Two-generation analysis of pollen flow across a landscape. III. Impact of adult population structure

FRÉDÉRIC AUSTERLITZ[1,2]* AND PETER E. SMOUSE[1]

[1] *Department of Ecology, Evolution & Natural Resources, Cook College, Rutgers University, New Brunswick, NJ 08901-8551, USA*
[2] *Laboratoire de Génétique et d'Amélioration des Arbres Forestiers, INRA, Pierroton, 33611 Gazinet cedex, France*

Summary

The rate and distance of instantaneous pollen flow in a population are parameters of considerable current interest for plant population geneticists and conservation biologists. We have recently developed an estimator (Φ_{ft}) of differentiation between the inferred pollen clouds that fertilize several females, sampled within a single population. We have shown that there is a simple relation between Φ_{ft} and the average pollen dispersal distance (δ) for the case of a population with no geographic structure. Though forest trees usually show considerable pollen flow, assuming an absence of spatially distributed genetic structure is not always wise. Here, we develop analytical theory for the relation between Φ_{ft} and δ, for the case where the probability of Identity by Descent (IBD) for two individuals decreases with the physical distance between them. This analytical theory allows us to provide an effective method for estimating pollen dispersal distance in a population with adult genetic structure. Using real examples, we show that estimation errors can be large if genetic structure is not taken into account, so it is wise to evaluate adult genetic structure simultaneously with estimation of Φ_{ft} for the pollen clouds. We show that the results are only moderately affected by changes in the decay function, a result of some importance since no completely established theory is available for this function.

1. Introduction

Estimating the distance of real-time pollen dispersal is important, both in predicting evolutionary dynamics of a regional plant population and in designing conservation strategies for anthropogenically fragmented species. With these needs in mind, Smouse *et al.* (2001) devised an estimator of the mean pollen dispersal distance (δ), extracted from an analysis of variance of the pollen clouds that fertilizes several females, spaced out across the landscape. The technique uses only the genotypes of the mothers and those of progeny (seedlings) collected from them. Male gametic genotypes within each of the pollen clouds are estimated from these genotypes using parentage analysis methods. Using an Analysis of Molecular Variance (AMOVA; Excoffier *et al.*, 1992), we estimate the intraclass correlation coefficient, Φ_{ft}, representing the proportion of the pollen cloud variance attributable to differences among the pollen clouds of different mothers. Thus, Φ_{ft} is analogous to Wright's (1951) F_{st} coefficient measuring differentiation among adult populations.

More recently (Austerlitz & Smouse, 2001), we dealt with the properties of that estimator for the case of an adult population within which there was no spatially distributed genetic structure. We showed that, as long as the average distance between mothers ($\bar{x}_1$) was large relative to the average distance of pollen dispersal (δ), Φ_{ft} was inversely proportional to δ^2. The approximation was reasonable for $\bar{x}_1 > 3\delta$ and very close for $\bar{x}_1 > 5\delta$. The exact relation between Φ_{ft} and δ also depends on the density (d) of trees in the population, but is not strongly affected by the shape of the dispersal function. Results were gratifyingly similar when either the bivariate normal or exponential distributions were used for pollen dispersal.

* Corresponding author. Laboratoire de Génétique et d'Amélioration des Arbres Forestiers, INRA – Domaine de l'Hermitage, B.P. 45, Pierroton, F-33611 Gazinet cedex, France. Tel: +33 5 57 97 90 83. Fax: +33 5 57 97 90 88. e-mail: austerli@pierroton.inra.fr

For that initial theoretical study, we assumed that adults were non-inbred and that they were unrelated, or more precisely that near neighbours were no more related than the average pair of adults drawn at random from the population. Since little genetic structure, measured either as F_{IS} or F_{ST}, or as decay of genetic relatedness with distance, has been shown in most temperate forest trees (e.g. Epperson & Allard, 1989; Knowles *et al.*, 1992; Berg & Hamrick, 1995; Leonardi & Menozzi, 1996; Leonardi *et al.*, 1996), that is a reasonable first-approximation assumption for many studies. There are, however, cases where significant genetic structure has been shown within populations. Bacilieri *et al.* (1994) estimated that $F_{IS} \sim 0.2$ within a single population of sessile oak (*Quercus petraea*), and also demonstrated declining autocorrelation of allelic frequencies over a distance of < 100 m; similar values of F_{IS} have been observed in other populations of *Q. petraea* (Le Corre *et al.*, 1997). Similarly, *Atherosperma moschatum*, an Australian temperate rainforest tree, showed a high level of inbreeding ($F_{IS} \sim 0.6$) and a decrease in autocorrelation with distance, but on a much shorter spatial scale < 15 m (Shapcott, 1994, 1995). Ueno *et al.* (2000) demonstrated micro-spatial structure within a 4 ha stand of *Camellia japonica*. Spatial autocorrelation was significant, though quite small, on a scale of < 50 m.

While spatially distributed genetic structure is not universal, departures from panmixia are a reality in at least some tree populations. If neighbouring individuals are more closely related than distant individuals, and if the pollen cloud that fertilizes a particular female is drawn mostly from nearby individuals, we should expect some impact on the relationship between Φ_{ft} and pollen dispersal distance, δ. The object of this paper is to explore the impact of inbreeding, and of declining probability of Identity by Descent (IBD) with distance, on the expected value of Φ_{ft} and, as a consequence, our estimate of average pollen dispersal distance. We provide an analytical solution for the case of a population that is subject to inbreeding, but for which there is no spatial structure, and also for the case in which the probability of IBD between two individuals decreases exponentially with the square of their distance. We also show numerically that results are similar for more realistic (but mathematically less tractable) decay functions.

2. The model

(i) *General context*

Assume first that we have an infinite population, with individuals randomly distributed across the landscape, at a density (d) per squared unit of distance. All individuals are monoecious and self-fertile, but do not practise self-reproduction more than would be expected at random, i.e. the probability of selfing is merely the probability of drawing a male gamete at distance $z = 0$. We assume that the probability distribution of pollen dispersal is a bivariate normal distribution, with parameter σ:

$$p(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x^2+y^2)}{2\sigma^2}}, \tag{1}$$

for which the average dispersal distance is (Austerlitz & Smouse, 2001)

$$\delta = \sigma \sqrt{\frac{\pi}{2}}. \tag{2}$$

We consider a sample of mothers drawn from among these individuals, situated (on average) at a distance $\bar{x}_1$ one from another. Mothers are selected on the basis of spatial position but not on the basis of genotype. The difference between this and our previous study (Austerlitz & Smouse, 2001) is that here we assume genetic structure among the adults, spatially distributed across the landscape, as described below.

(ii) *Inbreeding among the adults*

First, assume that we have a certain level of departure from panmixia (F) within the population, i.e. that the probability (f_i) of IBD for the two alleles of a single individual is higher than the probability (f_p) of IBD for two genes, chosen at random from the population. Or, conversely, we can say that the diversity (h_i) within individuals is lower than that (h_p) within the population as a whole. In general, $h_x = 1 - f_x$ is the relationship between IBD and diversity. Wright's F-statistic for the population is classically defined as

$$F = 1 - \frac{h_i}{h_p}, \tag{3}$$

which can also be written as

$$h_i = h_p(1 - F). \tag{4}$$

To compute Φ_{ft}, we use the relation given by equation (26) in Austerlitz & Smouse (2001):

$$\Phi_{ft} = \frac{f_f - f_t}{1 - f_t} = 1 - \frac{h_f}{h_t}, \tag{5}$$

where h_f is the probability that two genes within the pollen cloud of a single female are different, and h_t is the same probability for two genes sampled within the pollen cloud of all mothers. To compute h_f, equation (21) of Austerlitz & Smouse (2001) is still valid:
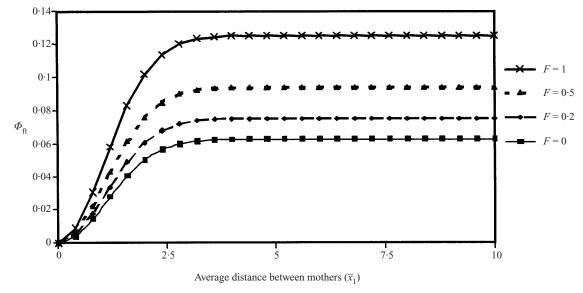
$$h_f = \frac{Q_0}{2} h_i + (1 - Q_0) h_p, \tag{6}$$

Fig. 1. Impact of the adult inbreeding coefficient F on Φ_{ft}. The intraclass correlation, Φ_{ft}, obtained with (14), is plotted against the average distance between mothers $(\bar{x}_1)$, for average pollen dispersal distance of $\delta = 1$ and adult density $d = 1$, for various values of F.

where Q_0 is the probability of two male gametes within the pollen cloud of a single female being drawn from the same father. Substituting (4) into (6) yields:

$$h_f = h_p \left[1 - \frac{Q_0}{2}(1+F) \right]. \tag{7}$$

h_t is given by equation (22) of Austerlitz & Smouse (2001):

$$h_t = \frac{\overline{Q}}{2}h_i + (1-\overline{Q})h_p, \tag{8}$$

where $\overline{Q}$ is the average probability of two male gametes, from the pollen clouds of two different females, being drawn from the same father. We neglect here the possibility that two gametes sampled from the total pollen cloud come, in fact, from the pollen cloud of the same female. That small probability is the inverse of the total number of females receiving pollen. Substituting (4) into (8) yields:

$$h_t = h_p \left[1 - \frac{\overline{Q}}{2}(1+F) \right]. \tag{9}$$

Substituting (7) and (9) into (5) yields:

$$\Phi_{ft} = \frac{(Q_0 - \overline{Q})(1+F)}{2-(1+F)\overline{Q}}. \tag{10}$$

As in our previous paper (Austerlitz & Smouse, 2001), Φ_{ft} does not depend on the actual level of h_p, which cancels out of the argument. When the mothers are far enough apart (say, $\bar{x}_1 > 5\delta$), then, $\overline{Q} = Q(\bar{x}_1) = 0$ and:

$$\Phi_{ft} = \frac{Q_0(1+F)}{2}. \tag{11}$$

At maximum $(F = 1)$, Φ_{ft} is twice its value with no inbreeding $(F = 0)$. Since we assume here that the pollen dispersal distribution is bivariate normal, we can use the analytical solutions derived previously for Q_0 and $\overline{Q}$ (equations (15) and (18) in Austerlitz & Smouse, 2001):

$$Q_0 = \frac{1}{4\pi\sigma^2 d} = \frac{1}{8\delta^2 d}, \tag{12}$$

and

$$\overline{Q} = Q_0 e^{-\bar{x}^2_1/4\sigma^2}. \tag{13}$$
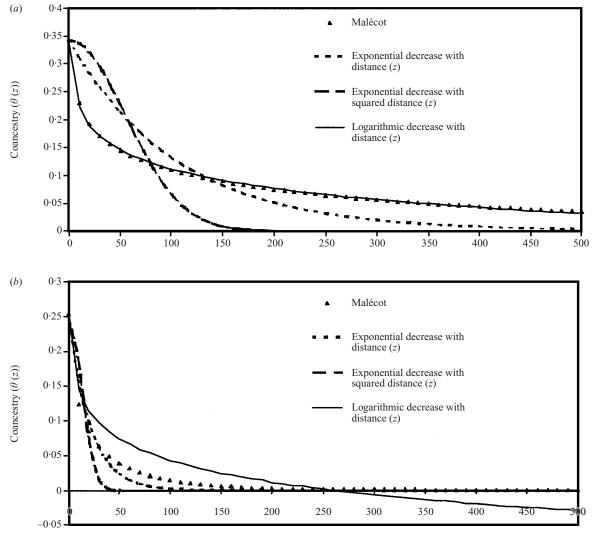
Substituting (12) and (13) into (10) yields:

$$\Phi_{ft} = \frac{[1-\exp(-\bar{x}_1^2/4\sigma^2)](1+F)}{8\pi\sigma^2 d-(1+F)\exp(-\bar{x}_1^2/4\sigma^2)}$$
$$= \frac{[1-\exp(-\pi\bar{x}_1^2/8\delta^2)](1+F)}{16\delta^2 d-(1+F)\exp(-\pi\bar{x}_1^2/8\delta^2)}, \tag{14}$$

which can be rewritten in terms of the average distance of pollen flow, using the relationship between σ and δ, established in (2). If there is no inbreeding $(F = 0)$, we recover the formula from Austerlitz & Smouse (2001):

$$\Phi_{ft} = \frac{Q_0 - \overline{Q}}{2 - \overline{Q}}. \tag{15}$$

Fig. 1 portrays the impact of distance between mothers on Φ_{ft} for various average distances between mothers $(\bar{x}_1)$ and various values of the inbreeding coefficient of the adults, F. Clearly, Φ_{ft} increases with F for any given value of $\bar{x}_1$, but the rate of convergence of Φ_{ft}

Fig. 2. Decay of coancestry $\theta(z)$ with distance z, calculated using Malécot's (1969) formula (text equation (16)): (*a*) with $k = 10^{-6}$ and $\sigma_m = 1$; or (*b*) with $k = 10^{-4}$ and $\sigma_m = 1$; compared with the best least-square fit, obtained for an exponential decrease of $\theta(z)$ with z or z^2, or a logarithmic decay, $\theta_0 - b \log(1 + z/\sigma)$.

towards its asymptote remains the same for all values of F.

(iii) *Impact of relatedness among fathers*

Now, assume instead that we have genetic population structure among the adults themselves, i.e. that the probability of IBD for two genes declines with the distance between them. We define $f_d(z)$ as the probability of IBD for two genes drawn from adult individuals that are a distance z apart. As pointed out by Epperson *et al.* (1999), for a continuous population, where the probability of migration decays with distance, there does not seem to be any completely accepted theory on the decline of $f_d(z)$ with distance. In his models, Wright (1943, 1946, 1969) gives general results on F_{IS} and F_{ST} for the whole population, but does not give anything on the decay of probability of IBD (or any related quantity) with distance.

Malécot (1969) did provide a formula for coancestry as a declining function of distance in a continuous population. Malécot denoted that quantity as $\phi(z)$, but we have changed the notation to $\theta(z)$, to avoid confusion with our Φ_{ft}. For the case where gametes migrate following a zero-centred bivariate normal distribution, with standard deviation σ_m (Malécot 1969, equation 3.3.4, p. 72), for a biallelic locus (Aa), $\theta(z)$ follows:

$$\theta(z) = \theta_0 \frac{\sum_{p=1}^{\infty} (1-k)^{2p} \left(\dfrac{1}{4\pi p \sigma^2_m} \right) e^{-z^2/4p\sigma^2_m}}{\sum_{p=1}^{\infty} (1-k)^{2p} \left(\dfrac{1}{4\pi p \sigma^2_m} \right)}, \tag{16}$$

where $k = (u+v)$ is the mutation pressure (u and v are the probabilities of mutation from 'a' to 'A' and from 'A' to 'a', respectively), and θ_0 is the coancestry at distance zero, calculated as:

$$\theta_0 = \frac{1}{1 - \dfrac{8\pi\sigma^2 d}{\log(2k-k^2)}}. \tag{17}$$

Felsenstein (1975) has pointed out that some of Malécot's hypotheses are inconsistent, if (16) was intended to describe a continuous population. In the dialogue that followed, it was argued that Malecot's model was not describing a continuous population but rather the limiting case of a discrete subdivision model (Lalouel, 1977; Morton, 1977; Felsenstein, 1979; Lalouel, 1979). Suffice it that while the closed-form theory is still not settled, the general idea is that $f_d(z)$ should decrease from the value f_i, the probability of IBD for the two alleles from a single individual (at distance $z = 0$), to f_p, i.e. the probability of IBD in the total population for distances that are sufficiently large, formally as $z \to \infty$. The probability of IBD and the autocorrelation approach have been shown to be formally equivalent, and their decay rates are identical (Hardy & Vekemans, 1999).

Since Malécot's function is computationally too intensive to be used in the quadruple integrals we use below, we need an approximation. Three possibilities, for example, are to assume that the probability of IBD, $f_d(z)$, decreases exponentially with distance:

$$f_d(z) = f_p + (f_i - f_p)\,e^{-\alpha_e z}, \tag{18}$$

exponentially with squared distance:

$$f_d(z) = f_p + (f_i - f_p)\,e^{-\alpha_n z^2}, \tag{19}$$

or logarithmically with distance:

$$\begin{cases} f_d(z) = f_i - (f_i - f_p)\log(1 + \alpha_l z) & \text{for } \alpha_l z < 20 \\ f_d(z) = f_i & \text{for } \alpha_l z > 20 \end{cases} \tag{20}$$

where α_e, α_n and α_l are the decay parameters of (18), (19) and (20), respectively. The last function has to be truncated; otherwise, $f_d(z) \to -\infty$ as $z \to \infty$. The function given in (20) is closer to Malécot's result for low values of the mutation pressure (k) (see Fig. 2a), and (18) is closer for larger values of k (see Fig. 2b), but neither yields any tractable analytical results for the derivations that follow. Conveniently, the function given in (19) does lead to clear analytical results. In the interest of tractability, we will deal here first with that function, afterwards providing some numerical results for (18) and (20). The model with inbreeding ($F > 0$, but no spatial structure among adults) is the limiting case of (18), with $\alpha_e \to \infty$, or (19), with $\alpha_n \to \infty$; in either case, IBD drops instantaneously from f_i, at $z = 0$, to f_p for any $z > 0$.

Concentrating now just on the case where the probability of IBD declines exponentially with the square of the distance, we note that (19) can also be expressed in terms of diversity, h_d, where $h_d = 1 - f_d$:

$$h_d(z) = h_p - (h_p - h_i)\,e^{-\alpha_n z^2}. \tag{21}$$

As before, to compute Φ_{ft} we first compute the probability h_f that two genes within the pollen cloud of a single mother are different, and the same probability h_t for two genes sampled randomly from the pollen cloud of all mothers. To compute h_f, we can still use the fact that two genes in the pollen cloud of a single mother have probability Q_0 of being derived from the same father and $(1 - Q_0)$ of being from two different fathers. If they come from the same father, they will be different with probability $h_i/2$, but if they come from two different fathers, they will be different with probability h_n, the integration of $h_d(z)$ for all pairs of potential fathers. This yields:

$$h_f = \frac{Q_0}{2}h_i + (1 - Q_0)\,h_n. \tag{22}$$

To compute h_n, we must consider all potential pairs of fathers that might have contributed pollen to a focal female. We denote by (x, y) and (x', y') the physical coordinates of these two fathers. Both are drawn from the distribution $p(x, y)$. The distance between the two males will be $\sqrt{(x-x')^2 + (y-y')^2}$, and their probability of being different will be $h_d(\sqrt{(x-x')^2 + (y-y')^2})$. Thus, the average probability of being different (h_n) for two genes coming from different fathers, integrated over the pollen cloud of a given mother is:

$$h_n = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h_d(\sqrt{(x-x')^2 + (y-y')^2})$$
$$p(x, y)\,p(x', y')\,dx\,dy\,dx'\,dy' \tag{23}$$

which, upon integration, yields:

$$h_n = h_p - (h_p - h_i)\beta_0, \tag{24}$$

where

$$\beta_0 = \frac{1}{1 + 4\alpha_n \sigma^2}. \tag{25}$$

Substituting (24) into (22), we obtain:

$$h_f = \frac{Q_0}{2}h_i + (1 - Q_0)(h_p - (h_p - h_i)\beta_0). \tag{26}$$

To compute h_t, we can still use the fact that genes sampled from the pollen clouds of two different mothers have a probability $\overline{Q}$ of being from the same father and $(1 - \overline{Q})$ of being from two different fathers, where $\overline{Q} = Q(\overline{x}_1)$, and $\overline{x}_1$ is the average distance

between sampled mothers in the population. If the gametes are from the same father, they will be different, as before, with probability $h_i/2$, but if they come from two different fathers, they will be different with probability h_{n2}, the integration of $h_d(z)$ for all pairs of potential fathers. This yields:

$$h_t = \frac{\overline{Q}}{2} h_i + (1 - \overline{Q}) h_{n2}. \tag{27}$$

To compute h_{n2}, we use a system of Cartesian coordinates, where the first female is at position $(0, 0)$, and the second at position $(x_1, 0)$, as in our previous paper (Austerlitz & Smouse, 2001). For the first female, the fertilizing male is at position (x, y), drawn from the distribution $p(x, y)$. For the second female, the fertilizing male is at position (x', y'), drawn from the distribution $p_2(x, y) = p(x - x_1, y)$. As above, the distance between the two males is $\sqrt{(x-x')^2+(y-y')^2}$, and their probability of being different is $h_d(\sqrt{(x-x')^2+(y-y')^2})$. Integrating over all possible pairs of fathers, the average probability of being different (h_{n2}) for genes coming from different fathers, and sampled from the pollen clouds of different mothers, is:

$$h_{n2} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h_d(\sqrt{(x-x')^2+(y-y')^2})$$
$$\times p(x, y)\, p(x' - x_1, y')\,dxdydx'dy' \tag{28}$$

which yields:

$$h_{n2} = h_p - (h_p - h_i)\beta, \tag{29}$$

where

$$\beta = \beta_0 e^{-\alpha_n \beta_0 \overline{x}^2_1}. \tag{30}$$

Substituting (29) into (27) yields:

$$h_t = \frac{h_i \overline{Q}}{2} + (1 - \overline{Q})[h_p - (h_p - h_i)\beta], \tag{31}$$

Substituting (26) and (31) into (5) and replacing h_i with its value in (4) yields:

$$\Phi_{ft} = 1 - \frac{(1-F)\dfrac{Q_0}{2} + (1-Q_0)(1-F\beta_0)}{(1-F)\dfrac{\overline{Q}}{2} + (1-\overline{Q})(1-F\beta)}. \tag{32}$$

where Q_0, $\overline{Q}$, β_0 and β are given by (12), (13), (25) and (30), respectively.

Fig. 3 illustrates the impact of α_n on Φ_{ft} for various values of $\overline{x}_1$. Φ_{ft} first increases, to reach a peak, and then decreases again. When $\alpha_n \to \infty$, the limit of Φ_{ft} is the value given in (14), since, as we have stated before, it corresponds to the limiting case. On the other hand,
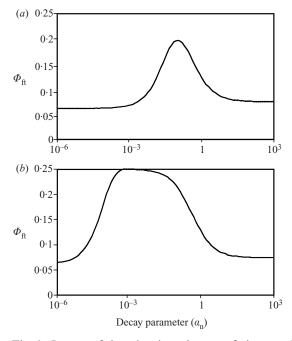

Fig. 3. Impact of the relatedness between fathers on Φ_{ft}, obtained with (32) and plotted against the decay parameter (α_n) of the probability function relating IBD to distance. Results are given for the normal distribution of pollen dispersal, with average pollen dispersal distance of $\delta = 1$, adult density of $d = 1$, departure from panmixia at long distance, $F = 0.2$, and an average distance between mothers of: (a) $\overline{x}_1 = 5$, or (b) $\overline{x}_1 = 100$.

when $\alpha_n \to 0$, Φ_{ft} converges to the case where there is no inbreeding, i.e. (15) with $F = 0$. This can be explained by the fact that, as $\alpha_n \to 0$, coancestry decays very slowly with distance, so even two individuals that are very far apart are almost as close genetically as two proximal individuals; we end up here with the case of a genetically homogeneous population. Thus, it is only for intermediate values of α_n, when the decrease in the probability of IBD with distance can be detected within the spatial scale of the population under study, that two mothers are likely to sample genetically differentiated pollen pools.

Fig. 4 illustrates the dependence of Φ_{ft} on $\overline{x}_1$ (the average distance between females) for different values of α_n. Φ_{ft} increases with $\overline{x}_1$ towards an equilibrium value. The rate of convergence to this equilibrium value decreases with α_n. In practice, the approximation given in Austerlitz & Smouse (2001) that the equilibrium has been reached for $\overline{x}_1 > 5\delta$ remains true as long as $\alpha_n \geqslant 1$. The equilibrium value is obtained by setting $\overline{x}_1 \to \infty$ in (32):

$$\Phi_{ft} = 1 - [(1-F)\frac{Q_0}{2} + (1-Q_0)(1-F\beta_0)]. \tag{33}$$

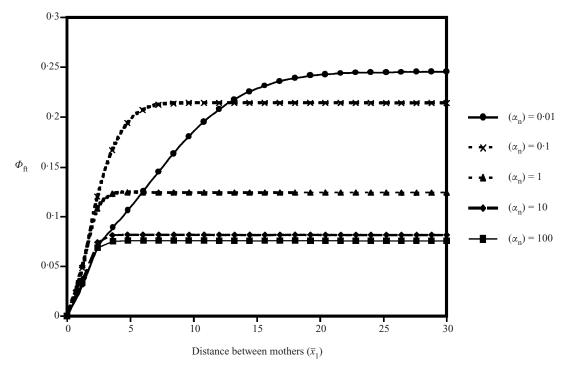The equilibrium value always decreases with α_n, but rapidly converges to an upper limit when α_n becomes

Fig. 4. Impact of the relatedness between fathers on Φ_{ft}, obtained with (32) and plotted against the average distance between mothers $(\overline{x}_1)$, for average pollen dispersal distance of $\delta = 1$, adult density $d = 1$, departure from panmixia at long distance, $F = 0.2$, and various values of the decay parameter (α_n) for the decline of probability of IBD with distance.

very small, derivable by setting $\alpha_n = 0$, tantamount to setting $\beta_0 = 1$ in (33):

$$\Phi_{ft} = \frac{Q_0}{2} + \left(1 - \frac{Q_0}{2}\right) F. \quad (34)$$

This limit might seem inconsistent with that provided above, when $\alpha_n \to 0$, as $\overline{x}_1$ remains finite, but the result comes from the presence of the product $\alpha_n \overline{x}^2_1$ in (30). There is no unique limit when $\alpha_n \to 0$, and $\overline{x}_1 \to \infty$, simultaneously, but that corresponds to rather unrealistic situations.

(iv) *Other decay functions for the probability of IBD*

The results above were based on an approximation of Malécot's (1969) equation, reproduced here as (16), by an exponential decay of IBD with *squared* distance, as given in (19), the motivation for which was (at least in part) tractability. We pointed out, at the outset, that an exponential decay with distance, as given in (18), or the logarithmic decay, as given in (20), were better approximations, but hard to deal with analytically. How much difference does it make? For these cases, the integrals given in (23) and (28) have to be computed numerically. We can reduce h_n to triple integral form with polar coordinates:

$$h_n = \int_0^{2\pi} \int_0^{2\pi} \int_0^{\infty} \int_0^{\infty} h_d$$
$$\times \left(\sqrt{(z \cos(\theta) - z' \cos(\theta'))^2 + (z \sin(\theta) - z' \sin(\theta'))^2}\right)$$
$$\times p_p(z) p_p(z') zz' \, dz \, dz' \, d\theta' \, d\theta,$$

which simplifies to

$$h_n = \int_0^{2\pi} \int_0^{2\pi} \int_0^{\infty} \int_0^{\infty} h_d$$
$$\times \left(\sqrt{z^2 + z'^2 - 2zz' \cos(\theta - \theta')}\right)$$
$$\times p(z) p(z') zz' \, dz \, dz' \, d\theta \, d\theta',$$

and using the transformation $\phi = \theta - \theta'$, we can write

$$h_n = 2\pi \int_0^{2\pi} \int_0^{\infty} \int_0^{\infty} h_d$$
$$\times \left(\sqrt{z^2 + z'^2 - 2zz' \cos(\phi)}\right)$$
$$\times p_p(z) p_p(z') zz' \, dz \, dz' \, d\phi. \quad (35)$$

Thus, to compute Φ_{ft}, we must first compute h_n and h_{n2} numerically, using (35) and (28), respectively, and then compute h_f and h_t, using (22) and (26), respectively, which leads to the value of Φ_{ft}, using (5).

To make a valid comparison with the previous section, we chose matching values of α_e, α_l and α_n (the decay rates of (18), (20) and (19), respectively). For a
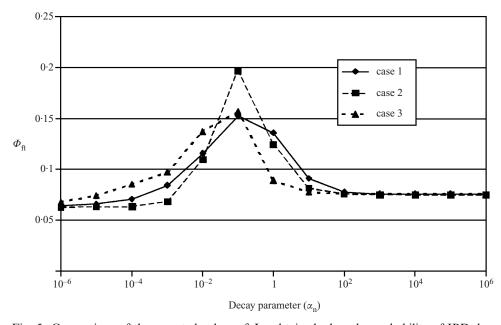
Fig. 5. Comparison of the expected values of Φ_{ft} obtained when the probability of IBD decreases exponentially with distance, as in (18) (case 1), or exponentially with the square of distance, as in (19) (case 2), or logarithmically with distance as in (20) (case 3). In all cases, parameters are average pollen dispersal distance $\delta = 1$, adult density $d = 1$, average distance between mothers $\overline{x}_1 = 5$ and departure from panmixia at long distance, $F = 0.2$. For the exponential decline with squared distance, Φ_{ft} is calculated directly for a given value of α_n, using (32); for exponential and logarithmic decline with distance, it is calculated numerically, using a rate of decay $\alpha_e = \sqrt{\alpha_n \log(2)}$ or $\alpha_1 = \sqrt{\frac{\alpha_n}{\log(2)}}(\sqrt{e-1})$, respectively, so that the decrease in probability of IBD with distance is similar in all cases (see text for details).

given value of α_n, we used values for α_e and α_1 such that the $z_{0.5}$ value was the same. The $z_{0.5}$ value is defined as the distance (z) between individuals for which the probability of IBD is halfway between f_i and the long-distance probability, f_p. That calibration yields $\alpha_e = \sqrt{\alpha_n \log(2)}$ and $\alpha_1 = \sqrt{\frac{\alpha_n}{\log(2)}}(\sqrt{e-1})$, and given that standardization, the relationship between Φ_{ft} and average distance between mothers $(\overline{x}_1)$ is qualitatively the same for all three decay functions, and even the quantitative differences are small (Fig. 5). It basically makes little difference which decay function is used, although the exponential decay with squared distance shows a sharper rise in the middle of the α_n range than do either the exponential or logarithmic functions. Exponential decay with squared distance, which is more tractable, should be sufficient for future analytical work.

3. Discussion

Inbreeding somewhat increases the level of differentiation between pollen clouds (Φ_{ft}). Genetic diversity within the pollination neighbourhood of each female is reduced, due to the consanguinity of the males, which increases the probability of finding a given allele more frequently in a given part of the landscape. Since different alleles will increase in frequency in

different neighbourhoods, overall differentiation will increase. This increase cannot be higher than a factor of 2, since when males are completely homozygous, it is as if they were haploid; effective neighbourhood size is half that expected in the case of no inbreeding.

A slow decay of the probability of IBD with distance has a much stronger impact when this decay can be detected on the spatial scale of the population under consideration. The discrepancy between the two Φ_{ft} values obtained by ignoring or taking into account the genetic structure can be high, even for reasonable values of the parameters. When relatedness is high within the pollination neighbourhood of a given female, the effective neighbourhood size will be very low. Conversely, differentiation will increase as females are sampled at larger distances from each other, since the probability that different alleles are fixed within different neighbourhoods will increase with the distance between them.

As originally pointed out by Felsenstein (1975, 1979), and more recently by Epperson *et al.* (1999), there is still no accepted theory for the decay function of the probability of IBD with distance within a continuous population; only simulation results are available (Rohlf & Schnell, 1971; Doligez *et al.*, 1998; Epperson *et al.*, 1999). We were forced to make an arbitrary choice of decay functions, and it is reassuring that a change in that function makes little difference,

although the need for proper (closed form) theory remains.

The theory tells us what could happen, but what does happen? For the many species showing 'no genetic structure' at the local scale ($F \approx 0$ and either $\alpha \to 0$ or $\alpha \to \infty$), we need not be concerned with the issue. For the few species that do show 'local genetic structure', how big is the impact of F_{IS} on Φ_{ft}? In *Quercus petraea*, for example, Bacilieri *et al.* (1994), in a study area of 5 ha, showed that genetic auto-correlation decreased slowly with distance and remained significant for up to approximately 100 m. Empirically, that corresponds to a decay parameter of $\alpha_n \sim 0.0005$, assuming that the probability of IBD decreases exponentially with squared distance, as in (19). The density in this plot is $d = 0.0059$ trees/m². The departure from panmixia is $F = 0.22$. Had we sampled mothers at an average pairwise distance of $\bar{x}_1 = 50$ m, and assuming an average pollen dispersal distance of $\delta = 10$ m, i.e. an effective pollen pool size $N_{ep} = 8\delta^2 d = 4.72$ effective males, or a pollen neighbourhood area $A_{ep} = 8\delta^2 = 800$ m² (Austerlitz & Smouse, 2001; Smouse *et al.*, 2001), the value of Φ_{ft} obtained with (32) would have been $\Phi_{ft} = 0.234$. An experimenter who ignored the adult genetic structure would estimate a value, using (15), of $\hat{\delta} = 6.73$ m ($\hat{N}_{ep} = 2.14$; $\hat{A}_{ep} = 362$ m²), a serious underestimate of the true values. If the real value of δ were 50 m ($N_{ep} = 118$; $A_{ep} = 20,000$ m²), Φ_{ft} obtained with (32) would be 0.016, and the experimenter, again in ignorance of adult genetic structure, would obtain $\hat{\delta} = 23.5$ m ($\hat{N}_{ep} = 26.1$; $\hat{A}_{ep} = 4418$ m²), an even more serious underestimate of the true values. Thus, the expected estimate of pollen dispersal distance (and therefore of neighbourhood size) can be less than the half of the parametric value if adult genetic structure within the population is ignored.

In the case of *Atherosperma moschatum* (Shapcott, 1994, 1995), $d \sim 0.04$ trees/m², on average. The departure from panmixia is $F = 0.598$, and the decay of genetic affinity with distance is sharper; auto-correlation becomes non-significant within about 15 m, which means $\alpha_n \sim 0.02$. Assume again that we had sampled mothers at an average pairwise distance of $\bar{x}_1 = 50$ m, and assume again that the average distance of dispersal was $\delta = 10$ m ($N_{ep} = 32$; $A_{ep} = 800$ m²). The expected estimate of Φ_{ft}, using (32), would have been $\Phi_{ft} = 0.016$. An experimenter who ignored genetic structure among the adults themselves would estimate a value $\hat{\delta} = 3.61$ m ($\hat{N} = 4.17$; $\hat{A}_{ep} = 104$ m²), using (15), only 36 % of the true value for δ, and thus only 13 % for N_{ep} and A_{ep}, because of the relation between these quantities and δ^2.

Thus, if the adult genetic structure is strong, as it is for these striking examples, there can be a rather large error in estimating demographic parameters from Φ_{ft} (the estimation of Φ_{ft} itself is not affected) if adult genetic structure is not taken into account. Because it is relatively easy to assess adult genetic structure in the study area, it is generally a good idea to do so. When genetic structure is detected, it is obviously better to infer δ from Φ_{ft}, using (32), which takes genetic structure into account, rather than from (15), which does not. In either case, numerical estimation can be performed within a few seconds. Thus, it would seem generally advisable to characterize the spatial genetic structure of the adults within the study population. This can be performed by studying either the increase in differentiation (Bacilieri *et al.*, 1994) or the decrease in autocorrelation (Sokal & Oden, 1978; Smouse & Peakall, 1999) with distance, these two approaches having being shown to be equivalent (Hardy & Vekemans, 1999).

We must be mindful, however, that the two cases described above are rather extreme, the first in terms of the distance at which genetic structure can be detected, the second for the level of inbreeding. For most temperate forest tree species, the error in using (15) will be small, since inbreeding is usually low (minimal F_{IS}) and there is minimal genetic affinity out to no more than 15 m (e.g. Epperson & Allard, 1989; Knowles *et al.*, 1992; Berg & Hamrick, 1995; Leonardi & Menozzi, 1996; Leonardi *et al.*, 1996), suggesting very small (or very large) values of α_n in most populations.

One might expect more local adult structure for tropical forest tree species, which are often insect pollinated, but gene flow is often substantial in tropical forest trees (Hamrick & Murawski, 1990; Chase *et al.*, 1996; Nason *et al.*, 1996; Stacy *et al.*, 1996; Loveless *et al.*, 1998). Data are scarce on decay of genetic affinity with distance for these species, but in one case, autocorrelation was detected over a distance of 100 m (Boshier *et al.*, 1995), whereas in another case (Doligez & Joly, 1997), none was found. Hamrick *et al.* (1993) studied three cases. While they found limited structure in two of them, strong structure, over more than 250 m, was found only among the juveniles for the third case, almost disappearing among reproductive adults, and thus not a problem for the estimation of 'pollen structure'.

Several other features will have to be integrated into the model in the future. For instance, selfing could occur more often than expected at random, or self-incompatibility could reduce the effective selfing rate, relative to random pollination at distance $z = 0$, either of which is likely to affect neighbourhood size. Anisotropic pollen flow might be expected in some cases, due to preferential wind flow directions. This will also affect the differentiation in pollen clouds of the females and their relative positions, not just the distances between them, and will have to be taken into account. We have shown here that complicated features can be integrated into the model, and within

certain limits will allow us to determine the features that really matter for the estimation of Φ_{ft}.

References

Austerlitz, F. & Smouse, P. E. (2001). Two-generation analysis of pollen flow across a landscape. II. Relation between Φ_{ft}, pollen dispersal and inter-females distance. *Genetics* **157**, 851–857.

Bacilieri, R., Labbe, T. & Kremer, A. (1994). Intraspecific genetic structure in a mixed population of *Quercus petraea* (Matt.) Leibl. and *Q. robur* L. *Heredity* **73**, 130–141.

Berg, E. E. & Hamrick, J. L. (1995). Fine-scale genetic structure of a turkey oak forest. *Evolution* **49**, 110–120.

Boshier, D. H., Chase, M. R. & Bawa, K. S. (1995). Population genetics of *Cordia alliodora* (Boraginaceae), a neotropical tree. 3. Gene flow, neighborhood, and population substructure. *American Journal of Botany* **82**, 484–490.

Chase, M. R., Moller, C., Kessell, R. & Bawa, K. S. (1996). Distant gene flow in tropical trees. *Nature* **383**, 398–399.

Doligez, A. & Joly, H. I. (1997). Genetic diversity and spatial structure within a natural stand of a tropical forest tree species, *Carapa procera* (Meliaceae), in French Guiana. *Heredity* **79**, 72–82.

Doligez, A., Baril, C. & Joly, H. I. (1998). Fine-scale spatial genetic structure with nonuniform distribution of individuals. *Genetics* **148**, 905–919.

Epperson, B. K. & Allard, R. W. (1989). Spatial autocorrelation analysis of the distribution of genotypes within populations of lodgepole pine. *Genetics* **121**, 369–378.

Epperson, B. K., Huang, Z. & Li., T. Q. (1999). Measures of spatial structure in samples of genotypes for multiallelic loci. *Genetical Research* **73**, 251–261.

Excoffier, L., Smouse, P. E. & Quattro, J. M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**, 479–491.

Felsenstein, J. (1975). A pain in the torus: some difficulties with models of isolation by distance. *American Naturalist* **109**, 359–368.

Felsenstein, J. (1979). Isolation by distance: reply to Lalouel and Morton. *Annals of Human Genetics* **42**, 523–527.

Hamrick, J. L. & Murawski, D. A. (1990). The breeding structure of tropical tree populations. *Plant Species Biology* **5**, 157–166.

Hamrick, J. L., Murawski, D. A. & Nason, J. D. (1993). The influence of seed dispersal mechanisms on the genetic structure of tropical tree populations. *Vegetatio* **107/108**, 281–297.

Hardy, O. J. & Vekemans, X. (1999). Isolation by distance in a continuous population: reconciliation between spatial autocorrelation analysis and population genetics models. *Heredity* **83**, 145–154.

Knowles, P., Perry, D. J. & Foster, H. A. (1992). Spatial genetic structure in two Tamarack [*Larix laricina* (Du Roi) K. Koch] populations with differing establishment histories. *Evolution* **46**, 572–576.

Lalouel, J. M. (1977). The conceptual framework of Malécot's model of isolation by distance. *Annals of Human Genetics* **40**, 355–360.

Lalouel, J. M. (1979). Comment on Felsenstein's reply to Lalouel and Morton. *Annals of Human Genetics* **42**, 529.

Le Corre, V., Dumolin-Lapegue, S. & Kremer, A. (1997). Genetic variation at allozyme and RAPD loci in sessile oak *Quercus petraea* (Matt.) Liebl.: the role of history and geography. *Molecular Ecology* **6**, 519–529.

Leonardi, S. & Menozzi, P. (1996). Spatial structure of genetic variability in natural stands of *Fagus sylvatica* L. (beech) in Italy. *Heredity* **77**, 359–368.

Leonardi, S., Raddi, S. & Borghetti, M. (1996). Spatial autocorrelation of allozyme traits in a Norway spruce (*Picea abies*) population. *Canadian Journal of Forest Research* **26**, 63–71.

Loveless, M. D., Hamrick, J. L. & Foster, R. B. (1998). Population structure and mating system in *Tachigali versicolor*, a monocarpic neotropical tree. *Heredity* **81**, 134–143.

Malécot, G. (1969). *The Mathematics of Heredity*. San Francisco: Freeman.

Morton, N. E. (1977). Isolation by distance in human populations. *Annals of Human Genetics* **40**, 361–365.

Nason, J. D., Herre, E. A. & Hamrick, J. L. (1996). Paternity analysis of the breeding structure of strangler fig populations: evidence for substantial long-distance wasp dispersal. *Journal of Biogeography* **23**, 501–512.

Rohlf, F. J. & Schnell, G. D. (1971). An investigation of the isolation-by-distance model. *American Naturalist* **105**, 295–324.

Shapcott, A. (1994). Genetic and ecological variation in *Atherosperma moschatum* and the implications for conservation of its biodiversity. *Australian Journal of Botany* **42**, 663–686.

Shapcott, A. (1995). The spatial genetic structure in natural populations of the Australian temperate rainforest tree *Atherosperma moschatum* (Labill.) (Monimiaceae). *Heredity* **74**, 28–38.

Smouse, P. E. & Peakall, R. (1999). Spatial autocorrelation analysis of individual multiallele and multilocus genetic structure. *Heredity* **82**, 561–573.

Smouse, P. E., Dyer, R. J., Westfall, R. D. & Sork, V. L. (2001). Two-generation analysis of pollen flow across a landscape. I. Male gamete heterogeneity among females. *Evolution* **55**, 260–271.

Sokal, R. R. & Oden, N. L. (1978). Spatial autocorrelation analysis in biology. 1. Methodology. *Biological Journal of the Linnean Society* **10**, 199–228.

Stacy, E. A., Hamrick, J. L., Nason, J. D., Hubbell, S. P., Foster, R. B. & Condit, R. (1996). Pollen dispersal in low-density populations of three Neotropical tree species. *American Naturalist* **148**, 275–298.

Ueno, S., Tomaru, N., Yoshimaru, H., Manabe, T. & Yamamoto, S. (2000). Genetic structure of *Camellia japonica* L. in an old-growth evergreen forest, Tsushima, Japan. *Molecular Ecology* **9**, 647–656.

Wright, S. (1943). Isolation by distance. *Genetics* **28**, 114–138.

Wright, S. (1946). Isolation by distance under diverse systems of mating. *Genetics* **31**, 39–59.

Wright, S. (1951). The genetical structure of populations. *Annals of Eugenics* **15**, 323–354.

Wright, S. (1969). *Evolution and the Genetics of Populations*. vol. 2, *The Theory of Gene Frequencies*. Chicago: University of Chicago Press.