

# Pollen flow in the wildservice tree, *Sorbus torminalis* (L.) Crantz. II. Pollen dispersal and heterogeneity in mating success inferred from parent–offspring analysis

SYLVIE ODDOU-MURATORIO,\* ETIENNE K. KLEIN† and FRÉDÉRIC AUSTERLITZ‡

\*Conservatoire Génétique des Arbres Forestiers, Office National des Forêts, Campus INRA, F-45160 Ardon, France, †Unité de recherche Biométrie, INRA Domaine Saint-Paul – Site Agroparc, F-84914 Avignon cedex 9, France, ‡Laboratoire Evolution et Systématique, Université Paris-Sud, Bât 362, F-91405 Orsay cedex, France

## Abstract

Knowing the extent of gene movements from parents to offspring is essential to understand the potential of a species to adapt rapidly to a changing environment, and to design appropriate conservation strategies. In this study, we develop a nonlinear statistical model to jointly estimate the pollen dispersal kernel and the heterogeneity in fecundity among phenotypically or environmentally defined groups of males. This model uses genotype data from a sample of fruiting plants, a sample of seeds harvested on each of these plants, and all males within a circumscribed area. We apply this model to a scattered, entomophilous woody species, *Sorbus torminalis* (L.) Crantz, within a natural population covering more than 470 ha. We estimate a high heterogeneity in male fecundity among ecological groups, both due to phenotype (size of the trees and flowering intensity) and landscape factors (stand density within the neighbourhood). We also show that fat-tailed kernels are the most appropriate to depict the important abilities of long-distance pollen dispersal for this species. Finally, our results reveal that the spatial position of a male with respect to females affects as much its mating success as ecological determinants of male fecundity. Our study thus stresses the interest to account for the dispersal kernel when estimating heterogeneity in male fecundity, and reciprocally.

**Keywords:** fat-tailed dispersal, maximum likelihood, paternity analysis, *Sorbus torminalis* L. Crantz, TWOGENER

Received 27 April 2005; revision accepted 25 July 2005

## Introduction

Recent developments have brought significant new insights on patterns of gene movements from parents to offspring in sexually reproducing plants. The idea of generally restricted pollen/seed flow (Levin & Kerster 1974) has been contradicted by various experimental studies, which allowed rare long-distance events of dispersal to be traced (e.g. Streiff *et al.* 1999; Dick 2001 for pollen dispersal; Godoy & Jordano 2001 for seed dispersal). These rare events will in particular directly favour the spread of advantageous mutation over the landscape and reduce the level of inbreeding. However, these long-distance dispersal

events are exceptional. The great majority of the pollen and seeds disperses only locally and among the reproductive male/female individuals, the contributions to reproduction are highly uneven (Godoy & Jordano 2001; Burczyk *et al.* 2002; Wright & Meagher 2004). This local dispersion and this uneven participation to reproduction, by reducing effective population size ( $N_e$ ) can be dangerous for the populations, since it increases the risk of loss of advantageous allele, of fixation of deleterious alleles and of extinction.

Thus, the future of a population will depend in part upon the balance between long- and short-distance dispersal and the level of heterogeneity in male or female fecundity. Estimating these different processes is therefore a key step to understand the evolution of populations, in particular to predict the reaction of species to environmental changes, like global warming or forest fragmentation (e.g. Sork *et al.*

Correspondence: S. Oddou-Muratorio, Present address: INRA–URFM, 20, Avenue Vivaldi F-84000 Avignon, France. Fax: +33-490 135 959; E-mail: oddou@avignon.inra.fr

2002; Dick *et al.* 2003). This issue requires to account not only for the mean dispersal distance, but also for the shape of the dispersal curve, because the frequency of long-distance dispersal has strong consequences, for instance by connecting habitat plots across the landscape (Dick 2001). Moreover, since the reproductive success of an individual is both the result of its intrinsic fertility and of the capacity of its propagules to disperse, it is of interest to jointly estimate these two quantities (Smouse *et al.* 1999; Morgan & Conner 2001; Burczyk *et al.* 2002; Wright & Meagher 2004). In both cases, the estimation procedure relies on the study of a sample of adults and offspring in a given population. In the case of gene flow studies, one will focus on the distribution of the distances between parents and offspring, while in the case of the study of heterogeneity in fecundity, one will focus on the number of offspring that an adult will produce.

Focusing first on gene flow, the natural approach is to estimate the dispersal kernel (e.g. Wright 1943). It is defined as the probability density function of the final position of the propagules carrying the genes (seeds or pollen grain) relative to the position of the adult from which it was issued. The dispersal kernel enables to represent the fine variations of gene flow with distance. This is quite important, since the impact of the general shape of the gene dispersal kernel (leptokurtic vs. platykurtic kernels) and of the specific shape of its tail (fat-tailed vs. thin-tailed kernels) on major processes in population biology has been highlighted by various theoretical and experimental studies. For instance, the pollen dispersal kernel allows one to gauge the risk of contamination of seed crops by other fields (Bateman 1947), while the seed dispersal kernel affects strongly the rate of colonization and the diversity in newly founded populations (e.g. Le Corre *et al.* 1997; Clark 1998; Nathan & Muller-Landau 2000).

Parentage/paternity analyses based on genetic markers provide a powerful approach to infer particular events of dispersal by seed/pollen, and thus to obtain direct estimate of dispersal. We focus here on the use of paternity analysis for examining pollen dispersal, which consists in using a set of polymorphic markers to genotype a sample of fruiting plants, a sample of seeds harvested on each of these plants, and all males within a circumscribed area, in order to detect the father of each seed (categorical assignment; Meagher 1986) or to evaluate the likelihood of each male as the father of the considered seed (fractional assignment; Devlin & Ellstrand 1990). Thanks to recent advances in the development of genetic markers, pollen dispersal events can be efficiently traced in natural populations (Streiff *et al.* 1999; Dick 2001). Yet, most dispersal studies based on paternity analysis use the inferred mating events only to compute the distribution of the distances between male and female parents, without translating this distribution of distances into a dispersal kernel (e.g. Streiff *et al.*

1999). In addition, this inferred distribution of dispersal distance often does not account for heterogeneity in male fecundity (HMF); yet, if this heterogeneity is not randomly distributed with respect to sampled mother trees, the estimation of dispersal may be biased.

On the other hand, paternity analyses based on genetic markers also provide a powerful approach to infer the relative number of offspring sired by each male and thus to estimate the heterogeneity of male mating success (MMS), and the relationship between phenotype and MMS (Roeder *et al.* 1989; Meagher 1991; Nielsen *et al.* 2001). Most existing models account for quantitative traits as determinant of MMS, and attempt to estimate one or two parameters relying a given trait value to MMS (the so-called selection gradient: Smouse *et al.* 1999; Morgan & Conner 2001; Wright & Meagher 2004). These models usually account for the effect of distance between males and females on MMS. However, they most often assume an exponential model of pollen dispersal (Burczyk *et al.* 2002; Jones 2003). Consequently, these approaches cannot provide estimates of the shape of the dispersal kernel, in particular of its tail. Additionally, if dispersal patterns strongly differ from what is predicted under exponential distribution (Austerlitz *et al.* 2004), they may be significantly biased.

To overcome these limitations, we propose a method that takes advantage of highly efficient markers for paternity analysis to estimate the shape and scale of pollen dispersal and the impact of different ecological factors on male fecundity simultaneously. We developed a nonlinear statistical model that allows the joint estimation of the pollen dispersal kernel and of HMF among ecological groups of individuals from genotypic data of adults and seeds collected on the mother trees. Besides estimation, this model also allows gauging the respective impact of spatial processes and heterogeneity in fecundity on MMS. This method is related to the fractional attribution of paternity, avoiding the step of categorical paternity assignment, which leads necessarily to wrong assignments and to miss some of the true paternities (Oddou-Muratorio *et al.* 2003). We chose a family of dispersal kernels that contains both thin- and fat-tailed curves, depending on the value of one of the parameters controlling the tail. Regarding the HMF, we translated it into an estimate of effective density, which is the equivalent for a spatially structured population to the effective population size in classical population genetics.

Our new method is related to the approach of Burczyk *et al.* (2002). We use the same neighbourhood model originally developed by Adams & Birkes (1991), which considers that offspring of individual mother plants are the product of either self-fertilization, outcrossing with males within a circumscribed area around each mother plant (the neighbourhood), or outcrossing with males outside the neighbourhood. However, whereas Burczyk *et al.* (2002) explicitly model pollen dispersal through an exponential

distribution in order to evaluate the effect of male–female distance on mating success, our approach consists in directly estimating the shape and the scale of the dispersal kernel. This difference may be of importance, as the exponential distribution may not be the most appropriate function to depict pollen dispersal in plants (Austerlitz *et al.* 2004).

We apply this new method to study mating patterns in a population of a European entomophilous forest tree, *Sorbus torminalis* (L.) Crantz (the wildservice tree). This scattered species combines extensive range with low local density (usually less than one individual per hectare), and is thus an interesting biological model to investigate pollen flow over a large area. Additionally, low-density entomophilous species like *S. torminalis* may be particularly vulnerable to landscape changes induced by forest management. In particular, trees left standing in low-density area opened by logging may become reproductively isolated, due to the disruption of pollinators (Rocha & Aguilar 2001).

To assess the importance of jointly estimating the dispersal kernel and the HMF, we first performed a joint estimation of these two vectors of parameters. Then, we estimated the dispersal kernel without assuming any HMF. Similarly, we estimated the heterogeneity of male fecundity without assuming any limitation of gene dispersal in the population. Then we compared the different estimates and gauged the increase in precision obtained when all parameters were estimated simultaneously.

This study addresses the following biological and methodological issues: (i) What is the extent of pollen dispersal and what is the proportion of long-distance dispersal? (ii) What is the relative impact of ecological factors on MMS, in comparison to spatial processes, and what will be their impact on the effective density of reproducing trees? (iii) Is it important to jointly estimate the dispersal kernel and the HMF? (iv) To which extent does our method, which requires exhaustive genotyping of the potential fathers within a given area, perform better than less work-intensive methods, such as the TWOGENER method used in a previous study on the same data set (Austerlitz *et al.* 2004)? (v) What are the consequences for conservation and long-term population evolution of the interaction between the dispersal process and patterns of heterogeneity in fecundity?

## Materials and methods

### Biological material

*Studied organism.* *Sorbus torminalis* (L.) Crantz ('the wildservice tree') is a monoecious woody species of the Rosaceae family (subfamily Maloideae). It produces hermaphroditic flowers, pollinated by a wide range of Hymenoptera, Diptera and Coleoptera (Griffin & Sedgley 1989; S. O.-M., personal observation).

*Study site and adult mapping.* The study site covers 472 ha, of mixed broadleaved stands (all details and a map of the study area are given in Oddou-Muratorio *et al.* (2003). It is a part of the managed Rambouillet forest in France (total area: 15 000 ha). Since 1960, forest management consists in a regular cycle of thinning (through selective logging, sparing *S. torminalis*) ending with clear-cutting. Within this site, we found 185 potentially flowering wildservice trees.

For all adult trees within the study site, we measured two phenotypic variables: (i) the diameter at breast height (DBH), which we translated into a class variable with six levels (from < 10 cm to < 60 cm), and (ii) the flowering intensity as a class variable with five roughly assessed levels (absent, anecdotal, significant, massive and missing data). The partition into discrete classes is necessary in our approach (see below). In any case, for flowering intensity, it is very difficult to quantify the number of flowers for trees, so other authors have also treated it as an ordered qualitative variable (Burczyk & Prat 1997). Additionally, we characterized the neighbourhood density (independently of species) through a class variable with five levels (isolated tree, hedge tree, dominant stand-tree, codominant stand-tree, suppressed stand-tree). The three last levels aimed at accounting for differences in light exposition among trees with crown, respectively, above, within, or below the surrounding canopy. Following Nielsen *et al.* (2001), the two phenotypic variables and the neighbourhood density will be denoted ecological factors in the following.

*Offspring sampling.* We genotyped 653 seeds collected in autumn 1999 in the crown of 14 fruiting trees (11–100 seeds/tree, mean = 46.6, SD = 21.1). The distances between the sampled mother trees ranged from 1.34 m to 2737 m, with a mean of 803 m.

We also genotyped 1075 seeds (2–27 seeds/tree) collected in autumn 2000 on 60 fruiting trees, including the 14 mothers sampled in 1999 (average distance between mother trees = 1016 m, max. = 2747 m). Thus, compared to 1999, the sampling design in 2000 involved many more mother trees, with smaller but more regular progeny arrays per mother (mean number of seeds/tree = 17.9, SD = 3.9). The objectives for both years were to sample all possible distances between mother trees, and to maximize the number of mother trees in the middle part of the study site.

*Microsatellite typing.* The genotypes of all the adults within the studied population and of the offspring were scored at six microsatellite loci: MSS1, MSS5, MSS6, MSS9, MSS13 and MSS16 (Oddou-Muratorio *et al.* 2001b). The theoretic exclusion probability over all six loci was 0.9834 for the 149 flowering adults of 1999 and 0.9869 for the 172 flowering adults of 2000 (Oddou-Muratorio *et al.* 2003).

*Modelling the dispersal kernel and heterogeneity in fecundity*

*Model for dispersal kernel.* We modelled pollen dispersal using a dispersal kernel  $p(\cdot; x,y)$  describing the probability for a pollen grain emitted at position (0,0) to participate to the pollen cloud at position (x,y) (Wright 1943; Clark 1998). We considered here the family of exponential power functions:

$$p(a,b; x,y) = \frac{b}{2\pi a^2 \Gamma(2/b)} \exp \left[ - \left( \frac{\sqrt{x^2 + y^2}}{a} \right)^b \right], \quad (\text{eqn 1})$$

where  $\Gamma$  is the classically defined gamma function (see, e.g. Abramowitz & Stegun 1964). The parameter  $b$  is the shape parameter affecting the tail of the dispersal function and  $a$  is a scale parameter homogeneous to a distance (Clark 1998). This function reduces to the bivariate normal distribution for  $b = 2$ , and to the bivariate exponential distribution for  $b = 1$  (see Austerlitz *et al.* 2004). The mean distance ( $\delta$ ) travelled by a pollen grain under the kernel  $p(a,b)$  is given by:

$$\delta = a \frac{\Gamma(3/b)}{\Gamma(2/b)} \quad (\text{eqn 2})$$

It increases with  $a$  and decreases with  $b$ . When  $b < 1$ , the dispersal kernel is fat-tailed (Clark 1998), i.e. the long-range decrease is slow (at least slower than an exponential of the distance), increasing the probability of long-distance dispersal events. Conversely, when  $b > 1$  (for instance the Gaussian model, for which  $b = 2$ ) the dispersal is thin-tailed, with a rapid decrease of the dispersal function, implying much less long-distance dispersal events than when  $b < 1$  (see Austerlitz *et al.* 2004 for further details). When  $b > 2$ , the dispersal kernel is platykurtic, meaning less long-distance dispersal events than the Gaussian kernel with the same variance. When  $b < 2$ , the dispersal kernel is leptokurtic.

*Models for male fecundity.* Unlike Burczyk *et al.* (2002), who considered only either binary traits or quantitative variables as determinant of MMS, our model groups the plants into several discrete classes for a given trait, and allows the estimation of relative fecundities of each class. Here, we partitioned the adults into the several ecological groups described previously. Denote  $f_c$  the relative male fecundity of a male belonging to the  $c$ th class of a given factor. The fecundity  $\Phi_k$  for a given male  $k$  is then given by  $f_{c(k)}$  where  $c(k)$  is the class to which male  $k$  belongs. For instance, to model the impact of flowering intensity on male fecundity, we set  $f_1 = 1$  for trees having shown anecdotal blooming corymbs,  $f_2 = u_2$  for abundant blooming corymbs and  $f_3 = u_3$  for massive blooming corymbs.

To take into account jointly several effects on the male fecundity, we did not consider any interaction effect between ecological variables. For instance, to take into account simultaneously flowering intensity and diameter, we set  $\Phi_k = u_{c_k} \times t_{d_k}$  for a tree  $k$  belonging to class  $c_k$  of flowering intensity and class  $d_k$  of diameter.

*Estimation of pollen dispersal kernel and HMF*

In this section, we derive the general maximum-likelihood estimator of the vector of dispersal parameters ( $\theta_d$ ), of the vector of the relative fecundities ( $\mathbf{F}$ ), and of selfing ( $s$ ) and pollen immigration ( $m$ ) rates. Here,  $\theta_d$  consists only in the parameter  $a$  if we consider a normal/exponential dispersal kernel; or in  $a$  and  $b$  if we consider an exponential power dispersal kernel.  $\mathbf{F}$  contains the relative fecundities of the different phenotypic/landscape classes of males (see above).

Our mating model is a modified version of the neighbourhood model of Burczyk *et al.* (2002). The main differences are that (i) we considered that all sampled adults are a single neighbourhood for all mother trees; (ii) within neighbourhood, we assumed a two-parameter dispersal kernel; and (iii) we assumed that males belonged to discrete fecundity classes according to ecological variables.

We considered that each offspring  $o$  sampled on mother  $j_o$  may be the result either of self-pollination (with probability  $s$ ), or of pollen flow from outside the investigated area (with probability  $m$ ), or of pollen from a sampled male (with probability  $1 - m - s$ ). We obtained a maximum-likelihood estimate of the dispersal parameters ( $\theta_d$ ), of the vector of male fecundities ( $\mathbf{F}$ ), of the selfing rate ( $s$ ) and of the pollen immigration rate ( $m$ ) by maximizing the log-likelihood function:

$$\log L(\theta_d, \mathbf{F}, s, m) = \sum_{o=1}^O \log \left[ sT(g_o | g_{j_o}, g_{j_o}) + (1 - s - m) \sum_{l \in \mathbf{F}} \pi_{j_o, l}(\theta_d, \mathbf{F}) T(g_o | g_{j_o}, g_l) + mT(g_o | g_{j_o}, AF) \right] \quad (\text{eqn 3})$$

where  $\pi_{jk}(\theta_d, \mathbf{F})$  is the expected proportion of the pollen of a given male  $k$  in the local pollen cloud of mother  $j$  for a given set of dispersal parameters and relative male fecundities, and where  $O$  is the total number of genotyped offspring.  $T(g_o | g_{j_o}, X)$  is the Mendelian segregation probability (e.g. Meagher 1986) of the offspring genotype ( $g_o$ ) given the genotype of the mother ( $g_{j_o}$ ) and  $X$ .  $X$  corresponds to the genotype of the mother in case of selfing; to the genotype of the considered male ( $g_l$ ) in case of outcrossing with a sampled male; or to the allelic frequencies in the pollen cloud external to the neighbourhood ( $AF$ ) in case of outcrossing with a nonsampled male. These  $AF$  were computed from the inferred paternal gametes of offspring finding no compatible male parent within the study site (Oddou-Muratorio *et al.* 2003). Genotypes with missing

data were not excluded and missing alleles were allowed to match with any allele at the specific locus. Equation (3) assumes that the father of each offspring is drawn at random according to the  $\pi_{jk}'s$ , i.e. that fertilization events are independent (no joint dispersal of pollen grains). It also assumes that the pollen of each sampled male tree disperses according to the same dispersal kernel  $p(\theta_d; x, y)$ .

We used this general estimator under three alternative models for the pollen clouds. First, we estimated simultaneously  $s, m, \theta_d$  and  $\mathbf{F}$  (complete model, or model 1). Then, we estimated  $s, m$ , and  $\theta_d$  assuming no differences in male fecundities (model 2). Finally, we estimated  $s, m$  and  $\mathbf{F}$  assuming random pollen dispersal within the neighbourhood (model 3). These three models differ only in the modelling of the pollen cloud coming from within the neighbourhood, i.e. the expression of the  $\pi_{jk}(\theta_d, \mathbf{F})$ . The expected proportions of the pollen of a given sampled male  $k$  in the local pollen cloud of female  $j$  are, respectively, given by:

$$\text{Model 1: } \pi_{jk} = \frac{\Phi_k p_{jk}}{\sum_{l \in M} \Phi_l p_{jl}} \quad (\text{eqn 4})$$

$$\text{Model 2: } \pi_{jk} = \frac{p_{jk}}{\sum_{l \in M} p_{jl}} \quad (\text{eqn 5})$$

$$\text{Model 3: } \pi_{jk} = \frac{\Phi_k}{\sum_{l \in M} \Phi_l} \quad (\text{eqn 6})$$

*Statistical methods*

In this section, we detail the statistical methods used to fit the model to the data, to evaluate the quality of these fits, and to compare the underlying models against each other (see also Supplementary material 2 for details). MATHEMATICA notebooks to implement the methods presented here are available upon request from E.K.K.

*Fits.* The fits were achieved by maximizing the log-likelihood described by equation (3) with respect to the parameters  $s, m$  and/or  $\theta_d$  and  $\mathbf{F}$  (depending on the model used) following a quasi-Newton algorithm (using MATHEMATICA 5.x, Wolfram Research). For the dispersal parameters, we estimated directly  $\delta$  and  $b$ , which was numerically easier than estimating parameters  $a$  and  $b$ . The parameters  $m$  and  $s$  were transformed through a logit function to ensure numerical convergence.

*Confidence intervals.* For the dispersal parameters  $b$  and  $\delta$ , we derived 95% likelihood-profile confidence intervals (e.g. Coles 2001) by plotting contour plots of the likelihood function. For the fecundity parameters  $\mathbf{F}$ , and for parameters  $m$  and  $s$ , we derived 95% confidence intervals by computing the asymptotic variance-covariance matrix following

classical techniques associated to likelihood (e.g. Coles 2001).

*Tests.* For the models that were nested into each other (e.g. normal within exponential power or model 2 or 3 within model 1), we used a likelihood-ratio test (LRT) to test whether the more complete model achieved a significantly better fit (Coles 2001): the deviance (i.e. twice the difference of log-likelihood between the complete model and the nested model) was compared to a chi-squared distribution with a number of degrees of freedom equal to the difference in the number of parameters between the complete model and the nested model.

For the models 1 and 3, the effect of each ecological factor was tested with a LRT in a type III approach, i.e. removing each tested effect separately from the complete model.

*Goodness of fit.* Graphical information about the goodness of fit of each model was obtained by comparing predicted vs. observed distributions of pollination events. The predicted distributions of pollination distances were constructed by computing for all mother trees the predicted proportions of offspring on mother  $j$  that should be sired by a male  $k$  in the studied site according to (4). The observed distribution was constructed using the, respectively, 216 (in 1999) and 308 (2000) father-offspring pairs reconstituted by a most-likely paternity analysis performed with CERVUS (Marshall *et al.* 1998) at 95% confidence level (see Oddou-Muratorio *et al.* 2003 for details). These categorical paternity assignments were only used as observed reference for this graphical representation. All estimations performed here were based directly on parent and offspring genotypes, as described by equation (3).

Additionally, we computed the proportion of variability in MMS explained by each model. Denote  $n_{jk}$  the number of offspring collected on mother  $j$  and categorically assigned to a sampled male  $k$  by the paternity analysis method described in (Oddou-Muratorio *et al.* 2003). The observed MMS of male  $k$  over all females was estimated by  $n_k = \sum_j n_{jk}$ . The predicted MMS of male  $k$  under each models was computed as:  $\hat{n}_{k} = \sum_j \hat{\pi}_{jk} \times n_{j\cdot}$  where the  $\pi_{jk}$  were computed using equations (4), (5) or (6) (depending on the model), and where  $n_{j\cdot}$  is the total number of offspring collected on mother  $j$ . The proportion ( $r^2$ ) of variability in MMS explained by the model under consideration was then estimated as:

$$r^2 = 1 - \frac{\sum_{k \in F} (n_k - \hat{n}_{k})^2}{\sum_{k \in F} (n_k - \bar{n})^2} \quad (\text{eqn 7})$$

where  $\bar{n}$  is the average MMS (estimated as the mean number of offspring per male). We used these  $r^2$  values in a comparative purpose to gauge the respective importance of spatial vs. ecological determinants of MMS.

*Effective male population density*

As soon as an ecological variable affects male fecundity, this effect can be translated into a reduction of effective population size. Define the effective male population size ( $N_{em}$ ) as the inverse of the probability that two pollen grains from the pollen cloud of all females come from the same male parent. In a spatially structured population, it is more convenient to consider instead the effective male population density  $d_{em}$  defined by  $d_{em} = N_{em}/A$ , where  $A$  is the area covered by the study population.

Translating the classical formula for effective population size yields:

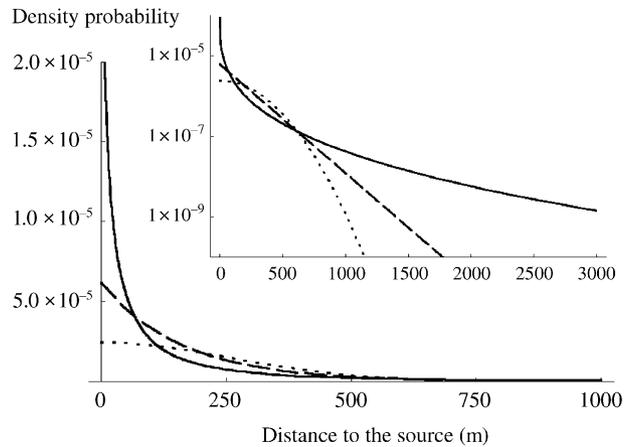
$$\frac{d_{em}}{d_{obs}} = \frac{\left(\sum_{c=1}^n \alpha_c f_c\right)^2}{\sum_{c=1}^n \alpha_c f_c^2} \tag{eqn 8}$$

where  $d_{obs}$  is the observed male density and  $\alpha_c$  the proportion of males belonging to the  $c$ th class of the factor and  $n$  the number of classes (Supplementary material 1).

**Results**

*Joint estimation of the dispersal kernel and the HMF*

Consider first the pollen dispersal kernels obtained with the complete model (model 1) including all the ecological determinants of male fecundity (Fig. 1). The exponential power kernel described by equation (1) provided a better fit than the normal (Table 1), and the exponential power



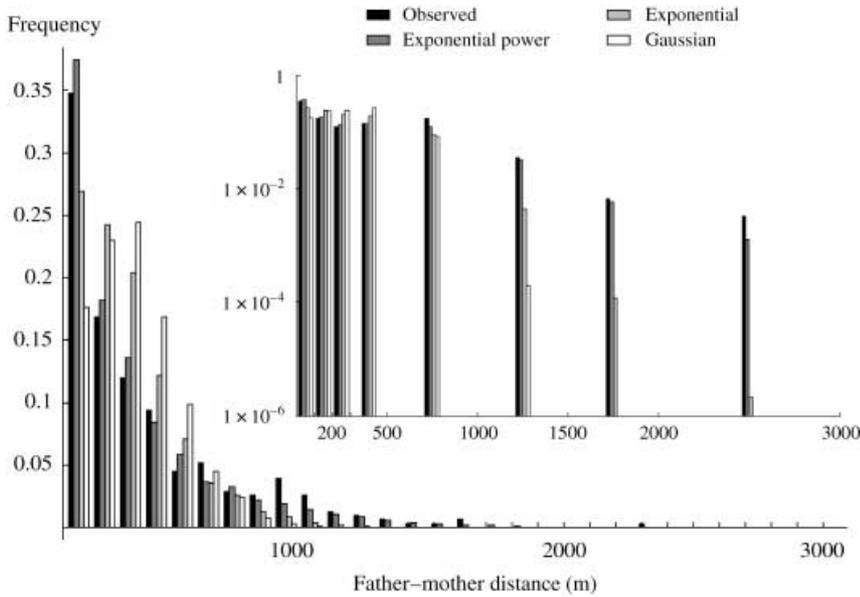
**Fig. 1** Dispersal kernels estimated under the different models: normal (dotted lines), exponential (black traits) and exponential power (black line) under the complete model (model 1), which accounts for HMF. In the up-right corner, log-plot version of the same curves, providing a better representation of their tails.

distribution provided an even significantly better fit ( $P$  value  $< 10^{-4}$  for all likelihood-ratio tests over both years). This better fit is clearly visible when the predicted patterns of pollination distances are compared to the observed ones as inferred by categorical paternity assignment performed with CERVUS (Marshall *et al.* 1998). The exponential power model performs better (Fig. 2), as further detailed in Supplementary material 2.

These fits yielded estimated values of  $b$  of 0.21 in 1999 and 0.33 for 2000 (significantly lower than 1.0 at the 5% level for both years). Thus, over both years, we observed a general tendency of marked fat-tailed dispersal curves in

**Table 1** Selfing rate ( $s$ ), immigration rate ( $m$ ) and pollen dispersal parameters inferred under the different models. Model 1 refers to the complete model where we estimated simultaneously  $s$ ,  $m$ , and the dispersal and fecundity parameters. Under model 2, we estimated  $s$ ,  $m$ , and the dispersal parameters assuming no differences in male fecundities. Under model 3, we estimated  $s$ ,  $m$  and the fecundity parameters assuming random pollen dispersal within the neighbourhood. Fit's quality was evaluated through the log-likelihood ( $L$ ) of the data set under each model (characterized by  $n$  parameters). Dispersal parameter  $a$  can be deduced from  $b$  and  $\delta$  using equation (2). Standard errors (SE) for  $m$  and  $s$  and confidence intervals (CI) at 95% for  $b$  and  $\delta$  are also provided (see text)

Year	Dispersal distribution	Model	$n$	$-L$	$\hat{m}$ (%)	$\hat{m}$ 2SE (%)	$\hat{s}$ (%)	$\hat{s}$ 2SE (%)	$\hat{b}$	$\hat{b}$ CI	$\hat{\delta}$ (m)	$\hat{\delta}$ CI
1999	Normal	1	15	8242	42.6	4.2	1.2	0.86	—	—	267	[236, 298]
	Exponential	1	15	8149	40.9	4.1	1.2	0.86	—	—	270	[243, 300]
	Exp. power	1	16	8080	39.1	4.1	1.2	0.86	0.21	[0.14, 0.30]	1077	[558, 3500]
	Exp. power	2	4	8213	39.4	4.1	1.2	0.86	0.17	[0.09; 0.24]	873	[446, 4500]
	Panmixia	3	14	8321	36.8	4.1	1.2	0.86	—	—	—	—
2000	Normal	1	15	14 183	46.7	3.8	0.43	0.40	—	—	322	[279, 365]
	Exponential	1	15	14 046	46.5	3.3	0.44	0.41	—	—	331	[297, 357]
	Exp. power	1	16	13 955	43.7	3.2	0.52	0.47	0.33	[0.27, 0.40]	743	[540, 1160]
	Exp. power	2	4	14 095	43.8	3.1	0.55	0.49	0.38	[0.31, 0.44]	589	[457, 826]
	Panmixia	3	14	14 760	43.7	3.3	0.33	0.53	—	—	—	—



**Fig. 2** Histograms of pollination distances (results shown for year 2000). The observed distribution of pollination distance (black) was plotted beside the expected histograms under the following models (from dark grey to white, and from left to right: Exponential power, Exponential, Gaussian). In the upright corner, in order to better visualize the long-distance dispersal events, we log-plotted the histogram with intervals of varying widths (ticks indicate the boundaries of the intervals and bars are located in their centres). Note that the heights of the bars that indicate the frequencies of pollination distances in the intervals are not scaled by the bar widths, leading to possible increases with distances when increasing the width.

**Table 2** Effect of phenotypic/neighbourhood factors on fecundity assessed, respectively, with (model 1) and without (model 3) accounting for spatial processes. *K* is the likelihood-ratio test statistics and d.f. the number of degrees of freedom (see text for details)

Year	Phenotypical factor	Model 1 (spatial)			Model 2 (nonspatial)		
		d.f.	<i>K</i>	<i>P</i> value	d.f.	<i>K</i>	<i>P</i> value
1999	Flowering intensity	3	56.0	< 10 <sup>-4</sup>	3	172.0	< 10 <sup>-4</sup>
	DBH	5	67.6	< 10 <sup>-4</sup>	5	170.1	< 10 <sup>-4</sup>
	Neighbourhood density	4	28.9	< 10 <sup>-4</sup>	4	23.0	10 <sup>-4</sup>
2000	Flowering intensity	3	74.4	< 10 <sup>-4</sup>	3	107.6	< 10 <sup>-4</sup>
	DBH	5	82.1	< 10 <sup>-4</sup>	5	61.8	< 10 <sup>-4</sup>
	Neighbourhood density	4	18.0	10 <sup>-3</sup>	4	33.0	< 10 <sup>-4</sup>

*Sorbus torminalis*. The estimated mean distance of pollen dispersal was high for the exponential power distribution ( $\hat{\delta} \approx 1077$  m in 1999 and  $\hat{\delta} \approx 743$  m in 2000). These mean dispersal distance computed from the dispersal kernel were significantly higher than the apparent mean mother-to-father distances as inferred by categorical paternity assignment performed with CERVUS (267 m in 1999 and 306 m in 2000).

Additionally, the selfing rate was estimated to 1.2% in 1999 and 0.5% in 2000. The best models yielded estimated values for the immigration rate of 39.1% in 1999 and 43.7% in 2000.

The three ecological factors investigated had significant effects on relative male fecundity (Table 2). The estimated male fecundity of phenotypic classes increased strongly with flowering intensity (Table 3A): from 1 to 6.4 in 1999 (*P* value < 10<sup>-4</sup>) and from 1 to 4.4 in 2000 (*P* value < 10<sup>-4</sup>). Similarly, it increased strongly with diameter (Table 3B): from 1 to 47.5 in 1999 (*P* value < 10<sup>-4</sup>) and from 1 to 13.9 in 2000 (*P* value < 10<sup>-4</sup>). The variations in fecundity due to

flowering intensity and tree diameter were globally larger in 1999 than in 2000.

Considering neighbourhood density, it had also a significant effect on fecundity (*P* value < 10<sup>-4</sup> in 1999 and *P* value = 10<sup>-3</sup> in 2000), with increasing fecundity for trees under weak neighbourhood density (e.g. dominant stand trees, or edge trees). This effect was nevertheless smaller than the effect of the two other factors: relative fecundities ranged only from 1 to 1.9 in 1999 and from 1 to 2.5 in 2000. However, completely isolated trees tended to have lower fecundities than dominant stand trees, or edge trees.

We estimated the reduction in effective density due to each ecological factor using equation (8) (Table 3). When considering simultaneously all the classes for the three investigated factors, the ratio  $d_{em}/d_{obs}$  was estimated to 0.3 in 1999 and 0.5 in 2000. Finally, in terms of the part of variability explained by the complete model, it turned out to be  $r^2 = 72.9\%$  in 1999 and  $45.2\%$  in 2000.

A

Year	Flowering intensity	$n_c$	Spatial model (model 1)		Nonspatial model (model 3)	
			$f_c$ (2SE)	$d_{em}/d_{obs}$	$f_c$ (2 SE)	$d_{em}/d_{obs}$
1999	Anecdotal	28	1.0	0.83	1.0	1.7
	Abundant	42	4.3 (2.5)		3.2 (1.8)	
	Massive	19	6.4 (3.9)		13.4 (7.6)	
	Unknown	60	3.7 (2.7)		4.1 (2.7)	
2000	Anecdotal	34	1.0	0.83	1.0	0.59
	Abundant	64	2.3 (0.9)		2.5 (0.9)	
	Massive	22	4.4 (1.7)		4.4 (1.6)	
	Unknown	52	2.0 (1.1)		1.1 (0.6)	

B

Year	Tree diameter	$n_c$	Spatial model		Nonspatial model	
			$f_c$ (2 SE)	$d_{em}/d_{obs}$	$f_c$ (2 SE)	$d_{em}/d_{obs}$
1999	< 10 cm	15	1.0	0.63	1.0	0.50
	< 20 cm	49	7.8 (15.5)		4.4 (10.3)	
	< 30 cm	17	15.8 (33.6)		9.6 (24.3)	
	< 40 cm	28	21.5 (43.3)		21.8 (51.1)	
	< 50 cm	25	19.4 (38.9)		23.7 (55.9)	
	≥ 50 cm	15	47.5 (92.2)		59.6 (142.0)	
2000	< 10 cm	16	1.0	0.71	1.0	0.83
	< 20 cm	45	1.9 (1.8)		0.7 (0.5)	
	< 30 cm	22	4.9 (4.5)		1.4 (1.0)	
	< 40 cm	39	7.4 (6.8)		2.2 (1.5)	
	< 50 cm	33	9.5 (9.1)		2.5 (1.8)	
	≥ 50 cm	17	13.9 (13.0)		3.1 (2.1)	

C

Year	Neighbourhood density	$n_c$	Spatial model		Nonspatial model	
			$f_c$ (2 SE)	$d_{em}/d_{obs}$	$f_c$ (2 SE)	$d_{em}/d_{obs}$
1999	Suppressed stand tree	54	1.0	0.83	1.0	0.90
	Codominant stand tree	13	0.7 (0.4)		0.6 (0.3)	
	Dominant stand tree	22	1.8 (0.9)		0.6 (0.3)	
	Edge tree	9	1.9 (1.1)		1.5 (0.8)	
	Isolated tree	51	0.7 (0.4)		0.9 (0.3)	
2000	Suppressed stand tree	65	1.0	0.90	1.0	0.90
	Codominant stand tree	15	1.3 (0.4)		1.7 (0.5)	
	Dominant stand tree	25	1.4 (0.5)		1.8 (0.6)	
	Edge tree	9	2.5 (1.2)		2.0 (1.0)	
	Isolated tree	58	0.9 (0.3)		0.9 (0.2)	

#### Separate estimation of the dispersal kernel and the HMF

When estimating the dispersal kernel without accounting for HMF (model 2), the fits were significantly worse ( $P$  value  $< 10^{-4}$ ) than with the complete model, and the estimated dispersal parameters slightly differed (Table 1). The  $b$  values were 0.17 instead of 0.21 in 1999 and 0.38 instead of

0.33 in 2000. The  $\delta$  values were of 873 m instead of 1077 m in 1999 and of 589 m instead of 753 m in 2000. Under model 2, the proportion of variability in MMS was lower than under model 1, both in 1999 ( $r^2 = 50\%$  instead of 72.9%) and in 2000 ( $r^2 = 22.3\%$  instead of 45.2%).

Consider now the model estimating the HMF assuming random pollen dispersal within the neighbourhood

**Table 3** Estimated relative male fecundity ( $f_c$ 's) for ecological groups of males defined by flowering intensity (A), diameter (B) or neighbourhood density (C).  $n_c$  is the number of males per class for each factor. The  $f_c$ 's were estimated with (model 1) or without (model 3) accounting for spatial processes, but including systematically all the ecological factors. The effects of each factor on effective population density ( $d_{em}$ ) were evaluated by computing the ratio  $d_{em}/d_{obs}$  where  $d_{obs}$  is the observed density of reproductive adults

(model 3). Again, the fit with this model was significantly worse than the fit with model 1 ( $P$  value  $< 10^{-4}$ ). Nevertheless, the factors that had the greater impact on male fecundities were the same as those identified by the complete model. Additionally, the classes for each factor were usually ranked in the same order with both models, except for neighbourhood density in 1999, for which fits under model 3 yielded a lower fecundity for dominant stand tree than for suppressed stand trees. The estimated fecundities of the different classes substantially differed between models 1 and 3, in particular for flowering intensity in 1999 and diameter in 2000. Moreover, estimations for model 1 were always more consistent between years than those of model 3. Finally, under this model 3, the proportion of variability in MMS was lower than the complete model in 1999 ( $r^2 = 48.9\%$  instead of  $72.9\%$ ) and in 2000 ( $r^2 = 25.5\%$  instead of  $45.2\%$ ).

## Discussion

### *Mating patterns in Sorbus torminalis*

Our study provides new insights on the mating system and dispersal pattern of this scattered, entomophilous tree species. First, we identified some of the major factors affecting male fecundity. Among these factors, the individual phenotypic characteristics (tree diameter and flowering intensity) had a strong effect, the larger individuals and the highly flowering ones showing a higher fecundity than the others, as expected. Stand density within the neighbourhood (independently of species) also strongly affected male fecundity: first, individuals under moderate competition (dominant stand trees or hedge trees) showed a higher fecundity than those under strong competition (suppressed or codominant trees). This may be the consequence of a higher attraction of pollinators for trees with an important crown development and better exposition to light, for instance through a more regular distribution of flowering corymbs in this crown. On another hand, individuals under no competition (isolated trees) had lower fecundity than trees within or in border of stand, suggesting a disruption of pollinators among stand and opened landscape. While some (such as bumblebees) can forage across various landscapes (Chapman *et al.* 2003; Kreyer *et al.* 2004), others (wild bees, coleopterans) may prefer closed landscapes. This is consistent with patterns of bee movements documented in apple or pear orchards, which show that pollinators forage preferentially from one tree to an adjacent tree in the same row, rather than switching rows (e.g. Monzón *et al.* 2004). This will decrease the amount of pollinators' visits for trees isolated in low stand-density areas, like for example areas opened by regeneration logging. Detailed observation of the pollinators' behaviour for this species would be required to confirm this hypothesis.

Concerning the pollen dispersal kernel, our results confirmed the important abilities of long-distance pollen dispersal in *S. torminalis* (Austerlitz *et al.* 2004). Indeed, the best-fitting dispersal curves for pollen movements within the studied site were always the fat-tailed exponential power dispersal kernels (Figs 1 and 2), with an estimated shape parameter ( $b$ ) that was always strongly below 1.0 ( $\hat{b} = 0.21$  in 1999 and  $\hat{b} = 0.33$  in 2000). This was associated with rather high estimates of the mean dispersal distances ( $\hat{\delta} = \sim 1100$  m in 1999 and  $\hat{\delta} = \sim 750$  m in 2000). The confidence interval for  $\hat{\delta}$  were wider in 1999 than in 2000, which can be explained by the high sensitivity of  $\hat{\delta}$  to small errors on  $\hat{b}$  when  $b$  is very small. Our results thus indicate that mating distances are highly variable: while most mating events occurred at close distance, a few of them occurred at long and even very long distances (up to 2.2 km, as estimated by paternity analysis; Oddou-Muratorio *et al.* 2003). Consistently, the estimated rate of external pollen flow was very high ( $\hat{m} = \sim 40\%$  for each year). Note that there is no functional link in the model between the dispersal kernel and the immigration rate, as  $m$  is estimated from the contribution of the individuals from the outside (mainly through seeds that have no compatible fathers in the inside).

Moreover, our model allowed us gauging the respective importance of spatial processes and of HMF on MMS. The complete model including both the dispersal kernel and the HMF explained 73% of the observed variation in MMS in 1999, and 45% in 2000. Among the explained variation in MMS in 2000, one-half was explained by the heterogeneity of male fecundity and the other half by the spatial distribution of males with respect to females. In 1999, the sampling design did not allow such a clear distinction: each process considered separately explained about 50% of the observed variability in MMS, vs. 73% when taking both processes into account (see below). These results however, show that the position of a tree may affect its MMS as much as its relative fecundity, as determined by its phenotype and environment.

It is worth noting that the effect of male and female spatial positions include both the dispersal processes and the geometry of the population (Meagher & Vassiliadis 2003). The effect of the geometry of the population on MMS is probably particularly high for this scattered species with individuals clearly aggregated in space, as shown by the spatial auto-correlation profile of a Ripley-based statistics (Oddou-Muratorio *et al.* 2004). In particular, it explains why the shape of the kernel (Fig. 1) is not similar to the shape of the father–mother distances distribution (Fig. 2). It is also the reason of the differences between the mean dispersal distance computed from the dispersal kernel ( $\hat{\delta} \sim 1100$  m in 1999 and  $\hat{\delta} \sim 750$  m in 2000) and the apparent mean mother-to-father distance ( $MFD_{\text{obs}} = 267$  m in 1999 and  $MFD_{\text{obs}} = 306$  m in 2000). Whereas  $\delta$  would be the

mean distance travelled by a pollen grain if it could land everywhere,  $MFD_{\text{obs}}$  accounts for the critical role of females as pollen receptor in the realization of MMS. However,  $MFD_{\text{obs}}$  is also likely to be an underestimate of realized pollination distance, as it is based on a truncated distribution, whereas the dispersal kernel modelling attempts to account for long-distance dispersal.

#### *Methodological insights for the estimation of the dispersal kernel and of the HMF*

The comparative approach used here first shows the increase in precision reached when accounting for HMF when estimating pollen dispersal parameters and reciprocally. This is especially true when the individuals are not distributed randomly in space regarding to their ecological status, which is often the case in natural population.

Overall, a higher proportion of the observed heterogeneity in MMS was explained in 1999 than in 2000 (73% vs. 45%). When estimating the dispersal kernel without accounting for HMF (model 2), the proportion of explained variation decreases from 73% to 50% in 1999 and from 45% to 22% in 2000. However, the estimated  $b$  and  $\delta$  values were only slightly affected. This low sensitivity is reassuring, considering that in most pollen flow studies, the phenotypic factors potentially affecting MMS are not evaluated; even when such factors are investigated, some of them may be missed.

When estimating the HMF without accounting for spatial processes (model 3), the proportion of explained variation decreases from 73% to 49% in 1999 and from 45% to 26% in 2000. Additionally, and by contrast with model 2, ignoring the spatial processes strongly affect the estimation of the relative fecundities of the phenotypic classes. This may lead to unrealistic results, for instance for the effect of neighbourhood density in 1999, where the dominant stand trees had lower fecundities than the suppressed ones. Also, the higher among-years variation in fecundity estimates obtained with the nonspatial model as compared to the spatial one showed that ignoring spatial processes induces a higher sensitivity to sampling design.

Finally, this study also stressed the role of the sampling design. Concerning dispersal kernel estimate, the sampling design was clearly more adequate in 2000 (60 females, ~18 offspring/female) than in 1999 (14 females, ~47 offspring/female), as depicted for instance by wider confidence intervals for  $\delta$  in 1999 as compared to 2000 (see Supplementary material 2). However, sampling large progeny arrays on few mother trees remains interesting in other contexts like the estimation of selfing rate. Concerning male fecundity estimates, our results suggest that increasing the number of females allows a better separation of the respective effects of spatial processes and of HMF on MMS: in 1999, some of the variability in MMS explained by spatial position

was also accounted for by ecological factors, and reciprocally. This is likely to result from the nonindependent distribution of ecological classes of males and of their spatial positions with respect to the sampled mother trees.

#### *Comparison with other methods*

Our results show the interest of our new estimates of pollen dispersal curve and effective density, as compared the TWOGENER method (Smouse *et al.* 2001), which differs in its approach of the problem. The TWOGENER method consists in estimating the differentiation in the estimated allelic frequencies among the pollen clouds of a sample of mothers spread across the landscape, and to infer from this level of differentiation the pollen dispersal kernel and the effective population density (Austerlitz *et al.* 2004). This indirect method thus allows the user to gauge pollen dispersal for a reasonable field and laboratory investment, as it uses only the genotypes of mother and offspring as an input, without the need of genotyping the potential fathers.

This TWOGENER method yielded also low estimated values for the shape parameter  $b$  ( $\hat{b} = 0.565$  in 1999 and  $\hat{b} = 0.285$  in 2000) and conversely high estimated values for  $\hat{\delta}$  (see Tables 4 and 5 in Austerlitz *et al.* 2004). The value of  $\hat{b}$  obtained with TWOGENER in 2000 is more consistent with the estimates based on our method. This might be due in part to the better configuration of the sampling design in 2000 (greater number of mother trees), which is expected to increase the precision of the estimation process (Smouse *et al.* 2001).

The ratio between observed and estimated effective density ( $\hat{d}_{\text{em}}/d_{\text{obs}} \sim 0.07$  in 1999 and  $\sim 0.03$  in 2000) is much higher for TWOGENER than that for our method ( $\hat{d}_{\text{em}}/d_{\text{obs}} \sim 0.3$  in 1999 and  $\sim 0.5$  in 2000). This may in part be the consequence of the level of imprecision of the TWOGENER method, which uses a lesser quantity of information. Moreover, the TWOGENER method assumes a Poisson distribution of males around females (Austerlitz & Smouse 2001; Smouse *et al.* 2001), while trees are clearly aggregated in our study population. However, it is also clear that our method takes only identified factors into account. Other hidden factors, either genetic or ecological, may increase the HMF, and contribute to the ~33%/55% of the variation in MMS that remained unexplained (respectively in 1999 and 2000). For instance, some individuals may be more inbred than others and thus have a lower fecundity. Also, some individuals produce flowers but no fruits (~33% in 1999 and ~19% in 2000), thus behaving as functionally male individuals (Oddou-Muratorio *et al.* 2001a); these functional males may invest more in pollen production than hermaphroditic individuals, increasing the HMF. Additionally, we have not taken into account several ecological parameters like soil quality or partially nonoverlapping flowering phenologies. All this contributes to explain the

discrepancy between both methods. Our results illustrate the classical methodological trade-off between measuring (but not explaining) all the variability as in the TWOGENER method and explaining a part of it with controlled/measured factors as in our method or Burczyk *et al.*'s (2002) method.

#### *Evolutionary consequences of fat-tailed dispersal kernel combined to low effective density*

Our results on *S. torminalis* are consistent with the data that are beginning to accumulate on other species, and that highlight a general trend of fat-tailed dispersal kernel (Austerlitz *et al.* 2004). Even if we clearly need more data to get a global picture, this general trend implies that pollen is readily able to travel at large distance, in anemophilous as well as in entomophilous species. It will ease for example the spread of a favourable gene within a landscape, and also allow isolated groups of individuals to remain connected to larger populations, which may reduce the risk of inbreeding depression within these groups. Also, it explains the high level of contamination usually observed in forest seed orchards (Burczyk & Prat 1997; Burczyk *et al.* 2002).

While this long dispersal distance might be advantageous for the species, the low effective density induced by the HMF could increase drift in natural populations. For instance, TWOGENER results in *S. torminalis* yield an estimated effective number of pollinators ( $N_{ep}$ ) of a given female between 7 and 12. This may seem in contradiction with the general observation of high within-population diversity and low among-population differentiation observed in tree populations at large scales (Hamrick *et al.* 1992), which are considered as signs of rather high effective population sizes. This discrepancy may be in part explained by the fact that even if effective density is low, the extensive movements of pollen will make that individuals will be involved in large reproductive units. Moreover, long-distance pollen flow connects more differentiated gene pools than does short-distance pollen flow. Cumulated over several reproduction seasons, this might thus increase significantly the population effective density. Also, because trees are long-living species, their lifetime reproductive success may show less variance among individuals than their instantaneous reproductive success. Finally, despite low  $N_{ep}$  at the scale of a single pollen cloud,  $N_{ep}$  cumulated over all females may be much larger (in our case, ~60 effective males in 2000).

Thus, only the following of populations through time will allow one to determine the changes over years of pollen dispersal patterns. Moreover, this following is also mandatory to assess whether the heterogeneity in instantaneous male fecundity between individuals will be compensated by the averaging process. In parallel, it will be also necessary to investigate the heterogeneity in female

reproductive success, to determine whether this will act synergically with the male reproductive success or in the opposite direction (Meagher & Vassiliadis 2003). Only this will yield detailed understanding of the interaction between evolutionary process and spatial patterns occurring in plant populations. From a conservation biology point of view, this will help to make predictions regarding the impact of habitat fragmentation or other human interventions. For instance, *S. torminalis* isolated in regeneration units have lower male fecundity than trees within stands, but as females, since they will not be massively pollinated by one or a few close neighbours, they may still sample a higher diversity of pollen gametes and thus increase significantly the population effective density of the next generation.

#### Acknowledgements

This work is part of S.O.-M. PhD thesis, which was funded by the French National Forest Services (ONF). This study was supported by the Bureau des Ressources Génétiques. We thank Victoria Sork, Cyril Dutech, Francois Lefèvre, Chris Dick, Peter Smouse, Jarek Burczyk and three anonymous reviewers for helpful comments on successive versions of this manuscript.

#### Supplementary material

The supplementary material is available from <http://www.blackwellpublishing.com/products/journals/suppmat/MEC/MEC2720/MEC2720sm.htm>

**Supplementary material 1:** Effective vs. observed pollination size in a species with heterogeneity in male mating success

**Supplementary material 2:** Confidence intervals of the parameters and goodness-of-fit of the model

#### References

- Abramowitz M, Stegun IA (1964) *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. US Government Printing Office, Washington, D.C.
- Adams W, Birkes D (1991) Estimating mating patterns in forest tree populations. In: *Biochemical Markers in the Population Genetics of Forest Trees* (eds Fineschi S, Malvolti M, Cannata F, Hattemer H), pp. 157–172. SPB Academic Publishing, The Hague, The Netherlands.
- Austerlitz F, Smouse PE (2001) Two-generation analysis of pollen flow across a landscape. II. Relation between  $\Phi_{it}$ , pollen dispersal and interfemale distance. *Genetics*, **157**, 851–857.
- Austerlitz F, Dick CW, Dutech C *et al.* (2004) Using genetic markers to estimate the pollen dispersal curve. *Molecular Ecology*, **13**, 937–954.
- Bateman AJ (1947) Contamination in seed crops. III. Relation with isolation distance. *Heredity*, **1**, 303–336.
- Burczyk J, Prat D (1997) Male reproductive success in *Pseudotsuga menziesii* (Mirb.) Franco: the effects of spatial structure and flowering characteristics. *Heredity*, **79**, 638–647.

- Burczyk J, Adams WT, Moran GF, Griffin AR (2002) Complex patterns of mating revealed in a *Eucalyptus regnans* seed orchard using allozyme markers and the neighborhood model. *Molecular Ecology*, **11**, 2379–2391.
- Chapman RE, Wang J, Bourke AFG (2003) Genetic analysis of spatial foraging patterns and resource sharing in bumble bee pollinators. *Molecular Ecology*, **12**, 2801–2808.
- Clark JS (1998) Why trees migrate so fast: confronting theory with dispersal biology and the paleorecord. *American Naturalist*, **152**, 204–224.
- Coles S (2001) *An Introduction to Statistical Modeling of Extreme Values*. Springer-Verlag, London.
- Devlin B, Ellstrand NC (1990) Male and female fertility variation in wild radish, a hermaphrodite. *The American Naturalist*, **136**, 87–107.
- Dick CW (2001) Genetic rescue of remnant tropical trees by an alien pollinator. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, **268**, 2391–2396.
- Dick CW, Etchelecu G, Austerlitz F (2003) Pollen dispersal of Neotropical trees (*Dinizia excelsa*: Fabaceae) by native insects and African honeybees in pristine and fragmented Amazonian rainforest. *Molecular Ecology*, **12**, 753–764.
- Godoy JA, Jordano P (2001) Seed dispersal by animals: exact identification of source tree with endocarp DNA microsatellites. *Molecular Ecology*, **10**, 2275–2283.
- Griffin AR, Sedgley M (1989) *Sexual Reproduction of Tree Crops*. London Academic Press, London.
- Hamrick JL, Godt MJL, Sherman-Broyles SL (1992) Factors influencing levels of genetic diversity in woody plant species. *New Forests*, **6**, 95–124.
- Jones B (2003) Maximum likelihood inference for seed and pollen dispersal distributions. *Journal of Agricultural, Biological and Environmental Statistics*, **8**, 170–183.
- Kreyer D, Oed A, Walther-Hellwig K, Frankl R (2004) Are forests potential landscape barriers for foraging bumblebees? Landscape scale experiments with *Bombus terrestris* agg. and *Bombus pascuorum* (Hymenoptera, Apidae). *Biological Conservation*, **116**, 111–118.
- Le Corre V, Machon N, Petit RJ, Kremer A (1997) Colonisation with long-distance seed dispersal and genetic structure of maternally inherited genes in forest trees: a simulation study. *Genetical Research*, **69**, 117–125.
- Levin DA, Kerster HW (1974) Gene flow in seed plants. *Evolutionary Biology*, **7**, 139–220.
- Marshall TC, Slate J, Kruuk LEB, Pemberton JM (1998) Statistical confidence for likelihood-based paternity inference in natural populations. *Molecular Ecology*, **7**, 639–655.
- Meagher TR (1986) Analysis of paternity within a natural population of *Chamaelirium luteum*. I. Identification of most-likely male parents. *American Naturalist*, **128**, 199–215.
- Meagher TR (1991) Analysis of paternity within a natural population of *Chamaelirium luteum*. II. Patterns of male reproductive success. *American Naturalist*, **137**, 738–752.
- Meagher TR, Vassiliadis C (2003) Spatial geometry determines gene flow in plant population. In: *Genes in the Environment* (eds Hails R, Beringer J, Godfray HCJ), pp. 76–90. Blackwell Publishing, Oxford.
- Monzón VH, Bosch J, Retana J (2004) Foraging behavior and pollinating effectiveness of *Osmia cornuta* (Hymenoptera: Megachilidae) and *Apis mellifera* (Hymenoptera: Apidae) on 'Comice' pear. *Apidologie*, **35**, 575–585.
- Morgan MT, Conner JK (2001) Using genetic markers to directly estimate male selection gradients. *Evolution*, **55**, 272–281.
- Nathan R, Muller-Landau H (2000) Spatial patterns of seed dispersal, their determinants and consequences for recruitment. *Trends in Ecology & Evolution*, **15**, 278–285.
- Nielsen R, Mattila DK, Clapham PJ, Palsboll PJ (2001) Statistical approaches to paternity analysis in natural populations and application to the North Atlantic humpback whale. *Genetics*, **157**, 1673–1682.
- Oddou-Muratorio S, Petit RJ, Le Guerroué B, Guesnet D, Demesure B (2001a) Pollen- versus seed-mediated gene flow in a scattered woody species. *Evolution*, **55**, 1123–1135.
- Oddou-Muratorio S, Aligon C, Decroocq S *et al.* (2001b) Microsatellite primers for *Sorbus torminalis* and related species. *Molecular Ecology Notes*, **1**, 297–299.
- Oddou-Muratorio S, Houot M-L, Demesure-Musch 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.
- Oddou-Muratorio S, Demesure-Musch B, Pélissier R, Gouyon P-H (2004) The roles of colonization and history in the local genetic structure of a scattered temperate tree species. *Molecular Ecology*, **13**, 3689–3702.
- Rocha OJ, Aguilar G (2001) Variation in the breeding behavior of the dry forest tree *Enterolobium cyclocarpum* (guanacaste) in Costa Rica. *American Journal of Botany*, **88**, 1600–1606.
- Roeder K, Devlin B, Lindsay BG (1989) Application of maximum likelihood methods to population genetic data for the estimation of individual fertilities. *Biometrics*, **45**, 363–379.
- Smouse PE, Meagher TR, Kobak CJ (1999) Parentage analysis in *Chamaelirium luteum* (L.) Gray (*Liliaceae*): why do some males have higher reproductive contributions? *Journal of Evolutionary Biology*, **12**, 1069–1077.
- Smouse PE, Dyer RJ, Westfall RD, Sork VL (2001) Two-generation analysis of pollen flow across a landscape. I. Male gamete heterogeneity among female. *Evolution*, **55**, 260–271.
- Sork VL, Davis FW, Smouse PE *et al.* (2002) Pollen movement in declining populations of California Valley oak, *Quercus lobata*: where have all the fathers gone? *Molecular Ecology*, **11**, 1657–1668.
- Streiff R, Ducouso A, Lexer C *et al.* (1999) Pollen dispersal inferred from paternity analysis in a mixed oak stand of *Quercus robur* L. and *Q. petraea* (Matt.) Liebl. *Molecular Ecology*, **8**, 831–841.
- Wright S (1943) Isolation by distance. *Genetics*, **28**, 114–138.
- Wright JW, Meagher TR (2004) Selection on floral characters in natural Spanish populations of *Silene latifolia*. *Journal of Evolutionary Biology*, **17**, 382–395.

---

The experimental part of this study was performed at the Genetic Conservatory of Forest Trees of the French National Forest Office, where Sylvie Oddou-Muratorio completed her PhD. As researcher at INRA, Sylvie Oddou-Muratorio is now studying the population genetics of expanding tree populations. Etienne Klein is concerned by modelling and estimating pollen dispersal at the landscape scale. Frédéric Austerlitz has a long-standing interest in theoretical population genetics, including the estimation of gene flow.

---