National Library of Canada
Bibliothèque nationale du Canada

Canadian Theses Service
Service des thèses canadiennes

Ottawa, Canada
K1A 0N4

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, tests publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30.

NL-339 (r. 88/04)

Canada

THE UNIVERSITY OF ALBERTA

Mating system dynamics and population substructure in natural stands of black spruce (*Picea mariana* (Mill.) B.S.P.).

by

Albert T. Sproule

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE

OF Doctor of Philosophy

Forest Science

EDMONTON, ALBERTA

Spring 1988

Permission has been granted to the National Library of Canada to microfilm this thesis and to lend or sell copies of the film.

The author (copyright owner) has reserved other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without his/her written permission.

L'autorisation a été accordée à la Bibliothèque nationale du Canada de microfilmer cette thèse et de prêter ou de vendre des exemplaires du film.

L'auteur (titulaire du droit d'auteur) se réserve les autres droits de publication; ni la thèse ni de longs extraits de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation écrite.

ISBN 0-315-42860-0

THE UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR Albert T. Sproule

TITLE OF THESIS Mating system dynamics and population substructure in natural stands of black spruce (*Picea mariana* (Mill.) B.S.P.).

DEGREE FOR WHICH THESIS WAS PRESENTED Doctor of Philosophy

YEAR THIS DEGREE GRANTED Spring 1988

Permission is hereby granted to THE UNIVERSITY OF ALBERTA LIBRARY to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

(SIGNED) A. Sproule

PERMANENT ADDRESS:

10933 - 73 Ave
EDMONTON
ALTA. T6G 0C3

DATED Jan 25th 1988

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research, for acceptance, a thesis entitled Mating system dynamics and population substructure in natural stands of black spruce (*Picea mariana* (Mill.) B.S.P.). submitted by Albert T. Sproule in partial fulfilment of the requirements for the degree of Doctor of Philosophy.

Bruce P. Dancik
.....
Supervisor

Kenneth A. Higginbottom
.....

S. J. Titus
.....

.....

.....

External Examiner

Date JAN. 16th 1986

Abstract

Mating system parameters for an upland and a lowland black spruce population were estimated simultaneously for the 1983, 1982, 1981 seed years, and for bulk seed collections from the 1976-78 seed years. The population multilocus outcrossing estimates increased from 0.62 in the 1983 seed crop, to 0.85 in the 1976-78 bulk collections. There were no significant differences in outcrossing estimates among years or between populations.

Values of Wright's Fixation index showed an excess of homozygotes in the embryo populations. The excess was greatest in the 1983 embryo population and decreased, but remained statistically significant, in the 1982 and 1981 embryo populations. The maternal population was in Hardy-Weinberg equilibrium. It is suggested that homozygotes are removed both by pre-germination selection and also by post-germination selection during the life of the stand.

Three factors suggested that these populations were substructured: (i) Single-locus estimates of outcrossing were consistently lower than multilocus estimates (ii) There was significant spatial heterogeneity of the outcross pollen pool and (iii) Wright's Index of Fixation was higher than the equilibrium inbreeding coefficient F_e.

There was no significant correlation between outcrossing estimates and stand density, so that substructure in these populations is probably the result of clustering of related trees rather than being due to spatial

separation of trees.

The lowland population was also investigated for substructure. Trees were sampled along four intersecting transects, each 120 meters long. Cluster analysis of genotypes, distribution patterns of rare alleles and the runs test for non-random distribution of allelic sequences all indicated the presence of spatial clusters of related trees as well as some larger genealogic clusters. There was no significant correlation between spatial distance and the level of genotypic dissimilarity among trees, probably because of the irregular shape and size of the clusters. Gene frequency differences among clusters at the *Gdh*, *Pgi2*, *Pgm* and *6pgl* loci (used in the mating system study) supported the model of significant familial structure.

Frequency of marker alleles in progeny of sampled trees implied a high rate of selfing and nearest neighbour pollination, and relatively short pollen dispersal distances.

It is suggested that the population is subdivided into groups related trees, possibly resulting from continued nearest neighbour pollination. It is further suggested that subdivision of the population in this way depressed the single-locus estimates of outcrossing, and possibly also the multilocus estimates.

Acknowledgements

I would like to express my gratitude to the many people who helped during the different phases of the project.

Dr. N.K. Dhir from Alberta Forest Service provided excellent background information on black spruce in Alberta which was a great help in defining the study areas. My thanks to Dan Perry and Greg Lee for help with the cone collections and to Mary Aleksiuk for her assistance in the laboratory portion of the study. Much appreciated technical and editorial help and comment came from Dr. B.P. Dancik, Dr. A. Flanagan, Dr. G. Furnier, Dr. K. Higginbotham and Dr. V. Lieffers. In addition to arranging financing through the Canadian Forestry Service block grant to the University of Alberta, Dr. Dancik was a constant help in dealing with the problems that arose from time to time during the study. I would like to thank Dr. K. Morgan, who could not be present at the at the thesis defense, Dr. S. Titus and Dr. A. Yanchuk, for their counsel on statistical matters.

It was indeed a pleasure to have such a well-respected figure as Professor E.K. Morgenstern from the University of New Brunswick as the external examiner on my committee.

Most importantly, I would like to thank my wife Cynthia for enduring what must have seemed an endless disruption of our family life. I wish to dedicate this thesis to her and to my children Kevin, Julia, Melanie and Shannon.

Table of Contents

List of Tables

List of Figures

1. GENERAL INTRODUCTION

Black Spruce (*Picea mariana* (Mill.) B.S.P.) is a characteristic species of the boreal forest, ranging from Alaska to Labrador and south to the Lake States (Hosie 1973). In the southern part of its range it is confined mostly to sphagnum bogs and swamps, where it is generally regarded as a pioneer (Dansereau and Segadas-Vienna 1952). It is often preceded by tamarack (*Larix laricina* (Du Roi) K.Koch) as it invades the sedge mat in filled-lake bogs (Dansereau and Segadas-Vienna 1952). Although classed as a pioneer, stand history shows that in these habitats, it may also, for practical purposes be regarded as an edaphic or physiographic climax (Moss 1953). It is the only tree, other than tamarack, which can survive the cold microclimates, unfavourable soils and high water tables of the bogs (Heinselman 1957; Moss 1953). In the north and west of its range it may be found on drier slopes and hillsides, in pure stands or in mixtures with white spruce (*Picea glauca* Moench), balsam fir (*Abies balsamea* (L.) Mill.), jack pine (*Pinus banksiana* Lamb.), white birch (*Betula papyrifera* Marsh.) and trembling aspen (*Populus tremuloides* Michx.). On these sites it is believed to establish following fire and is succeeded by white spruce and balsam fir in the absence of fire or other disturbance (Moss 1953). In central Alberta black spruce may be found on both types of habitat, mostly occurring in small discrete populations (Moss 1953).

Because of its extensive and diverse natural distribution, black spruce shows considerable phenotypic variation (Khalil 1975) in response to microevolution under widely differing environmental conditions. Lowland populations are typically short, slow growing and generally inferior to the more robust upland populations (Moss 1953). Taxonomically, however, the stunted lowland black spruce is not recognised as a separate variety. Only in the extreme north where its needles are consistently shorter has a form (*Picea mariana* var. *brevifolia* (Peck) Rehd.) been recognised as a taxonomically separate variety (Hosie 1978).

Investigations into the nature of variation between upland and lowland black spruce have been reported by several authors. Based on vegetative morphological characters there is no evidence of edaphic ecotypes (Morgenstern 1969, 1978; Fowler and Mullin 1977; Parker *et al.* 1983). However, in an isozyme study, lowland populations were reported to have significantly higher levels of polymorphic loci and average heterozygosities (O'Reilly *et al.* 1985).

Black spruce is a moderately long-lived species with individuals attaining ages of over 250 years. Flowering begins at about age six (Morgenstern and Fowler 1969), and mature stands bear cones for up to at least 150 years (Heinselman 1957). It is unique in the genus *Picea* in that its seeds are borne in semi-serotinous cones and can be released gradually over a number of years (Heinselman 1957). Cone opening is promoted by heat, suggesting that many

even aged stands of pure black spruce are of fire origin (Vincent 1965). Another unique feature of black spruce is its capacity to reproduce vegetatively. Two methods of vegetative reproduction have been reported. Layering is the rooting of lower pendant branches or sub-branches with each growing into an individual tree, where moss or duff build-up is fast enough to cover the lower branches (Horton and Lees 1961; Stanek 1975). A second type of vegetative reproduction called 'rooting' was described by Horton and Lees (1961). Rooting involves the development of a root sprout, which emerges to form a vegetative stem called a rootling. Both layering and rooting were more prevalent on wetter sites (Horton and Lees 1961; Stanek 1975).

The strength derived from its long fibres make black spruce the most valuable pulpwood species in the boreal forest region, and it is widely utilised and planted from Manitoba eastwards to the Maritime Provinces (Morgenstern 1975). Its importance in the pulp and paper industry has given impetus to breeding programs designed to improve the material used in reforestation. Nursery and field experiments have been established in many regions (Morgenstern 1975; Morgenstern and Kokocinski 1976; Nienstaedt 1984).

Information from these genetic studies indicates that black spruce is highly variable at the provenance level (Park and Fowler 1983). Optimisation of tree improvement strategies, however, requires information not only on the genetic variation present in the species, but also on those

factors which affect the amount and organisation of this variation. One such factor is the degree of outcrossing and selfing which occurs in a species, i.e., the type of mating system (Allard 1975). In turn, the estimation of these mating system components can itself be affected by the division of a population into smaller breeding units, (i.e. population substructure) (Ellstrand and Foster 1983). Apart from studies in central New Brunswick (Boyle and Morgenstern 1984, 1986), there has not been any published report on the mating system or population substructure in natural populations of black spruce. In this study I estimate, simultaneously, the mating system parameters for seed crops of four different ages in two natural black spruce populations. Population A is a lowland black spruce population growing on at least one meter of peat. Population B is an upland population growing on a thin layer of peat overlying mineral soil. In addition, I examine the population substructure in population A.

The formal objectives of this study, described in two sections, were firstly to estimate the mating system parameters, outcrossing and selfing, and secondly to investigate population substructure.

2. MATING SYSTEM

2.1 INTRODUCTION

The mating system is "the pattern of mating in sexually reproducing organisms" (Ayala 1982). It forms the essential link in transmission of genetic material between generations. By controlling the amount of assortative and disassortative mating which takes place, it determines the degree of relatedness among offspring.

Virtually all estimators used in conifer mating system studies are based on the mixed mating model, which assumes each mating event to be either a self-fertilisation or an outcross (Fyfe and Bailey 1951). Recently, however, a one-pollen parent model was formulated (Schoen and Clegg 1986). Both models were used to estimate outcrossing rates in morning glory (*Ipomoea purpurea* Roth) and white spruce (*Picea glauca* (Moench) Voss.). Comparison of the results showed that the mixed mating model was more suitable for estimation of outcrossing rates in anemophilous (wind-pollinated) species, which receive pollen from more than one pollen parent (Schoen and Clegg 1986).

Knowledge of the proportion of selfed seed produced by a species is of practical significance. Selfed progeny often exhibit marked inbreeding depression affecting many aspects of growth and survival (Franklin 1970; Sorensen and Miles 1974, 1982; Ying 1978). Inbreeding violates a basic assumption of most open-pollinated tree breeding programs -

that the parents of wind-pollinated progeny are unrelated. Violation of this assumption causes a bias in estimates of additive genetic variance, heritability and genetic gain (Namkoong 1966; Fujishima and Fredeen 1972; Squillace 1974).

Traditionally, mating system studies were based on morphological markers, such as albino seedlings (Fowler 1965; Morgenstern 1972), percentage of filled seed (Morgenstern 1972; Coles and Fowler 1976) and relative germination rates (Morgenstern 1972). Recently gel electrophoresis has been used almost exclusively to separate polymorphic allozymes as genetic markers in mating system studies (Yeh *et al.* 1983; King *et al.* 1984; Farris and Mitton 1984; Boyle and Morgenstern 1986; Furnier and Adams 1986; Perry and Dancik 1986; Yeh and Morgan 1987). Gymnosperms are particularly suitable for allozyme study, since the megagametophyte is haploid, representing the maternal contribution to the embryo, while the embryo contains both maternal and paternal contributions. Comparison of allozyme variants in the megagametophyte and diploid embryo permits direct determination of the genetic contribution from the male (pollen) parent, assuming that the paternal allele is not a null allele. Direct genotyping of diploid conifer tissue has been used in isozyme studies (Cheliak *et al.* 1985). With diploid tissue the maternal contribution can be directly observed, rather than being inferred, as is the case in the genotyping of haploid tissue. However, in the present situation, results of direct genotyping of diploid needle

tissue were not consistent enough to be used as a basis for the study.

Several statistical procedures have been developed to estimate mating system parameters. These include methods that use information from single loci (Brown *et al.* 1975; Shaw and Allard 1982; Cheliak *et al.* 1983) and those which combine information from multiple loci (Green *et al.* 1980; Ritland and Jain 1981; Shaw *et al.* 1981; Morgan and Yeh 1985; Neale and Adams 1985). All give estimates of the mating system parameters t, the proportion of viable progeny resulting from outcrossing, and s (=1-t), the proportion resulting from self-fertilisation. They also share the assumptions that gene frequency distribution in pollen is identical over all maternal plants and the rate of outcrossing is independent of the maternal genotype (Clegg 1980). Both types of estimator are based on 1) detection of outcrosses by direct observation of the phenotypes of progeny that carry non-maternal alleles and 2) compensation for outcrosses that are not directly observed. As more loci are observed the probability of directly identifying an outcross increases and the importance of compensation in step 2 decreases. The multilocus estimator, therefore, is a more efficient estimator of t, since it recovers information from multilocus data sets that is not recovered in single-locus estimation (Shaw *et al.* 1981). The single-locus estimators are, in general, more sensitive to violations of the assumptions of the mixed mating model (Epperson and Allard 1984). Because of this,

comparison of multilocus and single-locus t estimates can yield information concerning the phenomena that caused the single-locus estimate to depart from its parametric values (Shaw and Allard 1982).

The semi-serotinous cones of black spruce permit simultaneous collection of seed crops from different years and allow one to examine variation in the mating system of viable seeds from different seed crops. Estimation of mating system parameters in current seed crops, i.e. soon after the actual mating event, can provide information on the contribution of the mating system to genetic variation in the population. Comparison of mating system parameters in different seed crops could indicate differential survival of selfed and outcrossed seeds. If survival is random, one would expect mating system parameters to remain relatively constant from year to year, apart from fluctuations due to environmental effects, and assuming no significant temporal variation in selfing events themselves, and other violations of assumptions of the model. In a species which shows periodicity of cone crops, and where older seed may be collected and used in reforestation, information on the nature of older seed crops is extremely useful.

In this section of the study an isozyme analysis was conducted on seed from the 1983, 1982 and 1981 cone crops, plus a bulked seed collection from the 1976-78 cone crops. Results of the isozyme study were examined for variation in pollen pool allele frequencies, which can affect the amount

of gene exchange in the population. A possible cause of spatial heterogeneity in the pollen pool, which has been documented in white spruce (Cheliak 1985) and *Eucalyptus obliqua* L'Herit (Brown *et al.* 1975), is spatial variability in the genetic composition of the population. This can be caused by the Wahlund effect, i.e. clustering of relatives, with each group of related trees comprising a subpopulation (Li 1955), or by wide spacing between individual trees (Cheliak *et al.* 1985). The objective of this portion of the study was to estimate t for seed crops of different ages in two black spruce populations and to examine variation in pollen pool allele frequencies.

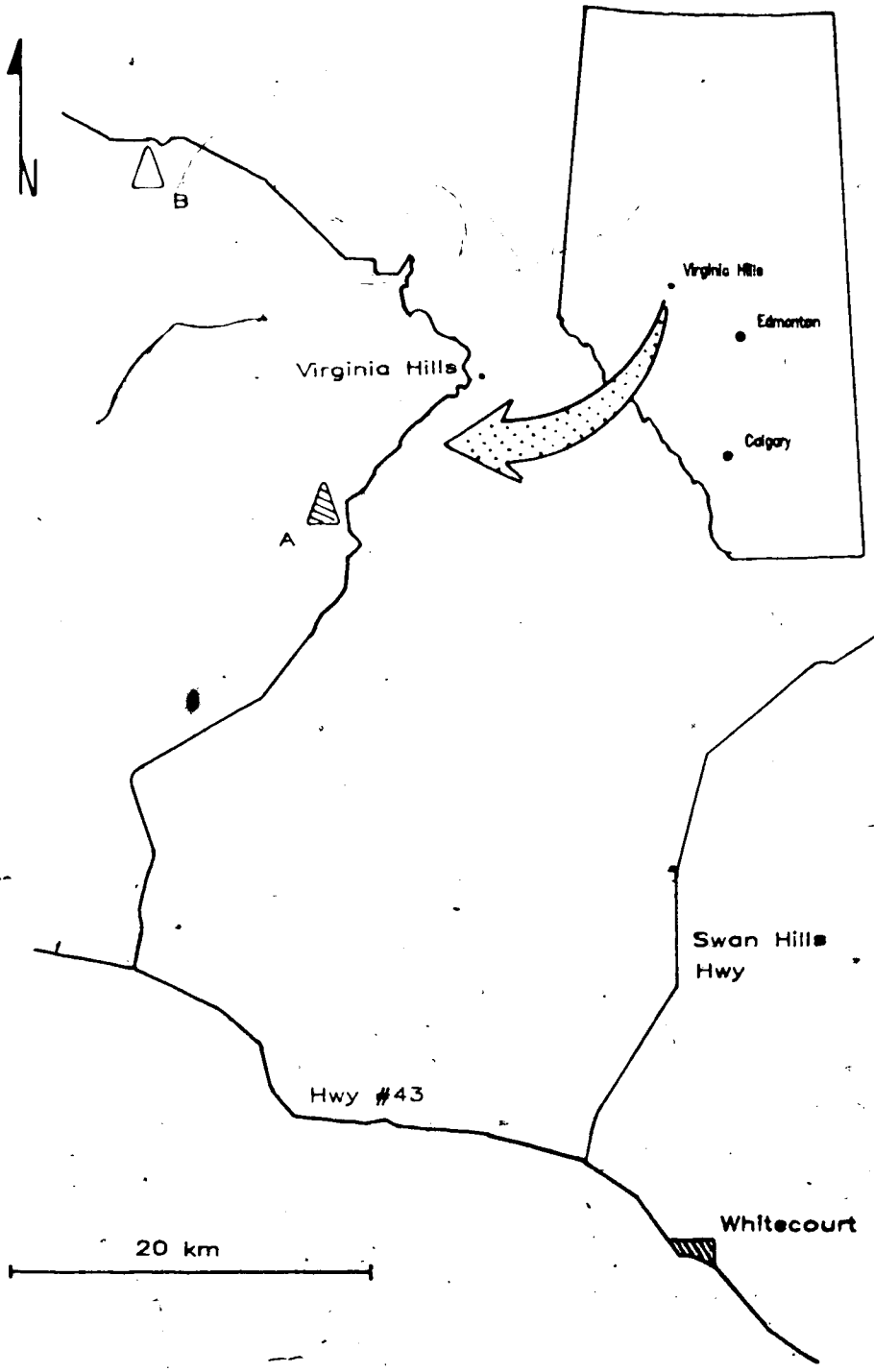
2.2 MATERIALS AND METHODS

2.2.1 Collections and Electrophoretic assay

Seed collections were made in October 1983, from two black spruce populations near Virginia Hills, 50 km northwest of Whitecourt, Alberta (Figure 1). The populations were about 20 km apart. Population A was on a poorly drained peaty site and population B was growing on a well-drained upland site. The average age of both populations was 116 years.

Forty-nine trees were sampled in each population on a 7 by 7 grid, with grid points being 30 metres apart. At each grid intersection the nearest cone-bearing tree was chosen as a sample tree. With a spacing of 30 metres between trees sampling was regularly distributed over a 180 metre square in each population. This should have been a large enough area to avoid clustering of samples in one or a few neighbourhoods, in the event that these populations contained any groups of related trees (Wright 1953). Each sample tree was felled. Sub-samples of cones were taken from the 1983, 1982 and 1981 crops, and a bulked sub-sample from the 1976-78 crops. Black spruce cones are borne on the previous year's branches (Heinselman 1957), so cone age can be identified by internode counts. A further aid to aging is the purple colour of current cones, which has turned to brown by the 3rd year (Vincent 1965). Bulked sub-samples were made from the 1978, 1977 and 1976 crops, since they were not expected to contain many viable seeds. Viable seeds per cone in a peatland stand

Figure 1: Location of black spruce population A (used in the mating system and population substructure study) and B (used in the mating system study).



in Ontario decreased from 7.8 for the current year's cones, to 1.3 in six-year-old cones (Haavisto 1980). While it must be recognized that there is regional variation in seed retention and disposal, in the absence of information on Albertan black spruce it was felt that the figures from Ontario would be a reasonable indication of seed availability in the older crops. Height, DBH' and age (determined by ring count at stump height) of each sample tree were recorded. At each grid intersection a circular fixed-area plot of radius five meters was marked out, with the sample point serving as plot centre. The number of trees in each plot was recorded. This provided an estimate of population density at each sample point. The exact location of each sample tree was mapped, to provide fixed reference points for comparison of results obtained.

All sub-samples were labelled and kept separate. In the laboratory, seeds were extracted by repeated cycles of wetting, drying and heating after the method of Wang (Safford 1974). The dewinged and cleaned seeds were stored at 4°C. until analysed to prevent loss of seed viability (Safford 1948). Seeds were germinated under uniform conditions on filter paper in a growth chamber, with 30°C day (16 hours) and 20°C night temperature (8 hours), until the radicle extended 4 mm beyond the seed coat (approx. 4-5 days). The megagametophyte and embryo tissues were prepared for electrophoresis following the methods of Conkle *et al.*

'Stem diameter, 1.3 m above ground level.

(1982). Ten seeds were randomly chosen from each sub-sample and assayed by horizontal starch gel electrophoresis for the four enzymes which could be successfully stained and scored in both megagametophyte and embryo tissue: glutamate dehydrogenase (GDH, E.C. 1.4.1.3.), phosphoglucose isomerase (PGI, E.C. 5.3.1.9.), phosphoglucomutase (PGM, E.C. 2.7.5.1.) and 6-phosphogluconate dehydrogenase (6PG, E.C. 1.1.1.44.) (Conkle *et al.* 1982). These enzyme systems are coded by four polymorphic loci (*Gdh*, *Pgi2*, *Pgm* and *6pg1*). All four loci have been shown to segregate independently (Boyle and Morgenstern 1985).

2.2.2 Statistical Analysis

Maternal genotypes were inferred from allozyme segregation of the four loci. The number of megagametophytes and embryos per maternal tree ranged from 10, where only one seed crop could be germinated, to 40, when all four seed crops germinated. The total germinants per year for each population are shown in Table 1.

Table 1. Numbers of embryos analysed by population and year, from two black spruce populations near Whitecourt, Alberta.

	Year			
Pop'n	1983	1982	1981	1978-76
A	410	390	420	180
B	360	350	340	40

The probability (p) of misclassifying a heterozygote at a

particular locus, assuming there are no 'nulls' and that segregation is normal, is given by the formula

$$p=(1/2)^{k-1},$$

where k is the number of megagametophytes sampled. With 10 megagametophytes p=0.0019.

Allele frequency distributions among years were compared by a contingency table, 2 x m for *Gdh*, *Pgi2* and *6pg1*, which each had two alleles, and 3 x m for *Pgm*, which had three alleles, where m represented the number of embryo populations, an embryo population being the germinants from a particular seed year.

Single-locus, t_s, and multilocus, t_m, outcrossing rates and the outcross pollen allele frequencies were estimated by the maximum likelihood procedure of Neale and Adams (1985) and the EM algorithm (Yeh and Morgan 1987). Both of these procedures are based on the mixed mating model (Fyfe and Bailey 1951) and assume that there is no selection between fertilization and progeny analysis, that the rate of outcrossing is independent of the maternal genotype and that the gene frequency distribution among pollen pools is identical over maternal plants (Clegg 1980).

A likelihood ratio test (Sokal and Rohlf 1981) was used to test the hypotheses Ho: $t_s=1$ and Ho: $t_m=1$ for the population estimates. The likelihood ratio test is a test for goodness of fit. It computes a ratio between the probability or likelihood of obtaining the observed results, on the hypothesis that the population parameter equals the observed

sample proportion, and the probability of observing the sample results as per the null hypothesis. When these two quantities are equal, the ratio between them is unity. The greater the difference between them, the higher the likelihood ratio will be (Sokal and Rohlf 1981).

Differences among outcrossing estimates by years and populations were tested by Fisher's chi-square homogeneity test (Rao 1973), with three degrees of freedom for the four loci tested.

Outcrossing estimates from the older subsamples were used to construct a graph of t_m vs seed age, but because of their large standard errors they were not used in any other analyses.

The deviation of observed heterozygosity from that expected under Hardy-Weinberg equilibrium in each embryo population and in the adult populations was calculated as

$$F_{is} = 1-H/h,$$

where F_{is} is Wright's fixation index (Wright 1951), i.e. the correlation of genes within individuals, within the population in question. H is the observed proportion of heterozygotes in the population and h is the expected proportion of heterozygotes under panmixia. Expected heterozygosity is calculated as

$$h = 1-\Sigma p_i^2,$$

where p_i is the estimated frequency of the ith allele at the locus.

The tree count obtained in the fixed-area plot at each sampling point was used as an estimate of population density in the neighbourhood of that sample tree. Sample trees were grouped into density classes and a multilocus outcrossing rate, t_m, estimated for each class in each population.

Since the nature of the distributions was not known, a non parametric test was used to compare density classes with multilocus outcrossing rates in each population. The values of each variable were ranked from lowest to highest and the correspondence between ranks of the paired variables was measured by Spearman's coefficient of rank correlation (r_s):

$$r_s = 1 - \frac{6\Sigma (R_1 - R_2)^2}{(n-1)n(n+1)},$$

where $(R_1 - R_2)$ computes the differences between the ranks of the paired variables (Steel and Torrie 1980). The correlation coefficient, r_s, can be tested by

$$t = r_s \sqrt{\frac{n-2}{1-r_s^2}},$$

which is distributed as Student's t with n-2 degrees of freedom.

In order to test local heterogeneity of gene frequencies in the outcross pollen pool, the number of detectable outcrosses (heterozygotes) compared with homozygotes for each homozygous mother was entered in a 2 x m contingency table, where m was the number of maternal trees of one homozygous genotype in the sample (Brown *et al.* 1975). In *Gdh*, *Pgm* and *6pg1*, which had more than one maternal tree homozygous for

more than one genotype, a chi-square value was obtained for each genotype and the population value was obtained by summation. All tests of significance were conducted at the $\alpha=0.05$ level.

2.3 RESULTS AND DISCUSSION

2.3.1 Multilocus outcrossing estimates

Allele frequencies of the four polymorphic loci in the mating system study are presented in Table 2 for the maternal population and the 1983, 1982 and 1981 pollen pools from sampling sites A and B. There was no significant heterogeneity in allele frequency distribution among populations. The contingency chi-square for population A was 9.99 with 10 degrees of freedom. In population B the contingency chi-square was 17.44 with 10 degrees of freedom. There was little difference in the multilocus outcrossing estimators obtained by the two procedures (Neale and Adams 1985; Yeh and Morgan 1987), although Yeh and Morgan's (1987) estimator gave slightly higher estimates in all but one sample (Table 3). To avoid needless complication, and since the two methods gave such similar results, all references to outcrossing estimates in the remainder of the study are to those obtained by the maximum likelihood method of Neale and Adams (1985).

Estimates of the outcrossing rate, t_m, increased significantly from the 1983 seed crop (0.626 in population A, 0.616 in population B) to the bulked sample from the 1976-78 seed crops (0.82 in population A, 0.856 in population B) (Table 3, Figure 2). Year to year estimates of t_m in both populations show a large, though not significant increase from the 1983 to the 1982 seed crop, going from 0.626 to

Table 2: Allele frequencies for four polymorphic enzyme loci in two black spruce populations, near Whitecourt, Alberta.

Pop.	Locus	Allele	Maternal tree frequency	Pollen pool frequency 1983	1982	1981
A			(n=45)[a]	(n=440)[a]	(n=389)[a]	(n=410)[a]
	Gdh	1	0.711	0.700	0.717	0.713
		2	0.289	0.300	0.283	0.287
	Pgi2	1	0.867	0.831	0.837	0.838
		2	0.133	0.169	0.163	0.162
	Pgm	1	0.611	0.601	0.617	0.589
		2	0.200	0.217	0.210	0.254
		3	0.189	0.182	0.174	0.157
	6pgl	1	0.811	0.850	0.875	0.848
		2	0.189	0.150	0.125	0.152
B			(n=37)	(n=340)	(n=350)	(n=360)
	Gdh	1	0.676	0.697	0.667	0.683
		2	0.324	0.303	0.333	0.317
	Pgi2	1	0.811	0.751	0.724	0.790
		2	0.189	0.249	0.276	0.210
	Pgm	1	0.568	0.538	0.517	0.567
		2	0.284	0.319	0.360	0.311
		3	0.149	0.143	0.123	0.122
	6pgl	1	0.676	0.781	0.753	0.769
		2	0.324	0.219	0.247	0.231

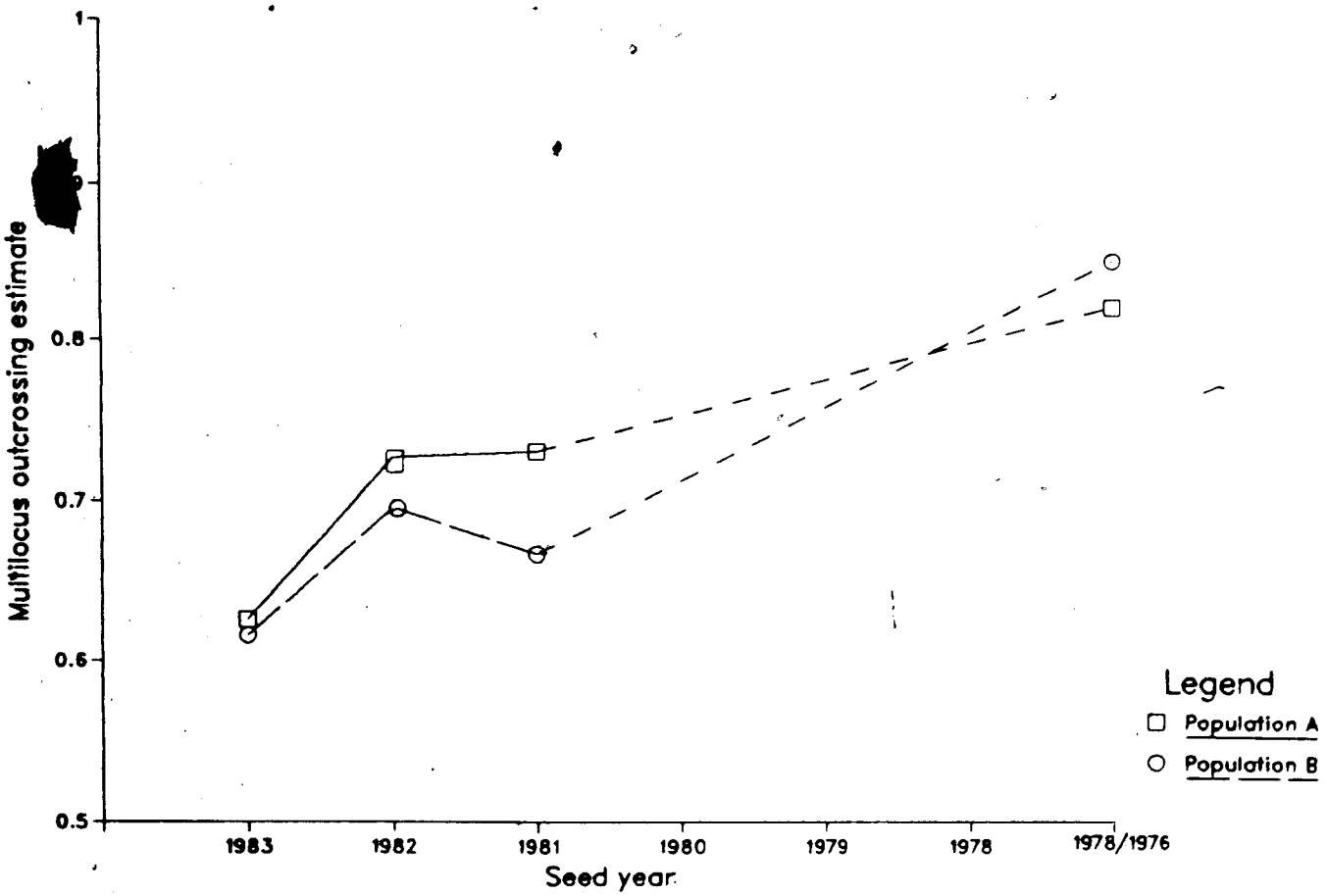
[a] n = number of individuals (trees or embryos)

Table 3: Comparison of maximum likelihood and E-M multilocus outcrossing estimates, t_m, for two black spruce populations near Whitecourt, Alberta, estimated by the E-M algorithm (Yeh and Morgan (1987) and the maximum likelihood method (Neale and Adams 1985).

Population	Year	t_m Estimator	
		Neale and Adams (1985)	Yeh and Morgan (1987)
A	1983	0.626 (0.036)[a]	0.617
	1982	0.727 (0.037)	0.729
	1981	0.731 (0.036)	0.743
	1976-78	0.833 (0.089)	0.876
B	1983	0.616 (0.035)	0.624
	1982	0.696 (0.032)	0.713
	1981	0.667 (0.033)	0.695
	1976-78	0.856 (0.046)	0.850

[a] Standard errors in parentheses.

Figure 2: Multilocus outcrossing estimates by seed-year in black spruce populations A and B near Whitecourt, Alberta..



0.727 in population A and from 0.616 to 0.696 in population B. From 1982 to 1981, however, the t_m of population A only increased from 0.727 to 0.731, while that of population B showed a slight, although not significant, decrease from 0.696 to 0.667. The estimate of t_m then increased in population A, to 0.833 in the bulked seed from 1976-78. The estimate in population B showed a parallel increase, to 0.856 in the bulked seed from 1976-78.

Amongst conifers with serotinous cones, the trend of increasing t_m estimates with age of seed, has also been reported for jack pine (Cheliak *et al.* 1985; Snyder *et al.* 1985), but not in lodgepole pine (*Pinus contorta* var. *latifolia* Dougl.) (Perry and Dancik 1986). While this trend could have been the result of temporal variation variation of the mating system in response to fluctuating environmental factors, it is unlikely that such a similar directional trend could have occurred simultaneously in two separate populations. Although the estimates of t_m were consistently lower in population B than in population A, the differences between the two populations were small and were not significant for any seed year. A more likely hypothesis is that the decrease in proportion of selfs was due to pre-germination selection against selfed seeds, so that a greater proportion of viable seeds from older samples arose from outcrossing events. Pregermination selection would eliminate some of those embryos which were homozygous for recessive lethal and deleterious alleles, a situation which

can arise after inbreeding in natural stands of wind-pollinated species (Sarvas 1962; Sorensen 1982). Some species may have up to 50% self pollination (Muller 1976; Sorensen 1982). However, a single, one-time analysis, as in this project, cannot distinguish between temporal heterogeneity and pregermination selection. Further exploration of the problem would require repeated estimates of mating system parameters on successive seed crops, along with estimates of various mating events over various storage times. Such studies have not yet been reported.

The anomaly for the 1981 seed year is of interest in both populations. Unseasonal weather in 1981 might have interfered with pollen production and/or dispersal during that year. Pollen dispersal in black spruce in the Whitecourt Forest in recent years has taken place in early to mid-June (personal observation since 1982). Pollen production can be affected by irregularities in the sensitive pollen mother cells if temperatures fall below a critical threshold value during meiosis (Eriksson 1968), which in black spruce precedes pollen dispersal by three to four weeks (Winton 1964). Weather records for the region do not show any temperature abnormalities for the period, but precipitation was more or less continuous during June of 1981 and 1983, while it was quite dry in 1982. This could have reduced the amount of pollen reaching the female strobili, thus leading to increased selfing in both 1981 and 1983. An alternative, and perhaps more likely explanation, is that so many selfed

seeds were eliminated by selection during the first year that t_m would then increase at a slower rate. A similar trend was reported for jack pine, with t increasing from 0.73 to 1.0 in one year (Snyder *et al.* 1985).

The estimates of t from this study are comparable to the estimate of t_m= 0.83 for black spruce in a seed orchard in Ontario (Barrett *et al.* 1987). However, they are lower than than the estimate of t_m=0.924, averaged over six black spruce populations in central New Brunswick (Boyle and Morgenstern 1986). Neither of these studies specified the age of the seeds used. The only published reports of temporal trends in estimates of outcrossing for serotinous-coned species are for jack pine (Cheliak *et al.* 1985; Snyder *et al.* 1985). Although the estimates of t_m for the black spruce in this study show a similar trend, they are generally lower than those reported for jack pine. In addition, the likelihood ratio test showed all estimates of t_m for black spruce in this study to be significantly lower than the null hypothesis of t_m=1, the estimate for a population in Hardy-Weinberg equilibrium.

Low estimates of t_m could be due to a number of factors. We would expect black spruce to be reasonably self-compatible because it is a pioneer species (Stebbins 1957; Lande and Schemske 1985). This allows new habitats to be colonised by isolated individuals. Park and Fowler (1983) concluded that black spruce had a relatively high self-fertility (47.2%) and carried a lower genetic load than late successional species.

Outcrossing estimates can also be depressed by spatial variation in the population due to wide spacing between trees or clustering of related trees (i.e., substructuring of the population). Stand density has been shown to be positively related to outcrossing levels in ponderosa pine (*Pinus ponderosa* Laws.) (Farris and Mitton 1984) and has been suggested as a possible explanation for low estimates of t in tamarack (Knowles *et al.* 1987). However, in this study, the result of the comparison of these paired variables was not significant in either population. Spearman's coefficient of rank correlation (r_s) gave values of -0.332 and 0.127 for populations A and B, respectively; i.e., the test did not reveal any relationship between stand density and the multilocus outcrossing estimate, t_m (Table 4).

2.3.2 Single-locus outcrossing estimates

The single-locus estimates of outcrossing ranged from 0.326 (*Pgm*) to 1.03 (*Pgi2*) (Table 5). There was significant heterogeneity amongst loci in all three years in population A. Chi-square values were 35.3 (1981), 19.8 (1982) and 54.6 (1983), all with three degrees of freedom. In population B chi-square values were 14.9 (1981), 7.0 (1982) and 2.0 (1983), all with three degrees of freedom, with only the 1981 population showing significant heterogeneity. Heterogeneity of single-locus estimates was also observed in *Eucalyptus obliqua* (Brown *et al.* 1975), *Eucalyptus delegatensis* (Moran and Brown 1980), Douglas-fir (*Pseudotsuga menziesii* (Mirb.)

Table 4: Values of the paired variables, stand density (tree count from fixed area plots), and multilocus outcrossing rate, t_m, in two black spruce populations near Whitecourt, Alberta.

Population A				Population B			
S.D.[a]		t_m[b]		S.D.[a]		t_m[b]	
7	(1)[c]	0.721	(13)[c]	5	(1)[c]	0.879	(16)[c]
8	(2)	0.416	(2)	6	(2)	0.725	(11)
9	(3)	0.824	(16)	7	(3)	0.492	(4)
10	(4)	0.542	(7)	9	(4)	0.527	(5)
11	(5)	0.557	(8)	10	(5)	0.865	(15)
12	(6)	0.855	(17)	11	(6)	0.538	(6)
15	(7)	0.937	(19)	12	(7)	0.661	(9)
16	(8)	0.618	(9)	13	(8)	0.423	(2)
18	(9)	0.744	(14)	14	(9)	0.636	(8)
19	(10)	0.637	(10)	16	(10)	0.431	(3)
20	(11)	0.684	(11)	17	(11)	0.781	(13)
21	(12)	0.875	(18)	18	(12)	0.221	(1)
22	(13)	0.801	(15)	20	(13)	0.560	(7)
23	(14)	0.430	(4)	21	(14)	0.698	(10)
24	(15)	0.510	(5)	22	(15)	0.796	(14)
26	(16)	0.512	(6)	23	(16)	0.969	(17)
28	(17)	0.418	(3)	26	(17)	0.753	(12)
30	(18)	0.698	(12)				
38	(19)	0.415	(1)				

[a] Stand density, represented by tree counts in fixed area plots.

[b] Multilocus outcrossing rate, t_m.

[c] Rankings of paired variables in parentheses.

Table 5: Single-locus estimates of outcrossing, t_s[a], by year, in two black spruce populations, near Whitecourt, Alberta.

Population	Enzyme	1983	1982	1981
A	*Gdh*	0.712 (0.072)[b]	0.831 (0.074)	0.728 (0.072)
	Pgi2	0.892 (0.077)	0.848 (0.090)	1.030 (0.069)
	Pgm	0.326 (0.042)	0.501 (0.050)	0.543 (0.042)
	6pgl	0.358 (0.075)	0.653 (0.071)	0.613 (0.072)
B	*Gdh*	0.480 (0.061)	0.666 (0.066)	0.638 (0.065)
	Pgi2	0.504 (0.086)	0.441 (0.089)	0.353 (0.086)
	Pgm	0.587 (0.051)	0.590 (0.052)	0.568 (0.051)
	6pgl	0.535 (0.056)	0.488 (0.064)	0.741 (0.057)

[a] Method of Neale and Adams (1985).

[b] Standard errors in parentheses.

Franco) (El-Kassaby *et al.* 1981; Shaw and Allard 1982; Yeh and Morgan 1987) and tamarack (Knowles *et al.* 1987). Variation amongst loci in single-locus outcrossing estimates could be due to spatial changes of genotypic frequencies as would result from clustering of related individuals (Shaw and Allard 1982).

Because t_s is more sensitive to violations of the assumptions of the mixed mating model, single-locus estimates of t are expected to be lower than estimates of t_m (Shaw and Allard 1982). Estimates of t_s in experimental populations of morning glory and sorghum were lower where clusters of related individuals occurred, than in populations without clustering (Ennos and Clegg 1982; Ellstrand and Foster 1983). Shaw and Allard (1982) suggested comparison of t_s and t_m as a possible indicator of family structure. To make this comparison, means for the single-locus estimates for each year were calculated by weighting individual estimates of t_s by the inverse of their standard errors. This adjustment compensated for the difference in precision among estimates made with different marker loci. The results (Table 6) show that with the exception of the 1981 estimate in population A, the mean values of t_s are consistently lower than the t_m estimates. Further evidence of spatial variation in the genetic composition of the population is presented in the results of the chi-square test for spatial heterogeneity in the pollen pool (Table 7).

Table 6: Comparison of multilocus, t_m, and mean of single-locus, t_s, outcrossing estimates for two black spruce populations near Whitecourt, Alberta, method of Neale and Adams (1985).

Population	Year	t_m	t_s [a]
A	1983	0.626 (0.036) [b]	0.565
	1982	0.727 (0.037)	0.713
	1981	0.731 (0.036)	0.733
	1978-76	0.833 (0.089)	0.906
B	1983	0.616 (0.035)	0.552
	1982	0.696 (0.032)	0.598
	1981	0.667 (0.033)	0.615
	1978-76	0.856 (0.046)	0.795

[a] Mean of the single-locus, t_s, outcrossing estimates over four loci (*Gdh Pgi2 Pgm* and *6pg1*) weighted by their standard errors.

[b] Standard errors in parentheses.

Table 7: Chi-square test for spatial heterogeneity in the pollen pools of two black spruce populations near Whitecourt, Alberta.

Enzyme	Pop. A	Pop. B
Gdh	34.05[a] (n=22)[b]	31.58 (n=24)
Pgi2	27.45 (n=31)	32.50 (n=22)
Pgm	54.47[a] (n=19)	74.41[a] (n=18)
6pg1	44.67[a] (n=26)	55.24[a] (n=20)

[a] Significant heterogeneity at the 0.05 level.

[b] n = the number of inferred homozygous trees at that locus.

Pgm and *6pg1*, which had low estimates of t_s in population A, and *Pgm*, which had a low estimate in population B, showed significant spatial heterogeneity in the pollen pool. Spatial changes in the genetic composition of the population are the most likely cause of pollen pool heterogeneity (Cheliak *et al.* 1985). Given that low estimates of t_s can be due to spatial variation of gene frequencies, one could expect that loci with low estimates of t_s might also show significant spatial heterogeneity in the pollen pool. One qualification that should be made, however, is that maintenance of such spatial variation would imply limited gene flow within the populations, i.e., limited dispersal distances of both pollen and seed.

Another reason for the low estimates of both single-locus and multilocus outcrossing rates could be that both estimates were depressed, although t_m less than t_s, by population substructure and consequent heterogeneity of pollen pool allele frequency. Multilocus estimates of outcrossing in a Douglas-fir seed orchard were felt to have been depressed by matings between relatives (Ritland and El Kassaby 1985). Investigation of the relationship between the multilocus outcrossing estimate and the number of loci studied was not one of the objectives of this study. However, this type of information would be useful in mating system studies.

2.3.3 Wright's Index of Fixation

Estimates of Wright's Index of Fixation (Wright 1951) showed a significant deviation from Hardy Weinberg equilibrium at all loci in the embryo populations. This was due to an excess of homozygotes in the embryos of both populations. The homozygote excess, as indicated by the Fixation Index, decreases with increasing age of the seed crop. In the adult populations there is a slight, non-significant excess of homozygotes (Table 8). This suggests that natural selection during the life of the stand may be acting to remove the excess homozygotes. Since inbred seeds are more likely to be homozygous, this could possibly be interpreted as selection against selfs and other inbreds. Poor performance of selfed individuals has been observed in many species, and a significantly lower survival rate has been documented for selfed seedlings of Douglas-fir and ponderosa pine (Sorensen and Miles 1974), white spruce (Ying 1978) and black spruce (Park and Fowler 1982).

An excess of homozygotes can result from (i) inbreeding, (ii) selection against heterozygotes, or (iii) the Wahlund effect - subdivision of the population into groups, each forming a breeding unit by itself, with allelic frequencies differing between groups (Li 1955). If the population is in inbreeding equilibrium, an equilibrium inbreeding coefficient, F_e, can be calculated from the outcrossing estimate t_m (Spiess 1977):

Table 8: Wright's Fixation Indices, F_{IS}, by enzyme and year, for two black spruce populations near Whitecourt, Alberta.

Pop.	Enzyme	Filial Populations			Maternal Population
		1983	1982	1981	
A	*Gdh*	0.380[a]	0.296[a]	0.383[a]	-0.048
	Pgi2	0.084[a]	0.088[a]	0.040[a]	0.082
	Pgm	0.455[a]	0.278[a]	0.241[a]	0.157
	6pg1	0.591[a]	0.345[a]	0.432[a]	-0.045
	Mean	0.377	0.252	0.274	0.0365
B	*Gdh*	0.525[a]	0.312[a]	0.415[a]	0.297
	Pgi2	0.422[a]	0.320[a]	0.346[a]	-0.064
	Pgm	0.285[a]	0.290[a]	0.250[a]	0.238
	6pg1	0.436[a]	0.424[a]	0.252[a]	0.169
	Mean	0.417[a]	0.336	0.315	0.160

[a] Significant (α =0.05) deviation from Hardy-Weinberg.

$$F_e = \frac{1-t_m}{1+t_m}$$

Given the situation that any deficiency of heterozygotes was due solely to the mating system, F_e for any seed year would then be equal to the mean of F_{is}, over the four loci, for that seed year. Calculation of F_e for the most recent seed crop (1983) minimises the effects of selection, but still shows F_{is} to be much larger than F_e in both populations (Table 9). This could be explained by variations in local gene frequency that would bias F_{is} upwards and t_m downwards (i.e., the Wahlund Effect).

Table 9. Comparison of F_{is} (arithmetic mean of the four loci) and F_e for the 1983 seed crop from two black spruce populations near Whitecourt, Alberta.

Year	Pop.	Mean F_{is}	F_e
1983	A	0.377	0.230
	B	0.417	0.238

The fact that t_s is consistently lower than t_m, that loci with the lowest t_s estimates show significant spatial heterogeneity of outcross pollen pool allele frequencies, and that F_{is} is larger than F_e suggests that there may be some substructuring of these populations.

The results of this study show that in natural stands of black spruce the excess of selfed seedlings in the embryo generations is probably reduced by natural selection as the

stand matures. However, the industry practice of collecting the most recent cones as a seed source for nursery production of seedlings may lead to germination of large numbers of selfed seed. To some extent this increase in the number of selfed seedlings may be counterbalanced by the nursery practice of culling the poorer seedlings, many of which are presumably selfed, from the seedbed. The time and effort spent in germinating selfed seedlings (which would later be culled from the seedbed) could be reduced by collecting only one-year-old cones which would mean that some of the selfed seeds had already been eliminated, while little seed viability would have been lost. However, since collection of new cones is the accepted practice, it would be instructive to examine directly the effects of different methods of storage of cones or seeds on the estimate of apparent outcrossing.

2.4 LITERATURE CITED

Allard, R.W. 1975. The mating system and evolution. Genetics 79:115-126.

Ayala, F.J. 1982. Population and evolutionary genetics: A primer. Benjamin/Cummings Publ. Co. Inc., Don Mills, Ont.

Barrett, J., Knowles, P. and W.M. Cheliak. 1987. The mating system in a black spruce clonal seed orchard. Can. J. For. Res. 17:379-382.

Boyle, T.J.B. and E.K. Morgenstern. 1986. Estimates of outcrossing rates in six populations of black spruce in central New Brunswick. Silv. Genet. 35:102-106.

Brown, A. H. D., Matheson, A.C. and K.G. Eldridge. 1975. Estimation of the mating system of *Eucalyptus obliqua* L'Herit. by using allozyme polymorphisms. Aust. J. Bot. 23:931-949.

Cheliak, W.M., Pitel, J.A. and G. Murray. 1985. Population structure and the mating system of white spruce. Can. J. For. Res. 15:301-308.

Cheliak, W., Morgan, K., Strobeck, C., Yeh, F.C.H. and B.P. Dancik. 1983. Estimation of mating system parameters in plant populations using the EM algorithm. Theor. Appl. Genet. 65:157-162.

Cheliak, W.M., Dancik, B.P., Morgan, K., Yeh, F.C. and C. Strobeck. 1985. Temporal variation of the mating system in a natural population of jack pine. Genetics 109:569-584.

Clegg, M.T. 1980. Measuring plant mating systems. Bioscience 30:814-818.

Coles, J.F. and D.P. Fowler. 1976. Inbreeding in neighbouring trees in two white spruce populations. Silv. Genet. 25:29-34.

Conkle, M.T., Hodgskiss, P.D., Nunnally, L.B. and S.C. Hunter. 1982. Starch gel electrophoresis of conifer seeds: a laboratory manual. USDA For. Serv. Gen. Tech. Rep. PSW-64.

Dansereau P. and F. Segadas-Vianna. 1952. Ecological study of the peat bogs of North America I. Structure and evolution of vegetation. Can. J. Bot. 30:490-520.

El-Kassaby, Y.A., Yeh, F.C. and O. Sziklai. 1981. Estimation of the outcrossing rate of Douglas-fir (*Pseudotsuga menziesii* (Mirb.) Franco) using allozyme polymorphisms. Silvae Genet. 30:182-184.

Ellstrand, N.C. and K.W. Foster. 1983. Impact of population structure on the apparent outcrossing rate of grain sorghum (*Sorghum bicolor*). Theor. Appl. Genet. 66:323-327.

Ennos, R.A. and M.T. Clegg. 1982. Effect of population substructuring on estimates of outcrossing rates in plant populations. Heredity 48:283-292.

Epperson, B.K. and R.W. Allard. 1984. Allozyme analysis of the mating system in lodgepole pine populations. J. Hered. 75:212-214.

Eriksson, G. 1968. Temperature response of pollen mother cells in *Larix* and it's importance for pollen formation. Stud. For. Suec. No. 63.

Farris, M.A. and J.B. Mitton. 1984. Population density, outcrossing rate, and heterozygote superiority in ponderosa pine. Evolution 38:1151-1154.

Fowler, D.P. 1965. Effects of inbreeding in red pine (*Pinus resinosa* Ait. III). Factors affecting maternal selfing. Silvae Genet. 14:37-46.

Fowler D.P. and R.E. Mullin. 1977. Upland-lowland ecotypes not well developed in black spruce in northern Ontario. Can. J. For. Res. 7:35-40.

Franklin, E.C. 1970. Survey of mutant forms and inbreeding depression in species of the family Pinaceae. USDA For. Serv. Res. Pap. SE-61 21p.

Fujishima, T. and H.T. Fredeen. 1972. General formula for estimating heritability in a population with related parents. Can. J. Genet. Cytol. 14:549-557.

Furnier, G. and W.T. Adams. 1986. Mating system in natural populations of Jeffrey pine. Amer. J. Bot. 73:1002-1008.

Fyfe, J.L. and N.T.J. Bailey. 1951. Plant breeding studies in leguminous forage crops. I. Natural cross-breeding in winter beans. J. Agric. Sci. 41:371-378.

Green, A.G., Brown, A.H.D. and R.N. Oram. 1980. Determination of outcrossing rate in a breeding population of (*Lupinus alba* L.). Z. Pflanzenzuchtg. 84:181-191.

Haavisto, V.F. 1980. The retention of seed viability in

semi-serotinous black spruce cones. Proc. IUFRO Int. Symp. Forest Tree Seed Storage, PNFI, Ont. Canada. pages 91-95.

Heinselman, M.L. 1957. Silvical characteristics of black spruce. USDA For. Serv. Lake States For. Exp. Stn. Stn Pap. 45.

Horton, K.W. and J.C. Lees. 1961. Black spruce in the foothills of Alberta. Can. For. Res. Br. Tech. Note 110.

Hosie. R.C. 1978. Native trees of Canada. Canadian Forestry Service. Dept. of the Environment. Ottawa. Canada.

Khalil, M.A.K. 1975. Genetic variation in black spruce (*Picea mariana* (Mill.) B.S.P.) in Newfoundland. Silvae Genet. 24:88-96.

King. J.N., Dancik, B.P. and N.K. Dhir. 1984. Genetic structure and the mating system of white spruce (*Picea glauca*) in a seed production area. Can. J. For. Res. 14:639-643.

Knowles, P., Furnier, G.R., Aleksiuk, M.A. and D.J. Perry. 1987. Significant levels of self-fertilisation in natural populations of tamarack. Can. J. Bot. 65:1087-1091.

Lande, R. and W.D. Schemske. 1985. The evolution of self-fertilisation and inbreeding depression in plants. I. Genetic models. Evolution 39:24-40.

Lande, R. and W.D. Schemske. 1985. The evolution of self-fertilisation and inbreeding depression in plants. II. Empirical observations. Evolution 39:41-52.

Li, C.C. 1955. Population genetics Uni. of Chicago Press. Chicago.

Moran, G.F. and A.H.D. Brown. 1980. Temporal heterogeneity of outcrossing rates in Alpine Ash (*Eucalyptus delegatensis* R.T. Bak). Theor. Appl. Genet. 57:101-105.

Morgenstern, E.K. 1969 a. Genetic variation in seedlings of *Picea mariana* (Mill.) B.S.P. I. Correlation with ecological factors. Silvae Genet. 18:151-161.

Morgenstern, E.K. 1969 b. Germination in seedlings of *Picea mariana* (Mill.) B.S.P. II. Variation patterns. Silvae Genet. 18:161-167.

Morgenstern, E.K. 1972. Preliminary estimates of inbreeding in natural populations of black spruce, (*Picea mariana* (Mill.) B.S.P.). Can. J. Genet. Cytol. 14:443-446.

Morgenstern, E.K. 1975. Red and black spruce genetics, Petawa, 1971-72. Proc. fourteenth meetg. Can. Tree Improv. Ass. (1973), Pt. 1, p.133-135.

Morgenstern, E.K. 1978. Range-wide variation of black spruce. Can. J. For. Res. 8:463-473.

Morgenstern, E.K. and D.P. Fowler. 1969. Genetics and breeding of black spruce and red spruce. For. Chron. 45:408-412.

Morgenstern, E.K., Holst, M.J., Teich, A.H. and C.W. Yeatman. 1975. Plus-tree selection: Review and outlook. Dep. of the Envir. CFS. Publ. No. 1347 pp.34-40.

Morgenstern, E.K. and G.H. Kokocinski. 1976. The range-wide cooperative black spruce provenance study. I. Establishment of field experiments in Ontario. CFS and Ont. Min. of Nat. Res. Special Joint Report No. 2.

Moss, E.H. 1953. Forest communities in North-western Alberta. Can. J. Bot. 31:212-252.

Muller, G. 1976. A simple method of estimating rates of self-fertilisation by analysing isozymes in tree seeds. Silvae Genet. 25:15-17.

Muller-Starck, G., Ziehe, M. and H.H. Hattemer. 1983. Reproductive systems in conifer seed orchards. 2. Reproductive selection monitored at an LAP locus in *Pinus sylvestris* L. Theor. Appl. Genet. 65:309-316.

Namkoong, G. 1966. Inbreeding effects on estimation of genetic additive variance. For. Sci. 12:8-13.

Neale, D.B. and W.T. Adams. 1985. The mating system in natural and shelterwood stands of Douglas-fir. Theor. Appl. Genet. 71:201-207.

Nienstaedt, H. 1984. Breeding implications of juvenile selection in a range-wide black spruce provenance test. Can. J. For. Res. 14:933-939.

O'Reilly, G.J., Parker, W.H. and W.M. Cheliak. 1985. Isozyme differentiation of upland and lowland *Picea mariana* stands in Northern Ontario. Silvae Genet. 34:214-221.

Park, Y.S. and D.P. Fowler. 1983. Inbreeding in black spruce (*Picea mariana* (Mill.) B.S.P.): self-fertility, genetic load and performance. Can. J. For. Res. 14:17-21.

Parker, W.H., Knowles, P., Bennett, F., Gray, A. and T. Krickl. 1983. Habitat-dependent morphological and

chemical variation in *Picea mariana* from Northwestern Ontario. Can. J. Bot. 61:1573-1579.

Perry, D.J. and B.P. Dancik 1986. Mating system dynamics of lodgepole pine in Alberta, Canada. Silvae Genet. 35:190-195.

Rao, C.R. 1973. Linear statistical inference and it's applications. John Wiley & Sons Inc., New. York.

Ritland, K. and S. Jain. 1981. A model for the estimation of outcrossing rate and gene frequencies using N independent loci. Heredity 47:35-52.

Safford, J. 1948. *in* Woody plant seed manual. USDA For. Serv. Misc. Publ. 654. p.255-259.

Sarvas, R. 1962. Investigations on the flowering and seed crop of *Pinus silvestris*. Comm. Inst. For. Fenn. 53.4, 198 pp.

Shaw, D.V., Kahler, A.L. and R.W. Allard. 1981. A multilocus estimator of mating system parameters in plant populations. Proc. Ntl. Acad. Sci. USA. 78:1298-1302.

Shaw, D.V. and R.W. Allard. 1982. Estimation of outcrossing rates in Douglas-fir using isozyme markers. Theor. Appl. Genet. 62:113-120.

Snyder, T.P., Stewart, D.A. and A.F. Strickler. 1985. Temporal analysis of breeding structure in jack pine (*Pinus banksiana* Lamb.). Can. J. For. Res. 15:1159-1166.

Sokal, R.R. and F.J. Rohlf. 1981. Biometry, second edition. W.H. Freeman and Company., San Francisco.

Sorensen, F.C. 1982. The role of polyembryony and embryo viability in the genetic system of conifers. Evolution 36:725-733.

Sorensen, F.C. and R.S. Miles. 1974. Self-pollination effects on Douglas-fir and ponderosa pine seedlings. Silvae Genet. 23:135-138.

Sorensen, F.C. and R.S. Miles. 1982. Inbreeding depression in height, height growth, and survival of Douglas-fir, ponderosa pine, and noble fir to 10 years of age. Forest Sci. 28:283-292.

Spiess, E.B. 1977. Genes in populations. John Wiley & Sons, New York.

Squillace, A.E. 1974. Average genetic correlations among offspring from open-pollinated forest trees. Silvae

Genet. 23:149-155.

Stanek, W. 1975. The role of layerings in black spruce forests on peatlands in the clay belt of northern Ontario. Proc. Black Spruce Symp. Gt. Lakes For. Res. Ctre. O-P-4 pp.242-249.

Stebbins, G.L. 1957. Self-fertilisation and population variability in the higher plants. Amer. Natl. 91:337-354.

Steel, R.G.D. and J.H. Torrie. 1980. Principles and procedures of statistics: a biometrical approach. McGraw-Hill Book Co. Toronto.

Vincent, A.B. 1965. Black spruce - a review of it's silvics, ecology and silviculture. Can. Dep. For., Publ. No. 1100.

Winton, L.L. 1964. Meiosis and pollen release in white and black spruce and their hybrid. Minn. For. Notes 154.

Wright, J.W. 1953. Pollen-dispersion studies: some practical applications. J. For. 51:114-118.

Wright, S. 1951. The genetical structure of populations. Ann. Eugenics 15:323-354.

Yeh, F.C., Brune, A., Cheliak, W.M. and D.C. Chipman. 1983. Mating system of *Eucalyptus citriodora* in a seed-production area. Can. J. For. Res. 13:1051-1055.

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

Ying, C.C. 1978. Performance of white spruce (*Picea glauca* (Moench) Voss) progenies after selfing. Silvae Genet. 27:214-215.

3. POPULATION SUBSTRUCTURE

3.1 INTRODUCTION

Mating system estimates may vary from species to species and both spatially and temporally within species (Hamrick 1982). One of the factors which can affect mating system parameters is departure from panmixia, or random outcrossing. This could happen if a population were substructured into groups or clusters of related trees, with random mating only occurring within each cluster. In the boreal forest, groups of related trees may develop from regeneration by serotinous-coned species after a forest fire, as a result of seed falling and germinating around the survivors of the fire (Heinselman 1957). Such clusters could obviously develop over a single generation. Clusters of related trees may also be the result of limited seed dispersal distance in conifers. Although conifer seeds are generally small and light, there is a tendency for seed to fall close to the maternal parent, leading to the development of groups of related trees within natural stands (Shaw and Allard 1982). This type of cluster may require several generations to develop. Development of clusters of related trees may also be aided by the process of layering, (Stanek 1975), and by rooting (Horton and Lees 1961). The progeny resulting from layering or rooting constitute a clone, and therefore have identical genotypes to the parent tree. However, in the natural stand, they represent a group of closely related trees with potential for

interbreeding. By whichever method a group of related trees is established, if pollen dispersal distance is limited, clustering leads to inbreeding, due to mating amongst related trees.

The area over which pollen is homogeneously distributed (i.e., over which mating is truly random) has been described as the neighbourhood (Wright 1946). The size of the neighbourhood is generally felt to be a function of pollen dispersal distance, seed dispersal distance and stand density (Wright 1946; Levin and Kerster 1968). In addition to these factors however one must also recognize that tree phenology can play an important role in determining neighbourhood size. The concept of the neighbourhood and the implication of mating amongst neighbouring trees is extremely important not only in natural regeneration, but also in tree breeding. This is especially true in superior tree selection where one wishes to avoid selecting related trees, and in the design of seed orchards, where the aim is to promote outcrossing.

The belief that conifer seeds generally fall close to the maternal parent is supported by observations of Douglas-fir (Isaac 1930) and engelmann spruce *Picea engelmannii* (Roe 1967; Ronco 1970), which showed few seeds were dispersed further than 60 metres from the maternal parent. However, Bannister (1968), reconstructing the spread of radiata pine (*Pinus radiata*) on South Island, New Zealand, reasoned that seeds had been wind-dispersed in cohorts over several kilometres in open country. Tigerstedt *et al.* (1982)

in a study of Scots pine (*Pinus sylvestris* L.) in Finland suggested that although seeds moved in cohorts, they often came to rest at some distance from the parent tree. A two-year study in Minnesota reported a maximum dispersal distance of 90 metres for black spruce seed in a cut-over area (Anon 1939), and in another Minnesota clearcut, dispersal of black spruce seed at 30 meters had dropped to 6% of dispersal at the edge of the stand (Le Barron 1939). Dispersal of seeds is greatly affected by the height of the maternal-tree and wind speed (McEwen 1971). Ground cover can also be important, especially in much of Canada where it can be snow-covered for much of the year. Up to 69% of black spruce seedfall occurs between October and April (Heinselman 1957; Vincent 1965; Howard 1962; Haavisto 1975). Greater dispersal distances could be expected during the winter months, since seeds are blown along the surface of crusted snow (Heinselman 1957).

Results of studies of pollen dispersal distance are also variable. Koski (1973), studied an areally continuous stand of Scots pine in Finland and concluded that pollen dispersal distance and the neighbourhood could be measured in kilometres. In North America, Wright (1953) reported dispersal distances ranging from a minimum of 90 m for white spruce to over 300 m for poplar (*Populus deltoides* L.) and elm (*Ulmus americana* L.). Norway spruce (*Picea abies* (L.) Karst.) and Douglas-fir showed similar patterns of pollen dispersal with the pollen capture for Norway spruce at 3.5

metres, 6.5 metres and 18.5 metres from the source tree being 20%, 10% and 4%, respectively, of the amount captured at source (Strand 1957). Douglas fir pollen dispersal declined rapidly from 15 to 90 m, but persisted in small amounts beyond 90 m (Silen 1962). More recently, Cheliak (1984) estimated a pollen dispersal distance of approximately eight metres in a Scots pine seed orchard in Ontario. Published research on anemophilous species is rather limited, with most recent work devoted to entomophilous herbs (Levin and Kerster 1968; Schaal 1975; Ennos and Clegg 1983; Brown and Clegg 1984).

Consanguineous matings, which might occur in a substructured population, depressed single-locus estimates of outcrossing, t_s, in experimental populations of and *Sorghum bicolor* (Ennos and Clegg 1982; Ellstrand and Foster 1983). In these studies an inverse relationship between the estimate of outcrossing and the degree of population substructure was also demonstrated. Reports in the literature tend to emphasise the effect of population substructure on single-locus estimates of outcrossing. However, in a multilocus study of a Douglas-fir seed orchard, both single-locus and multilocus (using five loci) estimates of outcrossing were considered to have been depressed by consanguineous matings (Ritland and El Kassaby 1985). Computer simulation of continued nearest-neighbour pollination showed the development and persistence of patches of homozygotes and microgeographic differentiation of the

population (Turner *et al.* 1982). From an initially random distribution of genotypes, population substructure had become evident by the fifth generation. Continued interbreeding between neighbours eventually should be reflected in the genealogy of the population. Meagher and Thompson (1986) developed a model which can be used to determine genealogical relationships in a population by jointly estimating the parent pair for each progeny, based on progeny genotype and gene frequencies of all possible parents. However, this method would only be applicable where all possible parent trees can be genotyped, as in a seed orchard or a small natural stand, and where there is no risk of pollen contamination from other stands.

The indirect evidence from the mating system estimates suggest that the black spruce study populations were substructured, possibly into groups of related trees. This section of the study attempts to determine the existence of groups of putative relatives in population A by comparing allozyme profiles from neighbouring trees for genetic similarities. The assumptions are that trees that are genetically similar are more likely to be related than trees which are genetically dissimilar, and that groupings of related trees would form patterns of genotypic distribution that would not be found under a purely random genotypic distribution.

Various methods have been employed to investigate population structure. Linhart *et al.* (1981) used Wright's F

statistics and Nei's concept of genetic distance to analyse allozyme data and demonstrate significant differences among spatially separated groups of ponderosa pine. The allozyme study was substantiated by comparison of reproductive and morphological features. The spatial separation, however, meant that the groups were already defined, prior to the study.

Multivariate statistics have become a popular method of investigating population structure. Discriminant analysis of allozyme data was used to confirm differences among populations of black spruce in Newfoundland (Yeh *et al.* 1986) and lodgepole pine in British Columbia (Yeh *et al.* 1985). Patterns of morphological and allozyme variation among the four subspecies of *Pinus contorta* Dougl. were examined by cluster analysis (Wheeler and Guries 1982a). The clustering technique used was the Unweighted Pair Group Means Analysis (UPGMA). The variables clustered were Nei's genetic distance between populations, and morphological data from cones and seeds combined into a single variable, for all possible pair-wise combinations of populations. Results confirmed earlier taxonomic treatment of the species. As part of the same study, distribution of rare alleles was summarised by population to provide evidence of common co-ancestry of populations (Wheeler and Guries 1982b). Populations sharing rare alleles were felt to have a recent co-ancestry. Yellow birch (*Betula alleghaniensis* Britton), bog birch (*Betula pumila* L.) and their hybrids were separated by multivariate

analysis of 19 morphological characters (Dancik and Barnes 1975). Five multivariate procedures were employed, with discriminant analysis and principal components analysis being the most helpful in separating the taxa. Their study resembled the present one on black spruce in that the study population did not contain any predefined groups. An investigation of intra-population structure of *Thujopsis dolobrata* (L.F.) Sieb and Zucc. (Sakai and Miyazakai 1972) was based on allelic differences, described as "disagreements". Sakai and Miyazakai also used the same method to investigate inter-population differences. In a New Brunswick study of within-population genetic diversity in black spruce, a coefficient of relationship was calculated as the expected number of allele differences, weighted by allele frequency for various levels of relationship, for each observed multilocus genotype. A regression was then computed for the coefficient of relationship on distance between trees (Boyle and Morgenstern 1984).

The present study utilised a number of different approaches to investigate population substructure. Following the philosophy of Mitton *et al.* (1977) and Smouse *et al.* (1982) that a large number of small differences at allozyme loci can be as useful in separating groups as a small number of large differences, a cluster analysis was performed, using n-1 alleles as the variables, to take account of the total variation over all loci for each tree. The clustering procedure chosen was the Complete Linkage Clustering Method

(CLINK) (Wishart 1978; Romesburg 1984). Its technique of merging clusters by comparing the most dissimilar individuals from each cluster made it particularly appropriate for this analysis, since the aim of the cluster analysis was to identify clusters whose members had a certain minimum resemblance to each other, based on their level of allelic similarity. Cluster analysis was supplemented by an examination of the distribution of rare alleles and by a "runs" test of randomness of the spatial distribution of each allele at each locus. The runs test is designed to uncover non-random sequences of events (Sokal and Rohlf 1981). For example, in the study population non-random sequences of alleles along a transect could be due to the transect passing through groups of related trees. In conjunction with these techniques, a sample-wide pairwise comparison of trees was made to compile levels of allelic difference and spatial distance between trees, which would help to interpret results of the cluster analysis.

In addition to the cluster analysis on the maternal trees, the pattern of pollen dispersal in the study population, a key factor in the establishment and maintenance of population substructure (Levin and Kerster 1968), was examined by analysing the frequency and distribution of selected alleles in the progeny of the maternal trees.

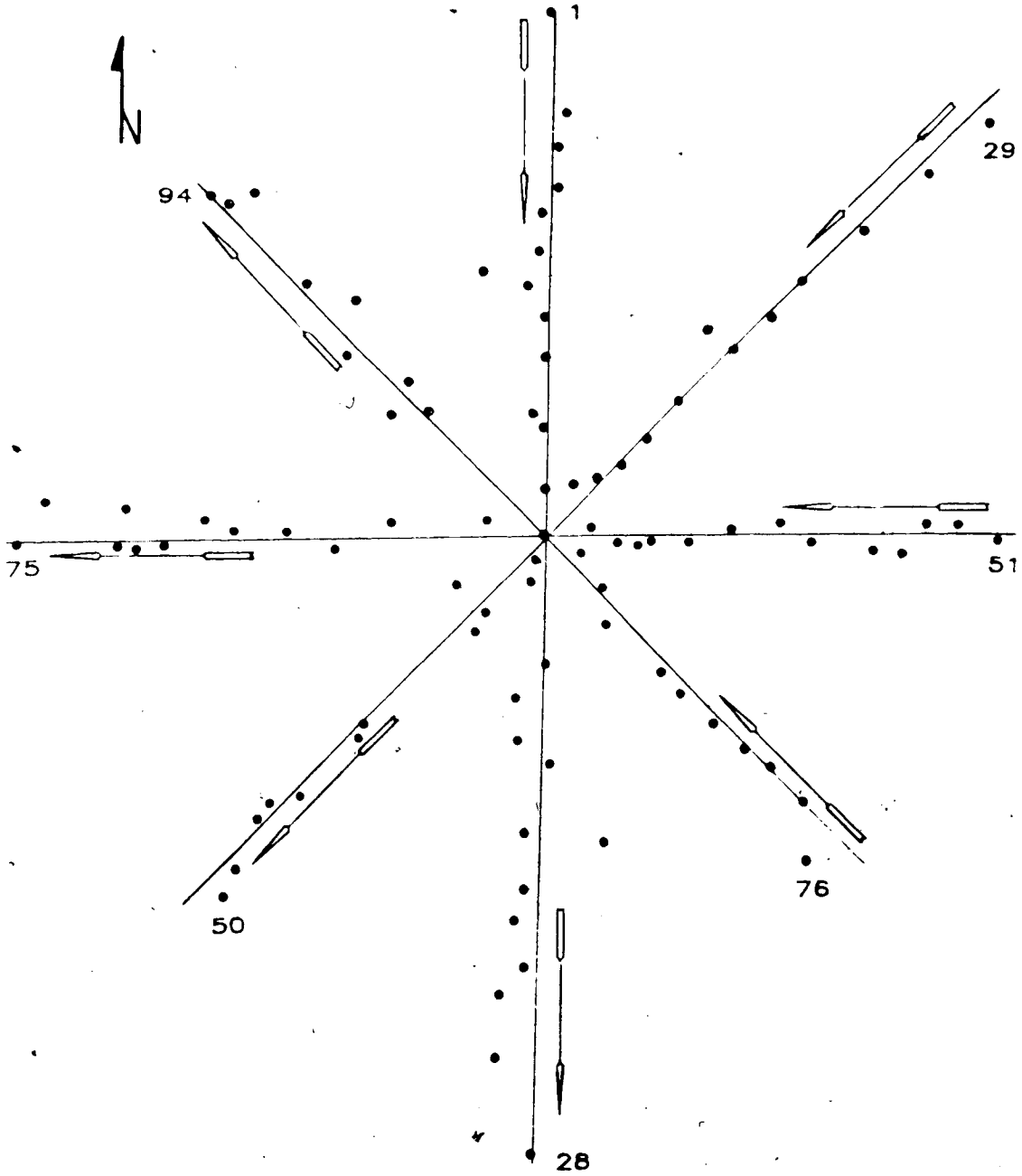
3.2 MATERIALS AND METHODS

3.2.1 Collections and Electrophoretic Assay

Population A (Figure 1) was chosen for this part of the study in the belief that layering, which could contribute to the subdivision of the population into groups of related trees, would be more likely to occur in the peaty soil.

Four intersecting transects were established in the stand: north-south, northeast-southwest, east-west and southeast-northwest, the sampling layout resembling the spokes of a wheel. All trees on or within approximately five metres of the transect (94 trees) were sampled, including ten trees used in the mating system study. The mean spacing between neighbouring trees was just over five meters (pairwise average over all trees). For the subsequent analyses, trees were numbered consecutively from the north end of the north-south transect, through the northeast-southwest, east-west and southeast-northwest transects (Figure 3). If the existence of clusters, and their size and shape had been known in advance, a sampling strategy could have been designed to provide the maximum information on the structure of the population. However, in the present situation where there was no prior information on the nature of population substructure, or indeed, if it existed, the sampling design used had two main advantages. Firstly, it provided total enumeration of the trees around the intersection of the transects, allowing a more complete study

Figure 3: Numbering system followed in sampled trees along four transects in a black spruce population near Whitecourt, Alberta.



of relatedness of neighbouring trees and patterns of pollen dispersal in this area. Secondly, comparison of neighbouring trees along the extremes of the transect would help to corroborate any pattern detected in the fully sampled area around the transect intersections, while minimising the number of samples collected and time required for laboratory analysis.

Each sample tree was felled, and sub-samples of cones were taken from the 1983 and 1982 crops. If 1983 and 1982 cones were absent, a bulked sub-sample of the next available cone year was taken. Height, DBH[2] and age (determined by ring count at stump height) were recorded for each tree, and the position of the tree along the transect was mapped. All sub-samples were labelled and kept separate. Seeds were extracted, stored and germinated as in the mating system study.

Sample trees were genotyped by electrophoretic assay of eight megagametophytes per maternal tree, after the method of Conkle *et al.* (1982). Genotype was based on the segregation of allozymes in the eight enzyme systems (ten loci) which could be scored: aspartate amino transferase (AAT, E.C. 2.0.1.1.), aconitase (ACO, 4.2.1.3.), glutamate dehydrogenase (GDH, 1.4.1.3.), malate dehydrogenase (MDH, 1.1.1.37.), menadione reductase (MNR, 1.6.99.2.), phosphoglucose isomerase (PGI, 5.3.1.9.), phosphoglucomutase (PGM, 2.7.5.1.) and 6-phosphogluconate (6PG, 1.1.1.44.) These enzymes were

---

[2]Stem diameter, 1.3 m above ground level.

coded by ten polymorphic loci (*Aat4*, *Aco*, *Gdh*, *Mdh1*, *Mdh2*, *Mnr1*, *Pgi1*, *Pgi2*, *Pgm* and *6pg1*). All ten loci have been shown to be independently inherited (Barrett *et al.* 1987). With eight megagametophytes the probability (p) of misclassifying a heterozygote at a particular locus is p=0.0078.

In addition, progeny of the maternal trees were assayed for polymorphic allozymes at the *Mdh1* and *Pgm* loci. The numbers of progeny analysed per maternal tree ranged from five for both *Mdh1* and *Pgm* in tree 12 of the east-west transect to 222 (*Mdh1*) and 214 (*Pgm*) in tree 8 in the north-south transect. *Mdh1* was used, because the rare allele *Mdh1-3* was present in the strategically located maternal tree at the intersection of the transects and therefore was useful in estimating pollen dispersal distance. However, to obtain separation and clear resolution of the *Mdh1* alleles, embryos had to be germinated and grown for eight to ten days, by which time *Pgm* was the only other locus that could be scored in the progeny, and enzymes from the megagametophytes could no longer be scored. This meant that maternal tree outcrossing rates could not be estimated for these loci. Maternal tree genotypes, however, were available for these loci, having been determined in the megagametophyte analysis.

3.2.2 Statistical procedures

3.2.2.1 Maternal Trees

Four techniques were used to investigate population substructure:

(i) investigation of genotypic similarity among trees versus physical distance between them;

(ii) cluster analysis of trees based on their level of genotypic similarity, as shown by the results of the isozyme study;

(iii) examination of the distribution of rare alleles;

(iv) the runs test to investigate allelic sequences;

(i) Genotypic similarity versus distance between trees

Since the concept of the degree of relatedness among trees versus the distance between them was central to this part of the study a program was written to address this question.

Each tree was compared with every other tree to determine their genotypic dissimilarity. The pairwise comparison of trees was performed for each allele at each locus. Each comparison showed the number of alleles at which the two trees differed i.e. a dissimilarity index for each possible pairwise comparison of trees. For example a pair of trees with five-locus genotypes 11, 1<u>2</u>, <u>1</u>2, 11, 1<u>2</u> and 11, 1<u>3</u>, <u>2</u>2, 11, 1<u>1</u> would be classed as differing by three alleles (underlined)), a dissimilarity index of three.

Allele-by-allele comparison has the advantage of compensating for the fact that although two individuals may differ at a particular locus, they could still have an allele in common at that locus.

The inter-tree distance for each comparison was calculated from the X and Y coordinates for each tree as the straight line distance between them. For example, tree number 30, which had X and Y coordinates of 52.7 and 48.7 meters, and tree number 78, which had X and Y coordinates of -30.7 and 31.3, meters were computed as 82.23 meters apart. The dissimilarity index for each pairwise comparison was assigned to a distance category, with distance intervals being set at 20 metres. A table was then compiled showing the observed and expected number of dissimilarity classes in each distance category. A table of observed comparisons by itself would be misleading, as the number of comparisons decreased with increasing distance between trees. For example, total comparisons decreased from 1220 in distance class 3 to 10 in distance class 8.

A Chi-square test of the null hypothesis, that the degree of dissimilarity of any two trees was independent of the distance between them, was conducted by entering the observed numbers of each dissimilarity level from each distance category in a cross-classification of distance (8) by dissimilarity (12) for pairwise comparisons, and computing the expected values as a function of the total observations made in that distance category and the total observations

made at that dissimilarity level. Distance classes 6, 7 and 8 were combined, as were dissimilarity classes 0 and 1, and classes 10, 11 and 12 to give a minimum expected value of 3 in any cell. If the expected value in the chi-square equation is too small it leads to large chi-square values which reflect only the smallness of the "expected" value rather than the departure of "observed from expected" (Ostle and Mensing 1975). The relationship between the variables, between-tree-distance and genotypic-dissimilarity, was investigated by the non-parametric Mantel Test of Matrix Correspondence (Mantel 1967, Smouse *et al.* 1986). Standard parametric tests of significance were not appropriate in this situation as the distance measures had unknown distributional properties and dependencies among the pairwise distances in each matrix violated the assumption of normal theory. The problem being that in a sample of size n there are $\frac{n(n-1)}{2}$ comparisons but not $\frac{n(n-1)}{2}$ independent items. Conventional parametric tests of significance have been used in some studies involving distance matrices, but they have been shown to over-estimate the significance of observed statistics (Dietz 1983). The test statistic for the null hypothesis that the measures of X and Y are independent is the sum of cross products:

$$\tilde{z}_{yx} = \sum_{ij} (X_{ij} Y_{ij})$$

where X and Y are the distance matrices, $\sum_{ij}$ indicates summation over all ij pairs other than i=j, and '$\sim$' indicates the observed value. This test criterion is then

compared with the distribution of Z_{yx} values obtained when the corresponding elements of the two matrices are not associated in any way, i.e. a null distribution. The null distribution in this study was obtained by holding the between-tree-distance matrix rigid and randomly permuting the rows and columns of the genotypic-dissimilarity matrix by Monte Carlo randomization. A total of 25 permutations were run. The number of permutations had to be curtailed due to the high computer cost of permuting a matrix of 4,371 observations. For the test statistic large values suggest positive association between the X_{ij}'s and the Y_{ij}'s. The P-value (p) of the observed statistic is therefore the proportion of the n! permutations for which Z_{yx} is greater than or equal to the observed value ($\tilde{Z}_{yx}$). For large values of n, where complete enumeration of the n! permutations is not feasible, random sampling of the n permutations allows estimation of p (Dietz 1983). For example, if $\tilde{Z}_{yx}$, the observed value, is exceeded only 5 times in 500 permutations, then

$$p = \frac{5}{500} = 0.01.$$

The mean, variance and standard deviation were also calculated for the distribution of allelic differences.

The correlation between genetic and geographic distance in a clumped population of whitebark pine (*Pinus albicaulis* Englm.) was studied by Furnier *et al.* (1987). They used

allelic comparison coupled with regression techniques.

(ii) Cluster analysis

The term cluster analysis covers a number of techniques that have the common purpose of separating cases or individuals described by multivariate data into constituent groups or clusters. In this analysis each allele represented a variable, the purpose being to group the trees into clusters based on the similarity of their isozyme profiles, with the underlying assumption that genotypically similar trees would be more likely to be related. Therefore, the required property of each cluster formed was that every member of the cluster would have a minimum specified resemblance to every other member of that cluster. The most appropriate clustering technique was found to be the Complete Linkage Clustering Method, known by the acronym CLINK. CLINK is not the most widely used clustering technique in research analyses (Romesburg 1984). However, its technique of forming clusters based on the largest single dissimilarity coefficient between two individuals, one from each proposed merging cluster (Wishart 1978), meant that it identified clusters whose members had a certain minimum resemblance to each other, based on their allelic dissimilarity. This was particularly suitable for the present study, where the aim was to identify clusters of genotypically similar individuals based on their level of allelic dissimilarity. When CLINK scans potential merging clusters, it computes a dissimilarity

matrix based on the the degree of resemblance between each cluster and every other cluster. The dissimilarity coefficient is actually based on a comparison of individual cases from each cluster, with the two most unlike cases providing the measure of dissimilarity. In this important respect it differs from more widely used procedures such as UPGMA and Ward's minimum variance method, which compute dissimilarity coefficients based on average values of each variable over all cases within a cluster (Romesburg 1984). Its technique of forming each new cluster on the basis of the two most unlike cases from the two combining clusters, allows one to set a maximum level of within-cluster dissimilarity. The Clustan package presents its results as a dendrogram, with the dissimilarity coefficient shown for each new cluster produced by a fusion cycle. Examination of the clusters formed at different levels in the clustering dendrogram allows one to ascertain the within-cluster level of dissimilarity at each stage of cluster formation. Consequently, a cut-off point can be set which will identify clusters whose members differ by no more than a pre-set level of dissimilarity. In this study the cut-off level for cluster acceptance could be set to identify clusters whose members differed by no more than a pre-set number of alleles. An early problem encountered in using the clustering procedure was that the clustering package used the literal value for each variable. This meant that in the present situation, where each allele was a variable, a '3' compared with a '1'

would obviously give a different dissimilarity coefficient than a '3' compared with a '2', or a '2' with a '1'. To avoid this involuntary weighting of variables allozyme profiles were coded, with each allele being assigned the value of 0.5 if present, and 0 if absent, following the method of Smouse and Neel (1977). Since the frequency of the alleles at a locus has to add to 1 there are n-1 independent dimensions to describe information at the locus. A diallelic locus with the alleles A_1 and A_2 would therefore be described by the vector Y = (1, 0.5, 0) for the genotypes (A_1A_1, A_1A_2, and A_2A_2). Similarly, the vector Y = (1 0, 0.5 0.5, 0.5 0, 0 1, 0 0.5, 0 0) represents the six possible genotypes A_1A_1, A_1A_2, A_1A_3, A_2A_2, A_2A_3 and A_3A_3, respectively, at a triallelic locus. The preliminary step in this clustering procedure is computation of a dissimilarity matrix to give a dissimilarity coefficient for each possible pair-wise comparison of cases (trees). The Clustan dissimilarity coefficient is actually a distance measure, the squared Euclidean distance between each of the two clusters being compared, summed over all variables. It is computed from the formula:

$$d^2 = \frac{1}{M} \sum (U_{jp} - U_{jq})^2,$$

where M is the number of variables, U_{jp} is the value of variable j for the cluster P and U_{jq} is the value of variable j for cluster q. As an example, cases 1 and 2 differed only at variables 12 and 13, with case 1 having values of 0.5 and 0.5 and case 2 having values of 1 and 0, respectively. Since

the two cases were identical at all other variables, the dissimilarity coefficient was actually computed over variables 12 and 13 as follows:

$$d^2 = \frac{1}{14}((0.5-1)^2 + (0.5-0)^2)$$

$$d^2 = 0.036$$

After each new cluster is formed by a fusion cycle, the dissimilarity matrix is recomputed and the subsequent scan to form a new cluster is based on the new dissimilarity matrix (Romesburg 1984).

A problem common to many types of cluster analysis, including CLINK, is that of the local optimum versus the global solution (Wishart 1978). Once a cluster is formed it remains intact throughout the analysis. This means that, with some data sets, different starting points can produce different clustering solutions. These solutions are known as local optima, as opposed to a global optimum, which is the perfect solution. As a check on the type of clustering solution reached, option Relocate in the Clustan package allows one to reform the clusters from predetermined or random starting points. With the present data set, procedure Relocate was run (a) using the 10-cluster stage computed by CLINK and (b) using 10 randomly formed clusters as a starting point. Relocate scans the clusters and compares the similarities of each case with all clusters, and if necessary moves cases to different clusters. Each of these analyses produces a local optimum. If the same classification results

from the three analyses, something better than a local optimum has been reached, although it could only be claimed to be a global solution if every possible random starting point had been used (Wishart 1978).

Interpretation of the clustering dendrogram was facilitated by reference to the dissimilarity indices resulting from the pairwise comparison of trees. Coefficients corresponding to different levels of allelic dissimilarity were found by examining the level of allelic difference among the members of the new cluster formed at each hierarchical step. This permitted identification of clusters based on the level of dissimilarity of their members. A decision had to be made on what constituted a meaningful cluster, i.e. what level of dissimilarity between cluster members could be tolerated. It was decided to set the cut-off point at five allelic differences, i.e. clusters whose members differed from each other by no more than five alleles were accepted as meaningful clusters. While the choice of a dissimilarity level is obviously arbitrary, the five-allele level was chosen for two reasons. It identifies clusters whose members show a high level of genotypic similarity, while, by not being too stringent, it also allows for the fact that genetic recombination and the input of different paternal genotypes could result in half-sibs and outcrossed full-sibs that need not have identical, or even very similar genotypes. This level would be adjusted for greater numbers of loci or a higher average number of alleles per locus.

(iii) Rare alleles

The comparison of allozyme profiles placed individuals into groups based on their similarity over all loci. Maternal trees were also studied for the occurrence of rare alleles in neighbouring trees. The presence of a rare allele, defined here as having a frequency of less than 0.05 in the total sample of 94 maternal trees, in neighbouring trees would suggest that these trees might comprise a related group. Five alleles had a frequency of less than 0.05. They were *Aat4-2* (0.048), *Mdh1-3* (0.032), *Mdh2-3* (0.021), *Pgi1-2* (0.016) and *Pgm-3* (0.042). Local occurrence of rare alleles was suggested as a possible indicator of relatedness amongst neighbouring trees in Jeffrey pine (*Pinus jeffreyi* Grev. and Balf.) (Furnier and Adams 1986). At the population level, the occurrence of the same rare alleles in different populations has been suggested as an indicator of recent co-ancestry of lodgepole pine populations (Wheeler and Guries 1982b).

(iv) Runs test of allelic sequences

Based on its frequency in a transect, one can test whether an allele is randomly distributed along the transect. Non-random distribution could be due to (a) the presence of too few groups or (b) the presence of too many groups, i.e. over-dispersion of the allele. For example, if the presence of the allele is signified by P and its absence by A, the sequence P, P, P, A, P, P, P, P, A, A, along a transect shows that the allele was present in seven and absent from three

individuals. Each sequence of like individuals, preceded and/or followed by an unlike individual is called a run. The above sequence therefore contains four runs. We can test whether this sequence represents a random distribution of the allele or whether the number of runs is fewer or greater than would be expected with a random distribution. Non-random distribution of an allele could be due to the clustering of related trees whose genotypes contain that allele. The standard test for random distribution of a sequence of alternatives is called a runs test. It tests whether events occur in a random sequence or whether the probability of a given event is a function of the outcome of a previous event (Sokal and Rohlf 1981). Runs tests are widely used in plant ecology to detect species associations in transect sampling (Knight 1974; Pielou 1977).

Mathematically, if there are "x" objects of one kind, e.g. x A's and y objects of another kind e.g. y B's in a sequence, the number of possible arrangements can be calculated from the binomial coefficient

$$\frac{(x+y)!}{x!y!} = \binom{x+y}{x}$$

To find the number of arrangements of A's and B's that will give a total of K runs, there are two possible situations. If K is even (e.g. K=2m), there are m runs of A's and m runs of B's. The number of arrangements that will give K runs is

$$\binom{x-1}{m-1}\binom{y-1}{m-1}$$

since there are $\binom{x-1}{m-1}$ ways of placing x objects into m cells so that each cell has at least one object, and similarly for the B's (Feller 1968). Since the sequence may start with a run of either A's or B's the number of ways of obtaining m runs of both A's and B's is therefore

$$2\binom{x-1}{m-1}\binom{y-1}{m-1}$$

and the probability of obtaining K runs P(K) is

$$P(K)=\frac{2\binom{x-1}{m-1}\binom{y-1}{m-1}}{\binom{x+y}{x}}$$

If K is odd there will be m+1 runs of A's and m runs of B's which can happen in

$$\binom{x-1}{m}\binom{y-1}{m-1}$$

ways or m runs of A's and m+1 runs of B's which can happen in

$$\binom{x-1}{m-1}\binom{y-1}{m}$$

ways.

The probability of K runs P(K) is then

$$P(K)=\frac{\binom{x-1}{m}\binom{y-1}{m-1}+\binom{x-1}{m-1}\binom{y-1}{m}}{\binom{x+y}{x}}$$

and the probability of obtaining R or fewer runs is

$$\sum_{K=2}^{R} P(K) ,$$

with K=2 since there must be at least two runs (Stevens 1939; Pielou 1969).

Tables are available that list the probabilities associated with different values of a and b (corresponding to x and y) and R, for sample sizes between 2 and 20 (Swed and Eisenhart 1943). For larger sample sizes the probabilities were calculated by hand calculator. A runs test was performed on allelic sequences in each of the four transects, based on presence or absence of each allele at each locus in each transect. The method of calculation therefore, was the same for both di-allelic and tri-allelic loci. The probability of obtaining each particular sequence in each transect was found by entering the tables (Swed and Eisenhart 1943) at the appropriate levels of a, b and R. Sequences with a probability of $p < 0.05$ were regarded as having a non-random distribution, in the sense of having too few groups.

The foregoing techniques were used to identify clusters of putative relatives. Once the clusters were assembled, comparison of their gene frequencies provided some insight on the degree to which substructuring had proceeded in the population. Spatial variation of gene frequencies among clusters is regarded as a possible indicator of substructure in a population (Hartl 1980).

Spatial variation of gene frequencies was tested by calculating frequencies for *Gdh*, *Pgi2*, *Pgm* and *6pg1*, the

enzymes used in the mating system study, for clusters with more than five members, and assessing the significance of gene frequency differences between clusters by a chi-square test of gene frequency heterogeneity (Snedecor and Irwin 1933)

$$X^2 = \frac{2N\sum \sigma_{p_i}^2}{\bar{p}_i},$$

where N is the total number of individuals, $\sigma_{p_i}^2$ is the variance in allele frequency and $\bar{p}_i$ is the weighted mean frequency of the i^{th} allele.

By using cluster gene frequencies, one can compute expected mean heterozygosities over all clusters and, for comparison, an expected heterozygosity for the entire population. These expected heterozygosities can be used to compute Wright's Index of Fixation, F_{st}, which is a measure of the extent to which population substructure has proceeded. It measures the differentiation of gene frequencies among neighbourhoods (Ritland 1985), and can also be described as the reduction in heterozygosity associated with division of the population into subpopulations. It is computed as

$$F_{st} = (H_t - H_s)/H_t ,$$

where H_t is the total heterozygosity in the population and H_s is the sub-population heterozygosity. As sub-population heterozygosity decreases in relation to total population heterozygosity the ratio F_{st} approaches unity. F_{st} was calculated for each cluster which had more than five members.

3.2.2.2 Progeny

A contributing factor to the formation and maintenance of population substructure is limited pollen dispersal distance (Turner 1982). To gain some insight into the patterns of pollen dispersal in this population, allele frequencies of *Mdh1* and *Pgm* for the progeny of each maternal tree were computed from the formula

$$f_a = \frac{2D+H}{2N},$$

where f_a is the frequency of the allele, D is the number of individuals homozygous for that allele, H is the number of individuals heterozygous for that allele and N is the total progeny from that maternal tree. The standard error was computed as

$$S.E. = \sqrt{\frac{pq}{2N}},$$

where p is the frequency of the common allele, q is the frequency of other alleles (q=1-p) and N is the total progeny from that maternal tree. Progeny sets (i.e. from the same maternal tree) with more than 30 members were included in the statistical analysis, and represented a total of 54 maternal trees, with a mean of 87 progeny per maternal tree. Standard errors ranged from a minimum of 0.10 to a maximum of 0.18 for *Mdh1* and from 0.017 to 0.058 for *Pgm*. The distribution of the 54 maternal trees is shown in Table 10.

Table 10. Distribution of maternal trees in black spruce population A whose progeny were statistically analysed for frequency of *Mdh1-2* and *Pgm-2*.

Transect	Enzyme *Mdh1-2*	*Pgm-2*
North-south (n=28)	17	17
Northeast-southwest (n=22)	7	7
East-west (n=25)	14	14
Southeast-northwest (n=19)	19	16

n=number of maternal trees along each transect.

Frequencies of the *Mdh1-2* and *Pgm-2* alleles were calculated for each progeny set. Both *Mdh1-2* (frequency 0.13 in all maternal trees along the transects) and *Pgm-2* (frequency 0.31 in all maternal trees along the transects) were well distributed along the transects. Because the alleles were present in some maternal trees the contribution from outcross pollen could not always be unambiguously ascertained. However, since none of these maternal trees carried the '3' allele, the progeny could be separated into three classes based on maternal genotype, for both the *Mdh1-2* and *Pgm-2* alleles:

1) Progeny of maternal trees with a '11' genotype at that locus (i.e. progeny sets whose maternal trees did not have the '2' allele).

2) Progeny of maternal trees with a '12' genotype at that locus (i.e. progeny sets whose maternal trees were heterozygous for the '2' allele).

3) Progeny of maternal trees with a '22' genotype at that locus (i.e. progeny sets whose maternal trees were homozygous for the '2' allele).

Comparison of the three classes of progeny allowed inferences to be made on the possible origin of *Mdh1-2* and *Pgm-2* alleles present in the progeny. *Mdh1-2* and *Pgm-2* alleles in the progeny of class 1 trees must have come from outcross pollen. Although the source of the alleles in the progeny of classes 2 and 3 could not be unambiguously determined, examination of the relative frequencies of *Mdh1-2* and *Pgm-2* in these progeny classes allowed inferences to be made on the possible origin of the alleles.

A 2 x m contingency table was set up for each class to test the null hypothesis that the '2' allele was randomly distributed throughout the progeny of that maternal class. For each progeny set the number of progeny with the allele was compared with the number which did not have the allele, where m was the number of progeny sets and maternal trees in the particular class. If the null hypothesis was accepted, classes 1 and 2, classes 1 and 3, and classes 2 and 3 were then compared by chi-square test, to test the null hypothesis that distribution of the '2' allele in the progeny was independent of parental genotype.

In addition, the distribution of the rare *Mdh1-3* allele (frequency 0.04 for all maternal trees along the transects) in the progeny of maternal trees in the fully sampled area around the transect intersections was examined. The allele

was present in the maternal tree at the intersection of the transects, but not in the neighbouring maternal trees. The neighbouring trees were placed into distance classes according to their distance from the maternal tree. Heterogeneity of distribution of the allele in the progeny was tested by comparing the number of progeny in each distance class whose genotype contained the allele with the number of progeny which did not have it. The respective numbers were entered in a chi-square 2 * m contingency table, where m was the number of distance classes in the fully sampled area that had a radius of approximately 17 meters. The test was repeated, with the maternal source tree excluded, to test the heterogeneity of distribution of the allele amongst the progeny of maternal trees whose genotype did not include the allele. Correlation of distance-from-source with frequency of *Mdh1-3* in the progeny at each distance class was examined by computing Spearman's Coefficient of Rank Correlation (r_s):

$$r_s = 1 - \frac{6\Sigma (R_1-R_2)^2}{(n-1)n(n+1)},$$

where (R_1-R_2) computes the differences between the ranks of the paired variables (Steel and Torrie 1980). As in the mating system study, r_s, the correlation coefficient, r_s, was tested by

$$t = r_s \sqrt{\frac{n-2}{1-r_s^2}},$$

which is distributed as Student's t with n-2 degrees of

freedom.

3.3 RESULTS AND DISCUSSION

3.3.1 Maternal Trees

The frequency distribution of the level of allelic difference observed in the 4,371 pairwise allelic comparisons was symmetrical about the modal value of 5, and mean of 5.3 (Figure 4 and Table 11). Standard deviation was 1.79. The listing of allelic differences between trees showed three genotypically identical trees, numbers 13, 28 and 73 (at the ten loci examined) in the sample. However, they could not have originated by either branch or root layering from a common parent as the distance between the closest pair, 13 and 73, was 60 meters. In addition, being identical at ten loci is not necessarily an indication of total genotypic identity. The level of allelic difference ranged from zero, among trees 13, 28 and 73, to twelve between trees 17 and 53.

A probability of p=0.48 for the Mantel Test of Matrix Correspondence indicated no significant correlation between genotypic-dissimilarity and between-tree-distance. In other words, one would expect to obtain the observed results 48 times in 100 trials by chance alone. Due to the considerable expense of permuting an array of 4371 observations it was only possible to run 25 permutations of the allelic-dissimilarity matrix for the Mantel test of Matrix Correspondence. However, the results of the test should be reliable as the probability of obtaining the observed results (p=0.48) is well removed from both tails of the distribution.

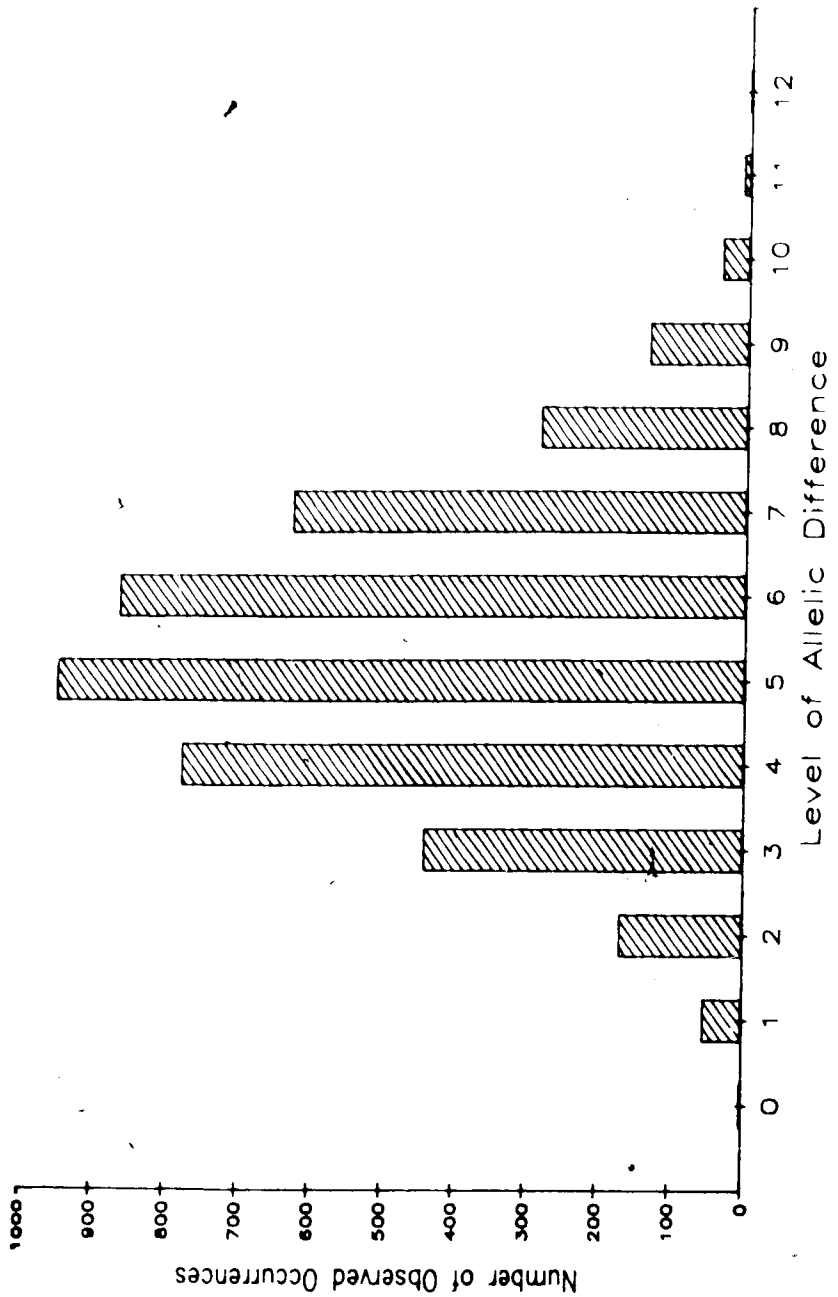


Figure 4: Distribution of allelic differences tabulated in pair-wise comparisons of trees along four transects in a black spruce population near Whitecourt, Alberta.

Table 11 : Comparison of the genotypic dissimilarity between trees versus the distance between them in pairwise comparisons in a black spruce population near Whitecourt, Alberta.

Genotypic Dissimilarity[a]	Average Distance Between Trees (meters)	No. of Comparisons
0	84.43	3
1	56.57	54
2	56.35	170
3	55.28	444
4	54.67	780
5	54.73	953
6	55.53	868
7	53.32	629
8	56.24	286
9	53.03	136
10	60.73	37
11	58.45	10
12	55.69	1

[a] measured as the number of allelic differences between trees in each pairwise comparison.

Results of the Mantel Test of Matrix Correspondence were confirmed by the chi-square analysis of the genotypic-dissimilarity by distance contingency table (Table 12). The chi-square value of 42.01 with 45 degrees of freedom was not significant, so the null hypothesis that genotypic dissimilarity between trees was independent of the distance between them could not be rejected.

Twenty-five groups of putative relatives were identified by the three clustering techniques. The Clustan procedure CLINK assigned a total of 90 trees to 20 clusters, (Table 13 and Figure 5), with members of individual clusters differing at no more than five alleles. The smallest cluster membership was two in clusters N, O, R and T. Cluster K had the largest membership with twelve. Cluster diameter ranged from sixteen meters in cluster N to 119 meters in cluster B.

Five rare alleles were scored in the isozyme analysis, *Aat4-2* (frequency 0.047 in the transect samples), *Mdh1-3* (0.032), *Mdh2-3* (0.021), *Pgi1-2* (0.016) and *Pgm-3* (0.043). Details of the clusters identified by the distribution of these alleles are shown (Table 14 and Figure 6). Of the clusters identified by the distribution of rare alleles, cluster X, based on the distribution of *Mdh1-3*, covered the largest area, with a maximum inter-tree distance of 84 meters. *Mdh1-3* did not show up in any neighbouring trees, the minimum distance between trees in the cluster being 26 meters. Each of the other four rare alleles was found in at least one pair of neighbouring trees, with the minimum

Table 12: Contingency table of observed and expected number of comparisons by dissimilarity and distance class for pairwise comparisons in a black spruce population near Whitecourt, Alberta.

Dist. Class[a]	Genotypic Dissimilarity[b]												
	0	1	2	3	4	5	6	7	8	9	10	11	12
1	0	5 (6)[c]	20 (17)	48 (44)	79 (77)	85 (94)	94 (86)	63 (62)	24 (28)	10 (13)	6 (4)	0	0
2	0	12 (12)	36 (37)	92 (98)	176 (172)	226 (210)	187 (191)	147 (138)	55 (63)	31 (30)	4 (8)	2	0
3	1	15 (15)	37 (48)	120 (124)	224 (218)	255 (267)	238 (238)	183 (176)	91 (80)	43 (38)	8 (10)	4	1
4	0	10 (12)	37 (36)	94 (95)	166 (166)	228 (203)	166 (185)	135 (134)	56 (61)	28 (29)	9 (10)	3	0
5	1	6 (7)	29 (20)	56 (53)	80 (94)	106 (114)	177 (104)	65 (75)	38 (34)	18 (16)	5 (4)	1	0
6	1	3	6 (8)	25 (22)	42 (39)	37 (48)	52 (43)	28 (32)	20 (14)	6 (7)	1	0	0
7	0	3	5	7 (7)	11 (12)	14 (14)	10 (13)	8 (9)	2	0	4	0	0
8	0	0	0	2	2	2	4	0	0	0	0	0	0

[a] class interval = 20 meters.

[b] measured as the number of allelic differences.

[c] expected numbers are in parentheses - not shown if less than 5 in any cell.

Table 13: Clusters of putative relatives from four intersecting transect samples in a black spruce population near Whitecourt, Alberta, identified by the CLINK clustering procedure.

Cluster	Tree #	Cluster Length (m	Ages (years	Mean Age (years)	Heights (meters)	Mean Ht. (meters)
A	1, 2, 4, 15, 21, 32, 51, 62, 72, 80	108	120, 99, 118, 120, 95, 125, 115, 90, 110, 90.	108.7	15, 11, 13, 17, 9, 16, 14, 9, 14, 11.	12.9
B	7, 11, 13, 28, 31, 73.	119	127, 112, 115, 106, 105, 112.	113.2	12, 11, 17, 9, 13, 14.	12.7
C	12, 46, 47.	56	105, 110, 130.	115.0	12, 13, 12.	12.3
D	5, 69, 70, 86.	97	105, 145, 105, 115.	117.5	12, 12, 13, 15.	13.0
E	9, 17, 82.	41	129, 105, 120.	118.0	12, 12, 13.	12.3
F	29, 35, 38, 40, 43, 45.	106	115, 128, 95, 111, 106, 135.	115.0	13, 16, 14, 14, 12, 11.	13.3
G	48, 49, 62, 84, 91.	71	134, 125, 140, 123, 120.	128.4	11, 14, 15, 15, 12.	13.4
H	20, 30, 64.	84	75, 127, 105.	102.3	7, 18, 11.	12.0
I	16, 46, 75, 78, 81, 93.	92	123, 142, 130, 115, 105, 99.	119.0	13, 16, 13, 13, 12, 11.	13.0
J	5, 6, 85.	40	109, 132, 125.	122.0	11, 12, 14.	12.3
K	10, 14, 22, 24, 26, 36, 55, 57, 60, 63, 78, 88.	87	127, 112, 107, 77, 85, 115, 125, 102, 124, 120, 100, 105.	109.2	12, 14, 13, 8, 7, 16, 16, 11, 14, 13, 11, 12.	12.2
L	44, 92, 94.	68	130, 115, 126.	123.7	13, 11, 15.	13.0
M	41, 65, 68.	32	95, 147, 145.	129.0	12, 15, 15.	14.0
N	39, 64.	16	103, 146.	124.5	12, 16.	14.0
O	27, 71.	76	90, 110.	100.0	11, 11.	11.0
P	3, 19, 53, 56, 89.	73	112, 75, 115, 100, 130.	106.4	12, 9, 12, 11, 14.	11.6
Q	18, 33, 83, 87.	56	105, 124, 130, 147.	126.5	11, 14, 15, 14.	13.5
R	23, 59.	45	85, 105.	95.0	6, 13.	9.5
S	8, 37, 50, 58, 66, 76.	86	106, 130, 118, 125, 134, 93.	117.7	11, 15, 12, 14, 15, 12.	13.2
T	74, 77.	101	145, 105.	125.0	13, 13.	13.0

Figure 5: Clusters of putative relatives along four transects in a black spruce population near Whitecourt, Alberta, identified by the CLINK clustering procedure.

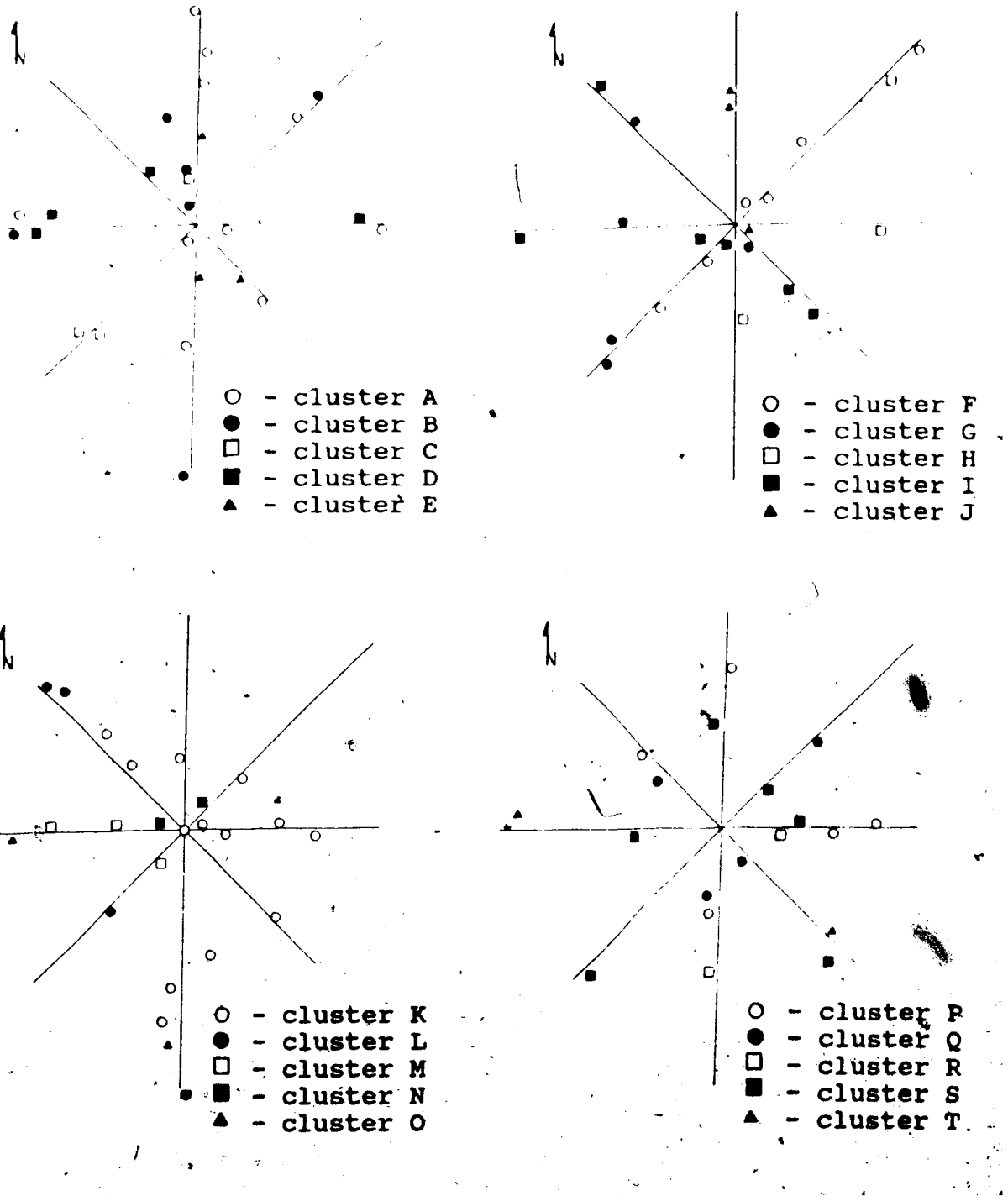
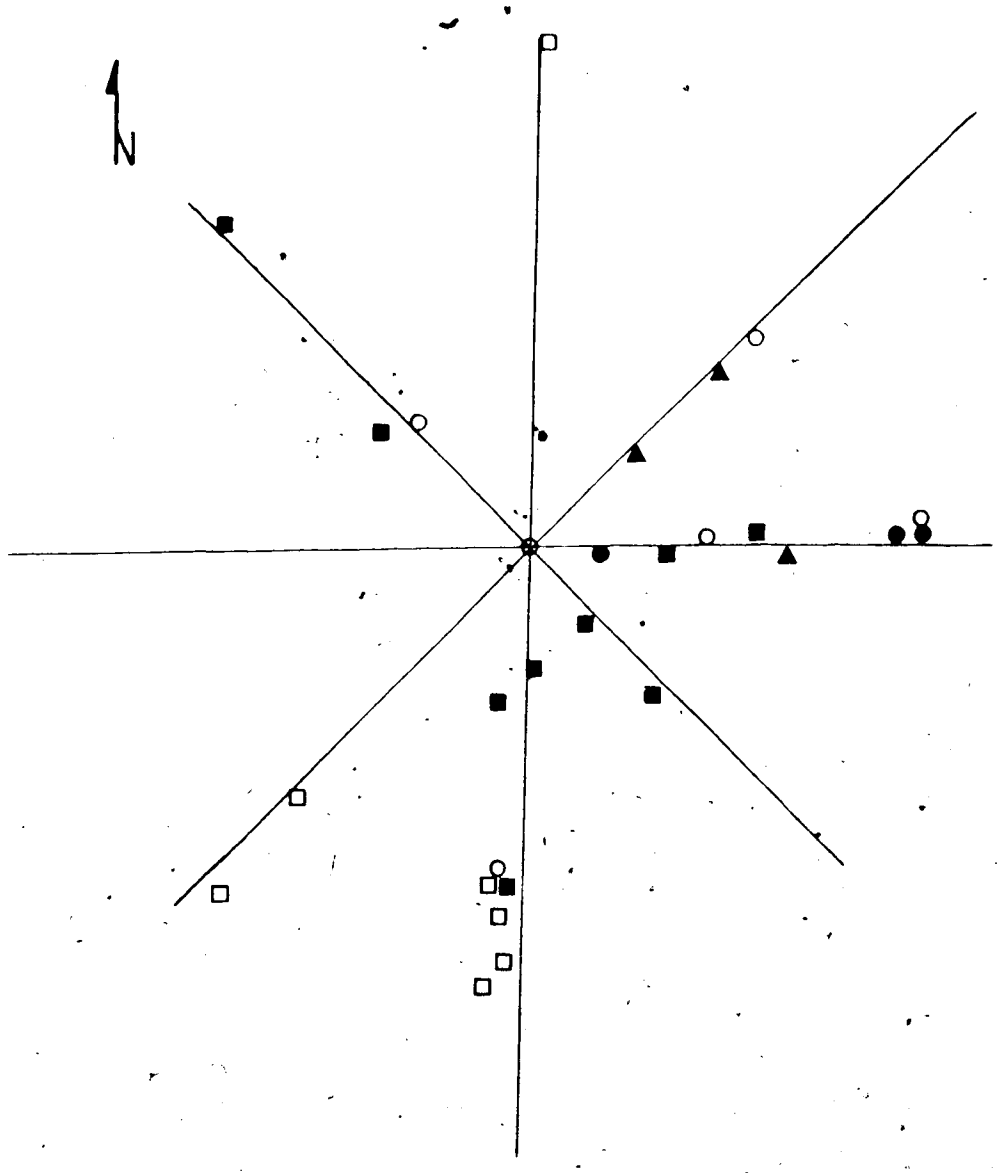


Table 14: Clusters of putative relatives from four intersecting transects in a black spruce population near Whitecourt, Alberta, identified by the distribution of rare alleles

Cluster	Tree #	Cluster Length (meters)	Tree Age (years)	Mean Age (years)	Tree Height (meters)	Mean Ht. (meters)
U (*Aat4-2*) [a]	17, 18, 57, 59, 81, 83	43	105, 105, 102, 105, 105, 130	108.7	12, 11, 11, 13, 12, 15	12.3
V (*Pgm-3*)	1, 23, 24, 25, 26, 45, 49	41	120, 85, 77, 125, 85, 135, 125	107.4	12, 6, 8, 9, 7, 13, 14	9.8
W (*Pgi1-2*)	52, 53, 62	40	105, 115, 90	110.0	12, 12, 9	11.0
X (*Mdh1-3*)	14, 23, 33, 52, 58, 86	84	112, 85, 124, 105, 125, 115	111.0	12, 6, 14, 12, 15, 15	12.3
Y (*Mdh2-3*)	34, 37, 56	29	120, 130, 100	116.7	15, 15, 11	13.7

[a] Rare alleles shown in parentheses.

Figure 6: Clusters of putative relatives along four transects in a black spruce population near Whitecourt, Alberta, identified by the distribution of rare alleles.



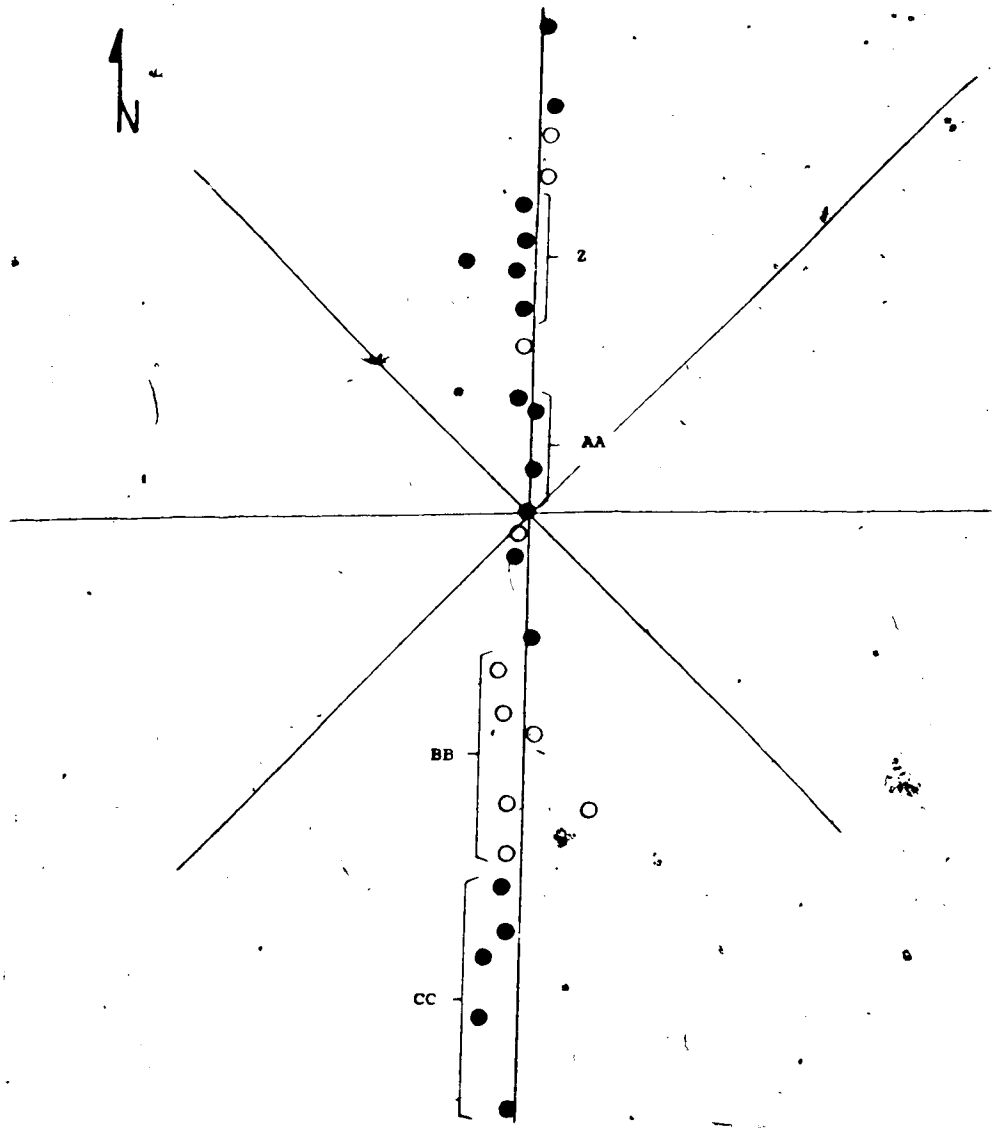
intertree distance being 5 meters between trees 52 and 53 (*Pgi1-2*), 23 and 24, and 25 and 26 (*Pgm-3*). Cluster U (*Aat4-2*), a six-member cluster, had members in three adjoining transects, north-south, east-west and southeast-northwest. Clusters V (*Pgm-3*) and Y (*Mdh2-3*) each had cluster members in two adjoining transects, north-south/northeast-southwest and northeast-southwest/east-west respectively. *Pgi1-2* (cluster W) was present in two neighbouring trees (52 and 53) in the east-west transect, but did not show up again until tree 62, forty one meters away in the same transect. Cluster membership in the clusters identified by the distribution of rare alleles ranged from three in cluster W (*Pgi1-2*) and Y (*Mdh2-3*) to seven in cluster V (*Pgm-3*).

The runs test identified one allele with a non-random distribution along a transect (Table 15 and Figure 7). Based purely on sequences of its presence and absence there was a probability of only p=0.034 of *6pg1-2* having its observed distribution along the north-south transect. The low probability indicated the presence of too few groups, i.e. it tended to be present or absent in sequences of neighbouring trees. There was no indication of over-dispersion of any allele, i.e. no allele had a non-random distribution in the sense of having too many groups. Taking account of the overall distribution of *6pg1-2* along the north-south transect there are four sequences suggestive of clustering. Cluster Z (trees 5 through 9, cluster AA (trees 11-14) and cluster CC

Table 15: Clusters of putative relatives from four intersecting transect samples in a black spruce population near Whitecourt, Alberta, identified by the runs test of allelic sequences. Tests based on each allele at each locus in turn.

Cluster	Allellic Sequence	Tree #	Cluster Length (meters)	Tree age (years)	Mean age (years)	Tree height (meters)	Mean Height (meters)
Z	Presence of *6pg1-2*	5, 6, 7, 8, 9	15	109, 132, 127 106, 129	120.6	11, 12, 12, 11, 12	11.6
AA	Presence of *6pg1-2*	11, 12, 13, 14	15	112, 105, 115, 112	111.0	12, 12, 17, 14	13.7
BB	Absence of *6pg1-2*	18, 19, 20, 21, 22, 23	27	105, 75, 75, 95, 107, 85	90.3	11, 9, 7, 9, 13, 6	9.2
CC	Presence of *6pg1-2*	24, 25, 26, 27, 28	32	77, 125, 85, 90, 108	97.0	8, 9, 16, 11, 9	10.6

Figure 7: Clusters of putative relatives along four transects in a black spruce population near Whitecourt, Alberta, identified by the runs test of allelic sequences.



● - cluster Z (presence of *6pg1-2*)

● - cluster AA (presence of *6pg1-2*)

O - cluster BB (absence of *6pg1-2*)

● - cluster CC (presence of *6pg1-2*)

(trees 24-28) are defined by the presence of *6pg1*. Cluster BB (trees 18 through 23) is defined by the absence of *6pg1*. The clusters defined by the runs test are, in general, shorter than those identified by the distribution of rare alleles or by the CLINK clustering procedure. This is not surprising as the runs test is based on unbroken sequences of the presence or absence of a particular allele - sequences which might not persevere through the entire length of a cluster identified by the distribution of rare alleles or by CLINK.

There was some overlap and inclusion of clusters formed by the different techniques. Cluster BB (Figure 7), identified by a sequence of the *6pg1-2* allele was also included in cluster V (Figure 6), defined by the presence of the rare allele *Pgm-3*. Clusters E and P (Figures 5 (a),(d)), defined as separate clusters by the CLINK technique were included in cluster U, identified by the distribution of rare allele *Aat4-2* (Figure 6).

The CLINK clustering procedure used information from both the rare allele distribution and significant non-random allelic sequences. The inclusion of the CLINK clusters E and P in the rare allele cluster U therefore, is not unexpected, and cannot be taken as an indication that these clusters have any special significance. However, the overlap of cluster CC (runs test) and cluster V (distribution of *Pgm-3*) confirms the reliability of identification of these clusters, since the techniques used to identify the clusters are independent of each other and each technique analysed a different allele.

Clusters occurring in a natural population are unlikely to be discrete entities with regular, well defined boundaries. Since each seed-bearing tree could be the source of a cluster, there is obviously the likelihood of overlap between clusters, and depending on factors such as the speed and direction of prevailing winds, and the transport and storage of cones and seeds by seed-eating animals, cluster size and shape may be expected to vary considerably. Also, some clusters may be the result of a chance phenomenon. Recombination among different genotypes may just happen to produce individuals with some alleles in common.

This is borne out by the results. Compact clusters include M and N (Figure 5(c)), Y (Figure 6) and clusters Z, AA, BB and CC (Figure 7). They have maximum inter-tree distances of 32, 16, 29, 15, 15, 27 and 32 meters respectively. At the other extreme, cluster B (Figure 5(a)) had cluster members ranging from the northeast to the southern part of the sampled area, with a maximum inter-tree distance of 119 meters.

A different cluster type is shown by clusters B (Figure 5 (a)), J (Figure 5 (b)), and U (Figure 6). They each have a clearly defined nucleus of neighbouring trees as well as a number of farther removed 'outlying' members. This type of cluster would probably be the most 'natural', in that one would intuitively expect a cluster of related trees to consist of a number of near-neighbours at the cluster centre, with cluster members becoming more scattered as one moves

away from the centre, since dispersal of pollen and/or seeds from the maternal tree would decrease with distance. However, while the transect sample shows the existence of this type of cluster structure, it would require a complete sample of the areas between the transects to fully demonstrate it.

One would expect clusters in a natural population to vary, both in size and shape. There is the likelihood that larger genealogic clusters are present with members randomly distributed throughout the stand. If the stand is substructured there would also be smaller clusters consisting of spatially clustered, related trees. It is also likely, however, that the sampling method may have contributed to the variation in cluster size, with the transects passing through the centre of some clusters while perhaps just catching the edge of others.

Assuming that many members of a cluster probably originated as half sibs from a common maternal parent, cluster size becomes a function of seed dispersal distance, implying effective dispersal distances of up to 100 meters. This compares with a maximum reported distance of 90 meters for black spruce (Anon 1939), which was, however, recorded over a clear cut. One would not expect longer dispersal distances in a reasonably well stocked stand. However, longer seed dispersal distances could be due to seeds being blown on the surface of crusted winter snow (Heinselman 1957). Another possible explanation is found in the number of squirrel caches encountered in the stand, suggesting that seed

dispersal, at least in the sampled area, may be aided considerably by squirrels.

Apart from varying size and shape, two overall patterns of clustering are apparent (Figures 5, 6 and 7). With the exception of clusters A, B, P and S (Figure 5 (a)(d)), which have members dispersed over most of the sampled area, all clusters are confined to a part of the sampled area. This is well demonstrated by clusters F, G (Figure 5 (b)). Cluster F is confined to the northeast-southwest transect, while cluster G is confined to the western part of the sampled area.

In the maternal trees used in the mating system study it was felt that the patchy occurrence of the *Pgm-3* allele might have been an artifact of the 30 meter spacing between sampled trees. However, apart from tree number 1 (Figure 3) in the north-south transect, its occurrence was limited to the southern and western part of the transect sample. This implies that it has only local distribution in the population. In fact *Pgm-3* was present in only seven of the maternal trees sampled, trees 1 and 23 through 26 in the north-south transect, and trees 45 and 47 in the northeast-southwest transect (Figure 3). Presumably it would also include trees from the unsampled area between the transects. The distribution of *Aat4-2* (Figure 6) showed a similar pattern, with trees from the north-south, southeast-northwest and east-west transects forming a single cluster. The occurrence of these rare alleles in the

genotypes of neighbouring trees suggests that these trees may in fact be related. Possession of rare alleles has been suggested as an indicator of clustering of related trees in a jeffrey pine population (Furnier and Adams 1986). The second pattern which can be observed (Figures 5 and 6) is that of intermingling of clusters. Clusters F and P, and I and S, for example, demonstrate this phenomenon, with cluster F being entirely surrounded by cluster P and cluster I almost enclosed by cluster S (Figures 5 (b),(d)). This type of overlapping and intermingling is not unexpected. Since all sampled trees were seed bearing, each tree was potentially the centre of a cluster, with the entire population probably consisting of a series of overlapping clusters of related trees. An interesting follow-up would be to conduct a similar study in a population with a wider tree spacing. Presumably with wider spacing, there would be less cluster overlap and more clearly defined cluster structure.

The concept of overlapping, intermingled clusters, does not agree with the results of computer simulation studies of continued nearest neighbour pollination (Turner *et al.* 1982). Their computer simulation showed the development, over several generations, of well defined clusters of highly homozygous related trees. Two factors could account for the difference between the computer simulation projections and the situation which exists in the study area. On the one hand, the study population may not have existed long enough to have developed well defined, distinct clusters. The fact

that upland black spruce usually occupies a site for just one generation, combined with its frequent fire origin, was suggested as a possible reason for absence of family structure in a central New Brunswick population (Boyle and Morgenstern 1984). It would seem more likely, however, that the proximity of maternal trees to each other would lead to a considerable degree of intermixing of seeds from different maternal trees. Also there is the possibility that many of the selfed seeds would be susceptible to pre- or post-germination selection, as suggested by the results of the mating system study. Thus the numerical advantage of having many seeds fall close to the maternal tree would tend to be nullified by the fact that many of these would be selfs and would not survive, while outcrossed seeds pollinated by neighbouring or more distant trees would have a better chance of survival. From the point of view of population substructure, the continuing loss of selfs and close relatives would reduce the level of relationship within the clusters. In a study such as this, which is really just a single window on the population, we may therefore be seeing a level of substructure which has in fact decreased steadily during the lifetime of this generation of maternal trees.

Results of the genotypic-dissimilarity versus distance study, at first glance, apparently contradict the results of the cluster analysis, since they show no correlation between genotypic similarity among trees and the spatial distance between them. However, examination of the size of clusters

and of the amount of intermingling among clusters shows that the results of the dissimilarity/distance study actually support the type of population substructure suggested by the three clustering techniques. The intermingling of clusters means that neighbouring trees are sometimes genotypically quite dissimilar. This, combined with the fact that in the larger clusters genotypically similar trees are widely separated, effectively masks any correlation between spatial distance and genotypic similarity. The distance between trees is, therefore, not a reliable indicator of genotypic similarity in the sampled area. Had the sampled area been larger, a relationship between distance and degree of similarity might have been revealed.

The seeming paradox of the results from the two approaches emphasizes the importance of using different methods of investigation. Had the study been based only on the investigation of genotypic similarity and spatial distance between trees one would have had to conclude that the population did not have any substructure. Results of the allelic comparison from this study do not agree with an investigation of *Thujopsis dolobrata* (Sakai and Miyazakai 1972). They found a negative correlation between spatial distance and degree of relationship and suggested that genetically related individuals occurred in circular clusters of 20 to 25 meters diameter. The statistical analysis of their observations is not detailed. *Thujopsis dolobrata* is an oriental species and has not been extensively covered in the

English language. However, those publications that have dealt with it (e.g., Rehder 1967) do not report any biological features unique to the species which might account for this type of clear-cut regular clustering of related trees. Had this type of clustering existed in black spruce population A, it would have been revealed by the pairwise comparisons of genotypic-dissimilarity versus physical-distance for all trees.

Clusters may arise by two methods. Isolated trees that survive a fire, insect infestation, or other natural disturbance can serve as cluster centres (Shaw and Allard 1982; Knowles *et al.* 1986) Due to the isolation of the maternal tree, the related seedlings have a much greater chance of survival. In this way a cluster could develop in a single generation. Alternatively, even in the absence of stand disturbance, continued nearest neighbour pollination, combined with restricted seed dispersal could lead, over several generations, to development of clusters of related trees (Turner *et al.* 1982). Since a disturbance such as fire, by reducing stand density, creates conditions favouring longer seed dispersal distances, one might expect larger clusters to develop following a natural disturbance than would develop as a result of continued natural regeneration in a fully or even moderately stocked stand. In addition, regeneration of groups of trees after a local disturbance would probably result in groups with a more even-aged structure than clusters resulting from continued natural

regeneration, as most regeneration occurs in the years immediately following the disturbance (Lutz 1956; Ahlgren 1959). The expected profile of a cluster developing after a local disturbance would be of a group of relatively even-aged trees covering a relatively large area. The results of this study, however, do not show any large even-aged clusters (Tables 13, 14 and 15). The study area contained one group of neighbouring even-aged trees, 64 through 69 from the east-west transect, 42 from the northeast-southwest transect and tree number 87 from the southeast-northwest transect. The ages of these eight trees ranged from 140 to 147 years. However, the cluster analysis assigned them to seven different clusters. If they did arise following a natural disturbance the seeds may have come from several different maternal trees.

Within-cluster age differences ranged from 15 years in cluster L, to 55 years in cluster P, suggesting that if the trees truly are related, these clusters are the result of continued natural regeneration rather than the aftermath of a local disturbance. In addition, there was no evidence, either in the form of charred tree remains or a charcoal lens in the soil, of fire having burned through the stand. Neither were there any remains to indicate the stand having been subject to insect infestation. The possible development of clusters of related trees following fire or other local disturbance has been suggested by many authors. However, apart from Ahlgren (1959) who documented the proportion of regeneration

occurring in the years immediately following a fire in a black spruce stand, there has not yet been a documented study of regeneration following a natural disturbance, detailing factors such as temporal and spatial patterns of regeneration that help to determine cluster size. With the present widespread use of electrophoresis, such a project would be feasible, and could provide insight into the effects of local disturbance on the development of population substructure. Interestingly, the severity of fires has been suggested as a possible reason for the lack of family structure in some upland black spruce stands in Central New Brunswick (Boyle and Morgenstern 1986).

Heights were recorded for all trees and are shown for the trees in each cluster (Tables 13, 14 and 15). However, within-cluster tree height could not be used to substantiate the family structure of any of these clusters because of the within-cluster age differences.

The development in isolation, of clusters of trees, i.e. the Wahlund Effect, is usually accompanied by two measurable phenomena: increased homozygosity within groups and the development of gene frequency differences between groups (Li 1955). Both processes can be caused by sampling error associated with small population size.

Since virtually all sampled trees were assigned to clusters a comparison could not be made between within-cluster homozygosity and homozygosity of non-clustered trees.

Gene frequency differences among the clusters at the *Gdh*, *Pgi2*, *Pgm* and *6pg1* loci were significantly heterogeneous, with chi-square values of 86.3, 113.7, 114.3 and 103.1, respectively, and ten degrees of freedom (Table 16). This sugggests that there is limited gene flow between different parts of the population. The significant gene frequency differences among clusters could be interpreted as supporting the model of significant familial structure. It could also be that the clusters are a response to natural selection, with certain genotypes responding to environmental factor(s) in different parts of the sampled area. In the absence of an exhaustive ecological survey of the stand, one cannot be certain that there are no microsite differences. However, both the peat cover and ground vegetation were uniform throughout the stand, so that limited dispersion of offspring may indeed be due to restricted pollen and seed dispersal, i.e. evidence of familial structure in the stand.

Given the fact that there is limited gene flow between different parts of the population one would expect alleles at some loci to be proceeding towards fixation. This progress towards fixation can be measured by Wright's Index of Fixation, F_{st}. The values of Wright's Index of Fixation were 0.26, 0.08, 0.13 and 0.23, for *Gdh*, *Pgi2*, *Pgm* and *6pg1*, respectively (the loci analysed in the mating system study), with a mean of 0.175. Comparable F_{st} values are not available for other tree species, but a value of F_{st}=0.069 (mean of 15 loci) was reported from a study of *Liatris cylindraceae*

Table 16: Chi-square test of spatial heterogeneity in gene frequencies of a black spruce population near Whitecourt, Alberta.

Cluster	*Gdh*[a]		*Pgi2*[a]		*Pgm*[a]			*6pgd1*[a]	
	1	2	1	2	1	2	3	1	2
A(n=10)[b]	0.60	0.40	0.95	0.05	0.95	0.00	0.05	0.85	0.15
B(n= 6)	1.00	0.00	0.92	0.08	0.50	0.50	0.00	0.50	0.50
F(n= 6)	0.00	1.00	0.67	0.33	0.50	0.42	0.08	1.00	0.00
I(n= 6)	0.42	0.58	0.42	0.58	1.00	0.00	0.00	0.67	0.33
K(n=12)	0.83	0.17	0.92	0.08	0.50	0.42	0.08	0.71	0.29
S(n= 6)	0.58	0.42	0.67	0.33	0.67	0.33	0.00	0.67	0.33
U(n= 6)	0.58	0.42	0.83	0.17	0.75	0.25	0.00	0.75	0.25
V(n= 7)	0.57	0.43	0.86	0.14	0.43	0.14	0.43	0.70	0.30
X(n= 6)	0.83	0.17	1.00	0.00	0.67	0.25	0.08	0.75	0.25
BB(n= 6)	0.83	0.17	0.67	0.33	0.67	0.25	0.08	1.00	0.00

[a] Significant (= 0.05) heterogeneity.

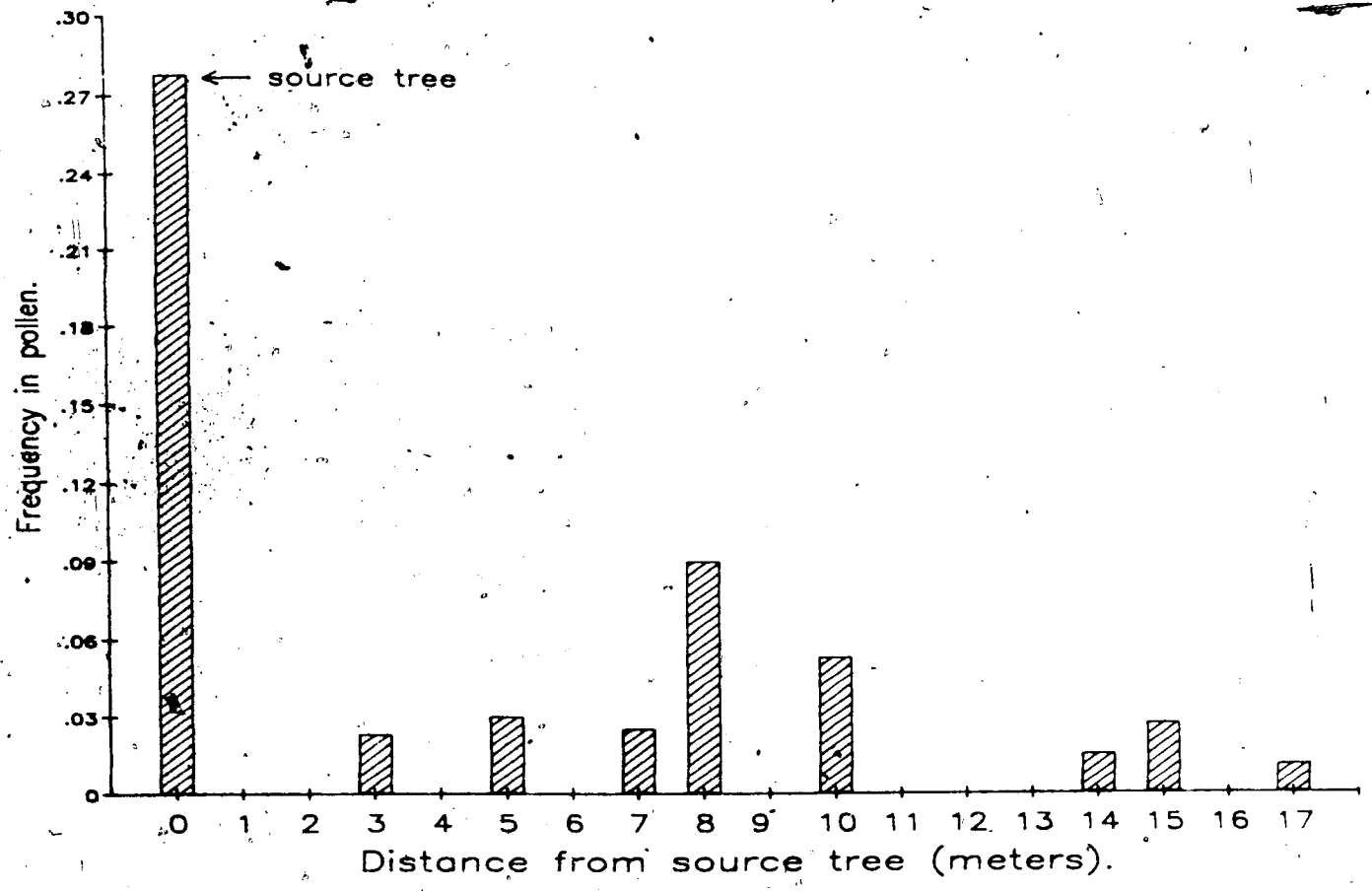
[b] n = number of genotyped trees in each cluster.

(Schaal 1975), and a value of 0.01 (for the *Esterase-5* locus) from a study of *Mus musculis* (Selander *et al.* 1969). Both authors felt that these values were too low to indicate differentiation among sub-populations relative to the total population. In more general comments on F_{st}, Hartl (1980) felt that values of 0.05 to 0.15 could indicate moderate differentiation among sub-populations, while Wright (1978) stated that "differentiation is by no means negligible if F_{st} is as small as 0.05 or even less". The mean of 0.175 for the four loci in the present study suggests a moderate to strong differentiation among clusters.

3.3.2 Progeny analysis

The frequency of the *Mdh1-3* allele in the progeny of trees surrounding the transect intersections declined rapidly with distance from the maternal source to approximately 10% of the source value at three meters. It then maintained a fluctuating, but low level, to the sampled distance of 17 meters from the source tree (Figure 8). There was significant heterogeneity of distribution of the *Mdh1-3* allele in the progeny of all trees around the transect intersections with $X^2=108.7$ with eight degrees of freedom. However, when the central source tree was excluded, chi-square was no longer significant with a value of 11.54 and seven degrees of freedom. The large chi-square value was obviously due to the comparison of zero distance versus all other classes. Spearman's Test of Rank Correlation gave a value of $r_s=0.4$,

Figure 8: Frequency of the *Mdh1-3* pollen allele in the progeny of maternal trees around transect intersections in a black spruce population near Whitecourt, Alberta.



showing that there was no significant correlation between distance from source and frequency of the *Mdh1 3* allele in the progeny of maternal trees around the transect intersections.

Since the *Mdh1 3* pollen was not present in the maternal trees around the transect intersections its presence in the progeny of these trees must be due to outcross pollen. Some of this outcross pollen undoubtedly came from nearby, unsampled, maternal trees. The results however, bear out what is apparent, in Figure 8, namely that pollen frequency declines rapidly with distance from the source tree, but persists at a low, fluctuating level, even with increasing distance from the source tree. These results are comparable to those obtained for Norway spruce, which showed a decrease of over 90% in pollen count at 6.5 meters from the pollen source (Strand 1957). They also concur with a pollen dispersal distance of eight meters estimated from an isozyme study in a Scots pine seed orchard (Cheliak 1984). However, Koski (1973) reported a pollen dispersal distance of several kilometers, calculated for an areally continuous stand of Scots pine. A problem with the interpretation of pollen dispersal studies is the question of total versus effective dispersal distance. Effective pollen fertilises the female ovule, leading to the development of an embryo. Most reports of pollen dispersal distance are based on total pollen count at different distances from the pollen source, which may not be a true reflection of the effective pollen count. With one

exception (Cheliak 1984), the above studies referred to total pollen count. Although one can compare trends, the comparison would be more valid if the proportion of effective pollen, especially in long distance dispersal, was known.

Results of the analysis of the frequency of the *Pgm-2* and *Mdh1-2* alleles in the progeny of the maternal trees were consistent for all transects. The profiles presented for the southeast-northwest transect (Figures 9 and 10) are typical of the other three transects.

There was no significant heterogeneity in distribution of *Mdh1-2* within the maternal genotypic classes. Chi-square values were 12.39 with 12 degrees of freedom, and 0.275 with two degrees of freedom, for progeny of the '11' and '12' maternal genotypic classes. The *Mdh1-2* allele, however, was present in significantly higher frequency in the progeny of the '12' maternal genotypic class than in progeny of the '11' class, with a chi-square value of 123.3 and one degree of freedom.

Results for *Pgm-2* showed a similar pattern, although *Pgm-2* had an extra maternal genotypic class - the '22' maternal genotype. Within-class heterogeneity chi-square values were 1.94 with four degrees of freedom, 15.26 with eight degrees of freedom and 1.28 with one degree of freedom for maternal genotypic classes '11', '12' and '22' respectively. None of these results was significant. Again, there were significant differences in frequency of the allele among progeny of the different maternal genotypic classes.

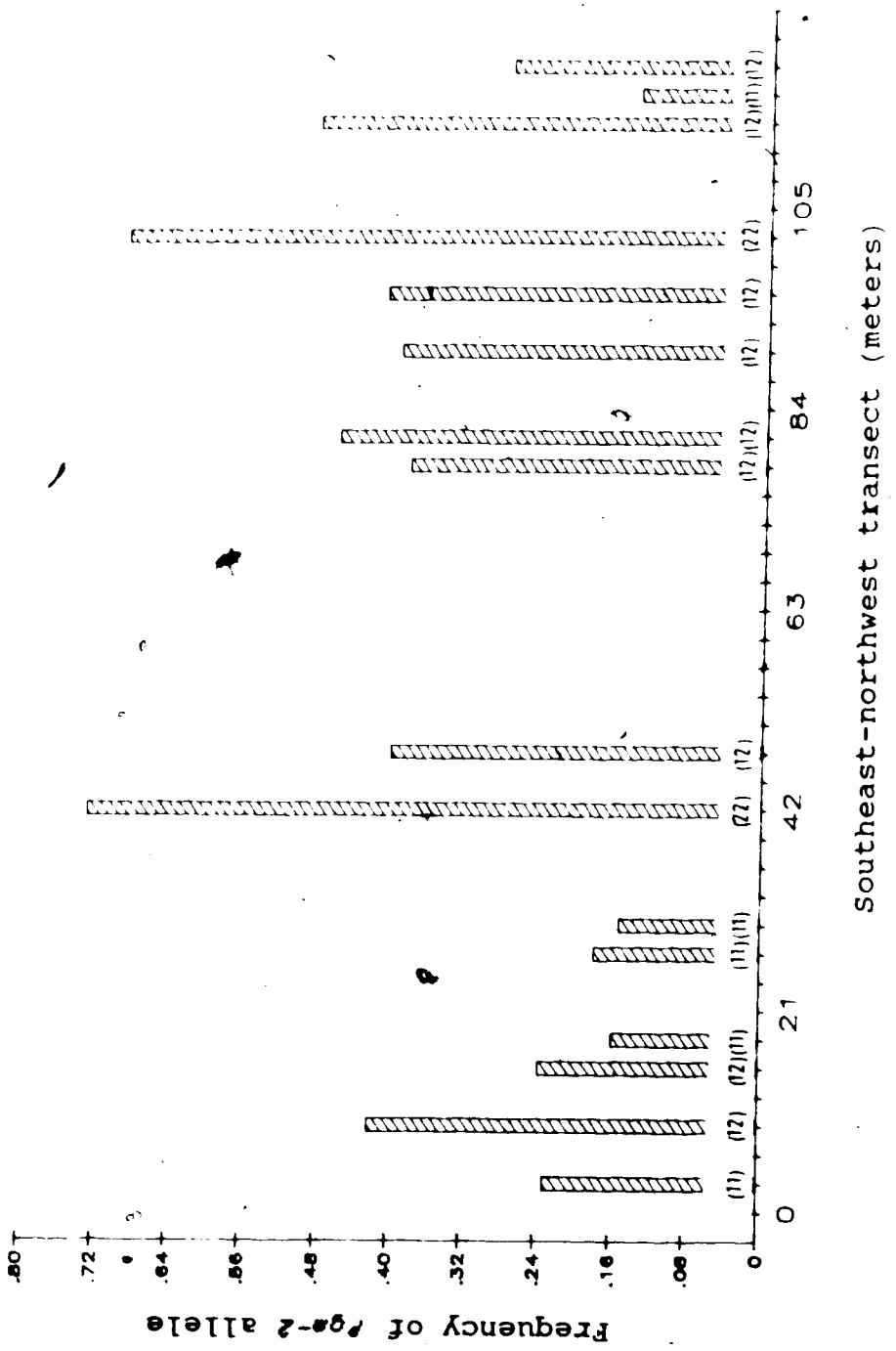


Figure 9: Frequency of the *Pgm-2* allele in the progeny of maternal trees along the southeast-northwest transect in a black spruce population near Whitecourt, Alberta. ( ) = Location and genotype of maternal trees.

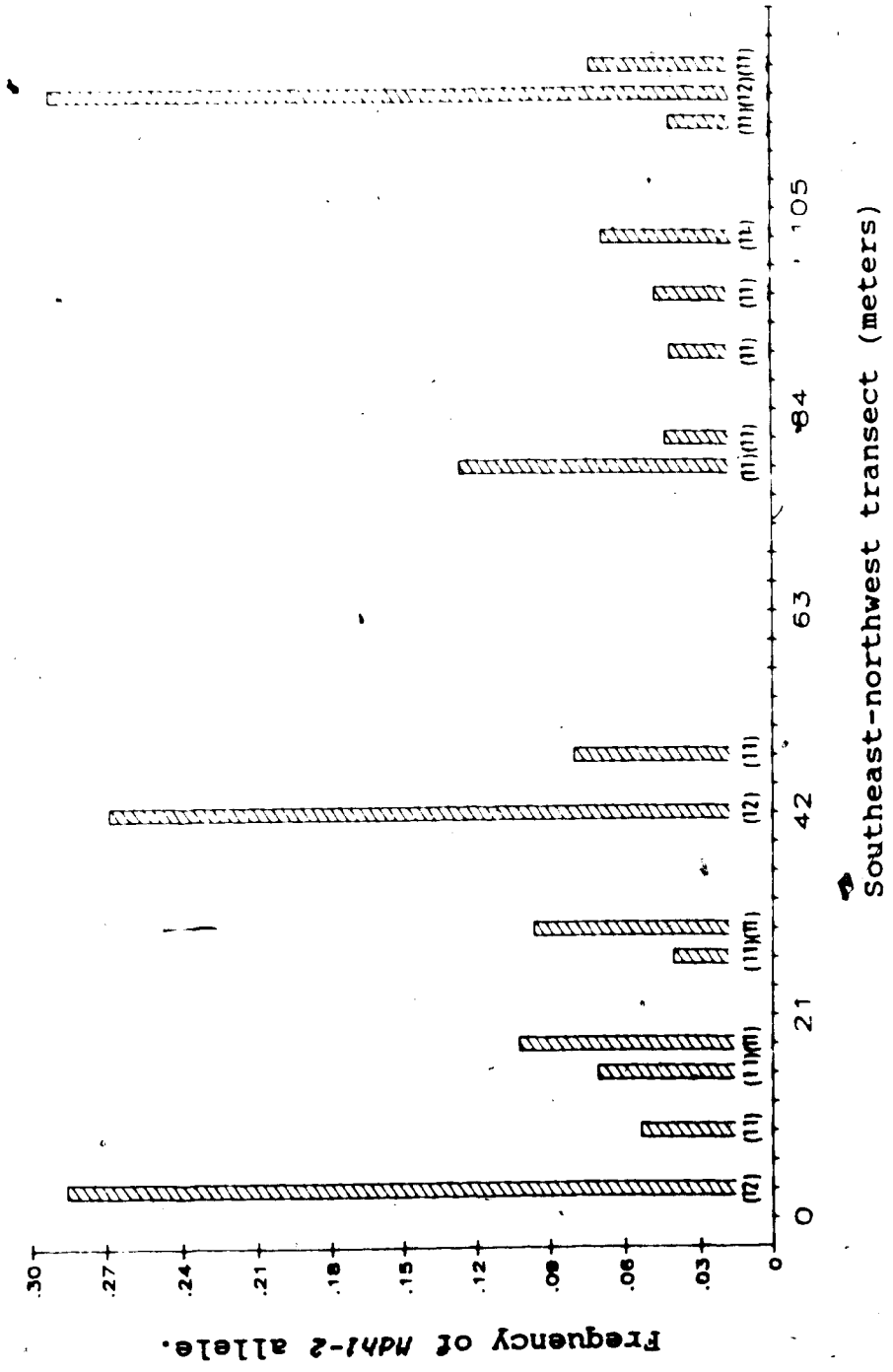


Figure 10: Frequency of the *Mdh1-2* allele in the progeny of maternal trees along the southeast-northwest transect in a black spruce population near Whitecourt, Alberta. ( ) = Location and genotype of maternal trees.

Frequency of *Pgm-2* in progeny of maternal genotypic class '22' was significantly higher than in class '12' (X^2=73.8 with one degree of freedom). It was also significantly higher than in class '11' (X^2=219.37 with one degree of freedom). As well, its frequency in progeny of maternal genotypic class '12' was significantly higher than in class '11' (X^2=78.5 with one degree of freedom).

The '2' alleles in the progeny of '11' maternal genotype must have come from outcross pollen. In the case of the '2' alleles present in the progeny of '12' and '22' maternal genotypes, those originating from outcross pollen cannot be distinguished from those originating from selfed pollen. However, the frequency of the '2' allele, both at the *Mdh* and *Pgm* loci, was significantly higher in the '12' than in the '11' maternal genotypic class. Also, at the *Pgm* locus, the frequency of *Pgm-2* was significantly higher in the progeny of the '22' maternal genotypic class than in the progeny of the '12' genotypic class.

The suggestion from both profiles is that the high frequency of the marker alleles in the progeny of the source trees is probably the result of self-pollination. This agrees with the pattern of pollen dispersal shown by *Mdh1-3* (Figure 8), with most pollen being dispersed within a few meters of the source tree. Some of the fluctuations in the frequencies of the *Mdh1-2* and *Pgm-2* marker alleles, particularly *Pgm-2*, which had a higher frequency in the population, are undoubtedly due to the presence of other sources of these

alleles However, one would expect this effect to be randomised over the transects, so that the pattern shown by the profiles should be a reasonable approximation of reality. Also, the evidence from the mating system study (Table 5) is supportive with *Pgm* having the lowest single-locus estimate of outcrossing, t_s, in population A for all three years studied.

The presence of clusters of related individuals (i.e. population substructure) differing in gene frequencies and with limited inter-cluster gene flow violates an assumption of the mixed mating model, that pollen is equally dispersed through the population. This violation was felt to be responsible for depressed single-locus estimates of outcrossing in several species (Ennos and Clegg 1982; Shaw and Allard 1983; Ellstrand and Foster 1983), and was also suggested as a reason for depressed multilocus outcrossing estimates (Ritland and El Kassaby 1985) in a Douglas-fir mating system study. It would appear that the single-locus estimates of outcrossing and possibly also the multilocus estimates in the present study were depressed by the presence of substructure in the population, combined with restricted pollen dispersal distances. While the overall picture is one of rapid decline with distance from source, there is a need for caution in the interpretation of the diagrams showing progeny frequencies of *Mdh1-2*, *Mdh1-3* and *Pgm*. Of all maternal trees and progeny analysed in both parts of the study, none was homozygous for *Mdh1-3*. This may have been due

3.4 CONCLUSION

Results of the mating system study, with single-locus estimates of outcrossing, t_s, being consistently lower than the multilocus estimates, t_m, suggest the possibility of substructure in these two black spruce populations. Further evidence is provided by spatial heterogeneity of the outcross pollen pool at the loci with the lowest t_s estimates, and the fact that Wright's Index of Fixation, F_{is}, is higher than the equilibrium inbreeding coefficient, F_e.

Results of the cluster analysis of population A by the CLINK clustering procedure provided further evidence of the presence of clusters of related trees in this population. Results of the investigation of the distribution of rare alleles and the runs test of allelic sequences are also supportive. Each identified new clusters and also confirmed some clusters previously identified by the CLINK clustering procedure.

While transect sampling can give an incomplete picture of cluster shape, it nevertheless appears that cluster size in this population varies from tight clusters of near neighbours, to clusters with a maximum inter-tree distance of over 100 meters. The results indicate that this type of cluster analysis is a suitable method of investigating population substructure. At the same time, however, they cast doubt on the usefulness of a simple comparison of genotypic similarity versus spatial distance between trees. Certainly the results of this investigation show that, in this case,

the two methods were better used separately.

Clusters are frequently thought to develop from dispersal of seeds close to the maternal parent after a fire or other local disturbance (Knowles *et al*. 1986; Shaw 1982). However, although this part of Alberta has a history of frequent fires, none of the usual evidence of fire, charcoal in the soil or charred remains of trees or stumps, was found. The clusters present in this population are, therefore, more likely to have arisen from continued natural regeneration, rather than regeneration around maternal trees after a fire or other local disturbance. Following this concept the presence of areally large clusters of genotypically similar individuals is not unexpected. Large clusters, such as A, B and F, probably represent genealogic clusters with members randomly distributed through the sampled area. It follows then, that if there is substructure in the stand, a mosaic of two types of cluster would be present:

(i) genealogic clusters with members randomly distributed over large areas of the stand. These are distantly related trees, related by virtue of being descended from a common, but distant, ancestral generation.

(ii) spatial clusters of related trees. This type of cluster might arise, e.g. from assortative mating with limited dispersion of offspring due to selection pressure, or from restricted dispersal of pollen and seeds.

In the present situation, the question to be addressed is whether the clusters representing genetically different

groups are simply genealogic clusters or whether groups of related, spatially clustered, trees are also present.

Intuitively one would feel that, of the clusters identified by the CLINK procedure, the larger clusters would represent genealogic groupings while the smaller could represent spatial clustering of relatives. However, the strongest evidence for spatial clustering comes from the distribution of rare alleles. All except *Mdh1-3* showed a very limited distribution. The fact that a group of neighbouring trees all possess the same rare allele suggests that they are probably related, while their limited distribution indicates limited dispersion of offspring. Limited dispersion of offspring could be due to local selective pressures or it could be the result of restricted pollen and/or seed dispersal. Habitat appeared uniform throughout the stand, so that limited dispersion of offspring may indeed be due to restricted pollen and seed dispersal. Results of the study of the frequency of pollen alleles in the progeny of maternal trees along the transects suggest that pollen flow is indeed limited. The high frequency of the *Mdh1-2*, *Mdh1-3* and *Pgm-3* alleles in the progeny of the source trees indicates a high level of selfing, also indicated by the results of the mating system study, emphasising the role of self-pollination in this population. Computer simulations show that with continued nearest neighbour pollination populations will evolve into clearly defined groups of closely related, highly homozygous trees (Turner *et al.* 1982). However, in this

population, while the evidence from the three clustering techniques indicates that the mature population contains clusters of relatives, the evidence from the Fixation Indices in the mating system study suggests that selfs are heavily selected against in both the pre- and post-germination phases. Thus, the related trees in the mature population are therefore probably half-sibs or outcrossed full-sibs.

There are two possible explanations for the divergence of population substructure in this population from that predicted by computer simulation studies. One possibility is that the population has not evolved over a long enough period to have generated the closely knit, discrete type of predicted cluster. A more likely explanation, however, is that computer simulations do not take into account some of the processes which occur in natural populations. Studies have shown that selfed seedlings are heavily selected against. If the intensity of selection was correlated with the level of inbreeding, it would effectively place an upper limit on the level of relatedness amongst neighbouring trees. This would mean that populations would not evolve into the discrete groups of highly related trees with well-defined boundaries, as is suggested by studies based only on pollen and seed dispersal distances. Also, computer simulations do not take into account occurrences such as the activities of seed-eating animals. They can be effective agents of seed dispersal.

One cannot be certain whether an equilibrium state has been reached in the development of population substructure in this stand. However, the evidence indicates that the population presently has a relatively high degree of substructure, so that a stage may have been reached where selection against inbreds counterbalances the development of more-closely related groups resulting from limited seed and pollen dispersal. As part of the evolutionary strategy of black spruce, self-compatibility, shown from the analysis of its mating system and from the study of the dispersion of pollen alleles, would be advantageous during the pioneer stage, while within-stand selection against highly inbred individuals would maintain a robust and competitive population once it had become established.

Since the study population is approximately ten kilometres from the nearest black spruce stand, there is probably little effective gene flow into the population by either pollen or seeds. The limited pollen dispersal distance suggests that there is mostly self-pollination or nearest-neighbour pollination. Assisted by limited seed-dispersal distance, this would maintain the level of substructure in the population.

Evidence from the second part of the study suggests that the level of substructure, combined with self-fertilisation and nearest neighbour pollination, probably depressed the single-locus, and possibly also the multilocus, outcrosssing estimates.

The results reinforce the typical tree breeding practice of avoiding selection of nearby trees as superior trees for the establishment of seed orchards. They also reaffirm the principles of effective seed orchard design. To fulfil the objective of producing seed which fully represents the genetic structure of the parents, panmixia (random mating) is required in a seed orchard (Van Buijtenen 1971). This requires a homogeneously distributed pollen pool through the entire orchard. With limited pollen dispersal there would be heterogeneity of the pollen pool and consequent violation of the assumption of panmixia.

The juxtaposing of inbred trees from such populations, or even the carrying out of controlled crosses between inbreds from small populations, could promote heterosis (Simmonds 1979), as is thought to have happened with radiata pine in New Zealand. Seed from many different shelterbelts and small populations was physically mixed and the resulting seedlings outplanted in forest blocks, effectively eliminating isolation by distance and promoting the greatest possible degree of outcrossing (Bannister 1968).

3.5 LITERATURE CITED

Ahlgren, C.E. 1959. Some effects of fire on forest reproduction in Northeastern Minnesota. J. For. 57:194-200.

Anon. 1939. Black spruce is a limited air traveller. USDA For. Serv. Lake States For. Expt. Stn. Tech. Note 147.

Bannister, M. H. 1968. Variation in the breeding system of *Pinus radiata*. p. 353-373. *In* H. G. Baker and G. L. Stebbins (eds.), The genetics of colonising species. Academic Press, New York.

Barrett, J., Cheliak, W.M. and P. Knowles. 1987. Variation at the PGI/AAT linkage group between populations of black spruce. Can. J. For. Res. 17:756-758.

Boyle, T.J.B. and E.K. Morgenstern. 1984. The population structure of black spruce in central New Brunswick. pages 142-156 *in* Demeritt, E.M. (Ed.), Proc. 29th Northeastern Forest Tree Improv. Conf. Div. For., W. Va. Univ., Morganstown.

Brown, A. H. D., Matheson, A.C. and K.G. Eldridge. 1975. Estimation of the mating system of *Eucalyptus obliqua* L'Herit. by using allozyme polymorphisms. Aust. J. Bot. 23:931-949.

Brown, B.A. and M.T. Clegg. 1984. Influence of flower color polymorphism on genetic transmission in a natural population of the common morning glory, *Ipomoea purpurea*. Evolution 38:796-803.

Cheliak, W.M. 1984. Mating system dynamics in a scots pine seed orchard. p. 107-117 *In* Proc. IUFRO Working Party on Ecological Population Genetics. Gottingen.

Conkle, M.T., Hodgskiss, P.D., Nunnally, L.B. and S.C. Hunter. 1982. Starch gel electrophoresis of conifer seeds: a laboratory manual. USDA For. Serv. Gen. Tech. Rep. PSW-64.

Dancik, B.P. and B.V. Barnes. 1975. Multivariate analysis of hybrid populations. Naturaliste Can. 102:835-843.

Ellstrand, N.C. and K.W. Foster. 1983. Impact of population structure on the apparent outcrossing rate of grain sorghum (*Sorghum bicolor*). Theor. Appl. Genet. 66:323-327.

Ennos, R.A. and M.T. Clegg. 1982. Effect of population

substructuring on estimates of outcrossing rates in plant populations. Heredity 48:283-292.

Everitt, B. 1974. Cluster Analysis. Heinemann Educational Books Ltd. London, England.

Feller, W. 1968. An introduction to probability theory and its applications. Vol. I. Wiley, New York.

Furnier, G.R. and W.T. Adams. 1987. The mating system in natural populations of Jeffrey pine. Amer. J. Bot. 73:1002-1008.

Furnier, G.R., Knowles, P., Clyde, M.A. and B.P. Dancik. 1987. Effects of avian seed dispersal on the genetic structure of whitebark pine populations. Evolution 41:607-612.

Haavisto. V.F. 1975. Peatland black spruce seed production and dispersal in Northeastern Ontario. Great Lakes Forest Research Centre, Ontario., Symp. Proc. O-P-4. p. 250-264.

Hamrick, J.L. 1982. Plant population genetics and evolution. Amer. J. Bot. 69:1685-1693.

Hartl, D. 1980. Principles of population genetics. Sinauer Assoc. Inc. Sunderland.

Heinselman, M.L. 1957. Silvical characteristics of black spruce. USDA For. Serv. Lake State Exp. Stn. Stn Pap. 45.

Howard, E.W. 1962. Establishment report: Black spruce seed trapping experiment, southwest Gander watershed, Newfoundland. Can. Dep. For., For. Res. Br., St. John's, Nfld. Mimeo. 62-65. 13 p.

Isaac, L.A. 1930. Seed flight in the Douglas-fir region. J. Forestry 28:492-499.

Knight, Wm. 1974. A run-like statistic for ecological transects. Biometrics 30:553-555.

Knowles, P., Furnier, G.R., Aleksiuk, M.A. and D.J. Perry. 1987. Significant levels of self-fertilisation in natural populations of tamarack. Can. J. Bot. 65:1087-1091.

Koski, V. 1973. On the effective population size in an areally continuous forest. Proc. Jt. IUFRO Mtg. S.02.04:1-3.

LeBarron, R.K. 1939. The role of forest fires in the reproduction of black spruce. Minn. Acad. Sci. Proc. 7:10-14.

Levin, D.A. and H.W. Kerster. 1968. Local gene dispersal in *Phlox*. Evolution. 22:130-139.

Levin, D.A. and H.W. Kerster. 1974. Gene flow in seed plants. Evol. Biol. 7:139-220.

Li, C.C. 1955. Population genetics Uni. of Chicago Press. Chicago.

Linhart, Y.B., Mitton, J.B., Sturgeon, K.B. and M.L. Davis. 1981. Genetic variation in space and time in a population of ponderosa pine. Heredity 46:407-426.

Lutz, H.J. 1956. Ecological effects of forest fire in the interior of Alaska. U.S. Dept. Agr. Tech. Bul. 1133.

McEwen, J.K. 1971. Aerial dispersion of black spruce seed. For. Chron. 47:161-162.

Meagher, T.R. and E. Thompson. 1986. The relationship between single parent and parent pair genetic likelihoods in genealogy reconstruction. Theor. Pop. Biol. 29:87-106.

Mitton, J.B. 1977. Genetic differentiation of races of man as judged by single-locus and multi-locus analyses. Am. Nat. 111:203-212.

Morgenstern, E.K. 1973. Heritability and genetic gain for height growth in a nursery experiment with black spruce. Can. For. Serv. Inf. Rep. PS-X-44. 10 pp.

Ostle, B. and R.W. Mensing. 1975. Statistics in Research. Iowa State Univ. Press. Ames.

Pielou, C.E. 1977. Mathematical ecology. Wiley, New York.

Rehder, A. 1967. Manual of Cultivated Trees and Shrubs. MacMillan Co. New York.

Ritland, K. 1985. The genetic mating structure of subdivided populations I. Open-mating model. Theor. Pop. Biol. 27:51-74.

Ritland, K. and Y.A. El Kassaby. 1985. The nature of inbreeding in a seed orchard of Douglas fir as shown by an efficient multilocus model. Theor. Appl. Genet. 71:375-384.

Roe, A.L. 1967. Seed dispersal in a bumper spruce seed year.

U.S.D.A. For. Serv. Research Paper INT-39.

Romesburg, H.C. 1984. Cluster Analysis for Researchers. Lifetime Learning Publications, Belmont Ca.

Ronco, F. 1970. Engelmann spruce seed dispersal and seedling establishment in clearcut forest openings in Colorado. USDA For. Serv. Res. Note RM-168.

Sakai, K-I. and Y. Miyazaki. 1972. Genetic studies in natural populations of forest trees. Silvae Genet. 21:149-154.

Schaal, B.A. 1975. Population structure and local differentiation in *Liatris cylindracea*. Am. Nat. 109:511-528.

Selander, R.K., Hunt, W.G. and S.Y. Yang. 1969. Protein polymorphism and genetic heterozygosity in two European sub-species of the house mouse. Evolution 23:379-390.

Shaw, D.V. and R.W. Allard. 1982. Estimation of outcrossing rates in Douglas-fir using isozyme markers. Theor. Appl. Genet. 62:113-120.

Silen, R.R. 1962. Pollen dispersal conditions for Douglas-fir. J. For. 60:790-795.

Simmonds, N.W. 1979. Principles of crop improvement. Longman Group Ltd., London.

Smouse, P.E. and J.V. Neel. 1977. Multivariate analysis of gametic disequilibrium in the Yanomama. Genetics. 85:733-752.

Smouse, P.E., Spielman, R.S. and M.H. Park. 1982. Multiple-locus allocation of individuals to groups as a function of the genetic variation within and differences among human populations. Amer. Nat. 119:445-463.

Snedecor, G.W. and M.R. Irwin. 1933. On the chi-square test for homogeneity. Iowa State College J. Sci. 8:75-81.

Sokal, R.R. and F.J. Rohlf. 1981. Biometry, 2nd Edition. W.H. Freeman and Co., San Francisco.

Stanek, W. 1975. The role of layerings in black spruce forests on peatlands in the claybelt of northern Ontario. Great Lakes Forest Research Centre, Ontario., Symp. Proc. O-P-4. p. 242-264.

Stevens, W.L. 1939. Distribution of groups in a sequence of alternatives. Annals of Eugenics. IX:10-17.

Strand, L. 1957. Pollen dispersal. Silvae Genet. 6:129-136.

Swed, F.S. and C. Eisenhart. 1943. Tables for testing the randomness of grouping in a sequence. Ann. Math. Statist. 14:66-85.

Tigerstedt, P.M.A., Rudin, D., Niemela, T. and J. Tammisola. 1982. Competition and neighbouring effect in a naturally regenerating population of scots pine. Silv. Fenn. 16:122-129.

Turner, M.E., Stephens, C. and W.W. Anderson. 1982. Homozygosity and patch structure in plant populations as a result of nearest-neighbour pollination. Proc. Natl. Acad. Sci. USA. 79:203-207.

Van Buijtenen, J.P. 1971. Seed orchard design, theory and practice. Proc. 11th Conf. S. Forest Tree Improvement. 197-206.

Vincent, A.B. 1965. Black spruce - a review of it's silvics, ecology and silviculture. Can. Dep. For., Publ. No. 1100.

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

Wheeler, N.C. and R.P. Guries. 1982b. Biogeography of lodgepole pine. Can. J. Bot. 60:1805-1814.

Wishart, D. 1978. Clustan user manual. Program Library Unit, Edinburgh Univ., Scotland.

Wright J.A. 1953. Pollen-dispersion studies: Some practical applications. J. For. 51:114-118.

Wright, S. 1946. Isolation by distance under diverse systems of mating. Genetics 31:39-59.

Wright, S. 1978. Evolution and the genetics of populations. Vol. 4. Variability within and among natural populations. Univ. of Chicago Press, Chicago.

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

Yeh, F.C., Khalil, M.A.K., Trust, D.C. and Y.A. El-Khassaby. Allozyme variation in *Picea mariana* from Newfoundland: genetic diversity, population structure and analysis of differentiation. Can. J. For. Res. 16:713-720.

Job number 845731 University of Alberta Computing Service (System EC'07) 17:04:01 MST .. Thu Jan 21 88

User: FGO6
Project no: FOGE

**** Obtained at 17:04:01 Thu Jan 21/88
**** Released at 17:04:40 Thu Jan 21/88
**** Lines printed 8235
**** Images printed 127
**** Pages printed 127

```
 8888888888        444     555555555555  777777777777   3333333333       11
888888888888      4444     555555555555  777777777777   3333333333      111
88        88     44 44     55            77       77   33        33     1111
88        88    44  44     55                     77                33    11
 88      88     44   44     55                    77                33    11
  88888888     44444444444  555555555           77              3333    11
  88888888    444444444444  5555555555          77              3333    11
 88      88          44            55           77                33    11
88        88        44             55           77                33    11
88        88        44             55           77       33       33    11
888888888888        44     555555555555         77      333333333333  1111111111
 8888888888         44     55555555555          77       3333333333   1111111111

FFFFFFFFFFFF   GGGGGGGGGG     OOOOOOOO     6666666666
FFFFFFFFFFFF  GGGGGGGGGGGG   OOOOOOOOOO   666666666666
FF            GG        GG   OO      OO   66      66
FF            GG             OO      OO   66
FF            GG             OO      OO   66
FFFFFFFF      GG             OO      OO   66666666666
FFFFFFFF      GG     GGGGG   OO      OO   666666666666
FF            GG     GGGGG   OO      OO   66      66
FF            GG        GG   OO      OO   66      66
FF            GG        GG   OO      OO   66      66
FF            GGGGGGGGGGGG   OOOOOOOOOO   666666666666
FF             GGGGGGGGGG     OOOOOOOO     6666666666
```