Theor Appl Genet (1985) 71:101–110

Genetic neighbourhood structure in a population of *Picea abies* L.

D. Brunel [1] and F. Rodolphe [2]

[1] INRA Station d'Amélioration des Arbres Forestiers, Ardon, F-45160 Olivet, France
[2] I.N.R.A. Laboratoire de Biométrie, Domaine de Vilvert, F-78350 Jouy-en-Josas, France

Received December 12, 1984; Accepted February 28, 1985
Communicated by P. M. A. Tigerstedt

Summary. Natural populations are currently the basic material for studying forest tree breeding, but little is known on the genotypic spatial structures in these stands. The use of gene markers, such as isozymes, leads to the determination of part of the allelic constitution of individuals. A method is presented here to estimate the degree of genetic relationship between any pair of genotypes. A French *Picea abies* population is analyzed by these means and a slight but significant correlation between estimated genetic relationship and topographic distance is found.

Key words: Genetic relationship – Population structure – *Picea abies* – Isozymes – Perennial

Introduction

Natural populations are at present the basic material for the forest tree breeder. Genetic variability is important in these large continuous perennial populations, yet the maintenance mechanisms of this variability are not clearly understood (Hamrick et al. 1979). Two conflicting opinions exist in the literature: a) pollen dispersal alone is sufficient for these large continuous populations to be panmictic and genetically quite homogeneous; b) small distances of pollen and seed dispersal generate clusters of genetically related individuals in a population.

So far, most of the studies on the genetic structure of a particular population of allogamic species have consisted of panmixia tests. Panmixia is often tested for and found to be present in most cases. The value of these tests can be questioned, however, because of the small sample size and methods of seedlot sampling.

For breeding forest trees, the two alternatives, panmixia or clusters, are important as most of the selection schemes start by evaluating the genetic quality of populations or trees by using offspring harvested from natural stands. In fact, one may wonder whether a population or a tree is discarded for its real genetic quality or because of a higher inbreeding coefficient.

Previously, many studies of inbreeding were, in fact, estimates of the proportion of self-fertilization as opposed to outcrossing in plant populations. Shaw et al. (1979), however, interpreted the observed discrepancies between outcrossing rates estimated with a single locus or with multilocus methods as a reflection of the family structure in a *Pseudotsuga menziesii* population. This impact of population structure on the apparent outcrossing rate of grain sorghum is clearly shown in a study by Ellstrand et al. (1983). However, the rare studies undertaken in order to detect family clusters seem to show that spatial genotypic structures exist in plant populations. This is revealed by different methods of investigation:

– pollen and seed dispersal, followed by radioactive tracers (Schmidt 1970), or by insect-pollinator flight studies, on *Lupinus texensis,* (Schaal 1980), and on *Lithospermum caraliniense* (Kester et al. 1968), etc.

– crosses between neighbouring plants (Coles and Fowler 1976). These authors found a clear relationship between inbreeding depression and closeness to the male parent in two white spruce populations. Trees separated by more than 100 m seemed to be unrelated.

– genetic markers such as isozymes, as they are easy to detect (Antlfinger 1982; Handel 1983). Schaal (1974, 1975) showed the existence of a strong genetic structure in a *Liatris cylindracea* population. She noted that genetic divergence over distance appears to be a function of random gene frequency fluctuations even if difference of allelic frequency between adjacent 3 m^2 squares are as high as 20%. Only when 15 polymorphic loci are considered together does the pattern of genetic variation in the population emerge. On forest material, Sakai et al. (1971) and Park (1972) showed, by the analysis of one locus, a high relationship between *Cryptomeria japonica* trees 25 m apart. Similar results have been found with *Pinus sylvestris* by studying one rare allele in offsprings (Müller 1977), or by studying spatial distribution of rare alleles (Rudin

et al. 1977). Considering each locus separately, Tigerstedt (1973) found a complete random gene distribution in *Picea abies*. While the prospected area was small a spatial structure seems to appear visually with the three loci together. Finally, working on *Pinus ponderosa*, Linhart et al. (1981) found that tree clusters 30 m apart can have as large Nei genetic distances as populations several kilometers distant from one another.

These studies indicate that the mating system of the forest material can be defined as a system of preferential mating between neighbouring trees, including self fertilization.

The present study concerns a *Picea abies* population. It was aimed at answering the following questions: does a spatial genetic structure exist between the mature trees of the population, and what is the scale. The genotype of each prospected tree is determined for several isoenzymatic characteristics and the genetic relationship between each pair of trees is calculated with a new method described in this paper.

Material and methods

1 Material

Picea abies is a monoecious species. Its pollen grains have large air sacs which enable them to travel long distances (several kilometers in a strong wind). The seeds fall from the cones remaining on the tree. If the reach of pollen and seeds is defined as being the radius of a circle within which 95% of the pollen and seeds emitting from the center land on the ground, the reach of the pollen amounts to 260 m and the reach of the seeds to 65 m (Müller 1976; Schmidt-Vogt 1978). Lundkvist (1979) found a self-fertilization rate as high as 26% but most authors consider that the proportion of trees which would grow to be less than 1%.

This study was realized with a natural French population of *Picea abies* L. The forest stand, located on the second plateau of the Jura at Bonnetage (N 47°10′, E 6°45′, altitude 875 m), is known as one of the best reforestation crops in France. The management archives indicate that in 1856 the stand constituted of one to thirty-year-old natural seedlings dominated by seed-trees of different ages, which were cut that year. We can consider this forest as a "futaie régulière" ("evenly aged high forest"): all trees of the generation G+1 have nearly the same age as they come from crosses between parental trees (generation G) chosen and kept by the foresters ("semenciers" = "seed trees").

Seeds from 104 parents were collected in 1968 over an area of 950×400 m (Fig. 1), with 35 seed lots, on a systematic transect from East to West. However, this population is a mixture between *Picea abies* and *Abies alba* in a ratio 50/50, with a lower density of *Picea abies* in the eastern part. This fact may introduce a sampling bias.

The main characteristics of the sample stand were: mean age = 150 years; height = 40–45 m, circumference = 160–220 cm.

2 Extraction and electrophoretic methods

For each parent, 20 endosperms were crushed together with a Kontes' potter in 1 ml of a 10 mM Tris, 25 mM KCl buffer (pH = 7.4) for 2 min in ice. After a refrigerated centrifugation (23,000 g, 20 min) the supernatants were preserved at −80 °C. With this method, genotypes can be determined with only one extraction.

For electrophoresis, each extract was placed in a 7.5% polyacrylamide slab gel well (BIORAD material). Both the gel and electrode buffer was 89 mM Tris, 89 mM Boric acid, 2.5 M EDTA pH = 8.3.

Eight enzyme systems were analysed: glutamate oxaloacetate transaminase – esterase – fluorescent esterase – peroxidase – phosphoglucomutase – malate dehydrogenase – amylase – glutamate dehydrogenase.

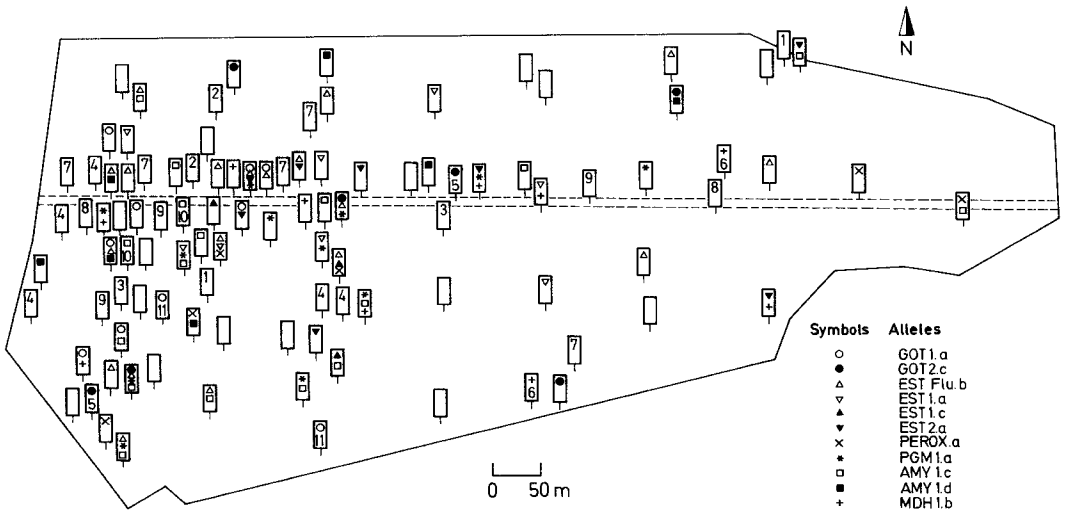


Fig. 1. Spatial distribution of the 104 prospected trees in the Bonnetage stand. Eleven rare alleles from the total sample size of 25 are indicated by different symbols (trees bearing the same number, from 1 to 11, have the same genotype)

First genetic analyses

Eleven loci were resolved with the electrophoretic techniques used: glutamate oxaloacetate transaminase (GOT1; GOT2), esterase (EST1; EST2), fluorescent esterase (EST FLUO 1), peroxidase (PEROX 1), phosphoglucomutase (PGM 1), malate dehydrogenase (MDH1–MDH2), amylase (AMY 1), glutamate dehydrogenase (GDH 1). Descriptions of the formal genetics are to be found in the literature for seven of them: GOT1, GOT2, EST1, EST2, MDH1, MDH2, GDH1 (Lundkvist 1978); others showed segregation patterns consistent with a single locus inheritance by endosperm analysis: EST FLUO 1, PEROX 1, PGM 1, AMY 1 (Brunel, unpublished data).

1 Spatial allelic distribution

A spatial allelic distribution can be drawn for each locus. Nine maps were realized but only eleven rare alleles (frequency less than 10%) are represented on Fig. 1. Visual analyses indicate a non-random distribution of alleles for some loci, but the irregular tree localisation makes any visual analysis subject to questioning.

2 Allelic frequencies, independence and panmixia tests

All the trees are homozygotes for the GDH1 and MDH2 loci. Table 1 indicates the allelic frequencies for the nine loci showing polymorphism. Because of the suspected sampling bias, two estimations were made: one with the 104 trees, the other with only 47 trees chosen in such a way that only one tree is found in each 50×50 m square. In fact, differences between the two samples seem to be small, but all the further analyses will be executed with both frequency estimations.

While very little data exists on loci linkage on *Picea abies,* Lundkvist (1979) showed EST1 and EST2 to be on the same chromosome.

The consistency of the observed genotype distribution conditionally to allelic frequencies with loci independence can be tested. Since all loci of a chromosome are ordered, independence of several loci is equivalent to independance within any pair of loci. At each locus we considered two sets of genotypes: the most frequent genotype, and groupings of all others. A Fisher's exact probability test has been performed on each of the 36 different 2×2 contingency tables obtained. No evidence appears of a discrepancy to the independence hypothesis.

In a large population, genotypic frequencies are predicted by the Hardy and Weinberg equilibrium under the assumption of random mating in the absence of migration, mutation, selection and genetic drift. All studied loci are considered to be markers (unselected by themselves and not linked to selected genes) and independent. Thus, if the number of parent trees was large enough, in the absence of genetic spatial structure, Hardy and Weinberg's equations should be verified. On the contrary, if a genetic spatial structure exists as a product of mating in one neighbourhood, Hardy and Weinberg's equations would not be verified and an excees of homozygotes can be expected. A chi-square test was performed on each locus. In order to observe a minimum number of trees of each genotype, rare alleles have been pooled together in such a way that only three genotype classes have been defined: two, one or zero copies of the most frequent allele. Thus, we could test only a subset of all Hardy and Weinberg's equations. These loci are considered independent, as are the chi-square test statistics. Two loci seem to have an excessively large number of homozygotes while one locus appears to have too many heterozygotes.

However, the probability of a first type error increases with the number of independent tests per-

Table 1. Allelic frequencies of the nine loci showing polymorphism, estimated with the 104 trees and 47 trees samples (see text)

Locus	Alleles	104 trees sample	47 trees sample
GOT 1	a	0.048	0.054
	b	0.952	0.946
GOT 2	a	0.543	0.564
	b	0.423	0.404
	c	0.034	0.032
EST FLUO 1	a	0.091	0.075
	b	0.909	0.925
EST 1	a	0.038	0.054
	b	0.924	0.915
	c	0.038	0.031
EST 2	a	0.014	0.075
	b	0.236	0.223
	c	0.750	0.702
PGM 1	a	0.082	0.054
	b	0.918	0.946
MDH 1	a	0.937	0.968
	b	0.058	0.032
	c	0.005	0.000
AMY 1	a	0.25	0.234
	b	0.620	0.627
	c	0.092	0.085
	d	0.038	0.054
PEROX 1	a	0.034	0.044
	b	0.822	0.776
	c	0.144	0.180

formed, as does, therefore, the probability of finding a surprising result. The final picture is not clear.

As a product of forest management, the actual stand might have been produced by a reduced number of seed trees. This fact may be the reason why one population falls outside the Hardy and Weinberg's equilibrium. The parent trees may not have the same fertility and they may vary in their individual ovule and pollen production (Müller-Starck et al. 1983). Therefore, allelic frequencies in male and female gametes can be quite different, producing a population with an excess of heterozygotes (see Ziehe 1982).

Another consequence of such a situation is that the observed genotypes, because of possible genealogical relationships, are not independent stochastic variables. This make the chi-square test no longer valid since there is not an underlying multinomial distribution on which it could be based.

The difficulty in obtaining a clear inerpretation of the genetic structure of this population led us to search for a new model for estimating genetic relationships which considered electrophoretic data and our knowledge about this stand. In fact, we are interested not only in the kinship rate of the population but rather in a characterization of genetic closeness between any two trees and the size of a possible spatial structure.

A simple genetic distance to imagine is to add each allelic difference between two trees and to compare this estimation with topographic distance. This was done but no relationship was found. Therefore, we tried to find a more efficient estimation with the assumption that two individuals have a higher probability to be related when they have in common alleles with low allelic frequencies. The model is presented in the sections.

Presentation of a genetic relationship estimation model

The parental stand from which seed trees have been selected must be regarded as the last generation of a collected forest which either followed Hardy and Weinberg's equilibrium or which presented an excess of homozygotes if there was a genetic spatial structure. No important excess of heterozygotes could be present, since at each generation the number of parents must have been large enough to discard any significant random effect of male and female differential contributions. As the parental population is considered to have been produced by many trees, the probability of a close genetic relationship between two seed trees is negligible and the only genetic relations to be considered in the sampled generation are:

- "non related" NR = A1
- "half sibs" HS = A2
- "full sibs" FS = A3

In this complete system of disjointed events, for any pair (x, y) of trees and with Ai the kinship degree, we can define the probability of each possible parental relationship conditionally to the pair genotype (Bayes' theorem):

$$P\{Ai/(x, y)\} = \frac{P\{(x, y)/Ai\} \cdot P(Ai)}{\sum_{i=1}^{3} P\{(x, y)/Ai\} \cdot P(Ai)}$$

The different terms of this equation can be estimated with the electrophoretic data.

1 Consequences of kinship on the distribution of genotypic pairs

Let $P\{(x, y)/Ai\}$ be the probability of observing the pair (x, y) under the hypothesis of the Ai type of kinship.

For one locus each pair of genotypes x and y can fall into one of the seven categories:

	Genotype x	Genotype y	
(1)	ij	kl	– both are heterozygotes, and they have no common allele
(2)	ij	ik	– both are heterozygotes and they have one common allele
(3)	ij	ij	– both are heterozygotes and they have two common alleles
(4)	ii	jk	– one is homozygote, the other heterozygote and they have no common allele
(5)	ii	ij	– one is homozygote, the other heterozygote and they have one common allele
(6)	ii	jj	– both are homozygotes; they have no common allele
(7)	ii	ii	– both are homozygotes; they have the same allele

Under the hypothesis that the parental population followed the Hardy and Weinberg equilibrium, we can calculate the probability of observing a pair (x, y) for each of the three relationships FS, HS and NR, as a function of allelic frequencies in the parental population.

Table 2 summarizes the probability $P\{(x, y)/Ai\}$ of observing each of the seven possible meetings between alleles of a same locus, considering three different relationships (one example is given there for a simpler explanation in Table 2 a). Obviously, each possible pair of genotypes has a different probability according to the kinship between the two observed individuals.

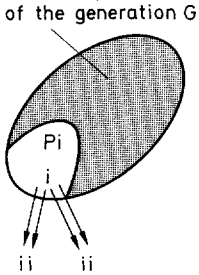
Table 2. Probability of a pair (x, y) above the 3 hypotheses of non related, half and full sib: P(x, y)/A_i, Pi, Pj, Pk, Pl are the frequencies in the parental population G of the alleles i, j, k, l of the locus L

Category	Non-related	Half sib	Full sib	Variation of P (x, y)/Ai
(1) ij kl	8 Pi Pj Pk Pl	4 Pi Pj Pk Pl	2 Pi Pj Pk Pl	NR → HS → FS
(2) ij ik	8 Pi^2 Pj Pk	Pi Pj Pk (1+4 Pi)	Pi Pj Pk (1+2 Pi)	if Pi > ¼: NR → HS → FS; if P_i < ¼: HS → NR, FS
(3) ij ij	4 Pi^2 Pj^2	Pi Pj (½ (Pi+Pj)+2 Pi Pj)	½ Pi Pj (1+Pi+Pj+2 Pi Pj)	NR → HS → FS
(4) ij jk	4 Pi^2 Pj Pk	2 Pi^2 Pj Pk	Pi^2 Pj Pk	NR → HS → FS
(5) ii ij	4 Pi^3 Pj	Pi^2 Pj (1+2 Pi)	Pi^2 Pj (1+Pi)	if Pi > ½: NR → FS; if Pi < ½: HS → NR, FS
(6) ii jj	2 Pi^2 Pj^2	Pi^2 Pj^2	½ Pi^2 Pj^2	NR → HS → FS
(7) ii ii	Pi^4	½ Pi^3 (1+Pi)	⅓ Pi^2 $(1+Pi)^2$	NR → HS → FS

Table 2a. Example of the estimation of $P\{(x,y)/Ai\}$ for the meeting (ii), (ii)

Ai = NR

Gametic pool of the generation G



$P\ [(x,y)/NR] = Pi^4$

Ai = HS

Common parent	Pi^2 ii	2Pi (1-P_i) ix
Other gamete Pi i	ii	ii

$P\ [(x,y)/HS] = \frac{1}{2} Pi^3 (1+Pi)$

Ai = FS

1st Common parent / 2nd Common parent	Pi^2 ii	2Pi (1-P_i) ix
Pi^2 ii	ii	ii
2Pi (1-P_i) ix	ii	ii

$P\ [(x,y)/FS] = \frac{1}{4} Pi^2 (1+Pi)^2$

Thus, examining the observed pairs of genotypes, it should be possible to infer the importance of each possible kinship in our sample.

For the nine polymorphic loci which are presumed to be independent:

$$P\{(x, y)/Ai\} = \prod_{L=1}^{9} P^{L}\{(x, y)/Ai\}$$

Allelic frequencies are estimated with the electrophoretic data of the G+1 generation, as we supposed the frequency of the different alleles not to have changed in a drastic manner between the G generation of the "seed trees" to the studied G+1 generation.

At each pair of genotypes the three normalized probabilities

$$P\{(x, y)/Ai\} \Big/ \sum_{i=1}^{3} P\{(x, y)/Ai\},$$

whose sum is unity can be associated.

Then, at each pair there is a coresponding point in an equilateral triangle. Two different pairs, whose images are close in this triangle, behave in the same way with respect to different kinship types. Figure 2 shows the distribution of the $N(N-1)/2 = 5356$ pairs of different sampled trees in 100 subtriangles.

As most of the 100 subtriangles defined are nearly empty, we group them into six classes (Fig. 2) in a manner that each class has an effective number higher than five hundred and that each class is as compact as possible (they shall thus be pertinent with respect to kinship examination).

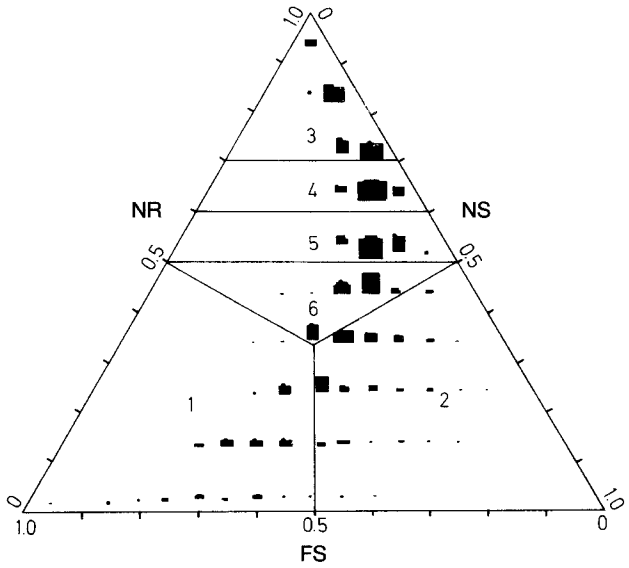


Fig. 2. Distribution of the 5356 pairs for their relative probabilities of $P\{(x, y)/Ai\}$ with the three genetic relationships Nr, HS and FS

The probability of a pair of genotypes to belong to each class can be calculated as a function of allelic frequencies in the parental population under the three possible genetic relationships within the pair. Let Pj, i be the probability that a pair of genotypes with Ai^{th} kinship type belongs to class j. Theoretically, Pj, i can be calculated by summing the probabilities associated to all possible pairs of genotypes belonging to class j. As it is practically impossible, we simulated 40,000 pairs of genotypes for each considered kniship type as follows: each time the right number of parents (4, 3 or 2) is drawn at random out of the parental population by using a uniformly distributed random variable U on (0, 1) for each of the nine loci and for each parent. Then, to constitute an offspring pair, random gene segregation is simulated using another independent uniformly distributed random variable V on (0, 1), for each locus and each gamete.

In order to accelerate convergence, these computations were made by independent blocks of four offspring pairs with a given genetic relationship. The four offspring pairs in a block are computed with the four sets of random variables (U, V), (1–U, V), (U, 1–V) and (1–U, 1–V) (see Fig. 3). Since a uniform variable and its antithesis have the same distribution, final estimation is unbiased; but since they are negatively correlated, we can expect "more than normal" different genotype pairs within a block and, consequently, much less variability between the blocks. Indeed, we observed that the probability estimator variances were about one ninth of their expected values if the same number of offspring pairs were simulated independently.

For each simulated offspring pair, its class is determined and the estimator $\widehat{Pj, i}$ of Pj, i is the proportion of offspring pairs simulated under the Ai^{th} genetic relationship type which fall into the j^{th} class. Figure 4 shows the comparison between the three theoretical population proportions and that of Bonnetage.

2 Kinship importance or estimation of P(Ai)

With a sample of N trees, it is possible to make N/2 disjoint pairs: each tree of odd rank is paired with the tree of following even rank. There are $N!/(N/2)! \cdot (2^{N/2})$ such different sets of N/2 disjoint pairs, each one corresponding to a certain permutation T of the ranks.

Kinship within a particular set of N/2 disjoint pairs can be approached by the proportions P_{NR}, P_{HS}, P_{FS} of respectively non related, half sibs and full sib pairs.

Let $M_j(T)$ be the number of pairs in this set belonging to class j. We have:

$$E(M_j(T)) = (N/2) \cdot \sum_{i=1}^{3} P_{ji} \cdot P(Ai)_T .$$

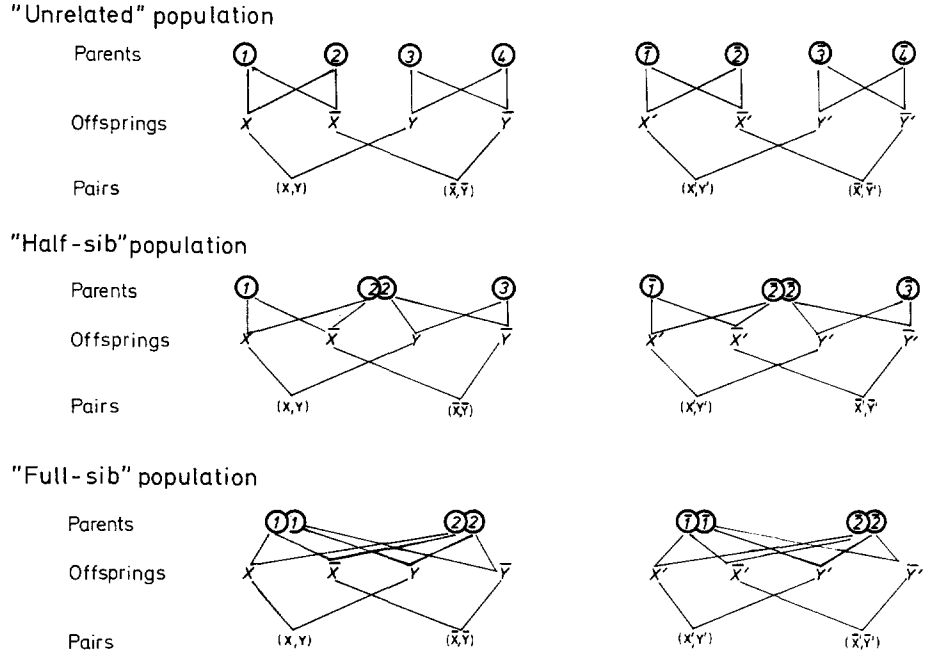


Fig. 3. Mating schemes of three simulated populations with three different genetic relationships between the offspring (bars indicate antithetics)

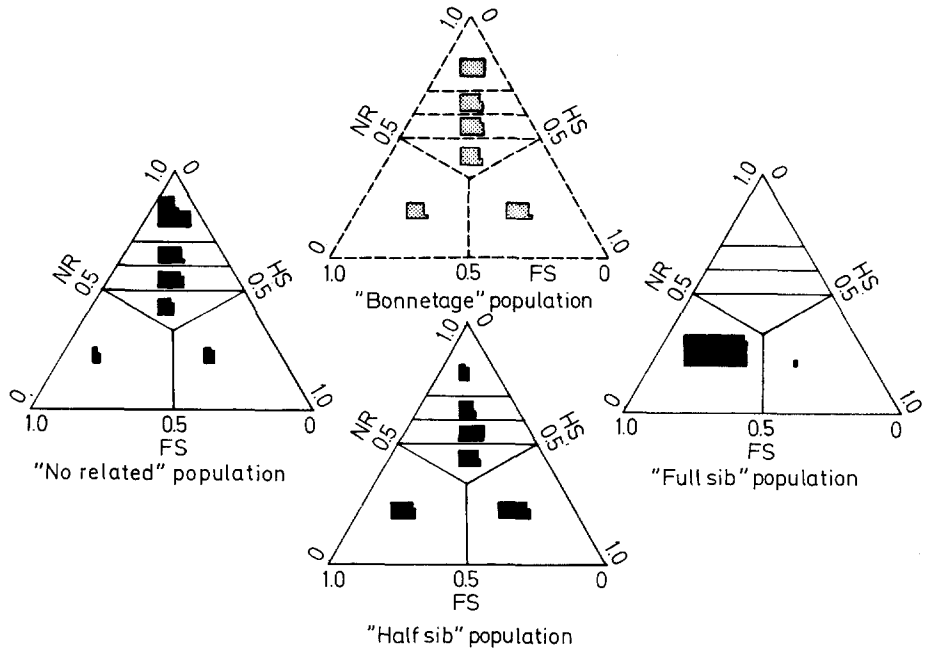


Fig. 4. Distributions of pairs according to their relative values of $P\{(x, y)/Ai\}$ for the three simulated populations and the observed one

If all N/2 disjoint pairs were independent, the proportions $P_{Ai}(T)$ could be estimated by several methods: minimum chi-square, maximum likelihood, least squares. We are interested in the proportions P_{Ai} not relative to a particular set of N/2 disjoint pairs but in the set of all pairs. These proportions can be estimated as the mean values over all permutations of all particular $P_{Ai}(T)$. Only the means of the least squares estimators can easily be calculated (as the solutions of linear equations).

But the variance-covariance matrix of these estimators is unknown since all N. (N–1)/2 different pairs are not independent; moreover since by hypothesis we cannot discard the kinship structure in the sampled population, all pairs of a set of N/2 disjoint pairs cannot be regarded as independent. Therefore, it is very difficult to get an idea about the parameter estimators precision.

The means' estimators $\widehat{P_{Ai}}$ must have a variance less than the mean variance of all $\widehat{P_{Ai}}(T)$. Therefore, we

considered the variance of $\widehat{P_{Ai}}$ (T) in the "median case" (where $M_j = 1/6 \times N/2$ calculated in all N/2 pairs were independent). This variance equals $N/2 \times 1/6 \times 5/6$ as a maximum for the variance of $\widehat{P_{Ai}}$.

3 Results

We found: $P_{NR} = 0.4$; $P_{HS} = 0.6$; $P_{FS} = 0.0$ with the standard deviations of P_{NR} and P_{HS} being less than 0.14 and 0.17. Similar results are found in the 47 tree sample. Under the hypotheses on which our model is based, P_{FS} is certainly very small. Thus, we can consider that there are nearly no full sibs in the prospected trees. Although the proportion of half sibs is surprisingly very high.

The absence of full sib pairs can be easily explained by the sampling method used and the small number of existing bibliographic references: 10% of the trees are less than 100 m apart, which is the average of the seed dispersal. In this area we could hope to find the full sibs and half sibs of mother.

On the other hand, 10% of the pairs are more than 600 m apart which correpsonds to the average of pollen dispersal, and so to genetically unrelated trees (Fig. 5). Because of the prospection method used, we could expect to see only paternal half sibs. Thus, considering that no full sibs are present, we can partition the sampled population into paternal half sibs families. If we suppose that these families are of the same size, a $P_{HS} = 0.2$ (under our hypotheses the proportion of half-sibs is certainly higher) means that there are only five such families! Of course there can be more families but at least one must be larger. For instance, the maximum number of such families is given by a lot of families

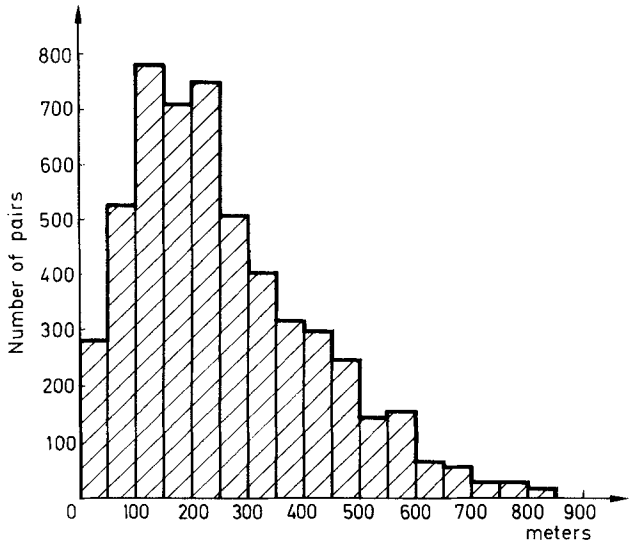


Fig. 5. Histogram of geographical distances between pairs of trees

reduced to one tree and only one large family which represents about 45% of the total sampled population!

This is rather difficult to believe, even if we suppose a low number of pollen producers in the parental generation as it means that either a low number of trees were kept in the stand 150 years ago or that a few trees produced a large amount of pollen. Nevertheless, the estimated proportion of half-sib pairs seems far too high.

Can this result be regarded as an evidence for rejecting the Hardy-Weinberg's equilibrium hypothesis in the parental population G?

The high value of P_{HS} is due to the relatively high proportion of class 1 and 2 tree pairs and a relatively low proportion of tree pairs in class 3. In other words (Table 2) there is an excess of pairs of trees sharing several common alleles, with respect to the hypothesis of a Hardy-Weinberg's equilibrium in the parental population (even with a reasonable proportion of half sibs).

We made three hypotheses for estimating the relative importance of each kinship type, and each one is questionable.

1. Allelic frequencies in the parental population were similar to those observed in the sampled population. In fact, allelic frequencies in the G generation could not be estimated in an other way since only the sampled generation had been studied. Even if a kinship structure exists in the analysed sample, these estimates are unbiased. Nevertheless, the way in which the trees were prospected could have had a negative effect on these estimations. That is the reason we made an other estimation with a sub-sample, but our results did not show a great sensitivity to this change.

2. All loci are independent. This hypothesis is very critical when one works with several loci as we did. Independence means, from a genetical point of view, no physical linkage of the loci on the chromosomes, and no epistatic effect between different loci. Independence between all pairs of loci was tested statistically (Fisher's exact probability test) but these tests could not be very powerful since we also had to pool together rare genotypes. Independence was not rejected even for EST1 and EST2 which we know to be linked (Lundkvist 1979). If independence does not exist then it leads to an increase in the number of pairs belonging to classes 1 and 2. So we made nine different analyses, avoiding one locus each time: identical results were found. But probably, the EST1 · EST2 linkage is screened by the high frequency of EST1–b (Table 1). This locus behaves as if there is only one allele.

3. Genotypic frequencies in the G parental population correspond to Hardy and Weinberg's proportions. It seems reasonable to suspect that Hardy and Weinberg's equilibrium in the parental population did not

exist, since our results would have been more understandable if preferential mating between neighbour trees could be supposed, associated with a genetic spatial structure (neighbour trees are supposed to be genetically closer).

If we look at the histogram of distances between sampled trees (Fig. 5), more than 40% of the pairs are less than 200 m apart and therefore a large proportion of them could be genetically "more alike" than normal, and, behave as "half sibs" in our model. Of course, to be really convincing this hypothesis should be more carefully studied in order to examine if in this frame the observed results are likely.

In fact, kinship in the parental population and a low number of pollen producers are not conflicting phenomena and both may explain the G+1 genotypic structure.

Test of a correlation between genetic and topographic proximities

For each pair of trees, the estimated probability to be half sibs, conditional to the observed genotypes, is an empirical genetical proximity (the probability to be full sibs is nearly zero for any pair and cannot be used).

As a first test, we plotted geographic distance against this probability. To keep the picture readable only the ranks of the mean values of all 50 m (topographic distance) tree pair classes were plotted (Fig. 6).

If there is no relationship between the genotype of a pair and its topographic distance, any permutation of the genotypes of the N observed trees, gives as likely a distribution as the observed one. Thus, any measure of correlation between both proximities has a probability distribution which is invariant by permutation under this hypothesis. On the contrary, if there is a genetic spatial structure in the sense that neighbour trees are genetically more alike than distant trees, the same measure of correlation is expected to be higher and has a distribution which is not invariant by tree genotype permutation.

We computed Spearman's rank correlation coefficient, r, between the probability to be HS and topographical distance, over all 5,356 pairs, and found $r=-0.016$.

This remark leads to a permutation test. If we compute the probability distribution of Spearman's rank correlation coefficient under random permutation of all genotypes we shall reject at level α the hypothesis of no linkage if the probability to be less than the observed value is less than α. This is a unilateral test of the hypothesis of no relation against the hypothesis of a spatial genetic structure.

r was computed each time after 2,600 independent random permutations (the 104 individual genotypes are permuted each time, not the pairs). We found only 54 permutations which gave a correlation coefficient less than the observed one. Thus, the estimated probability, under the hypothesis of no link, that r be less than the observed value, is 0.02.

Thus, with an α risk 0.05, a link exists between genetical and geographical closeness.

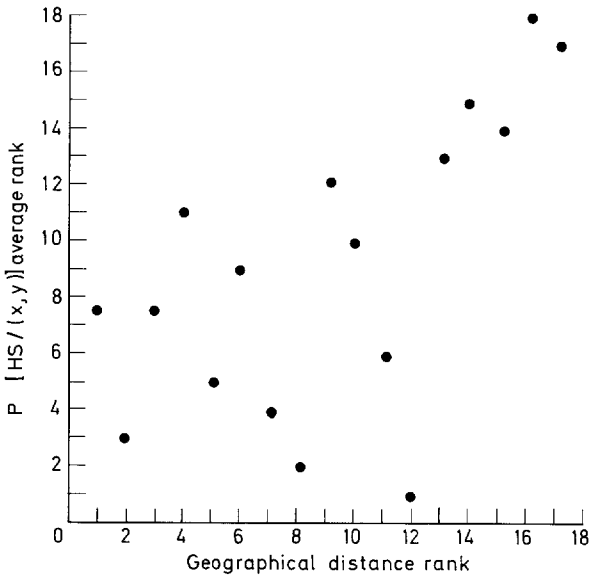


Fig. 6. $P\{HS/(x, y)\}$ average rank with respect to the 50 m distance classes

Remarks

1. $$P\{HS/(x, y)\} = \frac{P\{(x, y)/HS\} \times P_{HS}}{\sum_{i=1}^{3} p\{(x, y)/Ai\} \times P(Ai)}$$

let $P_{HS}=\psi$; if we consider that $P_{FS}=0$, then $P_{NR}=1-\psi$, and

$$P\{HS/(x, y)\} = \frac{P\{(x, y)/HS\}}{P\{(x, y)/HS\} + P\{(x, y)/NR\} \cdot \frac{1-\psi}{\psi}}$$

the ranks defined on all tree pairs by $P\{HS/(x, y)\}$ are the same as those defined by $P\{(x, y)/HS\}/P\{(x, y)/HS\} + P\{(x, y)/NR\}$. Consequently the ranks do not depend on the estimation of P_{Ai}, if we admit that $P_{FS}=0$.

2. The distribution of r under tree genotype permutation has nothing to do with the distribution of a Spearman's rank correlation coefficient on a sample of independant pairs. Over all 2,600 permutations, we found a mean value: +0.039 (with variance = 0.00065).

Conclusion

The genetic proximity was defined without any reference to tree localisation. Our test enables us to reject the hypothesis that, conditional to the observed genotypes and tree localisation, the way in which the geno-

types are attributed to the sampled trees is purely random: trees which are close are more alike than distant trees. In spite of the complex way in which this proximity was defined, it is a measure of genetic closeness.

If this result is a product of mating in a neighbourhood, it implies that the population doesn't follow the Hardy and Weinberg's equilibrium. However, testing directly this last hypothesis does not seem the best method (although it could be improved by using all loci simultaneously, but the hypothesis of loci independence would be in this case open to criticism).

Have we answered the questions expressed at the beginning of this study? We can not give an expression of the kinship estimator precision which would not allow us to use this method as a test of Hardy and Weinberg's equilibrium in the parental population.

Nothing is known about the size of these consanguinity circles: in examining Fig. 6, let us suspect a diameter about 600 m which gives a radius of the same order of magnitude as the pollen dispersion reported in the literature.

Several problems belong to the present methods of gene marker techniques and prospection:

- only eleven loci were available which should have been sufficient. However, in common with population genetic studies using such gene markers, the frequency of the allele has a higher frequency than 0.5, and usually only 2 or 3 alleles exist in the population. This masks the genetic relationship as all the individuals have nearly the same allelic composition. Increased precision would require more loci, or the use of other techniques like 2D electrophoresis.
- As we have seen, the help of literature data must lead to a better representation of the three different type of genetic relationships by increasing the low and high distance pairs. New prospecting methods have to be found, considering at the same time the improvement in the kinship knowledge but also, the limit of seedlots which can be reasonably collected.

Nevertheless, this method has a more general interest, for the natural population studies, as it allows, with more efficiency, a synthetic use of information from electrophoretic markers.

References

Antlfinger AE (1982) Genetic neighbourhood structure of the salt marsh composite *Borrichia frutescens.* J Hered 73: 128–132

Coles JF, Fowler DP (1976) Inbreeding in neighboring trees in two white spruce populations. Silvae Genet 25:29–34

Ellstrand NC, Foster KW (1983) Impact of population structure on the apparent outcrossing rate of grain sorghum (*Sorghum bicolor*). Theor Appl Genet 66:323–327

Hamrick JL, Linhart YB, Mitton JB (1979) Relationship between life history characters and electrophoretically detectable genetic variation in plants. Ann Rev Ecol Syst 1979:173–200

Handel SN (1983) Contrasting gene flow patterns and genetic subdivision in adjacent population of *Cucumis sativus* (Cucurbitaceae). Evolution 37:760–771

Kester HW, Levin DA (1968) Neighbourhood size in *Lithospernum caraliniense.* Genetics 60:577–587

Linhart YB, Mitton JB, Sturgeon KB, Davis ML (1981) Genetic variation in space and time in a population of *Ponderosa pine.* Heredity 46:407–426

Lundkvist K (1978) Allozymes in population genetic studies of Norway Spruce (*Picea abies*). These Akademisk Avhandling

Lundkvist K (1979) Allozyme frequency distributions in four Swedish populations of Norway Spruce (*Picea abies* K). 1. Estimation of genetic variation within and among populations, genetic linkage and a mating system parameter. Hereditas 90:127–143

Müller G (1976) Einschätzung genetischer Verwandschaftsverhältnisse anhand der Pollen- und Samenverbreitung bei Fichte (*Picea abies* (L) Kart.) und Kiefer (*Pinus silvestris*). Dissertation, University Göttingen

Müller G (1977) Cross-fertilization in conifer stand inferred from enzyme gene markers in seeds. Silvae Genet 26: 223–226

Müller-Starck G, Zhiehe M, Hattemer HH (1983). Reproductive systems in conifer seed orchards. 2. Reproductive selection monitored at an LAP gene locus in *Pinus silvestris* L. Theor Appl Genet 65:309–316

Park YG (1972) Gene flow in natural forest of *Cryptomeria japonica* by means of isozyme polymorphisms. Res Rep Inst Forest Genet, Suwon, Korea

Rudin D, Eriksson G, Rasmuson M (1977) Inbreeding in a seed tree stand of *Pinus silvestris* L. in Northern Sweden. A study by the aid of isozyme technique. Inst Skogsgenetik, Res notes 25

Sakai KI, Park YG (1971) Genetic studies in natural populations of forest trees. 3. Genetic differentiation within a forest of *Cryptomeria japonica.* Theor Appl Genet 41: 13–17

Schaal BA (1974) Isolation by distance in *Liatris cylindracea.* Nature 252:703

Schaal BA (1975) Population structure and local differentiation in *Liatris cylindracea.* Am Nat 109:511–528

Schaal BA (1980) Measurement of gene flow in *Lupinus texensis.* Nature 284:450–451

Shaw DV, Allard RW (1979) Analysis of mating system parameters and population structure in Douglas-fir using single locus and multilocus methods. In: Smith AB (ed) Proc Symp Isozymes North Am Forest Trees Forest Insects. Berkeley Calif, pp 18–22

Schmidt H (1970) Versuche über die Pollenverbreitung in einem Kiefernbestand. Dissertation, Forstl Fak, Universität Göttingen

Schmidt-Vogt H (1978) Genetics of *Picea abies* (L) Karst. In: Annales forestales. Acad Scien Art Slavorum Meridionalium, pp 146–182

Tigerstedt PMA (1973) Studies on isozyme variation in marginal and central populations of *Picea abies.* Hereditas 75:47–60

Ziehe M (1982) Sexually asymetric fertility selection and partial self-fertilization. 1. Population genetic impacts on the zygotic genotypic structure. Silva Fenn 16:94–98