Using genetic markers to estimate the pollen dispersal curve

FREDERIC AUSTERLITZ,* CHRISTOPHER W. DICK,† CYRIL DUTECH,‡§ ETIENNE K. KLEIN,*¶
SYLVIE ODDOU-MURATORIO,** PETER E. SMOUSE†† and VICTORIA L. SORK‡

*Laboratoire Ecologie, Systématique et Evolution, UMR CNRS 8079, Université Paris-Sud, F-91405 Orsay cedex, France,
†Smithsonian Tropical Research Institute, Unit 0948 APO AA 34002–0948 USA and Biological Dynamics of Forest Fragments Project,
Instituto Nacional de Pesquisas da Amazônia, C.P. 478, Manaus, AM 69011–970, Brazil, ‡Department of Organismic Biology, Ecology
& Evolution, University of California Los Angeles, Box 951786, Los Angeles CA 90095–1786, USA, §INRA-Bordeaux, UMR BioGEco,
Domaine de la Grande Ferrade, BP81, F-33883 Villenave d'Ornon, France, ¶UMR 518 Biométrie et Intelligence Artificielle, INRA/
INAPG/ENGREF, F-75231 Paris Cedex 05, France, **Conservatoire Génétique des Arbres Forestiers, Office National des Forêts,
Campus INRA, F-45160 Ardon, France, ††Department of Ecology, Evolution & Natural Resources, Rutgers University, New
Brunswick, New Jersey 08901–8551, USA

Abstract

**Pollen dispersal is a critical process that shapes genetic diversity in natural populations of
plants. Estimating the pollen dispersal curve can provide insight into the evolutionary
dynamics of populations and is essential background for making predictions about
changes induced by perturbations. Specifically, we would like to know whether the dispersal curve is exponential, thin-tailed (decreasing faster than exponential), or fat-tailed
(decreasing slower than the exponential). In the latter case, rare events of long-distance dispersal will be much more likely. Here we generalize the previously developed TWOGENER
method, assuming that the pollen dispersal curve belongs to particular one- or two-parameter
families of dispersal curves and estimating simultaneously the parameters of the dispersal
curve and the effective density of reproducing individuals in the population. We tested
this method on simulated data, using an exponential power distribution, under thin-tailed,
exponential and fat-tailed conditions. We find that even if our estimates show some bias
and large mean squared error (MSE), we are able to estimate correctly the general trend of
the curve — thin-tailed or fat-tailed — and the effective density. Moreover, the mean distance
of dispersal can be correctly estimated with low bias and MSE, even if another family of
dispersal curve is used for the estimation. Finally, we consider three case studies based on
forest tree species. We find that dispersal is fat-tailed in all cases, and that the effective
density estimated by our model is below the measured density in two of the cases. This
latter result may reflect the difficulty of estimating two parameters, or it may be a biological
consequence of variance in reproductive success of males in the population. Both the simulated and empirical findings demonstrate the strong potential of TWOGENER for evaluating
the shape of the dispersal curve and the effective density of the population (d_e).**

Keywords: gene flow, long-distance dispersal, microsatellites, plants, trees, TWOGENER

Received 20 July 2003; revision received 13 November 2003; accepted 13 November 2003

Introduction

Pollen dispersal is an important component of gene flow in
plants (Ennos 1994; Oddou-Muratorio *et al.* 2001), facilitating connections between individuals or populations.
Until recently, the standard method of estimating gene

Correspondence: Frederic Austerlitz.
E-mail: Frederic.Austerlitz@ese.u-psud.fr

flow from genetic data was through measures of genetic
differentiation among populations (Wright 1951) or individuals (Rousset 2000) sampled from a single generation. If
we assume discrete populations at evolutionary equilibrium
and an island model, Wright's (1951) F_{ST} provides an estimate of the product $N_e m_e$, where N_e is the effective population
size of each deme, and m_e is the effective migration rate among
populations. Alternatively, if the species exhibits isolation
by distance, linear regression can yield an estimate of $N_e \sigma_e^2$,

where σ_e^2 is the effective variance of gene flow (Rousset 1997). If the species constitutes a continuous population, where individuals mate preferentially with their neighbours, $d_e\sigma_e^2$ can be estimated instead, where d_e is the effective density of individuals (Rousset 2000).

Indirect methods have the drawback that they do not readily distinguish between seed and pollen flow. If nuclear loci are used, the estimated migration rate (m_e) in the island model will be a composite of the effective migration rates of seeds (m_{Se}) and pollen (m_{Pe}), with $m_e = m_{Se} + m_{Pe}/2$ (Ennos 1994). Similarly, in the isolation-by-distance model, Crawford (1984) showed that the dispersal parameter (σ_e^2) is comprised of a seed and pollen dispersal component $\sigma_e^2 = \sigma_{Se}^2 + \dfrac{\sigma_{pe}^2}{2}$. However, it is possible to gauge the relative contributions of seed and pollen flow by using maternally or paternally inherited cytoplasmic markers, which yield separate estimates of seed- and pollen-mediated gene flow. Paternally inherited cytoplasmic markers estimate the product $N_e m_{Pe}$ or $d_e\sigma_{pe}^2$ in gymnosperms. Studies of angiosperms utilize maternally inherited cytoplasmic markers, in combination with nuclear markers, to separate the seed and pollen contribution to gene flow (McCauley 1997; Oddou-Muratorio *et al.* 2001).

Using cytoplasmic markers to study gene flow is not without its problems. First, in plants, mutation rates are relatively low in organellar genomes (Wolfe *et al.* 1987), so there is often little within-population variation in chloroplast haplotypes, relative to nuclear markers. Second, since cytoplasmic markers are uniparentally inherited and nonrecombining, they must be treated as a single locus. Because cytoplasmic markers provide only one repetition of the process of genetic transmission, the estimation variance is large, whereas independently segregating nuclear loci yield several replications and a smaller estimation variance. Moreover, the absence of recombination accentuates any selective effects involving the entire cytoplasmic genome, which will strongly bias the dispersal estimates, all of which assume adaptive neutrality.

Indirect methods have a few additional drawbacks. They do not provide an unambiguous separation of the demographic parameter (N_e or d_e) from the dispersal parameter (m_e or σ_e^2). A recently proposed method provides separate estimates of these two parameters (Vitalis & Couvet 2001a,b), but this method requires incompletely linked loci, and thus cannot be applied to cytoplasmic genomes to disentangle seed from pollen flow. Finally, indirect methods provide a measure of the historical gene flow, i.e. an average of this value over at least a substantial portion of the population's past (Hudson 1998; Sork *et al.* 1999).

In cases where the population structure has been disrupted recently, historical structure may have such a short temporal memory (Smouse *et al.* 1991; Smouse & Long 1992) that it will be quickly erased by further evolution.

Moreover, the averaging process renders the estimation of more than one parameter almost hopeless. No method has been proposed thus far to estimate the shape of pollen or seed dispersal curves from genetic differentiation, something that would be very helpful for extrapolation or generalization. Indeed, key elements to the study of dispersal are the distance of dispersal and the variance in dispersal, since they have a high impact on the evolutionary process, conditioning (for instance) the rate of spread of a favourable gene or the effective population size.

A more direct approach for estimating pollen dispersal is provided by paternity analysis, which relies on a sampling of mother plants, along with a sample of their offspring, as well as an enumeration and genetic characterization of the surrounding males. Paternity analysis methods (e.g. Devlin & Ellstrand 1990; Marshall *et al.* 1998) attempt to detect (for each offspring) whether paternity of the seed can be attributed significantly to one of the males present at the study site. Then, for the subset of offspring for which a credible on-site father has been found, the position of the mothers and the fathers can be used to estimate the parameters of the pollen dispersal function chosen (S. Oddou-Muratorio, E. K. Klein and F. Austerlitz, in preparation). This method relies on an exhaustive sampling of the males in the vicinity of the sampled females, requiring substantial effort, since pollen can come from males that are far from the sampled site (for a review, see Slavov *et al.* 2002). Moreover, it is necessary to use highly polymorphic genetic markers to eliminate paternal ambiguity.

An alternative strategy is the TWOGENER analysis of Smouse *et al.* (2001), based on the differentiation among the inferred pollen pools of a sample of females, spread across the landscape, and encapsulated in a synthetic parameter Φ_{ft} that is analogous to F_{ST}, but which relates only to a single bout of pollination. The virtue of this method is that, unlike the paternity method, it does not require exhaustive sampling of the adults of the population. The global estimate of Φ_{ft}, computed from the entire collection of sampled mothers, is easily translated into an estimate of $d_e\sigma_{pe}^2$, from which we can infer the average distance of pollination (Austerlitz & Smouse 2001), provided that we use a one-parameter pollen dispersal distribution (e.g. a normal or exponential distribution) and that adult density is estimated independently from demographic data.

As an extension of TWOGENER, we can use the computation of pairwise Φ_{ft} between the pollen pools sampled by all pairs of sampled females to estimate multiple parameters jointly, among them the adult density and the average distance of pollen dispersal (Austerlitz & Smouse 2002). With this method, we can fit several families of two-parameter dispersal functions, and (at least in principle) estimate all the parameters simultaneously. So far, this approach has been performed for a single family of exponential power distributions, characterized by two

parameters, *a* and *b*, which determine the average pollination distance and the shape of the distribution, respectively (Tufto *et al.* 1997).

We used this method on empirical data from one temperate tree species, *Sorbus torminalis* (L.) Crantz (Rosaceae) (S. Oddou-Muratorio, E. K. Klein and F. Austerlitz, in preparation), and one tropical tree species, *Dinizia excelsa* Ducke (Fabaceae) (Dick *et al.* 2003). The estimated dispersal curve was strongly fat-tailed in the case of *S. torminalis*, but only moderately so in the case of *D. excelsa*. Dispersal distributions are considered to be fat-tailed when they decrease more slowly than would the exponential distribution at long distance (Clark 1998), and thin-tailed if they decrease more rapidly. The probability of long-distance dispersal is much higher for the fat-tailed than for the thin-tailed distributions. This tendency toward fat-tailed distributions is rather strong for *S. torminalis*, but is less pronounced for *D. excelsa*. In addition, high average pollination distances, greater than several hundred metres, were found for both species, relative to exponential predictions. Finally, these two studies revealed that when effective density is estimated jointly with the other parameters, that estimate is below the observed density of flowering males, probably a consequence of variance in male reproductive success, which we confirmed with paternity analysis in the case of *S. torminalis*.

In these two studies, we used a specific family of dispersal curves, the exponential power family. However, other families of dispersal curves can be used. For instance Tufto *et al.* (1997) conducted a study, based on Bateman's (1947) data on physical dispersal of pollen describing fat-tailed distributions, and concluded that the Weibull family of dispersal distributions (Weibull 1951) provided a closer fit to the data than did the exponential power family. The Bateman study, however, is based on the physical dispersal of pollen, rather than the analysis of successful pollination. Retrospective studies of pollination distance are all based on pollen that has both arrived and been successful at fertilization. Whether the better performance of the Weibull family of distributions will carry over to pollination studies remains unclear.

In this paper, we extend the previously designed method of estimation of the dispersal curve to other families of dispersal curves. These curves are the Weibull, geometric and bivariate Student's *t* (2*Dt*) (Clark 1998) distributions. While the Weibull distribution shows a decrease at long distance that mimics the exponential power distribution, the geometric and 2*Dt* are power-law functions that are fat-tailed for all the values of their parameter space. Our aim was to study the impact of the assumed dispersal family on the estimation of three parameters of interest: (i) the mean dispersal distance, (ii) the shape of the curve, and (iii) the adult reproductive density. First, using simulations that assume that pollen dispersal follows an exponential power function, for three sets of values of the parameters (*a*, *b*) of

the distribution, we determined the precision of the parameter estimates from the TWOGENER analysis, provided that the same family of dispersal curves was used. Then using the same data sets, simulated with the exponential power function, we performed the analysis under the assumptions of the other families of dispersal distributions, to determine the extent to which the estimation of these parameters was affected by the choice of the dispersal family.

Finally, we performed the same analysis on four experimental data sets, the two of *S. torminalis* and the one of *D. excelsa* mentioned above, and an additional data set developed on *Quercus lobata* Neé (Fagaceae) (C. Dutech and V. L. Sork, in preparation). We performed a TWOGENER analysis on all four data sets, and extracted estimates of both the effective densities and parameter(s) of the selected families of dispersal distributions. We then determined which family of distributions provided the best fit to the data and for the simulated data sets, whether the estimates of average pollen distance, shape of the dispersal curve and adult density were much affected by the choice of the dispersal family.

Materials and methods

Families of dispersal distributions

Normal family. One of the two reference groups of dispersal distributions was the normal family, characterized by a single parameter, *a*, such that

$$p_n(a; x, y) = \frac{1}{\pi a^2} \exp\left(-\left(\frac{r}{a}\right)^2\right), \qquad (1)$$

where $r = \sqrt{(x^2 + y^2)}$ is the pollination distance, and where $a = \sigma\sqrt{2}$, with σ defined as the standard deviation of the normal. This is the standard reference distribution used for TWOGENER analysis because the mathematics conveniently articulate with those of the least squares variance components methodology (Austerlitz & Smouse 2001, 2002; Irwin *et al.* 2003) and can easily be manipulated in closed form. In spite of its mathematical convenience, however, the emerging data (Dick *et al.* 2003; S. Oddou-Muratorio, E. K. Klein and F. Austerlitz, in preparation) suggest that it is much too thin-tailed to be ideal for natural populations.

Exponential family. The second group of distributions to be examined was that of the exponential family, also characterized by a single parameter, γ (Austerlitz & Smouse 2001)

$$p_e(a; x, y) = \frac{1}{2\pi a^2} \exp\left(-\left(\frac{r}{a}\right)\right), \qquad (2)$$

where *r* is defined as above and $a = \gamma$, the traditional exponential scale parameter. Although less convenient mathematically than the normal family, the exponential

family better captures the leptokurtotic features of the pollination distribution, and is analytically tractable. The variance of the distance of successful pollination is greater for the exponential than for the normal family (Austerlitz & Smouse 2001).

Exponential power family. The first two-parameter group of dispersal distributions to be examined with regard to the pollen dispersal curve was the exponential power family, characterized by the parameters *a* and *b* (Clark 1998)

$$p_{ep}(a, b; x, y) = \frac{b}{2\pi a^2 \Gamma(2/b)} \exp\left(-\left(\frac{r}{a}\right)^b\right), \quad (3)$$

where $r = \sqrt{(x^2 + y^2)}$ is the pollination distance, and Γ is the classically defined gamma function (Abramowitz & Stegun 1964). The parameter *b* is the shape parameter, affecting the 'fatness' of the tail of the dispersal distribution, and *a* is a scale parameter. For $b = 2$, equation 3 degenerates to the normal distribution, equation 1, with $a = \sigma \sqrt{2}$. For $b = 1$, equation 3 degenerates to the exponential distribution, equation 2, with $a = \gamma$. When $b < 1$, however, the dispersal kernel is fat-tailed (Clark 1998), i.e. the long-range decay of probability is slower than for the exponential distribution, and considerably slower than for the normal distribution. Conversely, when $b > 1$ (for sub-Gaussian and Gaussian models, for example), the dispersal is thin-tailed, with a rapid decrease of the dispersal function, implying few long-distance dispersal events.

Weibull family. We next considered the Weibull family (Weibull 1951; Tufto *et al*. 1997), also a two-parameter group of distributions, which takes the form

$$p_w(a, b; x, y) = \frac{a^{-b} b}{2\pi} \left(\frac{r}{a}\right)^{b-2} \exp\left(-\left(\frac{r}{a}\right)^b\right), \quad (4)$$

defined for any positive real numbers *a* and *b*. The distribution is fat-tailed when $b \leq 1$ and thin-tailed otherwise. As for the exponential power function, when $b = 2$, the Weibull degenerates to the normal distribution, but when $b = 1$, it does not degenerate to the exponential distribution.

Geometric and 2Dt families. We also considered the classical geometric family of dispersal distributions, defined by two parameters, *a* and *b*

$$p_g(a, b; x, y) = \frac{(b - 2)(b - 1)}{2\pi a} \left(1 + \frac{r}{a}\right)^{-b}, \quad (5)$$

which is defined for all values of $a > 0$ and of $b > 2$, and the bivariate Student's *t* (2Dt) distribution (Clark *et al*. 1999)

$$p_{2Dt}(a, b; x, y) = \frac{b - 1}{\pi a^2} \left(1 + \frac{r^2}{a^2}\right)^{-b}, \quad (6)$$
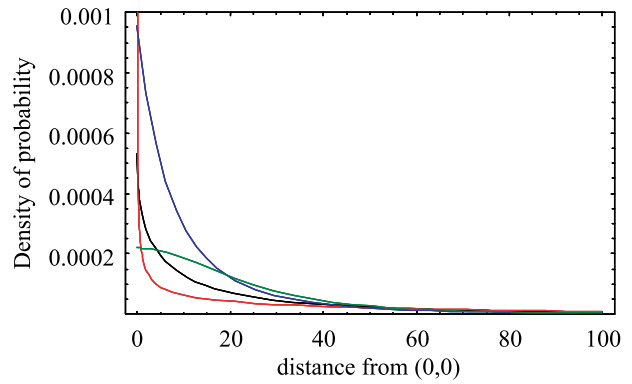

Fig. 1 Example of curves from the different family used, exponential power (black line), Weibull (red line), geometric (blue line) and 2*Dt* (green line), with shape parameter $b = 0.5, 1.5, 3.5$ and 2, and scale parameter (*a*) adjusted so that in all cases $\delta = 100$ m.

which is defined for all values of $a > 0$ and of $b > 1$. It is of interest to consider the geometric and 2*Dt* distributions, which, while both power-law functions, will behave quite differently from the exponential and Weibull distributions. They show a fat tail, whatever the value of the shape parameter (*b*), and the distributions become increasingly fat-tailed as *b* declines toward '1'. The 2*Dt* distribution was shown to be the best-fitting for seed dispersal for several tree species (Clark *et al*. 1999). The Weibull distribution shows the most leptokurtic shape, with a high peak at zero and a long tail; the exponential power and geometric distributions are intermediate, and the 2*Dt* distribution shows less leptokurtic shape, and is concave near the origin (Fig. 1).

The parameters *a* and *b* are specific for each family of curves. For a given data set, it is therefore meaningless to make a comparison for these parameters between the different families of curve. On the other hand, the moments of the different curves can be compared, the general formula for which is

$$\mu_n = \left(2\pi \int_0^\infty r^{n+1} p(r) dr\right)^{1/n} \quad (7)$$

as in Clark *et al*. (1999), except that we take the power $1/n$ so that every moment is expressed in metres. μ_1 corresponds to the mean dispersal distance (δ; see, e.g. Austerlitz & Smouse 2001) and μ_2 to the root-mean-square dispersal distance (Lande & Barrowclough 1987). These moments characterize the shape of the curve and thus allow us to compare the estimates obtained with different families of curves. They can be computed for the normal, exponential, exponential power, Weibull, geometric and 2*Dt* distributions, respectively.

$$\mu_n^n = a(\Gamma(1 + n/2))^{1/n}, \quad (8)$$

$$\mu_n^e = a(\Gamma(2 + n))^{1/n}, \quad (9)$$

$$\mu_n^{ep} = a \left(\frac{\Gamma((2+n)/b)}{\Gamma(2/b)} \right)^{1/n}, \tag{10}$$

$$\mu_n^{w} = a \, (\Gamma(1+n/b))^{1/n}, \tag{11}$$

$$\begin{cases} \mu_n^{g} = a \left(\dfrac{\Gamma(n+2)\Gamma(b-2-n)}{\Gamma(b-2)} \right)^{1/n} & \text{if } b > (n+2) \\ \mu_n^{g} = \infty & \text{if } 2 < b \le (n+2) \end{cases}, \tag{12}$$

$$\begin{cases} \mu_n^{2Dt} = a \left(\dfrac{\Gamma(n/2+1)\Gamma(b-1-n/2)}{\Gamma(b-1)} \right)^{1/n} & \text{if } b > (n/2+1) \\ \mu_n^{2Dt} = \infty & \text{if } 1 < b \le (n/2+1) \end{cases}. \tag{13}$$

While the moments of the normal, exponential, exponential power and Weibull distributions are always finite, they can be infinite for the geometric and 2*Dt* distributions.

TWOGENER *analysis*

The TWOGENER analysis is described in great detail by Smouse *et al*. (2001). In short, the method consists of genotyping a sample of mother plants, along with a sample of seeds harvested from each of these plants. Knowing the maternal genotype, it is possible to extract the paternal gametic genotype for each seedling. This can be done without ambiguity in all cases for gymnosperms, by assessing separately the megagametophyte. In the case of angiosperms, the absence of this megagametophyte makes the unambiguous assessment of the paternal gamete impossible in the case where both mother and seedling show the same heterozygous genotype at one or several loci. In this case, instead of categorically inferring the paternal gamete, the posterior likelihood of each possible male gamete is computed.

Once these male gametes have been inferred (either unambiguously or as posterior probabilities), it is possible to compute the differentiation among the pollen pools fertilizing the sampled females by performing an analysis of molecular variance (AMOVA, Excoffier *et al*. 1992), using the females as strata. This yields an estimate of the differentiation between the pollen clouds (Φ_{ft}), similar to the estimate of differentiation between populations (Φ_{st}), classically computed in an AMOVA.

The relation between Φ_{ft} and dispersal distance has been derived for normal and exponential dispersal distributions (Austerlitz & Smouse 2001), allowing the development of various estimates of average pollination distance (Austerlitz & Smouse 2002). Some of the estimates are based on the global Φ_{ft}, measured on all the females in the population. Since we only compute one Φ_{ft} value, we can only translate it into one parameter, usually the scale parameter of either the normal or exponential distribution, both of which are single-parameter models. That scale parameter can, in turn, be translated into the average distance of successful pollination, δ_e, via equations 8 or 9, respectively. For two-parameter families, we cannot jointly estimate both scale and shape parameters from a single value of Φ_{ft}.

We can, however, extend the pairwise estimation method that we designed for the normal distribution (Austerlitz & Smouse 2002) and the exponential power function (S. Oddou-Muratorio, E. K. Klein and F. Austerlitz, in preparation; Dick *et al*. 2003). This method is based on the computation of the observed Φ_{ft} values between each pair of females, *i* and *j*, denoted ϕ_{ij}^{obs}, a function of the physical distance (z_{ij}) between those females. Assume a particular family of dispersal curves, $p(\theta; x, y)$, where θ is the set of parameters for this family [e.g. $\theta = (a, b)$ for the exponential power family]. It is possible to derive the theoretical relation, $\Phi_{ft}(d, \theta; z)$, relating the Φ_{ft} value for two females, a distance *z* apart, where *d* is the adult density on the landscape.

The formula for $\Phi_{ft}(d, \theta; z)$ can be derived, whatever the chosen family of dispersal distributions, and it follows from the relation given in Austerlitz & Smouse (2002) that:

$$\Phi_{ft}(d, \theta, z) = \frac{Q_0(d, \theta) - Q(d, \theta, z)}{2 - Q(d, \theta, z)} \tag{14}$$

where $Q_0(d, \theta)$ is the probability that two male gametes sampled from the pollen cloud of the same female were drawn from the same father, and $Q(d, \theta, z)$ is the probability that two male gametes, sampled from two separate females (a distance *z* apart), were from the same father.

It follows that $Q_0(d, \theta)$ can be computed directly, using equation 10 from Austerlitz & Smouse (2001):

$$Q_0(d, \theta) = \frac{1}{d} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p^2(\theta, x, y) \, dx \, dy, \tag{15}$$

and the results for the six families are given by:

Normal $Q_0^n(d, a) = \dfrac{1}{2\pi \, da^2}$, \hfill (16)

Exponential $Q_0^e(d, a) = \dfrac{1}{8\pi \, da^2}$, \hfill (17)

Exponential power $Q_0^{ep}(d, a, b) = \dfrac{1}{4^{1/b} \pi \, da^2 \, \Gamma(1+2/b)}$, \hfill (18)

Weibull $Q_0^w(d, a, b) = \dfrac{b2^{2/b-3} \Gamma(2-2/b)}{\pi \, da^2}$ \quad for $b > 1$, \hfill (19)

Geometric $Q_0^g(d, a, b) = \dfrac{(b-1)(b-2)^2}{4(2b-1)\pi \, da^2}$, \hfill (20)

2*Dt* $Q_0^{2Dt}(d, a, b) = \dfrac{(b-1)^2}{(2b-1)\pi \, da^2}$. \hfill (21)

Now, $Q(d, \theta, z)$ can be computed analytically for the single-parameter families, by solving the numerical formula (Austerlitz & Smouse 2001)

$$Q(d, \theta, z) = \frac{1}{d} \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} p(x, y) \, p(x - z, y) dx \, dy, \qquad (22)$$

explicitly, where $p(x, y)$ stands for any of the formulae given in equations 1 to 6. Equation 21 can be evaluated analytically for the normal distribution but can only be evaluated numerically for the other distributions (the exponential and the two-parameter distributions).

We also note that $Q_0(d, \theta)$, and thus $Q(d, \theta, z)$, cannot be computed for the Weibull distribution over the entire range of its shape parameter (b). This difficulty follows from the fact that the approximation used in Austerlitz & Smouse (2001) to obtain equation 15, following Wright (1943), becomes invalid for this distribution when $b < 1$. As a practical matter then, we will only consider the Weibull distribution for $b \geq 1$.

We estimated the various parameters by minimizing the squared-error loss criterion for the choice of those parameters,

$$C(d, \theta) = \sum_{i<j}^{n_m} (\phi_{ij}^{obs} - \Phi_{ft}(d, \theta, z_{ij}))^2. \qquad (23)$$

For our first set of evaluations, we set the effective density, d_e, to the field-measured density of adults in the population, and for each of the dispersal families indicated above, found the best-fitting collection of pollen dispersal parameters ($\hat{\theta}$), choosing parameters for θ that minimized $C(d, \theta)$. For our second set of trials, we set the shape parameter (b) to the value estimated with fixed d_e, and then re-estimated the effective density (d_e), along with the scale parameter (a), to determine how far effective density diverged from measured density. For our final series of trials, we tried to estimate the three parameters (a, b and d) simultaneously. All these estimates have been included in the FAMOZ program available at http://www.pierroton.inra.fr/genetics/labo/Software/Famoz/index.html. Dos and Unix executables and sources codes are also available from F.A.

Simulations

Because these methods of estimation are extremely time consuming (several days for one replicate), it was impossible to perform an exhaustive series of simulations. We therefore restricted attention to a small number of contrasts, and used the simulation program developed by Austerlitz & Smouse (2002) (available from F.A.). We simulated a population of 10 000 individuals, with a density of 1.6 trees/ha, from among which we sampled 40 mothers, at distances

ranging from ~20 m to ~3 km. A population size of 10 000 individuals is large, but we used it to avoid border effects. The numbers of effective males expected to fertilize the sampled females was much smaller. We assumed three particular models of pollen dispersal: (i) a normal distribution, (ii) an exponential distribution, and (iii) an exponential power distribution with $b = 0.5$. In all cases, the parameter a was adjusted so that the mean dispersal distance (δ) was 100 m.

Under these conditions, we simulated data sets composed of mother and offspring genotypes. We assumed a sample size of 40 offspring per mother and that mothers and offspring were genotyped for 10 polymorphic loci, each locus with 10 alleles, all at frequency 1/10. Each parameter set was replicated 10 times. For each simulated replicate, we performed a TWOGENER analysis, assuming either an exponential power, Weibull, geometric, or 2Dt distribution of pollination distances. For each estimated curve, we computed the first four moments, as well as the bias and mean squared error (MSE) for these moments over the 10 replicates, as well as for the estimated density. For evaluation with the exponential power function, we also computed the bias and MSE of the estimated curve parameters a and b. We did this for the Weibull analysis when the normal dispersal was assumed because, as we pointed out, the normal distribution is a particular case of the Weibull distribution. For the other two cases, because the dispersal curves used here do not belong to the Weibull family, there was no true value for a and b when this family was used for analysis, so neither bias nor MSE could be computed. This was always the case for the geometric and 2Dt families.

Consider first the case where a normal distribution was used in the simulations (Table 1), but the TWOGENER procedure assumed an exponential power distribution, setting density at its real value of 1.6 trees/ha yielded slightly inflated estimates of a and more strongly inflated estimates of b, but, as a result, the average dispersal distance ($\hat{\delta}$) and the higher order moments were almost unbiased. Also, $\hat{a}$ and $\hat{b}$ showed quite large MSE values, but — again by a compensation process — MSE for $\hat{\delta}$ and higher order moments was more limited (compared to the true value of these moments).

When density (d) was estimated along with a, setting b at the previously estimated value, this yielded a downward bias for d but an upward bias for a. MSE-values were in all cases rather high, compared with the parametric mean. Again as a compensation process, the estimated moments showed much less bias than the estimated curve parameters. This was also true when all three parameters (d, a and b) were estimated jointly. They all showed some bias and a rather high MSE, but the moments were rather well estimated (limited bias and MSE).

Using the Weibull distribution in the analyses yielded much more limited bias, and smaller values of both standard deviation and MSE for the estimation of d, a and b. This

Table 1 Bias and mean squared error (MSE) of the parameter estimates and the estimates of the moments of the dispersal curve for the simulations performed with a normal distribution of pollination distance, but evaluated under exponential power, Weibull, and geometric models for TWOGENER analysis. The parameters used in the simulations were d = 1.6 trees/ha, a = 112.8 m, b = 2, which yields δ = 100 m, μ_2 = 112.8 m, μ_3 = 124.0 m, μ_4 = 134.1 m

Assumed dispersal function	Fixed parameter		Estimated parameters			Estimated moments			
			$\hat{d}$	$\bar{a}$	$\hat{b}$	$\hat{\delta}$	$\hat{\mu}_2$	$\hat{\mu}_3$	$\hat{\mu}_4$
Exponential power	d	bias	—	8.93	1.28	−3.29	−5.30	−7.22	−9.03
		√MSE	—	25.12	1.77	8.01	12.61	17.50	22.44
	b	bias	−0.18	17.17	—	2.97	1.61	0.24	−1.09
		√MSE	0.27	32.76	—	11.01	14.14	18.10	22.47
	none	bias	−0.22	5.89	1.04	8.79	11.08	13.55	16.22
		√MSE	0.36	38.28	1.89	21.29	31.09	41.73	52.88
Weibull	d	bias	—	−5.11	0.44	−4.37	−8.14	−11.5	−14.5
		√MSE	—	7.12	0.53	6.26	10.05	13.72	17.15
	b	bias	−0.24	6.84	—	6.24	−6.27	−19.3	−32.8
		√MSE	0.45	17.91	—	16.01	17.06	25.8	37.55
	none	bias	−0.28	12.87	0.33	12.02	−17.8	−31.7	−45.9
		√MSE	0.52	26.18	0.62	23.96	18.77	32.5	46.78
Geometric	d	bias	—	n.a.*	n.a.	10.95	−7.37	−32.3	−63.4
		√MSE	—	n.a.	n.a.	12.46	10.15	33.35	64.15
	b	bias	???†	???	—	???	???	???	???
		√MSE	???	???	—	???	???	???	???
	none	bias	−0.55	n.a.	n.a.	33.02	22.41	6.95	−11.8
		√MSE	0.70	n.a.	n.a.	37.62	34.00	39.07	60.00

*n.a., not applicable: the dispersal curve used in the simulations does not belong to the family of dispersal curves assumed for the TWOGENER analysis. Thus no bias and consequently no MSE can be computed for the shape parameters.
†???, the fit did not converge here, probably because of the high value taken by the b parameter.

remained also true to a lesser extent for the estimates of the moments of the distribution. When the geometric distribution was used, the estimates of the moments showed much more bias and MSE than when the Weibull or exponential power distributions were used. The estimation of d and the scale parameter a did not even converge. The 2Dt estimates did not converge in that case, even when density was fixed.

When the dispersal curve used in the simulation was either the exponential distribution (Table 2) or the exponential power distribution with b = 0.5 (Table 3), and the exponential power distribution was assumed for TWOGENER analysis, the parameters (d, a and b) showed some bias and MSE, while the average distance of dispersal (δ) was better estimated, except for the case when b = 0.5 and all three parameters were estimated jointly, which yielded large biases. The bias and MSE increased when higher order moments were considered, but remained lower, relative to their expected values, than the bias and MSE of the curve parameters a and b. When the true dispersal distribution was exponential, using the Weibull or geometric analyses yielded similar values of bias and MSE for the estimate of the moments (Table 2). However, when the true dispersal curve was the exponential power distribution with b = 0.5, the Weibull and geometric analyses

yielded much larger bias but slightly smaller MSE than when the exponential power distribution was used, except when all three parameters were jointly estimated, in which case they always behaved better. However, in several cases, the moments estimated by the geometric distribution were infinite and thus had to be removed from the computation of the bias and MSE. Again, 2Dt estimates did not converge at all when b = 1. It behaved like the geometric distribution when b = 0.5 (data not shown).

Looking at the simulations globally, when the exponential power distribution was used, the estimated values of the shape parameter (b) were below 1.0 when the true value was 0.5 in nine cases out of 10. Thus, even if the level of bias and MSE were rather high, the fat-tailed character of the curve was correctly detected. Conversely, when the true value of b was 2.0, the estimated values were always > 1, indicating that the thin-tailed character of the curve was also well detected.

Real world case studies

We applied these estimation methods to a trio of real world case studies in an effort to estimate the average distance of successful pollination and the shapes of the dispersal curves.

Table 2 Bias and MSE of the parameter estimates and the estimates of the moments of the dispersal curve for the simulations performed with an exponential distribution of pollination distance, but evaluated under exponential power and Weibull models for TWOGENER analysis. The parameters used in the simulations were d = 1.6 trees/ha, a = 50 m, b = 1, which yields δ = 100 m, μ_2 = 122.4 m, μ_3 = 144.2 m, μ_4 = 165.5 m

Assumed dispersal function	Fixed parameter		Estimated parameters			Estimated moments			
			$\hat{d}$	$\hat{a}$	$\hat{b}$	$\hat{\delta}$	$\hat{\mu}_2$	$\hat{\mu}_3$	$\hat{\mu}_4$
Exponential power	d	bias	—	20.84	1.01	−7.99	−13.18	−18.46	−23.78
		√MSE	—	36.46	1.90	13.91	23.70	34.16	45.08
	b	bias	0.11	23.47	—	−8.45	−11.85	−16.66	−21.52
		√MSE	0.56	42.45	—	14.96	19.99	27.02	34.88
	none	bias	0.16	28.68	1.44	12.79	29.65	48.59	73.64
		√MSE	0.88	43.80	2.22	86.98	156.73	240.03	342.18
Weibull	d	bias	—	n.a.*	n.a.	−8.32	−15.79	−24.20	−33.26
		√MSE	—	n.a.	n.a.	14.13	23.95	34.17	44.70
	b	bias	0.14	n.a.	—	−8.78	−16.62	−25.36	−34.74
		√MSE	0.63	n.a.	—	17.25	25.55	34.98	45.06
	none	bias	0.16	n.a.	n.a.	22.39	−15.79	−24.20	−33.26
		√MSE	0.91	n.a.	n.a.	114.63	155.45	193.78	230.49
Geometric	d	bias	—	n.a.	n.a.	−1.33	2.11	8.18	19.22
		√MSE	—	n.a.	n.a.	6.64	12.57	25.21	51.07
	b	bias	0.17	n.a.	—	1.83	−3.24	1.21	9.95
		√MSE	0.29	n.a.	—	22.02	7.00	14.16	34.30
	none	bias	0.09	n.a.	n.a.	0.09	4.96	7.85	11.38
		√MSE	0.66	n.a.	n.a.	17.04	20.84	25.44	30.41

*n.a., not applicable: the dispersal curve used in the simulations does not belong to the family of dispersal curves assumed for the TWOGENER analysis. Thus no bias and consequently no MSE can be computed for the shape parameters.

Table 3 Bias and MSE of the parameter estimates and the estimates of the moments of the dispersal curve for the simulations performed with an exponential power distribution of pollination distance, but evaluated under exponential power and Weibull models for TWOGENER analysis. The true values of the parameters used in the simulations were d = 1.6 trees/ha, a = 5.0, b = 0.5, which yields δ = 100 m, μ_2 = 144.9 m, μ_3 = 196.2 m, μ_4 = 253.9m

Assumed dispersal function	Fixed parameter		Estimated parameters			Estimated moments			
			$\hat{d}$	$\hat{a}$	$\hat{b}$	$\hat{\delta}$	$\hat{\mu}_2$	$\hat{\mu}_3$	$\hat{\mu}_4$
Exponential power	d	bias	—	16.21	0.24	−1.17	−4.75	−6.49	−5.04
		√MSE	—	20.90	0.33	37.70	81.48	145.54	233.40
	b	bias	−0.14	19.73	—	6.99	3.57	0.59	−0.87
		√MSE	0.70	25.27	—	30.80	60.34	106.52	171.06
	none	bias	−0.19	12.96	0.18	928.4	2542	5 859	11 989
		√MSE	1.60	17.32	0.46	1788	5040	11 850	24 622
Weibull	d	bias	—	n.a.*	n.a.	−13.74	−38.12	−70.83	−111.1
		√MSE	—	n.a.	n.a.	20.39	44.10	76.53	116.70
	b	bias	0.08	n.a.	—	−5.23	−27.95	−59.19	−98.17
		√MSE	1.10	n.a.	—	28.75	45.56	73.27	110.20
	none	bias	−0.12	n.a.	n.a.	240.21	307.93	307.93	362.82
		√MSE	1.54	n.a.	n.a.	586.47	795.60	795.60	991.81
Geometric†	d	bias	—	n.a.	n.a.	−4.19	107.20	−49.78	−99.01
		√MSE	—	n.a.	n.a.	25.36	292.73	70.46	102.54
	b	bias	−0.22	n.a.	—	8.51	136.78	−27.18	−70.25
		√MSE	0.50	n.a.	—	30.63	345.01	59.66	82.08
	none	bias	−0.16	n.a.	n.a.	138.24	−62.16	−97.40	−138.8
		√MSE	1.63	n.a.	n.a.	370.28	65.04	100.08	141.41

*n.a., not applicable: the dispersal curve used in the simulationsdoes not belong to the family of dispersal curves assumed for the TWOGENER analysis. Thus no bias and consequently no MSE can be computed for the shape parameters.
†In several cases one or several moments were infinite and could thus not be included in the averages.

Sorbus torminalis

Sorbus torminalis (L.) Crantz (Rosaceae), the wild service-tree, occurs in a scattered distribution across a landscape. It produces hermaphroditic flowers, visited by a wide range of generalist pollinators. The study site, located in the large Rambouillet forest near Paris, covers a 475 ha mixed stand of oak and other broad-leaved species. Within this site, we found 185 potentially reproducing wild service-trees, all of which were genotyped for six microsatellite loci. We sampled and genotyped seeds from mother-trees in two consecutive years. In 1999, fruits were collected in the crown of 14 identified fruiting trees. A total of 653 seeds were extracted from the harvested fruits (11–100 seeds/tree, mean = 46.6, SD = 21.1). In 2000, fruits were collected either in the crown or near the trunk of 60 identified fruiting trees (total = 1075 seeds; mean number of seeds/tree = 17.9, SD = 3.9).

For the 1999 data set, analysed with the TWOGENER method (Table 4), fixing the density at its observed value (d = 0.33 trees/ha), the exponential distribution (b = 1) provided a better fit than the normal distribution (b = 2), in terms of squared error $C(d, \theta)$, but when the exponential power distribution was used, the fit was even better, yielding an estimated b-value of 0.565, associated with an estimated dispersal distance of 209 m, slightly more than when the normal or exponential distributions were used, where the estimated distance was ~150 m. The fits with the other two parameter curves yielded estimated distances ~200 m, like the exponential power function.

When density was jointly estimated with the scale parameter (a), setting the shape parameter (b) at the value previously estimated, leaving aside the case of the Weibull distribution, the estimated density ($\hat{d}$) ranged from 0.021 to 0.093 trees per ha, substantially below the measured density of 0.33 trees/ha. This discrepancy was even stronger for the Weibull distribution, where the estimated density reached an excessively low value of 0.0005 trees/ha. Conversely, the estimated dispersal distance was higher than in the case where density was fixed: from 249 m to 713 m, except for the Weibull distribution where it reached a value as high as 3687 m. When the three parameters were jointly estimated, the algorithm converged to unrealistically low values of d and b, and conversely to unrealistically high values of δ. In some cases, the algorithm failed to converge.

The fits to the 2000 data set (Table 5) with density fixed at its observed value, yielded similar estimated dispersal distance for the normal ($\hat{\delta} \cong 150$ m) and exponential distribution ($\hat{\delta} \cong 170$ m), but higher values when the other dispersal models were used ($\hat{\delta}$ = 478 m for the exponential power distribution, 360 m for the Weibull distribution, and ∞ for the geometric and $2Dt$ distributions). The exponential power function showed a stronger tendency toward a fatter-tailed distribution ($\hat{b}$ = 0.285) than observed for the 1999

data set. When density was jointly estimated with the scale parameter (a), the estimated d_e values were again much below the observed density, except for the normal and exponential distributions, offset by a higher estimate of the average dispersal distance, except for the normal and exponential distributions. As for the 1999 data set, the estimation algorithm did not converge properly when all three parameters were included in the model.

Dinizia excelsa

Dinizia excelsa Ducke (Fabaceae), endemic to the Amazon basin, is one of the largest Neotropical rainforest trees, attaining heights of over 55 m (Ducke 1922). It has very small (calyx 1–1.5 mm) hermaphroditic flowers held in racemes (10–18 cm) that attract diverse small insects (Dick 2001b). Owing to its value for timber and shade, large *D. excelsa* are commonly left standing in pastures and forest fragments in Brazilian ranches, thus providing an experimental design with which to study the effects of habitat fragmentation on gene flow in rainforest trees. Our study was performed in the forest reserves of the Biological Dynamics of Forest Fragments Project (BDFFP) (Laurance *et al*. 2002) north of Manaus, Brazil (S 2°30′, W 60°). In the BDFFP reserves, stingless bees (tribe Meliponini) are important pollinators of *D. excelsa* in undisturbed forests, whereas exotic African honeybees (*Apis mellifera scutellata*) are the primary pollinators of *D. excelsa* located in pastures and forest fragments (Dick 2001b). A microsatellite-based paternity analysis of seeds from pasture trees previously showed extremely long distance pollination flow (≤ 3.2 km) mediated by the African honeybees (Dick 2001a). Dick *et al*. (2003) performed a TWOGENER analysis on the same ranch population and on an additional population in undisturbed forest (site km41).

The km41 site, on which we focus here, is part of a vast tract of biologically diverse rainforest, containing up to 300 tree species per hectare (De Oliveira & Mori 1999). The study population consisted of 50 adult individuals (≥ 40 cm diameter at breast height) at a population density of 0.17 individuals/ha. Field observations indicated that most of the mapped trees flowered synchronously (Dick 2001a). For the TWOGENER analysis, we scored microsatellite genotypes at five highly polymorphic loci in 241 seeds from 13 maternal trees, representing an average of 21 seeds per array.

When the normal or exponential distribution was applied to the km41 population, we obtained estimated values for the average dispersal distance ($\hat{\delta}$) of ~200 m (Table 6). When density was jointly estimated along with the dispersal parameter, we obtained estimated values for this effective density that were slightly below those observed. Conversely, the estimated pollen dispersal distance was slightly higher. When the exponential power function was used, the shape parameter was estimated as $\hat{b}$ = 0.821, indicating a distribution

Table 4 Analytical results of the TWOGENER analysis for the 1999 data set for *Sorbus torminalis*, for the various dispersal distributions and with the different estimation strategies, fixed vs. estimated parameters. The minimum squared error criterion, $C(\hat{d}_e, \hat{\theta})$, is also provided

Dispersal function	Fixed parameters		Estimated parameters			Estimated moments (m)				Error
	d*	b	$\hat{a}$	$\hat{a}$†	$\hat{b}$	$\hat{\delta}$	$\hat{\mu}_2$	$\hat{\mu}_3$	$\hat{\mu}_4$	
Normal	0.33	2	—	161	—	143	161	177	191	0.239
Normal	—	2	0.0930	284	—	252	284	312	338	0.225
Exponential	0.33	1	—	79.5	—	159	195	229	263	0.236
Exponential	—	1	0.0521	181	—	362	443	522	599	0.217
Exponential power	0.33	—	—	18.9	0.565	210	293	385	484	0.234
Exponential power	—	0.565	0.0212	64.4	—	717	999	1310	1649	0.208
Exponential power	—	—	1.33×10^{-7}	2.94×10^{-9}	0.0908	1.60×10^7	1.07×10^8	5.17×10^8	1.98×10^9	0.195
Weibull	0.33	—	—	238	1.23	223	287	347	404	0.231
Weibull	—	1.23	0.0005	3945	—	3688	4764	5758	6698	0.192
Weibull	—	—	0.0004	4363	1.22	4087	5295	6414	7473	0.192
Geometric	0.33	—	—	594	9.69	178	236	304	388	0.235
Geometric	—	9.69	0.0402	1523	—	455	605	780	995	0.215
Geometric	—	—	—	???‡	???	???	???	???	???	???
2Dt	0.33	—	—	153	2.29	177	284	∞	∞	0.237
2Dt	—	2.29	0.0569	338	—	392	628	∞	∞	0.220
2Dt	—	—	7.45×10^{-6}	147	1.00361	∞	∞	∞	∞	0.210

*units for d are trees/ha.

†units for $\hat{a}$ are metres.

‡???, estimation did not converge.

Table 5 Analytical results of the TWOGENER analysis for the 2000 data set for *Sorbus torminalis*, for the various dispersal distributions and with the different estimation strategies, fixed vs. estimated parameters. The minimum squared error criterion, $C(\hat{d}_e, \hat{\theta})$, is also provided

Dispersal function	Fixed parameters		Estimated parameters			Estimated moments (m)				
	$d*$	b	$\hat{d}$	$\hat{a}$†	$\hat{b}$	$\hat{\delta}$	$\hat{\mu}_2$	$\hat{\mu}_3$	$\hat{\mu}_4$	Error
Normal	0.33	2	—	170	—	151	170	187	202	7.13
Normal	—	2	22.8	66.35	—	58.8	66.3	73.0	78.9	7.04
Exponential	0.33	1	—	85.4	—	171	209	246	283	7.07
Exponential	—	1	15.3	40.35	—	80.7	98.8	116	134	7.03
Exponential power	0.33	—	—	0.298	0.285	482	902	1511	2345	6.95
Exponential power	—	0.285	0.0119	1.42	—	2298	4300	7199	11 173	6.85
Exponential power	—	—	???*	???	???	???	???	???	???	???a
Weibull	0.33	—	—	373	1.1	360	487	607	723	6.91
Weibull	—	1.1	1.28×10^{-4}	14 367	—	13 863	18 746	23 374	27 835	6.65
Weibull	—	—	7.23×10^{-5}	12 648	1.17	11 978	15 779	19 329	22 713	6.59
Geometric	0.33	—	—	57.5	2.76	∞	∞	∞	∞	6.98
Geometric	—	2.76	0.0859	110	—	∞	∞	∞	∞	6.96
Geometric	—	—	???	???	???	???	???	???	???	???
2Dt	0.33	—	—	1821	1.21	∞	∞	∞	∞	7.00
2Dt	—	1.21	0.167	3534	—	∞	∞	∞	∞	6.99
2Dt	—	—	7.11×10^{-5}	2791	1.00326	∞	∞	∞	∞	6.96

*units for d are trees/ha.

†units for $\hat{a}$ are metres.

‡???, estimation did not converge.

Table 6 Analytical results of the TWOGENER analysis for *Dinizia excelsa*, for the various dispersal distributions and with the different estimation strategies, fixed *vs.* estimated parameters. The minimum squared error criterion, $C(\hat{d}_e, \hat{\theta})$, is also provided

Dispersal function	Fixed parameters		Estimated parameters			Estimated moments (m)				Error
	$d*$	b	$\hat{a}$	$\hat{a}†$	$\hat{b}$	$\hat{\delta}$	$\hat{\mu}_2$	$\hat{\mu}_3$	$\hat{\mu}_4$	
Normal	0.17	2	—	203	—	185	209	230	249	0.0856
Normal	—	2	0.142	284	—	201	227	250	270	0.0854
Exponential	0.17	1	—	104	—	208	255	300	344	0.0851
Exponential	—	1	0.113	126	—	252	309	363	417	0.0842
Exponential power	0.17	—	—	72.3	0.821	225	285	347	408	0.0851
Exponential power	—	0.821	0.0995	92.85	—	289	367	445	524	0.0838
Exponential power	—	—	1.46×10^{-3}	8.20×10^{-4}	0.159	2.15×10^4	6.44×10^4	1.59×10^5	3.46×10^5	0.0817
Weibull	0.17	—	—	267	1.34	245	307	363	416	0.0847
Weibull	—	1.34	0.0802	379	—	348	436	516	590	0.0831
Weibull	—	—	0.00948	1518	1.14	1448	1929	2380	2812	0.0817
Geometric	0.17	—	—	4068	41.2	213	264	316	367	0.0851
Geometric	—	41.2	0.130	4614	—	242	300	358	416	0.0843
Geometric	—	—	1.90×10^{-10}	83.3	2.00345	∞	∞	∞	∞	0.0824
2Dt	0.17	—	—	277	3.15	206	258	332	511	0.0854
2Dt	—	3.15	0.121	32	—	241	302	388	598	0.0847
2Dt	—	—	3.05×10^{-4}	9096	1.00453	∞	∞	∞	∞	0.0828

*units for d are trees/ha.
†units for $\hat{a}$ are metres.

with a thinner tail curve than was the case for *S. torminalis*. This same tendency was true for all of the two-parameter families, which all showed a fatter tail for *S. torminalis*. For *D. excelsa*, the average pollination distances inferred from the two-parameter distributions were all consistent, ranging between 205 m and 245 m when density was fixed at its observed value, and between 240 m and 388 m when density was jointly estimated with the scale parameter (*a*). This estimated density was slightly lower than the observed density in all cases, but the discrepancy was much smaller than for *S. torminalis*. Again the joint estimation of the three parameters provided no realistic results.

Quercus lobata

Valley oak [*Quercus lobata* Neé (Fagaceae)] is one of California's most distinctive oak species. Its massive size and majestic canopy, combined with its longevity, make it a signature element of California's foothills, valleys and floodplains (Griffin 1971). It typically occurs in oak savannah, which is the habitat of our study site. The study described here was conducted at the Sedgwick Reserve (N 34°42′, W 120°2′), 10 km northeast of Santa Ynez, California, USA. Sedgwick Reserve is a 2380-ha area, managed for research, education and conservation of native biodiversity by the University of California Natural Reserve System and University of California Santa Barbara. The study trees were located on the valley floor and surrounding hill slopes in a broad, shallow basin, ranging in elevation from 360 m to 405 m above sea level. This area was sampled in 1999 (Sork *et al*. 2002a,b), but in 2001, acorns were collected from 40 Valley oak adults during a year of very low acorn production. The TWOGENER analysis is based on 33 adult trees, 288 progeny with 5–12 progeny

per adult, and six microsatellite loci (C. Dutech and V. L. Sork, in preparation).

The estimated average pollination distance (Table 7) was ~100 m for the normal and exponential distributions, if density was set at its observed value (1.19 trees/ha). The exponential power distribution showed a slightly fat tail, as for *D. excelsa*, with an estimate of $\hat{b} = 0.847$. All families of distributions yielded a similar estimate for the average pollination distance, $\hat{\delta} \cong 120$ m.

When density was estimated jointly with the shape parameter (*a*), setting the shape parameter to the estimated value, we obtained in all cases an effective density $\hat{d}_e \cong d/10$, except for the geometric distribution, for which the estimated effective density was quite a bit closer to the observed density. The estimated average pollination distance ($\hat{\delta}$ ~300 m for all distributions) was higher than we have reported elsewhere (Sork *et al*. 2002a), obtained with the assumption that $d_e \sim d$. For *Q. lobata*, it was also possible to estimate the three parameters jointly (last line for each distribution in Table 7), yielding $d_e \cong d/10$ average pollination distance of $\hat{\delta}$ ~300 m, and a mildly fat tail ($\hat{b} = 0.713$ for the exponential power distribution).

Distributional overview

Comparing the fits obtained with the different functions, we observed in all cases the same order, starting from the best fitting function: Weibull, exponential power, geometric, 2*Dt*, exponential, normal. The general shape of the dispersal curves was rather similar for *Q. lobata* and *D. excelsa*, but the curves were more divergent for the 1999 and 2000 collections of *S. torminalis* (Fig. 2). Similarly, the estimated moments were more similar in the cases of *Q. lobata* and *D. excelsa* than for *S. torminalis* (see Tables 4–7).
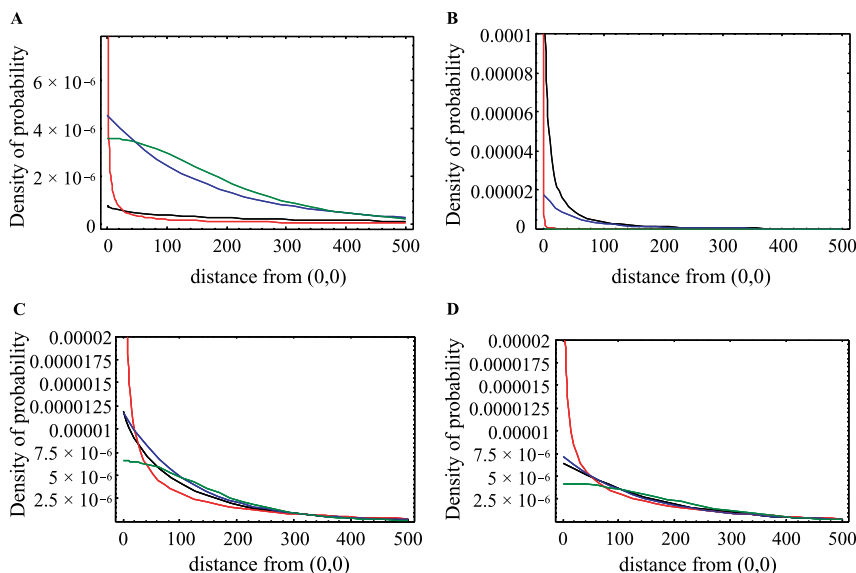

Fig. 2 Best fitting curves for the four experimental data sets, *Sorbus torminalis* 1999 (A) and 2000 (B), *Dinizia excelsa* (C) and *Quercus lobata* (D), for the four families of curves studied, exponential power (black line), Weibull (red line), geometric (blue line) and 2*Dt* (green line) They correspond to the case where the shape parameter (*b*) was set to a fixed value and density was jointly estimated (second line of Tables 4–7).

Table 7 Analytical results of the TWOGENER analysis for *Quercus lobata*, for the various dispersal distributions and with the different estimation strategies, fixed *vs.* estimated parameters. The minimum squared error criterion, $C(\hat{d}, \hat{\theta})$, is also provided

Dispersal function	Fixed parameters		Estimated parameters			Estimated moments (m)				Error
	$d*$	b	$\hat{d}$	$\hat{a}$†	$\hat{b}$	$\hat{\delta}$	$\hat{\mu}_2$	$\hat{\mu}_3$	$\hat{\mu}_4$	
Normal	1.19	2	—	—		101	114	125	136	1.763
Normal	—	2	0.177	—		241	272	299	323	1.715
Exponential	1.19	1	—	57.0	—	114	140	164	189	1.761
Exponential	—	1	0.161	143	—	286	350	412	473	1.711
Exponential power	1.19	—	—	42.0	0.847	121	152	184	216	1.760
Exponential power	—	0.847	0.152	108	—	310	392	473	555	1.710
Exponential power	—	—	0.137	75.8	0.713	353	462	576	693	1.710
Weibull	1.19	—	—	141	1.38	129	160	188	214	1.758
Weibull	—	1.38	0.138	379	—	346	429	505	575	1.705
Weibull	—	—	0.141	367	1.40	334	413	484	550	1.705
Geometric	1.19	—	—	2387	44.1	116	144	172	200	1.761
Geometric	—	44.1	0.986	2615	—	127	158	188	219	1.756
Geometric	—	—	0.157	5460	40.09	294	366	437	508	1.171
2Dt	1.19	—	—	230	6.77	90.9	105	119	133	1.827
2Dt	—	—	0.159	664	—	262	304	344	383	1.760
2Dt	—	—	0.158	630	6.23	264	306	348	390	1.760

*units for d are trees/ha.
†units for $\hat{a}$ are metres.

Table 8 Ratio of estimated density over the measured density ($\hat{d}/d$) for the three species studied and the six families of dispersal curves assumed

Dispersal function	*Sorbus torminalis* (1999)	*Sorbus torminalis* (2000)	*Dinizia excelsa*	*Quercus lobata*
Normal	0.282	6.909	0.835	0.101
Exponential	0.158	4.636	0.665	0.091
Exponential power	0.064	0.036	0.585	0.093
Weibull	0.002	0.00038	0.472	0.084
Geometric	0.122	0.260	0.765	0.981
2Dt	0.172	0.506	0.712	0.099

When density was estimated jointly with the scale parameter *a*, assuming a fixed value for the shape parameter (*b*), the estimated effective density ($\hat{d}$) was always lower than the field measured adult density (*d*), except for the normal and exponential distribution for *S. torminalis* in 2000. The ratio $\hat{d} : d$ varied between the species (Table 8): being ~0.1 for *Q. lobata*, whatever dispersal function was assumed (except the geometric); from 0.064 to 0.158 for *S. torminalis* in 1999, except for the Weibull distribution where it took a very low value of 0.0016; and from 0.472 to 0.835 for *D. excelsa*. The results were quite variable for *S. torminalis* in 2000, but for the best fitting functions, the Weibull and the exponential power families, $\hat{d} : d$ was also quite small, 0.00038 and 0.036, respectively.

Paternity analysis comparisons

We also performed paternity analyses on *Sorbus torminalis*, for which we have substantial genetic information on the males for a considerable distance around the sampled females (Oddou-Muratorio *et al*. in press), using the program CERVUS (Marshall *et al*. 1998). Then the best-fitting dispersal curve was determined, using a maximum-likelihood method (see details in S. Oddou-Muratorio, E. K. Klein and F. Austerlitz, in preparation). As with the less-detailed TWOGENER analysis, the fits show a tendency towards fat-tailed dispersal curves (Table 9) when the exponential power distribution was used with estimated values for *b* of 0.25 and 0.35, respectively. When the fits with other curves were compared, the Weibull distribution provided the poorest fit for both years, the best fitting curve was the geometric for 1999 and the exponential power for 2000, but the difference between the likelihood of the two curves was very small in both cases.

Discussion

Our results indicate that it is possible to approximate the shape of the dispersal curve, a critical element of the estimation of gene flow parameters. The simulation study showed that TWOGENER can infer the shape of the dispersal curve, even if the estimations show some bias and large MSE. We always used the exponential power distribution

Table 9 Fit obtained with the maximum likelihood method based on paternity analysis

Dispersal curve	-log Likelihood	$\hat{a}$	$\hat{b}$	$\hat{\delta}$
1999 data				
Normal	835.24	448.11		397
Exponential	753.68	158.45		317
Exponential power	700.40	0.10	0.25	648
Weibull	705.99	398.77	0.87	429
Geometric	699.50	22.35	2.06	∞
2Dt	703.35	27.56	1	∞
2000 data				
Exponential	1196.31	215.86		432
Exponential power	1150.42	4.21	0.35	847
Weibull	1159.72	588.72	1.21	553
Geometric	1150.99	67.17	2.10	∞
2Dt	1155.50	67.16	1.0000426	∞

for simulated data sets, and when this same family was used for estimation, all values of $\hat{b}$ but one were < 1.0, when the true value of *b* was 0.5, as expected of a fat-tailed distribution. Conversely, when the true value was *b* = 2, all estimated values of $\hat{b}$ were substantially above 1.0, as expected of a thin-tailed distribution. Thus, the exponential power distribution was effective at determining whether the dispersal curve was fat- or thin-tailed.

Of course, the accuracy of the estimates will increase with the amount of data available. In a previous study (Austerlitz & Smouse 2002), assuming a normal dispersal for the TWOGENER analysis, which is the only case for which many simulated data sets can be analysed, we have shown that increasing the number of sampled mothers or increasing the number of loci used in the TWOGENER analysis were the best way to reduce the MSE of the estimates. The impact of the amount of data available should also be studied in the future for the case of the more realistic curves that we studied here.

Another interesting result is that we obtained no more bias and no larger MSE in the estimation of average pollen dispersal distance and adult density if we assumed another family of dispersal curves (here the Weibull or the geometric distributions) than that used in the simulations (the

exponential power distribution). This remained also true for the higher order moments for the Weibull distribution, and also for the geometric distribution, except that we sometimes obtained dispersal curves with infinite moments. This general congruence is important because, for experimental data sets, the assumed dispersal curve is unknown and it is thus helpful that these estimated values are not strongly affected by the choice of dispersal curve. All things considered, we recommend the use of the exponential power distribution, which is sufficiently flexible and has the added attraction of encompassing the normal and exponential as degenerate special cases, and the moments are always finite. The Weibull distribution becomes advantageous only when the distribution is thin-tailed, because the shape parameter is estimated with more precision. Thus, the best strategy might be to start first with the exponential power, and in the case that a thin-tailed curve is obtained, to perform a subsequent analysis with the Weibull distribution.

The frequency of thin-tailed pollen-mediated dispersal may not be high in the real word. We need more studies of pollen dispersal to verify the overall tendency, but the cases studied so far suggest a trend towards fat-tailed distributions of pollen dispersal. Indeed, when the exponential power function was assumed, the TWOGENER method yielded estimates of b that were close to 1, or even smaller, i.e. a distribution that decreased more slowly than the exponential distribution at long distance. For the case of *Sorbus torminalis*, for which we had data sets for two consecutive years, it is interesting to note a consistent pattern of strongly fat-tailed curves over both years for that species, and this is confirmed by paternity analysis. However, the level of precision of our estimates does not allow us to draw any conclusion from the fact that we obtained a lower b-value in 2000 than in 1999, especially since the paternity analysis yielded the opposite result.

Another consistent pattern was that the estimated adult density on the experimental data sets was always below the observed density, but since our estimates are affected by bias and have large MSEs, we must be cautious with our interpretation. On the one hand, this result might indicate that it is preferable to have an estimate of adult density to insert into the model, stabilizing the other parameter estimates. Notably, we are estimating several parameters simultaneously, and the variance in the data around the curve is great. It is important to understand whether our estimate of low density has validity, because otherwise our method also estimates larger dispersal distances than indicated by estimates based on measured density.

On the other hand, it is not unexpected that the effective adult density is less than the demographic estimate, given the many factors (e.g. uneven reproductive output, asynchronous flowering phenology) that reduce the effective size of the reproductive population. For example, for *S. torminalis* and *Quercus lobata*, we obtained estimated

effective densities that were ~1/10 of the observed density. From what we know from the simulation study about the precision of the estimates, we can conclude here with reasonable certitude that the effective density is lower than that observed. By contrast, the estimated density was about half of the observed density for *Dinizia excelsa*. Considering the level of MSE observed in our simulation study, we cannot conclude clearly for that case that effective density is indeed lower than that observed in this last case.

Low effective density may be an indication of a variance in reproductive success among males, which is illustrated by paternity analysis in the case of *S. torminalis* (S. Oddou-Muratorio, E. K. Klein and F. Austerlitz, in preparation). This variance in reproductive success can be a consequence of variation in phenology of the species under study. Indeed, if the flowering periods of some individuals are completely nonoverlapping, the probability of a mating event becomes nil, and it is reduced considerably when these periods overlap only partially. It is interesting that we observed this phenomenon for *S. torminalis*, but not for *D. excelsa*. Also, large variations in pollen or pistillate flower production among trees could be a cause of variance in reproductive success.

We need many other experimental studies to confirm all the tendencies that have been observed in the experimental data sets studied here. The species studied here are scattered, with low-density spatial distributions; two are insect-pollinated (*S. torminalis* and *D. excelsa*), and the other is wind-pollinated. More data are needed, both on these types of species and on species with different life histories. In particular it would be interesting to study densely packed wind-pollinated species. A similar study was performed on *Fraxinus excelsior* (Morand-Prieur 2003), a temperate wind-pollinated species that occurs in much higher densities (48 trees/ha in the study plot) than studied here, but we found a similar pattern, a very fat-tailed dispersal curve ($b = 0.240$, if the exponential power function was assumed) and an effective density slightly smaller than that observed. Also, it would be interesting to repeat the study for the same species over several years, as we did for *S. torminalis*, which showed us that the same trend toward a fat-tailed curve was observed for both years. Also repetitions in different places, subject to different environmental conditions, would be useful to assess the extent to which the curve estimated during one year and in one place can be extrapolated to results observed in other years and other places.

From a theoretical point of view, it would be useful to perform a more detailed study to determine the best sampling strategy to estimate all the dispersal parameters. This information would probably help determine the optimal way to estimate all of the parameters of the dispersal curve effectively. More generally, it would provide useful information about the best strategy of how to choose mothers

on the landscape and the sample allocation needed to perform the estimation properly. Ultimately, information on the distance and shape of the dispersal function will yield insight into the scale of the evolutionary neighbourhood size and about the dynamics of gene movement in managed populations.

As already mentioned in previous works (Austerlitz & Smouse 2001) several factors may affect the estimation process. For instance, a degree of selfing higher than what is expected at random, will increase the differentiation of the pollen clouds (Φ_{ft}), by increasing the proportion of self-gametes in each pollen cloud. Our method, which does not account for this effect, would thus provide a downwardly biased estimate of the distance of dispersal for the outcrossing pollen. A reduction of selfing will have the opposite effect. Similarly, for animal-pollinated species, correlated dispersal of male gametes will decrease the effective number of pollinators of each female and thus increase Φ_{ft}, again yielding a downward bias in our estimates of dispersal distance. In all cases, the impact on the estimation of the shape of the curve is less predictable. Such processes should be integrated in future studies.

Acknowledgements

References

Abramowitz M, Stegun IA (1964) *Handbook of mathematical functions with formulas, graphs, and mathematical tables* U.S. Govt. Print. Off., Washington. D.C..

Austerlitz F, Smouse PE (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.

Austerlitz F, Smouse PE (2002) Two-generation analysis of pollen flow across a landscape. IV. Estimating the dispersal parameter. *Genetics*, **161**, 355–363.

Bateman AJ (1947) Contamination of seed crops. II. Wind pollination. *Heredity*, **1**, 235–246.

Clark JS (1998) Why trees migrate so fast: confronting theory with dispersal biology and the paleorecord. *American Naturalist*, **152**, 204–224.

Clark JS, Silman M, Kern R, Macklin E, HilleRisLambers J (1999) Seed dispersal near and far: patterns across temperate and tropical forests. *Ecology*, **80**, 1475–1494.

Crawford TJ (1984) The estimation of neighborhood parameters for plant populations. *Heredity*, **52**, 273–283.

De Oliveira AA, Mori SA (1999) A central Amazonian terra firme forest. I. High tree species richness on poor soils. *Biodiversity and Conservation*, **8**, 1219–1244.

Devlin B, Ellstrand NC (1990) The development and application of a refined method for estimating gene flow from angiosperm paternity analysis. *Evolution*, **44**, 258–259.

Dick CW (2001a) Genetic rescue of remnant tropical trees by an alien pollinator. *Proceedings of the Royal Society of London Series B: Biology Sciences*, **268**, 2391–2396.

Dick CW (2001b) Habitat change, African honeybees and fecundity in the Amazonian tree *Dinizia excelsa* (Fabaceae). In: *Lessons from Amazonia: the Ecology and Conservation of a Fragmented Forest* (eds Bierregaard RO, Gascon C, Lovejoy TE, Mesquita R), pp. 149–157. Yale University Press, New Haven.

Dick CW, Etchelecu G, Austerlitz F (2003) Pollen dispersal of tropical trees (*Dinizia excelsa*: Fabaceae) by native insects and African honeybees in pristine and fragmented Amazonian rainforest. *Molecular Ecology*, **12**, 753–764.

Ducke A (1922) Plantes nouvelles ou peu connues de la région amazonienne. *Archivos Do Jardim Botanico Do Rio de Janeiro*, **3**, 2–129.

Ennos RA (1994) Estimating the relative rates of pollen and seed migration among plant populations. *Heredity*, **72**, 250–259.

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

Griffin JR (1971) Oak regeneration in the upper Carmel Valley, California. *Ecology*, **52**, 862–868.

Hudson RR (1998) Island models and the coalescent process. *Molecular Ecology*, **7**, 413–418.

Irwin AJ, Hamrick JL, Godt MJ, Smouse PE (2003) A multiyear estimate of the effective pollen donor pool for *Albizia julibrissin*. *Heredity*, **90**, 187–194.

Lande R, Barrowclough GF (1987) Effective population size, genetic variation, and their use in population management. In: *Viable Populations for Conservation* (ed. Soulé ME), pp. 87–123. Cambridge University Press, Cambridge.

Laurance WF, Lovejoy TE, Vasconcelos HL *et al.* (2002) Ecosystem decay of Amazonian forest fragments: a 22-year investigation. *Conservation Biology*, **16**, 605–618.

Marshall TC, Slate J, Kruuk LE, Pemberton JM (1998) Statistical confidence for likelihood-based paternity inference in natural populations. *Molecular Ecology*, **7**, 639–655.

McCauley DE (1997) The relative contributions of seed and pollen movement to the local genetic structure of *Silene alba*. *Journal of Heredity*, **88**, 257–263.

Morand-Prieur M-E (2003) Evolution et maintien d'un système de reproduction polymorphe. Approche génétique et écologique de la polygamie chez le frêne commun, *Fraxinus excelsior* L. PhD Thesis, ENGREF, Paris, France.

Oddou-Muratorio S, Petit RJ, Le Guerroue B, Guesnet D, Demesure B (2001) Pollen- versus seed-mediated gene flow in a scattered forest tree species. *Evolution*, **55**, 1123–1135.

Oddou-Muratorio S, Houot M-L, Demesure-Mush B, Austerlitz F (2003) Pollen flow in the wildservice tree, *Sorbus torminalis* (L.) Crantz. I. Evaluating the paternity analysis procedure in continuous populations. *Molecular Ecology*, **12**, 3427–3439.

Rousset F (1997) Genetic differentiation and estimation of gene

flow from *F*-statistics under isolation by distance. *Genetics*, **145**, 1219–1228.

Rousset F (2000) Genetic differentiation between individuals. *Journal of Evolutionary Biology*, **13**, 58–62.

Slavov GT, DiFazio SP, Strauss SH (2002) Gene flow in forest trees: from empirical estimates to transgenic risk assessment. In: *Scientific Methods Workshop: Ecological and Agronomic Consequences of Gene Flow from Transgenic Crops to Wild Relatives* (eds Snow A, Mallory-Smith C, Ellstrand N, Holt J, Quemada H), pp. 113–133. Ohio State University, Colombus, OH.

Smouse PE, Long JC (1992) Matrix correlation analysis in anthropology and genetics. *Yearbook of Physical Anthropology*, **35**, 187–213.

Smouse PE, Dowling TE, Tworek JA, Hoeh WR, Brown WM (1991) Effects of intraspecific variation on phylogenetic inference: a likelihood analysis of mtDNA restriction site data in cyprinid fishes. *Systematic Zoology*, **40**, 393–409.

Smouse PE, Dyer RJ, Westfall RD, Sork VL (2001) Two-generation analysis of pollen flow across a landscape. I. Male gamete heterogeneity among females. *Evolution*, **55**, 260–271.

Sork VL, Nason J, Campbell DR, Fernandez JF (1999) Landscape approaches to the study of gene flow in plants. *Trends in Ecology and Evolution*, **142**, 219–224.

Sork VL, Davis FW, Smouse PE *et al.* (2002a) Pollen movement in declining populations of California Valley oak, *Quercus lobata*: where have all the fathers gone? *Molecular Ecology*, **11**, 1657–1668.

Sork VL, Dyer RJ, Davis FW, Smouse PE (2002b) Mating patterns in a savanna population of valley oak (*Quercus lobata* Neé). In: *Proceedings of the Fifth Symposium on Oak Woodlands: Oaks in California's Changing Landscape, October 22–25, 2001* (eds Standiford R, McCreary D, Purcell KL), pp. 427–439. Pacific SW Research Station, US Forest Service, USDA, San Diego, CA.

Tufto J, Engen S, Hindar K (1997) Stochastic dispersal processes in plant populations. *Theoretical Population Biology*, **52**, 16–26.

Vitalis R, Couvet D (2001a) Estimation of effective population size and migration rate from one- and two-locus identity measures. *Genetics*, **157**, 911–925.

Vitalis R, Couvet D (2001b) Two-locus identity probabilities and identity disequilibrium in a partially selfing subdivided population. *Genetical Research*, **77**, 67–81.

Weibull W (1951) A statistical distribution function of wide applicability. *Journal of Applied Mechanics*, **18**, 293–297.

Wolfe KH, Li WH, Sharp PM (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proceedings of the National Academy of Sciences of the USA*, **84**, 9054–9058.

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

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