of clustering. Significance testing is performed by calculating the standard normal variate  $c$ , which is:

$$c = |r_A - r_E|/\sigma(r_E).$$

Table 1 Isozyme loci used in the study

Enzyme	Abbreviation	EC No.	No. of loci
Phosphoglucosmutase	PGM	2.7.5.1	1
Isocitrate dehydrogenase	IDH	1.1.1.42	1
6-Phosphogluconate dehydrogenase	6-PGD	1.1.1.44	2
Glucose-6-phosphate dehydrogenase	G6PDH	1.1.1.49	1
Malate dehydrogenase	MDH	1.1.1.37	2
Shikimic acid dehydrogenase	SDH	1.1.1.25	1
Aspartate aminotransferase	AAT	2.6.1.1	2
Glutamate dehydrogenase	GDH	1.4.1.3	1
Phosphoglucose isomerase	PGI	5.3.1.9	1
Leucine aminopeptidase	LAP	3.4.11.1	1

For finite populations, edge effects invalidate this procedure, but simulation studies (Donnelly, 1978) have indicated that suitable adjustments can be made to the expectation and variance to produce unbiased estimates for populations of any size, such that:

$$r_E = 0.5\sqrt{(A/N)} + 0.0514P/N + 0.041P/N^{3/2}$$

and

$$\sigma(r_E) = 0.0703A/N^2 + 0.037P\sqrt{(A/N^5)}$$

where *A* is the total study area and *P* is the length of the perimeter (Upton and Fingleton, 1985).

For the spatial autocorrelation, each tree was assigned a value depending on the presence or absence of a specific allele. If the tree is a homozygote for the allele of interest, the value assigned (*z<sub>i</sub>*) is (1 - *p*) where *p* is the frequency of that allele. If the individual is a heterozygote the value assigned is (0.5 - *p*), and if the allele is absent the value assigned is -*p*. These values were chosen in order to produce a mean value of zero for *Z<sub>i</sub>*. Moran's *I*-statistic (Upton and Fingleton, 1985) was then calculated as:

$$I = n \sum_{i \neq j} w_{ij} z_i z_j / \left( S_0 \sum_i z_i^2 \right)_i$$

where *z<sub>i</sub>* is the value of tree *i* ([1 - *p*], [0.5 - *p*], or -*p*), *z<sub>j</sub>* is the value of tree *j*, *w<sub>ij</sub>* is a weighting applied to the product *z<sub>i</sub>z<sub>j</sub>*, *S<sub>0</sub>* is the sum of all weightings ( $\sum_{i \neq j} w_{ij}$ ) and *n* is the number of joins, i.e., the total number of pairwise comparisons among trees.

A "correlelogram" (Upton and Fingleton, 1985) was produced for each population by calculating *I* statistics for trees separated by different distances. A weighting (*w<sub>ij</sub>*) of 1 is applied to a pair of trees separated by less than a specified distance, and a weighting of 0 is assigned otherwise. For each of the two populations, five distance classes were selected such that approximately equal numbers of joins fell into each distance class, and Moran's *I* was calculated for each class.

RESULTS

The two populations were very different in terms of population density and age structure. The density of sampled trees on site 2 was 1292 per hectare compared with only 212 per hectare on site 1. The density of all trees on site 1 was much higher than this value, but the numbers of sub-dominants, saplings, and seedlings was not re-

corded. The average age of dominants and co-dominants on site 1 was 161 years, with a standard deviation (S.D.) of 25.9 and coefficient of variation (*cv.*) of 16.1 per cent. The equivalent values for site 2 were mean: 82 years, S.D.: 4.5, and *cv.*: 5.5 per cent.

Between two and four alleles were observed at each of the 13 loci (table 2), although many of

**Table 2** Allele frequencies, heterozygosities and fixation indices for the two populations

Locus	Allele	1 (Lowland)	2 (Upland)
PGM	1	0.630	0.498
	2	0.213	0.313
	3	0.145	0.188
	4	0.012	0.002
IDH	1	0.994	0.997
	2	0.006	0.003
6PGD-1	1	0.366	0.318
	2	0.631	0.675
	3	0.003	0.007
6PGD-2	1	0.875	0.982
	2	0.110	0.018
	3	0.012	—
	4	0.003	—
G6PDH	1	0.813	0.787
	2	0.152	0.187
	3	0.027	0.027
	4	0.009	—
MDH-1	1	0.994	0.994
	2	0.003	0.006
	3	0.003	—
MDH-3	1	0.882	0.872
	2	0.047	0.088
	3	0.071	0.039
	4	—	0.002
SDH	1	0.627	0.569
	2	0.373	0.431
AAT-1	1	0.932	0.939
	2	0.059	0.057
	3	0.009	0.004
AAT-2	1	0.944	0.961
	2	0.003	0.026
	3	0.003	0.013
GDH	1	0.920	0.866
	2	0.080	0.134
PGI	1	0.873	0.861
	2	0.107	0.126
	3	0.009	0.008
	4	0.012	0.004
LAP	1	0.828	0.710
	2	0.172	0.290
<i>h<sub>e</sub></i>		0.242	0.251
<i>h<sub>0</sub></i>		0.240	0.202
<i>F</i>		0.008	0.195

these alleles were extremely rare. Although more alleles were recorded from site 1, the expected heterozygosity was higher on site 2. Comparison of observed with expected heterozygosities (table 2) showed that the two values were almost identical on site 1, yielding a fixation index close to zero. Conversely, on site 2 the average observed heterozygosity was substantially less than the expected value, and the resulting estimate of the fixation index was 0.195.

Comparison of the allele frequencies observed on the two sites by means of Nei's  $G_{st}$  statistics (table 3), indicated that only small differences existed between the sites. The overall  $G_{st}$  value of 0.009 means that, of the total genetic variation observed, less than 1 per cent was due to differences in allele frequencies between the two sites. Figures for individual loci show that such differences as exist are largely due to only two loci, 6PGD-2 with a  $G_{st}$  value of 0.038, and LAP with a  $G_{st}$  value of 0.027. The only other locus with a  $G_{st}$  value greater than 0.01 was PGM ( $G_{st} = 0.013$ ). Nei's genetic distance between the two populations was 0.0051, which was not significantly greater than zero.

Values of the Clark-Evans ratio of observed to expected mean distance to nearest neighbour are shown in table 4. Values less than one indicate clustering of alleles, whilst values greater than one indicate overdispersion. Figures are only given for alleles occurring in at least two individuals. As can be seen from table 4, every allele in the population from site 1 shows evidence of clustering and, in most cases, the ratio of observed to expected distance is significantly less than one. In contrast, for the second population, 13 alleles yielded ratios greater than one.

Table 3 Genetic diversity measures

Locus	$H_t$	$H_s$	$G_{st}$
PGM	0.585	0.578	0.013
IDH	0.009	0.009	0.001
6PGD-1	0.456	0.455	0.002
6PGD-2	0.134	0.129	0.038
G6PDH	0.331	0.331	0.002
MDH-1	0.012	0.012	0.000
MDH-3	0.226	0.224	0.008
SDH	0.481	0.479	0.003
AAT-1	0.121	0.121	0.000
AAT-2	0.091	0.190	0.008
PGI	0.235	0.234	0.001
LAP	0.401	0.390	0.027
Overall	0.252	0.250	0.009
Genetic distance (between pops)	0.0051 (0.0045)		

Table 4 Clark-Evans' ratio of observed to expected mean distance to nearest neighbour. Values marked "\*" are significantly different from 1.0 at the 5 per cent level, and those marked "\*\*" are significant at the 1 per cent level.

Locus	Allele	Population 1	Population 2
PGM	1	0.76**	1.13**
	2	0.38**	0.92**
	3	0.41**	0.94**
6PGD-1	1	0.87*	1.11**
	2	0.42**	0.98**
6PGD-2	1	0.83*	1.23**
G6PDH	1	0.94	1.36**
	2	0.34	0.85
MDH-1	1	0.71**	1.12**
MDH-3	1	0.73**	1.14**
	2	0.42**	0.91**
	3	0.39	0.90
SDH	1	0.85*	1.25**
	2	0.38**	0.90**
AAT-1	1	0.73**	1.14**
	2	0.28**	0.77**
AAT-2	1	0.86*	1.22**
GDH	1	0.89	1.04
	2	0.22	0.56
PGI	1	0.82**	1.17**
	2	0.42*	1.08
LAP	1	0.84*	1.23**
	2	0.41**	1.00

Fig. 1 shows the distribution of spatial autocorrelation across distance classes. For site 1, there is evidently clustering over a distance defined by distance class 1 (<8 m), but by distance class 2 (<40 m) evidence of clustering has disappeared. Onsite 2, there is no pattern to the autocorrelation across distance classes.

## DISCUSSION

Clearly, the two populations are structurally very different, reflecting their different origin. Site 2 has a very high density of trees, all of approximately the same age, which is what would be expected to result from regeneration following fire. On the other hand, the population on site 1 has a much lower density of dominants and co-dominants, with a much larger range in age. Thus, it appears that this stand has regenerated without a major disturbance. These density differences are much more striking than those found by O'Reilly *et al.*

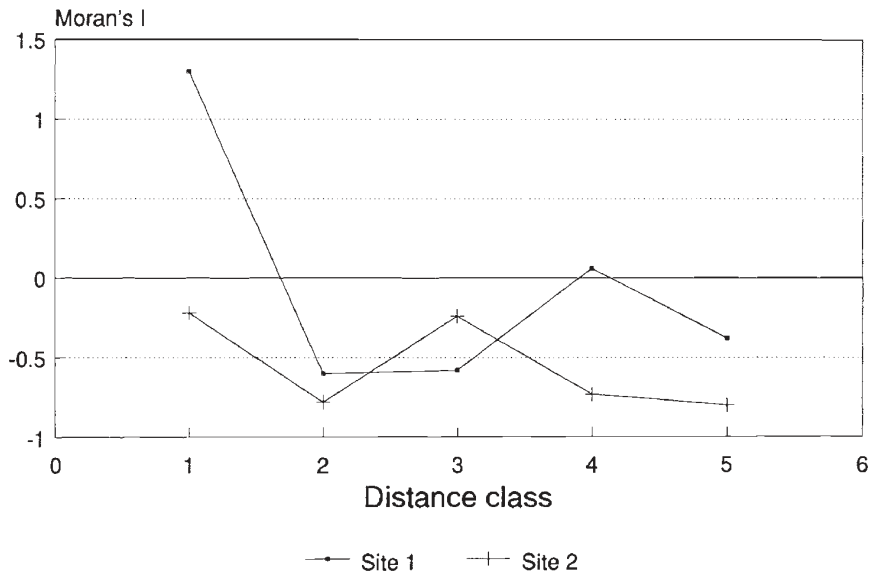


Figure 1 Moran's *I*-statistic plotted for trees separated by different distance classes.

(1985), who reported very similar stand densities for ecologically contrasting populations.

The expected heterozygosity on site 2 was slightly higher than for site 1, but observed heterozygosity on site 2 was considerably less than expected, yielding a large estimate of the fixation index (0.195). This estimate is considerably greater than those reported in other studies. For example, the largest value among six New Brunswick populations was only 0.054 (Boyle and Morgenstern, 1986).

O'Reilly *et al.* (1985) found mean observed heterozygosities between 0.16 and 0.27 for ten populations in Ontario. Specifically, for a lowland: upland comparison from the Cochrane area, the reported values were 0.25:0.23, compared with 0.24:0.20 for the site 1: site 2 comparison in this study. Although the differences were small, the consistently higher heterozygosity in lowland populations was interpreted by O'Reilly *et al.* (1985) to reflect ecological differences between the two site types, with more stringent selection pressures on upland sites and a greater requirement for vigorous growth on lowland sites in response to nutrient stress.

An alternative interpretation relates to the ecological history of the two populations. Clearly, site 2 regenerated after fire, whereas site 1 had remained relatively undisturbed over a considerable period. The mating system contributing to the regeneration on site 2 would therefore have a major

influence on the genetic organization of the resulting population. Specifically, if only a small number of trees survived the fire and a series of poor seed years followed, a high level of inbreeding would be expected among the regenerating seedlings. According to Heinzelman (1957), black spruce cones are semi-serotinous, retaining viable seed for several years. However, it appears that this characteristic is variable, and from second-year and older cones collected from these populations, almost no viable seed was recovered. The six black spruce populations in central New Brunswick studied by Boyle and Morgenstern (1986) were also believed to be of fire origin, but had much lower inbreeding levels. In fact, in four populations excesses of heterozygotes were observed compared with expectations. The level of inbreeding resulting from regeneration following fire will be strongly dependent on the size and intensity of the fire and the quality of subsequent seed years. A small fire, followed by heavy seed production, will yield populations with low levels of inbreeding.

Apart from this difference in the fixation index, the two populations are similar in terms of allozyme frequencies. The mean  $G_{st}$  value is low compared with many other conifer species (Hamrick *et al.*, 1979), but it is very similar to the estimate obtained by Boyle and Morgenstern (1987) for New Brunswick populations of black spruce.

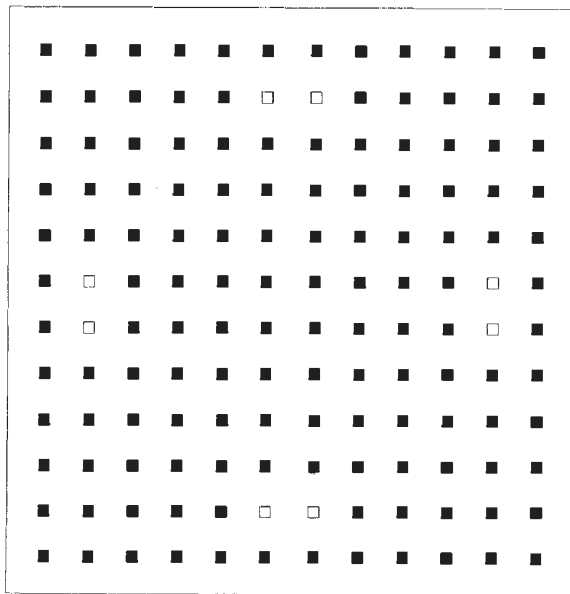
Although the population on site 2 in this comparison was more inbred, the spatial arrangement



of genotypes was more random. Significant deviations from a null hypothesis of random distribution were fewer than for site 1 by Clark-Evans nearest-neighbour analysis and, in fact, there were more alleles with ratios greater than one on site 2 than those with ratios less than one.

Interpretations of the two methods of spatial analysis differ. The nearest neighbour analysis only provides evidence for very localized clustering, whereas the spatial autocorrelation analysis results indicate a much broader pattern. The difference can be visualized by considering the hypothetical distribution shown in fig. 2. By the Clark-Evans analysis, there is an extreme and highly significant clustering of the "2" allele (ratio = 0.36\*\*). Spatial autocorrelation, however, indicates only a slight departure from expectation ( $I = -0.00012$ ;  $E(I) = -0.00833$ ). Thus, a comparison of results from the two methods suggests that on site 1 there is a localized clustering of similar genotypes indicative of a high coefficient of relationship among neighbouring trees. However, the development of large groups of related trees and the formation of genetically divergent neighbourhoods has not occurred. On site 2, neither effect is detectable, and the distribution of trees appears to be essentially random.

Knowles (1990) described the pattern of spatial arrangement in two black spruce populations in



**Figure 2** Hypothetical distribution of genotypes yielding very different results from nearest-neighbour and spatial autocorrelation analyses. Symbols ■ = 11; □ = 22.

northwestern Ontario, based on spatial autocorrelation analyses, as "patchy and subtle". She did, however, detect some highly localized clustering of clonally related trees as a result of layering. Multilocus genotype comparisons failed to detect clonally related trees in the lowland population. It appears that regeneration without disturbance has led to related individuals being located close to each other. Concern over the efficacy of the comparison-tree method of plus-tree selection has been expressed in conditions where family clustering is significant (Ledig, 1974). It appears that such concerns may be valid for undisturbed black spruce such as on site 1, but they are less likely to be valid for fire-regenerated populations.

The high fixation index on site 2 would, at first sight, appear to be contradictory to the observation of a lack of clustering. However, black spruce has a much greater potential for seed distribution compared with other species, especially in a large open area following a fire and if, as is typical, seed is released in winter when there is a snow cover. Therefore, although high levels of inbreeding could result from limited pollen production in small groups of surviving trees, efficient seed dispersal would eliminate clustering and result in the pattern of spatial autocorrelation across distance classes seen in fig. 1. In contrast, on site 1, seed dispersal would be less efficient in undisturbed, closed-stand conditions, a condition under which local groups of related trees would develop. Under such circumstances an asymptotic value of the fixation index would be approached whereas, with fire regeneration, the fixation index would fluctuate from generation to generation, depending on specific conditions of strobilus production following each fire.

Implications of these results for tree improvement programs include the following. Levels of inbreeding detectable in fire-generated populations are likely to be variable but, because of an absence of family clustering, open-pollinated seed collected from mature individuals will not be significantly inbred, as reported by Boyle and Morgenstern (1986). The comparison-tree method of plus-tree selection will not be fully effective in undisturbed populations, because the comparison assumes genetic independence (Ledig, 1974) which does not exist, at least locally. Based on these isozyme results, seed from local populations can safely be interchanged for planting on opposite site types. This conclusion remains to be confirmed for non-allozyme, selectively relevant, traits such as root growth and structure.

**Acknowledgements** The authors wish to acknowledge the valuable contributions of W. Selkirk, D. Hoelke, P. Copis, and J. Veen in mapping the populations and collecting the seed, the Ontario Ministry of Natural Resources for helping to locate suitable populations, H. Recoskie for allowing his commercial felling operations to be disturbed, E. Gilchrist for the seed extraction, D. Hoelke for the computer-generated plots, Dr S. Magnussen for discussions of the spatial analysis techniques, and Drs D. Joyce and P. Knowles for providing valuable suggestions on an earlier version of the manuscript.

## REFERENCES

- BOYLE, T. J. B. AND MORGENSTERN, E. K. 1985a. The population structure of black spruce in central New Brunswick. In *Proc. 29th Northeastern Forest Tree Improvement Conf.*, Morgantown, WV, pp. 142-156.
- BOYLE, T. J. B. AND MORGENSTERN, E. K. 1985b. Inheritance and linkage relationships of some isozymes of black spruce in New Brunswick. *Can. J. For. Res.*, 15, 992-996.
- BOYLE, T. J. B. AND MORGENSTERN, E. K. 1986. Estimates of outcrossing rates in six populations of black spruce in northern New Brunswick. *Silvae Genet.*, 35, 102-106.
- BOYLE, T. J. B. AND MORGENSTERN, E. K. 1987. Some aspects of the population structure of black spruce in central New Brunswick. *Silvae Genet.*, 36, 53-60.
- CLARK, P. J. AND EVANS, F. C. 1954. Distance to nearest neighbour as a measure of spatial relationships in populations. *Ecology*, 35, 445-453.
- CLIFF, A. D. AND ORD, J. K. 1981. *Spatial Process Models and Applications*. Pion Ltd., London.
- DONNELLY, K. 1978. Simulation to determine the variance and edge-effect of total nearest neighbour distance. In Hodder, I. R. (ed.) *Simulation Methods in Archaeology*, Cambridge University Press, London, pp. 91-95.
- HAMRICK, J. L., MITTON, J. B. AND LINHART, Y. B. 1979. Relationships between life history characteristics and electrophoretically detectable genetic variation in plants. *Ann. Rev. Ecol. Sys.*, 10, 173-200.
- HEINSELMAN, M. L. 1957. *Silvical Characteristics of Black Spruce*. USDA For. Serv., Lake States For. Expt. Sta. Paper 45.
- KNOWLES, P. 1990. Spatial genetic structure within two natural stands of black spruce (*Picea mariana* [Mill.] B.S.P.). *Silvae Genet.*, (In press).
- LEDIG, F. T. 1974. An analysis of methods for the selection of trees from wild stands. *Forest Sci.*, 20, 2-16.
- LIENGSI, C., PIEWLUANG, C. AND BOYLE, T. J. B. 1990. *Starch Gel Electrophoresis of Tropical Trees—A manual*. ASEAN-Canada Forest Tree Seed Centre Muak Lek, Thailand.
- LINHART, Y. B., MITTON, J. B., STURGEON, K. B. AND DAVIS, M. L. 1981. Genetic variation in space and time in a population of ponderosa pine. *Heredity*, 46, 407-426.
- NEI, M. 1972. Genetic distance between populations. *Amer. Nat.* 106, 282-292.
- NEI, M. 1973. Analysis of gene diversity in subdivided populations. *Proc. Natl Acad. Sci. USA*, 70, 3321-3323.
- NEI, M. AND ROYCHOUDHURY, A. K. 1974. Sampling variances of heterozygosity and genetic distance. *Genetics*, 76, 379-390.
- O'REILLY, G. J., PARKER, W. H. AND CHELIAK, W. M. 1985. Isozyme differentiation of upland and lowland *Picea mariana* stands in northern Ontario. *Silvae Genet.*, 34, 214-220.
- UPTON, G. J. G. AND FINGLETON, B. 1985. *Spatial Data Analysis by Example. Volume I. Point Pattern and Quantitative Data*. John Wiley & Sons, Chichester, U.K.
- WRIGHT, S. 1965. Interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution*, 19, 395-420.