



Pollen flow in the wildservice tree, *Sorbus torminalis* (L.) Crantz. IV. Whole interindividual variance of male fecundity estimated jointly with the dispersal kernel

Etienne Klein, Nicolas Desassis, Sylvie Oddou-Muratorio

► To cite this version:

Etienne Klein, Nicolas Desassis, Sylvie Oddou-Muratorio. Pollen flow in the wildservice tree, *Sorbus torminalis* (L.) Crantz. IV. Whole interindividual variance of male fecundity estimated jointly with the dispersal kernel. *Molecular Ecology*, 2008, 17 (14), pp.3323-3336. 10.1111/j.1365-294X.2008.03809.x . hal-02408994

HAL Id: hal-02408994

<https://hal.science/hal-02408994>

Submitted on 30 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Pollen flow in the wildservice tree, *Sorbus torminalis* (L.) Crantz. IV. Whole inter-individual variance of male fecundity estimated jointly with the dispersal kernel.

E. K. Klein^{1,2}, N. Desassis¹, S. Oddou-Muratorio²

¹ INRA, UR 546, Biostatistique et Processus Spatiaux, F-84000 Avignon, France

² INRA, UR 629, Recherches Forestières Méditerranéennes, F-84000 Avignon, France.

KEY WORDS: mating system, mixed mating models, long-distance dispersal, pollen dispersal, microsatellite markers, Bayesian approach

CORRESPONDING AUTHOR

Etienne Klein

INRA, UR 546, Biostatistique et Processus Spatiaux, Domaine Saint Paul, Site Agroparc, F-84914 Avignon Cedex 9, France

Fax Number: +33 4 32 72 21 82

E-Mail: etienne.klein@avignon.inra.fr

RUNNING TITLE: Variance of male fecundity & pollen dispersal

ABSTRACT

Inter-individual variance of male reproductive success (MRS) contributes to genetic drift, which in turn interacts with selection and migration to determine the short-term response of populations to rapid changes in their environment. Individual relative MRS can be estimated through paternity analysis and can be further dissected into fecundity and spatial components. Existing methods to achieve this decomposition either rely on the strong assumption of a random distribution of pollen donors (TwoGener) or estimate only the part of the variance of male fecundity that is explained by few covariates.

We developed here a method to estimate jointly the whole variance of male fecundity and the pollen dispersal curve from the genotypic information of sampled seeds and their putative fathers and geographic information of all individuals in the study area. We modelled the relative individual fecundities as a log-normally distributed random effect. We used a Bayesian approach, well suited to the hierarchical nature of the model, to estimate these fecundities.

When applied to *Sorbus torminalis*, the estimated variance of male fecundity corresponded to an effective density of trees 13 times lower than the observed density ($d_{obs}/d_{ep} \sim 13$). This value is between the value (~ 2) estimated with a classical mating model including three covariates (neighbourhood density, diameter, flowering intensity) that affect fecundity and the value (~ 30) estimated with TwoGener. The estimated dispersal kernel was close to previous results. This approach allows fine monitoring of ongoing genetic drift in natural populations, and quantitative dissection of the processes contributing to drift, including human actions.

INTRODUCTION

Populations of long-lived organisms such as trees may be subject, within only a few generations, to large environmental and demographic changes (Hamrick, 2004). For these species, adaptation and evolution of genetic diversity result from an intricate interaction of gene flow, genetic drift and natural selection (Petit & Hampe, 2006). Moreover the study of these evolutionary processes needs to be addressed at a contemporary time scale because tree populations are far from equilibrium (Sork *et al.*, 1999) and predictions and management guidelines are required at the scale of few generations (Lande & Barrowclough, 1987).

The first quantitative estimates of genetic drift at contemporary time scales relied on the evaluation of male or female reproductive successes of individuals within a population (Devlin & Ellstrand, 1990; Smouse & Meagher, 1994). These studies, based on neutral polymorphic genetic markers to infer paternity (or parentage) of sampled progenies, benefited from the development of microsatellites (Smouse & Sork, 2004). Many of these studies demonstrated highly uneven contributions among reproductive individuals, resulting in low effective numbers of both pollen donors (e.g. Smouse *et al.*, 1999 and review in Smouse & Sork, 2004) and mothers (e.g. Devlin & Ellstrand, 1990; Schnabel *et al.*, 1998). Now that the extensive variation in male and female reproductive success has been acknowledged, the challenge is to identify, quantify and model its causes in order to produce accurate predictions of male and female individual reproductive successes under different scenarios of a changing environment (fragmentation, changes in species assemblages...) or management actions.

The causal factors of this heterogeneity has usually been addressed, for the case of pollen-mediated gene flow, through explicit assessment of covariates effect on male reproductive success (MRS). The measurement of the relative contributions of putative fathers to the pollen pool of a small set of sampled mothers has shown that many environmental, ecological or individual variables affect MRS, such as diameter, flowering intensity, phenology, sub-species, genotype or neighbourhood density (Smouse & Sork, 2004). This pattern was shown either by partitioning individuals into distinct phenotypic classes and evaluating the relative MRS of the classes (Nielsen *et al.*, 2001) or by using selection gradients within mating/neighbourhood models (Adams & Birkes, 1991; Burczyk *et al.*, 2002). Selection gradients model the effect of quantitative traits on reproductive success through the regression of relative MRS on trait values measured in single plants (Lande & Arnold, 1983).

In population genetics, ongoing genetic drift is directly related to the inter-individual variance of relative MRS, i.e. number of seeds sired by a father relatively to another father or to the male population average. The absolute MRS, i.e. number of seeds sired by a given father, is of lower interest for population biology because pollen is generally not the principal determinant of seed-sets. And as a matter of fact, we can only get the relative MRS in classical experiments using mother plants as biological pollen traps, because we work on a set of seeds with a given sample size. This specificity of pollen dispersal and mating systems contrasts with studies on seed dispersal where seed traps provide information about the absolute numbers of seeds and where the dynamics of a population is governed by the absolute female reproductive success (Clark *et al.*, 1998).

The study of mating system requires a global framework to incorporate the different source of heterogeneity in relative MRS. The relative MRS of a tree is actually strongly dependent on (i) the phenotype and the micro-environment of the tree, (ii) the position of the tree relative to the mother-plants that produce seeds and (iii) the flowering phenology of the tree relative to the phenology of the sampled mother-plants (Smouse & Sork, 2004). A further step in understanding the MRS components would consist of a better distinction between the fecundity of a tree, a measure of its ability to produce offspring (proportional to the amount of efficient pollen emitted) and the male reproductive success, i.e. fertility, of a tree defined as

the actual number of offspring it sired (proportional to the amount of pollen that fertilised mother-trees). Under the assumption of random mating, MRS is proportional to fecundity and independent of the mother-trees sampled. Thus many of the studies based on this assumption estimated fertilities from the effective male reproductive success and interpreted them in terms of differences of fecundities among trees. This portrait is no longer true if pollen dispersal is limited by distance, or in presence of phenological variations among trees (e.g. (Gérard *et al.*, 2006). In these cases, the inter-individual variance of MRS (determining the effective number of pollen donors, N_{ep} , *sensu* (Austerlitz & Smouse, 2001)) can be decomposed into an inter-individual variance of male fecundity (determining the effective male reproductive density, d_{ep} , *sensu* (Austerlitz & Smouse, 2002)), a spatial component due to the positions of father-trees relatively to mother-trees and a temporal component due to the differences of phenologies among trees. Similarly to MRS, only relative fecundities are accessible in experiments and are of interest in population genetics.

Several recent studies on pollen dispersal showed the quantitative importance of this spatial component. Indeed the probability of mating for two plants typically decreases with distance between them (e.g. (Smouse *et al.*, 1999), and this pattern is generally modelled through a pollen dispersal kernel, describing the probability for a pollen grain to travel any distance from the release position to the ovule it pollinates (Klein *et al.*, 2006b; Lavigne *et al.*, 1998). The dispersal kernel has a crucial influence on the spatial distribution of genetic diversity, and also on the genetic connectivity among patches in a heterogeneous environment. Thus, many recent studies estimated the dispersal kernel from genetic data (Austerlitz *et al.*, 2004; Hardy *et al.*, 2004; Oddou-Muratorio *et al.*, 2005; Robledo-Arnuncio & Gil, 2005; Smouse & Sork, 2004) and characterised its tail, which controls long-distance dispersal (Austerlitz *et al.*, 2004; Devaux *et al.*, 2007; Klein *et al.*, 2006a; Klein *et al.*, 2006b).

A fine characterization of the effect of distance on MRS through the dispersal kernel is needed not only because of the large part of the variance of MRS it explains, but also because the spatial pattern of a population changes rapidly, either naturally (colonization, natural perturbations...) or because of human action (fragmentation, land-planning, silviculture ...). Predictions of the impact of future spatial patterns on the evolution of populations are thus needed, notably to investigate the efficiency of management scenarios. The calibration of dispersal kernels is a way to obtain these predictions (e.g. (Robledo-Arnuncio & Austerlitz, 2006), for the effects of aggregation).

Two contrasted approaches are used nowadays to investigate the components of mating systems (Burczyk & Koralewski, 2005; Smouse & Sork, 2004): either via the dissection of the variables affecting the male reproductive success, including position and phenology or via an overall robust estimate of the variance of male fecundity (d_{ep}) and variance of male reproductive success (N_{ep}). The first approach, comprising the paternity-based methods, needs extensive and explicit spatial and genetic information (positions and genotypes of all potential fathers in the experimental area) but only investigates the variance of male fecundity due to a few covariates individually measured on the putative fathers (Bacles *et al.*, 2005; Burczyk *et al.*, 2002; Oddou-Muratorio *et al.*, 2005). The second approach, relying on the TwoGener method (Austerlitz & Smouse, 2002; Smouse *et al.*, 2001), is an indirect approach based on the genetic differentiation among the pollen pools sampled by different mothers to estimate the whole variance of male fecundity (estimating directly the effective male reproductive density d_{ep}). It also covers a wider spatial range but assumes a Poisson spatial distribution of pollen donors, not realistic for some species.

The respective drawbacks of these two approaches are well illustrated by a case-study on *Sorbus torminalis* on which the neighbourhood and the TwoGener models were applied (Austerlitz *et al.*, 2004; Oddou-Muratorio *et al.*, 2005). They led to contradictory results for the variance of male fecundity and its correlate d_{ep} , the effective male reproductive density.

Using a mixed mating model and three major explanatory variables of male fecundity (diameter, flowering intensity and neighbourhood density) (Oddou-Muratorio *et al.*, 2005) estimated a d_{ep} -value twice lower than the observed density of flowering adults, d_{obs} . In contrast, the TwoGener method provided an effective density 30 times lower than the observed density (Austerlitz *et al.*, 2004), data for year 2000). The mating model ($d_{ep} = d_{obs}/2$) may have overestimated d_{ep} because several ecological variables that are also expected to decrease d_{ep} were not included. Alternatively, TwoGener ($d_{ep} = d_{obs}/30$) probably underestimated d_{ep} because the highly aggregated distribution of *Sorbus torminalis* (Oddou-Muratorio *et al.*, 2004) could not be considered, despite its potentially strong effect on the differentiation among pollen pools (Robledo-Arnuncio & Austerlitz, 2006).

The goal of the present study is to overcome the limitations of both available estimation methods, which can lead to highly divergent interpretations of the importance of genetic drift, as illustrated above. We propose here a new approach to estimate jointly the whole variance of male fecundity and the dispersal kernel into a spatially explicit model inspired from existing mixed mating models. The novelty is that (i) the male fecundity of each potential father is modelled as a random variable that follows a given distribution (here log-normal) and (ii) the variance of this distribution is estimated directly and jointly with the dispersal kernel. We use that approach to estimate the relative individual fecundities in the case-study of *Sorbus torminalis* and we evaluate how they depend on explanatory variables, whether they are spatially correlated and at which spatial scale. We found that the method we used here provided estimates of the effective male reproductive density that differ from the estimates obtained with current approaches.

MATERIAL AND METHODS

Studied species and experimental design

Sorbus torminalis (L. Crantz) is a scattered woody species. Individuals are hermaphrodites but mostly outcross, due to self-incompatibility and/or high abortion rates of inbred seeds. Flowers are pollinated by insects (Hymenoptera, Diptera and Coleoptera).

The study site covered 472 ha in the managed Rambouillet forest, France (Fig. 1). The study population consisted in 185 reproductive trees among which $n_p=172$ flowered in 2000 (see details in (Oddou-Muratorio *et al.*, 2006). Each reproductive tree was mapped and genotyped at six microsatellite loci MSS1, MSS5, MSS6, MSS9, MSS13, MSS16 revealing 6 to 21 alleles per locus (Oddou-Muratorio *et al.*, 2001). The theoretical exclusion probability for the study trees was 0.987. Additionally, we recorded for each reproductive tree (i) its diameter at breast height (6 classes from <10 cm to <60cm), (ii) its neighbourhood density (5 classes: suppressed, co-dominant, dominant, isolated, edge tree) and (iii) its flowering intensity (5 classes: absent, anecdotal, significant, massive and missing data). Finally, reproductive trees were classified into three groups (Fig. 1) corresponding to different silvicultural stages (Zone 1: dense and young oak-stand, Zone 2: old stand with moderate density of oaks, Zone 3: old stand with low density of oaks).

A total of 1075 seeds were sampled from 60 mother-trees (2-27 seeds per tree, mean=17.9, SD=3.9) evenly located throughout the site (Fig. 1). The pairwise distances between mother-trees varied from 2.5m to 2747m (mean=1016m, SD=611m).

Modelling variance of male fecundity, pollen dispersal and pollen pool composition

The mating model developed here relies on previous models of dispersal kernel and pollen pool composition (Oddou-Muratorio *et al.*, 2005) but differs in the way fecundity is modelled. Inter-individual variation of fecundity is expressed as a random individual effect. The different components of the model are detailed below (chart on Fig. 2).

Fecundity: The male fecundity of each mature individual k is denoted F_k and assumed to follow a log-normal distribution of mean 1 and variance Σ^2 :

$$E[F_k] = 1, \text{Var}[F_k] = \Sigma^2.$$

Because the observed data relate to proportions of pollen from potential fathers in sampled pollen pools and not to absolute fecundities, we arbitrarily set a reference level for absolute fecundity, which is biologically meaningless. The individual fecundities modelled are thus relative to this reference level. Because we choose 1 for the reference level, all components of the model must be invariant under multiplication by any constant.

By definition of a log-normal distribution, the logarithm of individual fecundities is normally distributed with variance σ^2 and mean $-\sigma^2/2$, both related to Σ^2 :

$$f_k = \log(F_k), \text{ with } f_k \sim N\left(-\frac{\sigma^2}{2}, \sigma^2\right),$$

and

$$\sigma^2 = \log(\Sigma^2 + 1) \text{ or } \Sigma^2 = e^{\sigma^2} - 1 \quad (\text{eqn 1})$$

We choose this distribution because it is L-shaped for large variances σ^2 (>1), with a long-tail leading to occasional high values of fecundity. Assuming this log-normal distribution, the likelihood function for the set of all fecundities $\mathbf{F} = \{F_k\}_{k=1 \dots n_p}$ is given by (Patel *et al.*, 1976)

$$\Pr(\mathbf{F}|\sigma^2) = \prod_{k=1 \dots n_p} \frac{1}{\sqrt{2\pi}\sigma F_k} \exp\left(-\frac{\left(f_k + \frac{\sigma^2}{2}\right)^2}{2\sigma^2}\right), \quad (\text{eqn 2})$$

where n_p is the number of pollen donors in the study site (here $n_p = 172$). Note that a log-normally distributed variable multiplied by any constant is still log-normally distributed.

The effective density of pollen donors (d_{ep}) is defined by the number of equi-fertile pollen donors per unit area that would provide a probability of co-paternity before dispersal equal to that observed (Austerlitz & Smouse, 2002). It is related to the variance of male fecundity Σ^2 through the relation (Oddou-Muratorio *et al.*, 2005), Supplementary material)

$$\frac{d_{obs}}{d_{ep}} = \frac{\text{Var}[F_k] + E[F_k]^2}{E[F_k]^2} = \Sigma^2 + 1 = e^{\sigma^2}, \quad (\text{eqn 3})$$

where d_{obs} is the observed density of male reproductive trees. The parameter d_{obs}/d_{ep} , by rescaling the variance of fecundity by the squared mean fecundity, is independent of the arbitrary choice of the mean fecundity.

Dispersal kernel: Following several studies (Hardy *et al.*, 2004; Oddou-Muratorio *et al.*, 2005; Robledo-Arnuncio & Gil, 2005) we used an exponential power dispersal kernel to model the proportion of pollen released at the punctual source (0,0) that contributes to the pollen pool at point (x,y):

$$p(a, b; x, y) = \frac{b}{2\pi a^2 \Gamma(2/b)} \exp\left[-\left(\frac{r}{a}\right)^b\right] \quad (\text{eqn 4})$$

with $r = \sqrt{x^2 + y^2}$ and Γ the classical gamma function.

For this family of functions b is a shape parameter providing fat-tailed kernels for $b < 1$ and thin-tailed for $b > 1$ and a is a scale parameter, it has the dimension of a distance and is proportional to mean dispersal distance δ equal to (Clark *et al.*, 1998)

$$\delta = a \frac{\Gamma(3/b)}{\Gamma(2/b)}.$$

Pollen pool composition: Given the fecundities $\mathbf{F} = \{F_k\}_{k=1 \dots n_p}$ and the dispersal kernel $p(a, b; x, y)$, the proportion of pollen from each father k in the pollen pool of each mother j originating from all known fathers is given by the mass-action law (Fig. 2B)

$$\pi_{jk} = \frac{F_k p_{jk}}{\sum_{l: \text{father}} F_l p_{jl}}, \quad (\text{eqn 5})$$

where $p_{jk} = p(a, b; x_j - x_k, y_j - y_k)$ and (x_j, y_j) , (x_k, y_k) are the coordinates of mother j and father k . Note from this expression that the pollen pool composition depends on the relative fecundities and not on the absolute fecundities.

Mating model and probabilities of the offspring genotypes: Following the classical mating models (Adams & Birkes, 1991; Burczyk *et al.*, 2002) we assume that an ovule of a mother j has the probabilities s , m and $(1-m-s)$ to be fertilised by a pollen grain from (i) the mother tree (self-fertilisation), (ii) an uncensored father tree from outside the study site (immigration) and (iii) any censored father within the study site, respectively. The probability (or likelihood) of the set of genotypes $\mathbf{g} = \{g_o\}_{o: \text{offspring}}$ of the sampled seeds is thus

$$L(\mathbf{g} | \mathbf{F}, \sigma^2, \delta, b, s, m) = \prod_{o: \text{offspring}} \left[s T(g_o | g_{j_o}, g_{j_o}) + m T(g_o | g_{j_o}, AF) + (1-s-m) \sum_{k: \text{father}} \pi_{j_o k} T(g_o | g_{j_o}, g_k) \right] \quad (\text{eqn 6})$$

where π_{jk} are the compositions of the pollen pools (eqn 5). $T(g_o | g_{j_o}, X)$ is the Mendelian segregation probability (Meagher, 1986) of the offspring genotype (g_o) given the genotype of the mother (g_{j_o}) and X , where X corresponds (i) to the genotype of the mother in the case of self-fertilisation (ii) to the allelic frequencies in the pollen pool external to the neighbourhood (AF) in the case of outcrossing with a non sampled father tree or (iii) to the genotype of the considered father tree (g_k) in the case of outcrossing with a sampled male k . The AF are computed from the inferred paternal gametes of 621 offspring that had no compatible male parent within the study site (including seeds sampled in 1999 and 2000).

Bayesian estimation through MCMC algorithm

Using the above model, our objectives were to estimate (i) the variance of (log-) fecundity (σ^2) to infer the ratio d_{obs}/d_{ep} , (ii) the dispersal parameters (δ , b) and the mating system parameters (s , m), and (iii) all individual relative fecundities F_k (see Figure 1 for the

relations between the parameters, the observations and the unobserved individual fecundities).

The standard methods based on maximum likelihood would either estimate all F_{ks} as parameters (and subsequently infer σ^2) or estimate directly σ^2 requiring the computation of a likelihood through the integration over all possible values for all F_{ks} . Both solutions appeared inadequate here because the first requires maximizing the likelihood relatively to a huge number of parameters, which is technically difficult (if possible), and the second makes any computation of the likelihood time-consuming and its maximization impossible. We thus developed a Bayesian framework for models and data that are similar to ours. The Bayesian analysis computes the posterior distributions for the parameters, i.e. their distributions conditional on the observations (Beaumont & Rannala, 2004; Clark & Gelfand, 2006). Prior distributions for the parameters are first defined and then modified according to the observations, here the genotypes of the sampled seeds. We used the following improper prior distributions for the parameters: $PDF(\sigma) \propto \sigma^{-1}$ for σ in (0,5); $PDF(\delta) \propto \delta^{-1}$ for δ in (0,100000); $PDF(b) \propto b^{-1}$ for b in (0,5); $PDF(m) \propto m^{-1}$ for m in (0,0.9), $PDF(s) \propto s^{-1}$ for s in (0,0.1) (these intervals guarantee that $m+s < 1$).

We used a MCMC approach based on the Metropolis-Hastings algorithm. The posterior distributions are reached at ‘equilibrium’ by a Markov chain in which all parameters are randomly modified at each iteration, following a proposal-acceptation-rejection scheme detailed in the Appendix (Gilks *et al.*, 1996). At each iteration t , where the current values of the parameters were $(\sigma_t, \delta_t, b_t, m_t, s_t)$ and F_t , we successively proposed new values for σ , all F_k ’s, δ , b , m and s respectively by drawing a random value centred on the current value of the parameter. The proposed value was accepted if it increased the log-likelihood. When it decreased it, we drew randomly whether we accepted the proposed value, with a probability decreasing with the loss of log-likelihood or with a small probability if it decreased it.

We initiated the Markov chain with $(\sigma_0, \delta_0, b_0, m_0, s_0) = (2,5000, 1, 0.5, 0.1)$ and $F_0 = (1,1,...,1)$ as starting values for the MCMC. We simulated 110 000 iterations and only kept the last 100 000 ones from which we computed the posterior distributions for each parameter and their means and medians.

Estimation of effective density and individual fecundities

At each iteration t , the effective density was computed through the ratio

$$\frac{d_{obs}}{d_{ep,t}} = e^{\sigma_t^2},$$

and the posterior distribution for this ratio was computed as for the other parameters.

Using the individual fecundities at each iteration t , F_t , we computed the posterior distribution for each individual fecundity F_k and considered its mean \bar{F}_k as an estimate of F_k .

We also computed the individual relative fecundities predicted from the three covariates (diameter, neighbourhood density, and flowering intensity). These predictions, \hat{F}_k , were computed using the effects of the covariates estimated by maximum likelihood in the spatial mating model ((Oddou-Muratorio *et al.*, 2005), Table 3). We analyzed the differences between the log-transformed estimated fecundities and the fecundities predicted from the covariates, $R_k = \log(\bar{F}_k) - \log(\hat{F}_k)$, to search for a spatial structure in these residuals. We first tested the difference in the residuals R_k among zones defined by different silvicultural stages using a one-way ANOVA (using R software). Additionally, to investigate whether the residuals R_k are correlated for neighboring trees, we estimated the empirical variogram for bins of distances equal to (0, 25, 50, 75, 100, 150, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900, 1000m) and we tested the significance of the estimated covariances through permutation tests (package geoR in R 2.1.1).

Computation of the effective number of pollen donors

The classical way to measure the variance of MRS is to compute the effective number of pollen donors, N_{ep} in the pollen pool within a single maternal sibship. We computed here the effective number of pollen donors, N_{ep} , in the whole pollen pool sampled by the 60 mother-trees. It was calculated as the inverse of the probability of paternal identity for two seeds drawn at random within the 1075 sampled seeds. It was computed using the model developed here and three nested models (see below) to investigate how N_{ep} depends on the spatial and fecundity components.

For this purpose, every 50 iterations we computed t of the MCMC the pollen pool compositions $\{\pi_{jk}\}_{k=1\dots n_p}$ for the 60 mothers j , following equation (5) with the individual fecundities, F_t , and the dispersal parameters, δ_t and b_t , and used them to calculate the effective number of pollen donors:

$$N_{ep} = \left(\sum_{k:\text{fathers}} \left(\sum_{j:\text{mothers}} \left(\frac{n_j}{n} \pi_{jk} \right) \right)^2 \right)^{-1}, \quad (\text{eqn 7})$$

where n_j is the number of seeds sampled on mother j and n the total number of seeds sampled. The posterior distribution for N_{ep} was computed from the set of 2000 values (=100 000/50).

The effective number of pollen donors was also computed with equation (7) in three nested models considering (i) spatial effects only (pollen pools computed with equation (5), all $F_k=1$ and $(\delta, b)=(743, 0.33)$), (ii) effects of covariates only (pollen pools computed with equation (5), all F_k equal to \hat{F}_k and $p(\delta, b; \cdot)=1$), and (iii) effects of covariates and spatial effects (pollen pools computed with equation (5), all F_k equal to \hat{F}_k and $(\delta, b)=(743, 0.33)$). These dispersal parameters and the fecundities \hat{F}_k predicted from the covariates come from (Oddou-Muratorio *et al.*, 2005).

We then used our model to investigate how the effective number of pollen donors N_{ep} in the whole pollen pool of the sampled mother-trees is sensitive to the choice of the sampled mothers and number of seeds genotyped for each. To do this, we recomputed the paternal contributions and deduced the values of N_{ep} for two other sets of seeds: (i) 308 seeds from the 60 mother-trees that were successfully categorically assigned to a father using the Cervus software (Oddou-Muratorio *et al.*, 2003) and (ii) the same number of seeds collected over all of the 172 reproductive trees of the population thus considered as equi-fertile mother-trees. Unfortunately, the true N_{ep} in the set of seeds that contribute to the next generation could not be evaluated because it strongly depends on individual female fertilities which are unknown.

RESULTS

Estimates of the dispersal kernel and mating system parameters

The Bayesian approach used here gave estimations of the dispersal parameters consistent with previous ones: the posterior distribution for the mean dispersal distance δ was concentrated on a 95%- credibility set (482, 1238) with a mean of 737.6m, a median of 699.1m, (Fig. 3C) and that for the shape parameter b was concentrated on (0.22, 0.37) with both mean and median at 0.29 (Fig. 3D). The estimation of the self-fertilization and immigration rates were also close to those previously obtained, with mean=0.43 for m with (0.40, 0.46) as a 95% credibility set (Fig. 3E) and mean=0.005 for s with (0.001, 0.011) as a

95% credibility set (Fig. 3F).

Estimates of the variance of male fecundity

The ratio d_{obs}/d_{ep} was high, revealing a large variance of the individual fecundities. 95% of the posterior distribution for d_{obs}/d_{ep} was concentrated on (5.1, 41.5) with a mean of 14.4, a median of 11.6 and a mode close to 10 (Figure 3B). This large variance was related to the L-shaped distribution of the individual fecundities: a few individuals were very fecund (especially one with a relative fecundity of 7.4) while the large majority of individuals had a low fecundity (Fig 4A-B).

The inter-individual variance of male fecundity was much higher than the variance explained by diameter, neighbourhood density and flowering intensity ($d_{obs}/d_{ep}=2.0$ in (Oddou-Muratorio *et al.*, 2005)). The data obtained here fit better the model based on random effects than on fixed effects. This was illustrated by the conditional likelihood (equation 6) that reached values as high as -13728 during the MCMC whereas the maximum likelihood was -13955 in the previous model including fixed effects (Fig 3A). Although it is not academic to compare Bayesian results with maximum likelihood approach, we still noticed the large increase (227) for the log-likelihood with a reduction of 160 degrees of freedom (172 individuals – (5+4+3) parameters in the fixed effects model).

The fecundities estimated in the random effect model were correlated with the fecundities predicted from diameters, neighbourhood densities and flowering intensities (Fig. 5A, P -value<0.0001 for the log-log regression), but only weakly ($R^2=0.20$). The residuals of individual fecundity, R_k , depended on the zone ($F_{2,169}=9.9$, $P<0.0001$): they were significantly larger in the first zone than in the other two (Fig. 5C). The residuals of individual fecundity, R_k , were also spatially structured, i.e. correlated within the first classes of distances (first three bins for distances < 75m, Fig. 5D). This correlation confirmed the visible aggregated pattern of positive and negative values of R_k (Fig. 5B).

Consequences for variance of male reproductive success

For the 1075 seeds collected from the 60 sampled mothers, the effective number of pollen donors inside the area was estimated as $N_{ep} = 53.1$ by the random effect model (Table 1). If we consider only the spatial pattern of pollen donors and the sampled mother-trees and no variation in male fecundities, this resulted in an N_{ep} (122.8) that was twice as large, whereas the effect of covariates on fecundity resulted in a smaller N_{ep} (86.2). Considering both the spatial pattern and the effects of neighbourhood density, diameter and flowering intensity lead to a higher N_{ep} than the random effect model (80.3 compared to 53.1). As expected, slightly higher values were obtained when computing the effective number of pollen donors for an even number of seeds collected per tree (Table 1). In particular, the effect of the spatial component was greatly reduced ($N_{ep}=157.7$ for 172 mother-trees). Finally, for the 308 seeds categorically assigned to a father by Cervus for the random effects model, we estimated $N_{ep}=46.3$, which is close to $N_{ep}=49.6$ directly estimated from the paternities categorically assigned to these seeds (Oddou-Muratorio *et al.* 2003).

DISCUSSION

We succeeded in estimating the variance of individual relative fecundities using a Bayesian scheme. Our model dissected the variance of male reproductive success (MRS) into a variance of male fecundity and a spatial component (through a dispersal kernel). The former accounted for variation in the relative amounts of efficient pollen emitted by individuals. The latter component accounted for the positions of the fathers relative to the 60 sampled mothers, which affect the MRS when dispersal is limited by distance. For *Sorbus torminalis* in the study area, accounting only for the distance among trees reduced the effective number of

pollen donors from 172 observed reproductive trees to 122 while the variance of fecundity also reduced this number to 53. These estimations proved the necessity of taking into account both fecundity and dispersal in the estimations of MRS in this species.

Our approach is the first to estimate the whole inter-individual variance of fecundity jointly with a spatial configuration effect, while previous studies only estimated the variance of fecundity due to a few covariates. To make a comparison with the linear-model theory, the model we built here considered fecundity as a random effect (e.g., a father effect) while the neighbourhood model from (Burczyk *et al.*, 2002) considered the effects of quantitative explanatory variables (analogous to a multiple regression) and the model of (Oddou-Muratorio *et al.*, 2005) considered the fixed effects of class variables (analogous to an ANOVA). Further extensions of the present study may associate fixed with random effects.

Methodological aspects

The Bayesian approach (BA) was chosen for its adequacy to hierarchical models (Banerjee *et al.*, 2003; Beaumont & Rannala, 2004), i.e. models where the likelihood of the observations (here the genotypes) can be written easily if it is conditional on an intermediate level of unobserved variables, here the random individual fecundities, themselves depending on parameters. Our results offer an additional example of the adequacy of Bayesian approaches for parentage analyses based on genetic data (e.g. (Emery *et al.*, 2001; Hadfield *et al.*, 2006). In particular, the MCMC procedure used here allowed easy exploration of a 172-dimensional space for the individual fecundity parameters. We checked that the Bayesian approach did not result in intrinsic differences with a maximum likelihood approach (MLA): we implemented a Bayesian analysis of the fixed effects model previously analysed through MLA (Oddou-Muratorio *et al.*, 2005). We obtained similar results with both approaches (algorithm and results in Supplementary material 2). The consistency of the dispersal parameters estimated by the random effect model (this study) and the fixed effects model (treated with either MLA or BA) also supports the robustness of our method.

More specifically, the choice of the prior distributions is a critical point in all Bayesian approaches, especially when used for convenience such as in the study. The similarity of the results obtained for the fixed effects model from the BA and the MLA indicates that the prior distributions did not affect the results for the dispersal and mating parameters. The more likely values of the posterior distributions from the BA were similar to those obtained with the MLA, which would not hold if the prior distribution largely affected the posterior one. However, this argument is true for the dispersal and mating parameters but not necessarily for the inter-individual variance of fecundity.

The low sensitivity of our results to the prior distributions is certainly due to the quantity and quality of information in our microsatellite data, compared to that contained in the prior distributions. More than a thousand seeds were analysed under a genetic system that provided a high exclusion power. Furthermore, the rate of mistyping is expected to be sufficiently low (Oddou-Muratorio *et al.*, 2003) and is assumed null in the transition probabilities (Slavov *et al.*, 2005). Situations of fewer seeds or a significant probability of mistyping (e.g. (Gérard *et al.*, 2006) could necessitate a more careful definition of the prior distributions. In these cases, defining an informative prior distribution from published data or from external expertise would improve the accuracy of the results (Clark & Gelfand, 2006).

The distribution used for the individual fecundities should also be discussed independently of the statistical approach implemented. We chose a log-normal distribution, which is long-tailed and strongly skewed and kurtosed, because we expected (and found) an L-shaped distribution for the individual fecundities. This results from a few individuals having an exceptionally high fecundity.

However, as other long-tailed distributions, the log-normal distributions produces

large difference between the estimated variance and the empirical variance (here, the variance of the individual fecundities F_k) (Huber, 1981). We found here a large difference between the posterior distribution of the variance parameter σ leading to a ratio $d_{obs}/d_{ep} \sim 13$ and the posterior mean of the empirical variances of individual fecundities (i.e. directly computed from the 172 estimated fecundities) providing $d_{obs}/d_{ep} = 3.3$ (credible set (2.7, 4.20)). This difference results from the sensitivity of the variance of a long-tailed distribution to extreme events (i.e. very fecund individuals) that are usually not sampled enough to reduce this effect. For instance, the largest F_k in the data set was ~ 7 whereas under the log-normal distribution we estimated, there is one chance in two to find an $F_k > 43$ with 1000 sampled individuals and an $F_k > 112$ with 10000 sampled individuals.

We thus face here the dilemma of estimating either the actual variance of fecundities among the individuals in the study area or the variance of fecundities in the whole population of pollen donors. In the latter case the estimation relies strongly on the tail of the distribution, i.e. an extrapolation to more extreme fecundities than those sampled. When the estimation of the variance of fecundity is compared with an estimate from TwoGener or used to simulate dispersal patterns in a larger area, the estimated variance (of the whole population) should be favoured. On the contrary, the empirical estimation from the individual fecundities is more pertinent when quantifying the departure from random mating in a given population, especially if isolated.

To investigate the impact of the distribution chosen for the random effect on the estimation of the variance of fecundity, we repeated the analysis with a gamma distribution for the individual fecundities, which has a shorter tail than the log-normal distribution. We estimated a ratio d_{obs}/d_{ep} concentrated on (2.65, 4.17) with a posterior mean of 3.34, which is smaller than that estimated with a log-normal distribution but closer to that from the estimated F_k values, as expected from the shorter tail of the gamma distribution. The ratio was also much closer to the ratio d_{obs}/d_{ep} obtained from the parentage approach with three covariates (densities, flowering, diameter), although still significantly higher, but greatly lower than the estimate of TwoGener. The Bayes factor computed from the likelihoods explored by the two MCMC (with the log-normal and with the gamma distributions) strongly supports the gamma distribution (BF $\sim 4 \cdot 10^5$) but other distributions with long tails could be worth investigating (e.g. Weibull) before drawing a definitive conclusion. These large differences confirm that the statistical properties of the estimator of the ratio d_{obs}/d_{ep} should be investigated under different scenarios for the determinants and distributions of the individual fecundities. Such investigation, based on simulated data, would also allow a more formal assessment of the performances of our method compared to previous ones in various spatial patterns for the pollen donors.

Consequences for Sorbus torminalis

Applying the new method to *Sorbus torminalis*, we estimated a variance of fecundity leading to ratio d_{obs}/d_{ep} of about 13 for the whole population and about 3.3 for the 172 actual pollen donors in the study area. Both estimates are large and fall between the values estimated from our three-covariates model (2) and from TwoGener (30). Interestingly, the ratio d_{obs}/d_{ep} for the whole population is closer to the TwoGener estimate, which represents the ratio over the whole forest, whereas the ratio for the actual 172 individuals is closer to that from the three-covariates model.

We expected the estimation of d_{obs}/d_{ep} to be below 30 because TwoGener assumes the random distribution of pollen donors whereas the mature trees in the experimental site are clustered. The clusters have radii ~ 100 m (Oddou-Muratorio *et al.*, 2004), 10 times lower than the axial dispersal standard deviation (~ 1200 m from Eq. 9 in (Robledo-Arnuncio & Austerlitz, 2006) and the within-cluster density is ~ 3 tree/ha, 10 times higher than the global

density. These values, associated with a fat-tailed dispersal kernel ($b=0.3$), are typical of a strong reduction of N_{ep} compared to the Poisson distribution case (Robledo-Arnuncio & Austerlitz, 2006), which should thus contribute to the overestimation of the ratio d_{obs}/d_{ep} by TwoGener.

We also expected the estimates of d_{obs}/d_{ep} to be higher than 2 because the three covariates used in the model were unlikely to explain the whole variance of fecundities. And indeed, despite a clear correlation between the individual fecundities F_k and the fecundities predicted from the covariates a large unexplained variation remained around the trend: some trees with “favourable traits” had actually a low F_k (e.g. one dominant tree with a diameter of 56 cm and massive flowering had a fecundity of 0.34) and conversely some trees had high fecundities despite “unfavourable traits” (e.g. one tree with a diameter of 10 cm and an anecdotal flowering had a fecundity of 3.64).

Finally, the better modelling of individual fecundities, although permitting much finer descriptions of the mating patterns observed in the *Sorbus torminalis* plot, did not result in a different estimation of the pollen dispersal kernel compared to previous studies (Oddou-Muratorio *et al.*, 2005). We are thus left with the tricky task of characterizing precisely the tail of the dispersal kernel, which has a strong influence on the population genetics at the landscape scale. New methods, such as linking of the migration rate parameter with the dispersal kernel parameters (Goto *et al.*, 2006; Shimatani *et al.*, 2007) are needed to improve this aspect.

The method we developed here enabled us to further investigate the ecological processes contributing to the variance of individual fecundities. We found a spatial structure in the residuals of the individual fecundities, with fecundities correlated for trees separated by ~75m. The spatial structure we found could result from silvicultural practices because the zone effect was significant and the three zones in the study area actually corresponded to three different silvicultural stages (Oddou-Muratorio *et al.*, 2006). Mature stands (Zones 2 and 3, south-east) with lower densities of wildservice trees and oaks had lower fecundities whereas young stands with higher densities had higher fecundities. The differences found among zones could be induced by the increase in the activity of pollinators in denser areas. However, the estimated range of the spatial structure does not completely overlap with the zones (i.e. it is smaller, ~ 500m) and it is still detected after accounting for the zone effect. This spatial structure probably results from other mechanisms acting at smaller scales: environmental covariates that are spatially structured (soil depth, exposition, inter-specific competition...), spatial genetic structure (reported in (Oddou-Muratorio *et al.*, 2004) associated with heritable fecundity, or localised foraging activity of pollinators associated with the presence of other attractive species pollinated by the same insects (apple/pear trees, herbaceous...).

Phenology is a characteristic contributing much to the individual fecundity of a flowering male tree, and more specifically synchrony with female trees (Gérard *et al.*, 2006). For some species at high density, with large differences of flowering dates among individuals and short flowering periods for each tree, temporal distance between trees can influence much more strongly the mating patterns than spatial distances. In these cases, measuring the flowering dates of all trees would allow estimation of the effect of flowering delay on mating probability with a formalism similar to the one used here (Smouse & Sork, 2004) by replacing physical distances by temporal distances in the algorithm developed in this study. Because *Sorbus torminalis* is a scattered species with very short and synchronous flowering periods, spatial distance was expected as the strongest determinant of mating patterns and we did not characterise the phenologies of all pollen donors. We thus could not simply use the available information about the flowering phenology of the 60 sampled mothers in the computation of the pollen pools (equation 5). However, alternatively, the Bayesian approach developed here

could also include the phenology of each pollen donor as an additional latent variable, such as fecundity, and estimate its effects jointly with distance on the composition of the pollen pool. Here, we did not include phenology but rather assumed that the individual fecundities F_k could be regarded as “effective” fecundities perceived by the 60 sampled mothers. Consequently, if asynchronous flowering was important the resulting effective density we found is not suitable when considering only one mother because a single mother receives the pollen from fewer males, being confronted to a smaller effective density.

Finally, the thorough investigation of the spatial component of the variations of MRS should be extended to contribute to a better understanding the inter-annual variations of fecundities (Irwin *et al.*, 2003). Because each mature tree participates in numerous mating episodes that define the reproductive success at the generation scale, the temporal variations of fecundities - and particularly the individual \times year interaction - strongly affect the intensity of genetic drift and subsequently its impact on evolution. The extension of our Bayesian approach to multi-annual data could lead to quantification of the inter-annual variations of fecundities together with the spatial component of mating success.

REFERENCES

- Adams WT, Birkes DS (1991) Estimating mating patterns in forest tree populations. In: *Biochemical markers in the population genetics of forest trees* (eds. Fineschi S, Malvolti ME, Cannata F, Hattemer HH), pp. 157-172. SPB Academic Publishing, The Hague.
- Austerlitz F, Dick CW, Dutech C, *et al.* (2004) Using genetic markers to estimate the pollen dispersal curve. *Molecular Ecology* **13**, 937-954.
- Austerlitz F, Smouse PE (2001) Two-generation analysis of pollen flow across a landscape. II. Relation between Φ (ft), pollen dispersal and interfemale 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.
- Bacles CFE, Burczyk J, Lowe AJ, Ennos RA (2005) Historical and contemporary mating patterns in remnant populations of the forest tree *Fraxinus excelsior* L. *Evolution* **59**, 979-990.
- Banerjee S, Carlin BP, Gelfand AE (2003) *Hierarchical modeling and analysis for spatial data* Chapman & Hall/CRC, Boca Raton.
- Beaumont MA, Rannala B (2004) The Bayesian revolution in genetics. *Nature Reviews Genetics* **5**, 251-261.
- 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 neighbourhood model. *Molecular Ecology* **11**, 2379-2391.
- Burczyk J, Koralewski TE (2005) Parentage versus two-generation analyses for estimating pollen-mediated gene flow in plant populations. *Molecular Ecology* **14**, 2525-2537.
- Clark JS, Gelfand AE (2006) A future for models and data in environmental science. *Trends in Ecology & Evolution* **21**, 375-380.
- Clark JS, Macklin E, Wood L (1998) Stages and spatial scales of recruitment limitation in southern Appalachian forests. *Ecological Monographs* **68**, 213-235.
- Devaux C, Lavigne C, Austerlitz F, Klein EK (2007) Modelling and estimating pollen movement in oilseed rape (*Brassica napus*) at the landscape scale using genetic markers. *Molecular Ecology* **16**, 487-499.
- Devlin B, Ellstrand NC (1990) Male and female fertility variation in wild radish, a hermaphrodite. *American Naturalist* **136**, 87-107.
- Emery AM, Wilson IJ, Craig S, Boyle PR, Noble LR (2001) Assignment of paternity groups

- without access to parental genotypes: multiple mating and developmental plasticity in squid. *Molecular Ecology* **10**, 1265-1278.
- Gilks WR, Richardson S, Spiegelhalter DJ (1996) Markov chain Monte Carlo in practice. In: *Interdisciplinary Statistics*. Chapman & Hall, London.
- Goto S, Shimatani K, Yoshimaru H, Takahashi Y (2006) Fat-tailed gene flow in the dioecious canopy tree species *Fraxinus mandshurica* var. *japonica* revealed by microsatellites. *Molecular Ecology* **15**, 2985-2996.
- Gérard P, Klein EK, Austerlitz F, Fernandez-Majarres JF, Frascaria-Lacoste N (2006) Assortative mating and differential male mating success in an ash hybrid zone population. *BMC Evolutionary Biology* **6**, 96.
- Hadfield JD, Richardson DS, Burke T (2006) Towards unbiased parentage assignment: combining genetic, behavioural and spatial data in a Bayesian framework. *Molecular Ecology* **15**, 3715-3730.
- Hamrick JL (2004) Response of forest trees to global environmental changes. *Forest Ecology And Management* **197**, 323-335.
- Hardy OJ, Gonzalez-Martinez SC, Freville H, *et al.* (2004) Fine-scale genetic structure and gene dispersal in *Centaurea corymbosa* (Asteraceae) I. Pattern of pollen dispersal. *Journal of Evolutionary Biology* **17**, 795-806.
- Huber PJ (1981) *Robust statistics* John Wiley & Sons, New York.
- Irwin AJ, Hamrick JL, Godt MJW, Smouse PE (2003) A multiyear estimate of the effective pollen donor pool for *Albizia julibrissin*. *Heredity* **90**, 187-194.
- Klein EK, Lavigne C, Gouyon PH (2006a) Mixing of propagules from discrete sources at long distance: comparing a dispersal tail to an exponential. *BMC Ecology* **6**, 3.
- Klein EK, Lavigne C, Renard M, Gouyon PH (2006b) Pollen dispersal of oilseed rape: estimation of the dispersal function and effects of field dimension. *Journal of Applied Ecology*.
- Lande R, Arnold SJ (1983) The measurement of selection on correlated characters. *Evolution* **37**, 1210-1226.
- 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.
- Lavigne C, Klein EK, Vallée P, *et al.* (1998) A pollen-dispersal experiment with transgenic oilseed rape. Estimation of the average pollen dispersal of an individual plant within a field. *Theoretical and Applied Genetics* **96**, 886-896.
- Meagher TR (1986) Analysis of paternity within a population of *Chamaelirium luteum*. I. Identification of the most-likely male parents. *American Naturalist* **128**, 199-215.
- Nielsen R, Mattila DK, Clapham PJ, Palsboll PJ (2001) Statistical approaches to paternity analysis in natural populations and applications to the North Atlantic humpback whale. *Genetics* **157**, 1673-1682.
- Oddou-Muratorio S, Aligon C, Decroocq S, *et al.* (2001) Microsatellite primers for *Sorbus torminalis* and related species. *Molecular Ecology Notes* **1**, 297-299.
- Oddou-Muratorio S, Demesure-Musch B, Pelissier R, Gouyon PH (2004) Impacts of gene flow and logging history on the local genetic structure of a scattered tree species, *Sorbus torminalis* L. Crantz. *Molecular Ecology* **13**, 3689-3702.
- Oddou-Muratorio S, Houot ML, 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, Klein EK, Austerlitz F (2005) Pollen flow in the wildservice tree, *Sorbus torminalis* (L.) Crantz. II. Pollen dispersal and heterogeneity in mating success inferred from parent-offspring analysis. *Molecular Ecology* **14**, 4441-4452.

- Oddou-Muratorio S, Klein EK, Austerlitz F (2006) Real-time patterns of pollen flow in the wildservice tree, *Sorbus torminalis* III. Mating patterns and the ecological maternal neighborhood. *American Journal of Botany* **93**, 1650-1659.
- Patel JK, Kapadia CH, Owen DB (1976) *Handbook of statistical distributions* Marcel Dekker, INC., New York.
- Petit RJ, Hampe A (2006) Some evolutionary consequences of being a tree. *Annual Review of Ecology Evolution and Systematics* **37**, 187-214.
- Robledo-Arnuncio JJ, Austerlitz F (2006) Pollen dispersal in spatially aggregated populations. *American Naturalist* **168**, 500-511.
- Robledo-Arnuncio JJ, Gil L (2005) Patterns of pollen dispersal in a small population of *Pinus sylvestris* L. revealed by total-exclusion paternity analysis. *Heredity* **94**, 13-22.
- Schnabel A, Nason JD, Hamrick JL (1998) Understanding the population genetic structure of *Gleditsia triacanthos* L.: seed dispersal and variation in female reproductive success. *Molecular Ecology* **7**, 819-832.
- Shimatani K, Kimura M, Kitamura K, *et al.* (2007) Determining the location of a deceased mother tree and estimating forest regeneration variables by use of microsatellites and spatial genetic models. *Population Ecology* **49**, 317-330.
- Slavov GT, Howe GT, Gyaourova AV, Birkes DS, Adams WT (2005) Estimating pollen flow using SSR markers and paternity exclusion: accounting for mistyping. *Molecular Ecology* **14**, 3109-3121.
- 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.
- Smouse PE, Meagher TR (1994) Genetic-analysis of male reproductive contributions in *Chamaelirium luteum* (L) Gray (Liliaceae). *Genetics* **136**, 313-322.
- 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, Sork VL (2004) Measuring pollen flow in forest trees: an exposition of alternative approaches. *Forest Ecology And Management* **197**, 21-38.
- Sork VL, Nason J, Campbell DR, Fernandez JF (1999) Landscape approaches to historical and contemporary gene flow in plants. *Trends in Ecology & Evolution* **14**, 219-224.

ACKNOWLEDGMENTS

We are grateful to Céline Devaux, Santiago Gonzalez-Martinez, Claire Lavigne, Cindy Morris, Kenichiro Shimatani, Peter Smouse, Samuel Soubeyrand, and three anonymous reviewers for comments on the successive versions of the manuscript. This work was supported by the program “ECologie pour la Gestion des Ecosystèmes et de leur Ressources” funded by INRA.

APPENDIX. Algorithm of the MCMC used to compute the posterior distributions.

At each iteration t of the MCMC, where the current values of the parameters were $(\sigma_t, \delta_t, b_t, m_t, s_t)$ and \mathbf{F}_t , we did the following steps

Step 1— We generated a candidate value for the parameter σ as $\sigma_t^* = \sigma_t \times LN(0,0.1)$, i.e. we multiplied the current value by a random number drawn from a log-normal distribution (LN stands for “log-normal”). This resulted in multiplying the current value by a number between 0.5 and 2 in 97% of the cases.

Using equation (2) we computed the acceptance ratio:

$$r_{\sigma} = \frac{\Pr(\mathbf{F}_t | \sigma_t^*) \Pr(\sigma_t | \sigma_t^*) \text{Prior}(\sigma_t^*)}{\Pr(\mathbf{F}_t | \sigma_t) \Pr(\sigma_t^* | \sigma_t) \text{Prior}(\sigma_t)},$$

which simplifies into

$$r_{\sigma} = \frac{\Pr(\mathbf{F}_t | \sigma_t^*)}{\Pr(\mathbf{F}_t | \sigma_t)}$$

because of the prior distribution chosen for σ , and the distribution used to draw the candidate (Supplementary material 1).

With probability $\min(1, r_{\sigma})$, the candidate value was accepted ($\sigma_{t+1}^2 = \sigma_t^{*2}$), otherwise it was rejected and the former value was kept ($\sigma_{t+1}^2 = \sigma_t^2$).

Step 2— For each successive individual k , we generated a candidate fecundity by drawing a random value for $F_{k,t}^*$ from a log-normal distribution

$$f_{k,t}^* \sim N\left(-\frac{\sigma_{t+1}^2}{2}, \sigma_{t+1}^2\right) \text{ and } F_{k,t}^* = \exp(f_{k,t}^*)$$

Using equation (6), we computed the acceptance ratio

$$r_{F,k} = \frac{L(\mathbf{g} | \mathbf{F}_t^{(k)*}, \sigma_{t+1}^2, \delta_t, b_t, m_t, s_t)}{L(\mathbf{g} | \mathbf{F}_t^{(k)}, \sigma_{t+1}^2, \delta_t, b_t, m_t, s_t)},$$

where $\mathbf{F}_t^{(k)*} = \{F_{1,t+1}, \dots, F_{k-1,t+1}, F_{k,t}^*, F_{k+1,t}, \dots, F_{n_p,t}\}$ and $\mathbf{F}_t^{(k)} = \{F_{1,t+1}, \dots, F_{k-1,t+1}, F_{k,t}, F_{k+1,t}, \dots, F_{n_p,t}\}$.

With probability $\min(1, r_{F,k})$, the candidate value was accepted ($F_{k,t+1} = F_{k,t}^*$), otherwise it was rejected ($F_{k,t+1} = F_{k,t}$).

Step 3— For each parameter δ , b , m , s successively, we generated a candidate value by multiplying the current value by a random number drawn in a $LN(0,0.1)$ distribution

We computed the ratio $r_{\delta} = \frac{L(\mathbf{g} | \mathbf{F}_{t+1}, \sigma_{t+1}^2, \delta_t^*, b_t, m_t, s_t)}{L(\mathbf{g} | \mathbf{F}_{t+1}, \sigma_{t+1}^2, \delta_t, b_t, m_t, s_t)}$ and accepted the candidate value δ_t^*

with probability $\min(1, r_{\delta})$.

We computed the ratio $r_b = \frac{L(\mathbf{g} | \mathbf{F}_{t+1}, \sigma_{t+1}^2, \delta_{t+1}, b_t^*, m_t, s_t)}{L(\mathbf{g} | \mathbf{F}_{t+1}, \sigma_{t+1}^2, \delta_{t+1}, b_t, m_t, s_t)}$ and accepted the candidate value b_t^*

with probability $\min(1, r_b)$.

We computed the ratio $r_m = \frac{L(\mathbf{g} | \mathbf{F}_{t+1}, \sigma_{t+1}^2, \delta_{t+1}, b_{t+1}, m_t^*, s_t)}{L(\mathbf{g} | \mathbf{F}_{t+1}, \sigma_{t+1}^2, \delta_{t+1}, b_{t+1}, m_t, s_t)}$ and accepted the candidate value m_t^*

with probability $\min(1, r_m)$.

We computed the ratio $r_s = \frac{L(\mathbf{g}|\mathbf{F}_{t+1}, \sigma_{t+1}^2, \delta_{t+1}, b_{t+1}, m_{t+1}, s_t^*)}{L(\mathbf{g}|\mathbf{F}_{t+1}, \sigma_{t+1}^2, \delta_{t+1}, b_{t+1}, m_{t+1}, s_t)}$ and accepted the candidate value s_t^* with probability $\min(1, r_s)$.

AUTHOR INFORMATION BOX

Etienne Klein is interested in modelling and estimating long-distance pollen dispersal. Nicolas Desassis has just finished his PhD in geostatistics, using in particular Bayesian statistics for spatial hierarchical models. Sylvie Oddou-Muratorio is now studying population genetics of expanding tree populations.

FIGURE LEGENDS

Fig. 1 Map of the study site. Reproductive trees (empty symbols) and sampled mother-trees (full symbols) are plotted within the three zones at different silvicultural stage (diamonds: dense and young oak-stand, triangles: moderately dense and old oak-stand, circles: sparse oak-stand).

Fig. 2 Flow chart of the hierarchical statistical model used to estimate fecundity and dispersal through a Bayesian approach. (A) Structure of the model presenting the strata of the model, the parameters used and the prior distributions. (B) Graphical 1D-representation of the dispersal kernel approach used to model the pollen pool composition over one mother tree. The information about positions and intensities (i.e. individual fecundities) of the pollen donors is used with the dispersal kernel to compute the relative contribution of each pollen donor to the pollen pool at any position. This relative contribution π_{jk} is the absolute contribution of the pollen donor (gray, dashed and dotted curves) divided by the total contribution of all pollen donors (black thick curve).

Fig. 3 Estimations of the parameters. (A) Conditional likelihood along the 100 000 iterations of the MCMC, together with the likelihood previously obtained under the fixed-effects model (dotted horizontal line). (B-F) Prior distribution (dashed line) and posterior distribution (histogram) for the ratio d_{obs}/d_{ep} , the mean dispersal distance δ , the shape parameter b , the immigration rate m and the self-fertilization rate s (B to F, respectively). The medians and means of the posterior distributions are given by the left and right (resp. and sometimes confounded) black vertical lines.

Fig. 4 Estimated individual fecundities. (A) Posterior means for the 172 individual fecundities of the 172 flowering mature trees. Different colours correspond to zones at different silvicultural stages (1: White, 2: Gray, 3: Black). Bars represent the 95%-credibility sets (B) Distribution of these 172 posterior means.

Fig. 5 Spatial pattern of the individual fecundities. (A) Log of individual fecundities (i.e. posterior means) estimated by the random effect approach developed here, plotted against the logarithm of individual fecundities predicted by the fixed effects model of Oddou-Muratorio *et al.* (2005) in which fecundities depended on (i) flowering intensity, (ii) diameter and (iii) neighbourhood density. The diagonal is also plotted (black line). (B) Spatial pattern of the residuals R_k (log-fecundity in the random effect model – log-fecundity in the fixed effects model). Negative/positive values are plotted as gray/black circles with diameters proportional to their absolute values. (C) Boxplots of the residual (log of) fecundities within each of the three zones at contrasted silvicultural stages. (D) Empirical semi-variogram of the residuals R_k

(log-fecundity in the random effect model – log-fecundity in the fixed effects model) (dots), and confidence intervals under the hypothesis of the absence of spatial structure (dashed lines).

Table 1 Effective number of pollen donors, N_{ep} , within the investigated area computed from different models and for different seed pools.

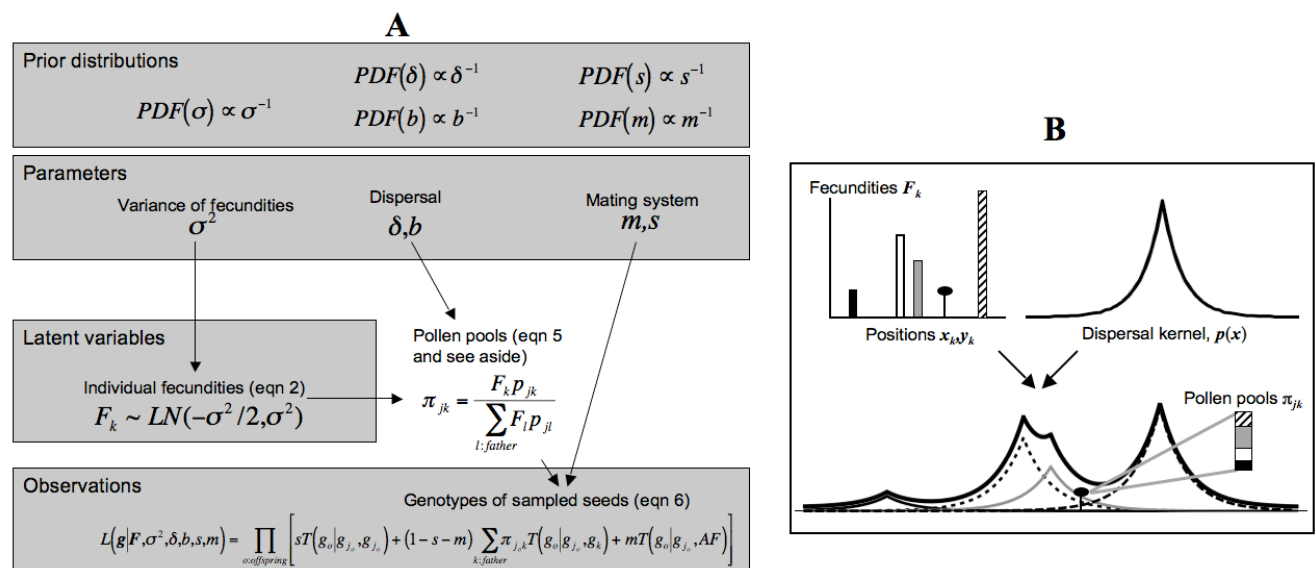
Hypothesis for the composition of pollen pools	Actual number of sampled seeds (min=2,max=27, mean=17.9)	Actual number of seeds categorically assigned by Cervus at 95% (min=0, max=11, mean=5.1)	Same number of seeds* for all of the 172 reproductive trees
Equal fecundity and random mating	172	172	172
Spatial only	122.8	110.5	157.7
Covariates only	86.2	86.2	86.2
Spatial and covariates	80.3	72.0	106.5
Spatial and random effects	53.1	46.3	74.1

* In this case, the actual number of seeds per tree does not affect the computation of N_{ep} because it is eliminated from equation (7).

500 m

Klein, E.K., Dessasis, N., Oddou-Muratorio, S. (2008). Pollen flow in the wildservice tree, *Sorbus terminalis* (L.) Crantz. IV. Whole inter-individual variance of male fecundity estimated jointly with the dispersal kernel. *Molecular Ecology*, 17 (14), 3323–3336. DOI : 10.1111/j.1365-294X.2008.03809.x

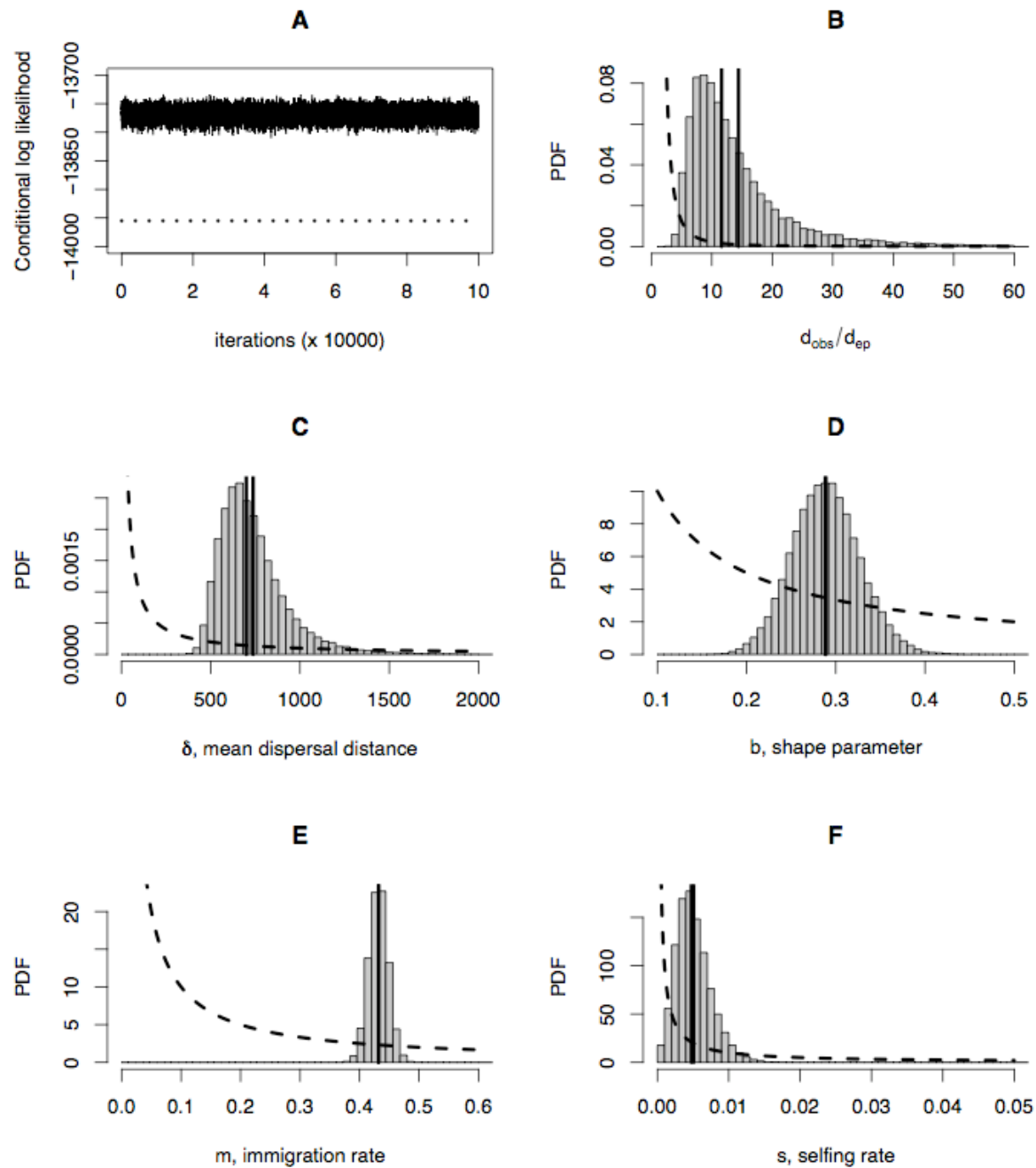
4 Fig. 2—



5

6

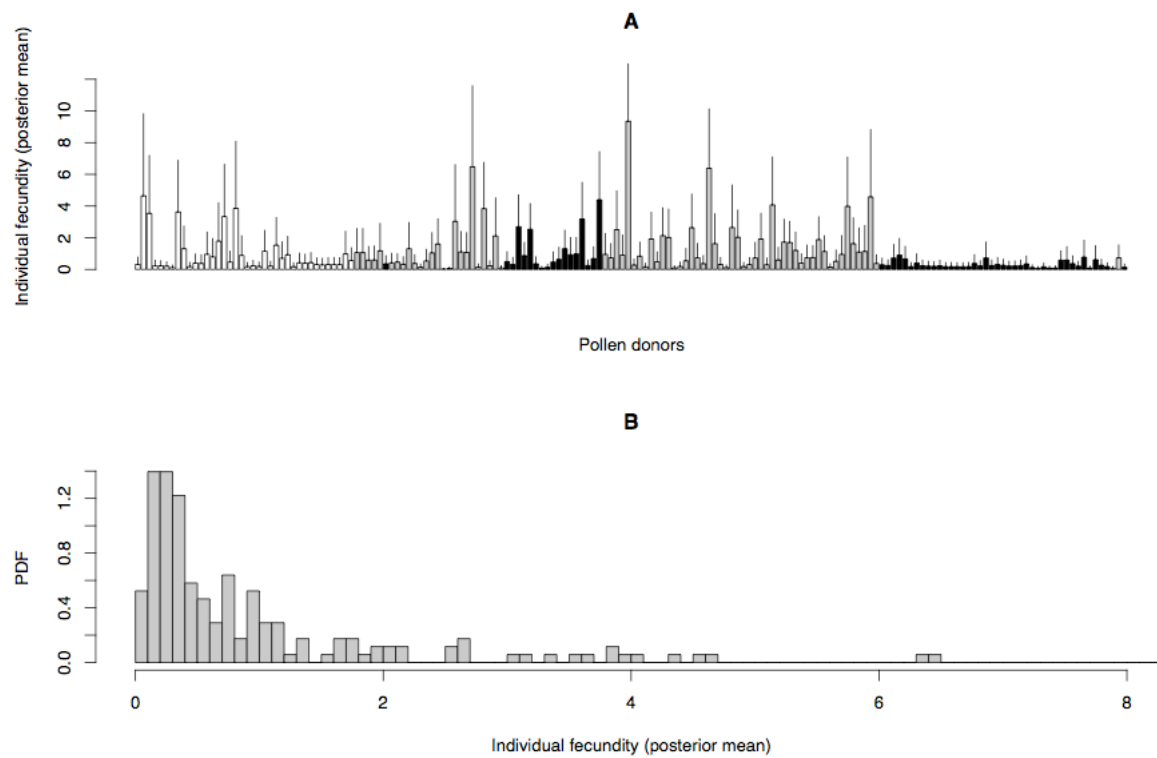
6 Fig. 3—



7

8

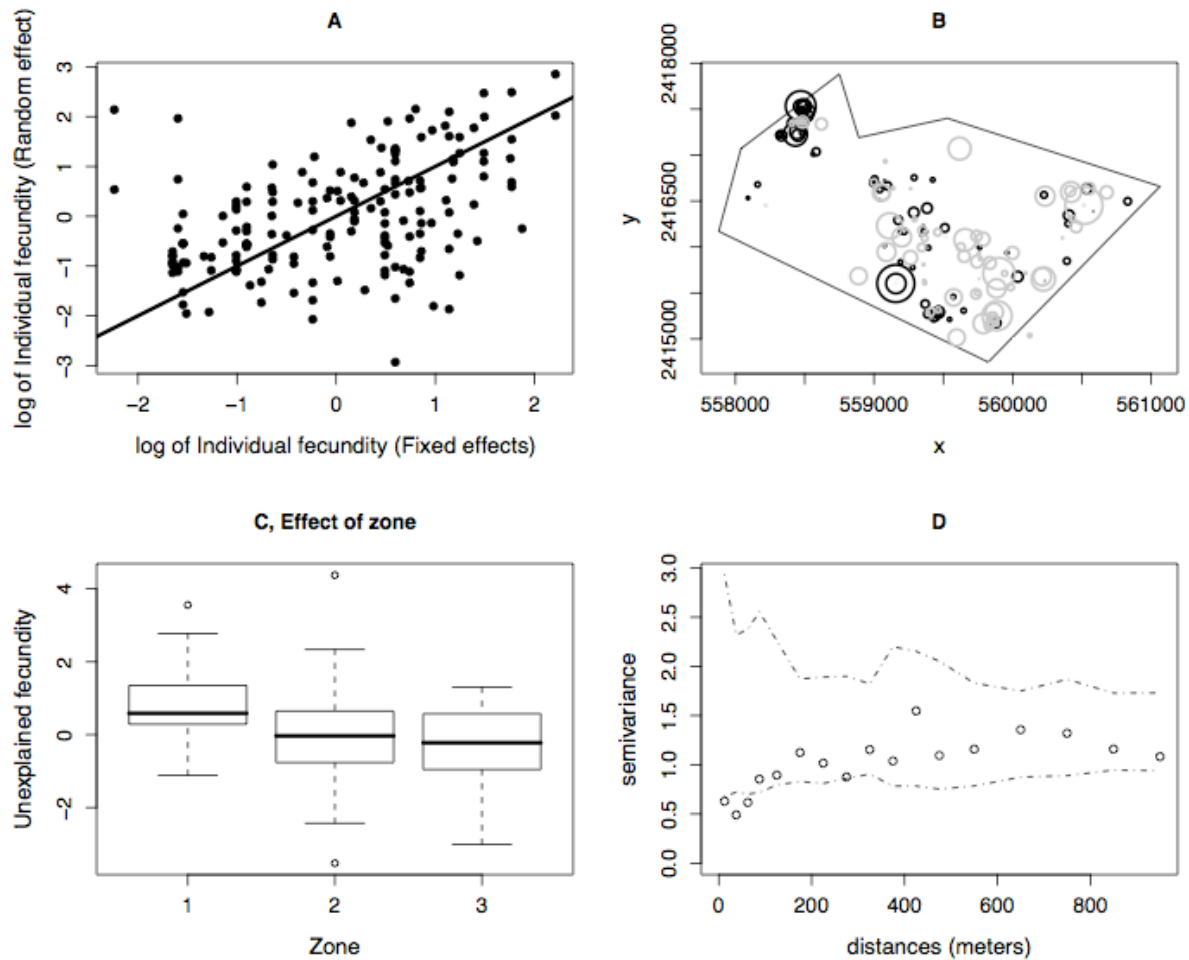
8 **Fig. 4—**



9

10

10 Fig. 5—



11