

**Supplementary material 2. Bayesian analysis of the fixed effects model developed in Oddou-Muratorio *et al.* 2005.**

*Fixed effect model*

The fecundities  $\mathbf{F} = \{F_k\}_{k=1\dots n_p}$  are modelled from three categorical explanatory variables. Each individual fertility is expressed as

$$F_k = \alpha_{\varnothing_k} \beta_{ND_k} \varphi_{FI_k}, \tag{eqn SM2.1}$$

where  $\varnothing_k$ ,  $ND_k$  and  $FI_k$  are the diameter, neighbourhood density and flowering intensity classes of the mature tree  $k$ .

Given the fecundities  $\mathbf{F} = \{F_k\}_{k=1\dots n_p}$  and the dispersal kernel  $p(a,b;.)$ , the proportion of pollen originating from each father  $k$  in the pollen pool of each mother  $j$  originating from all father-trees is given by

$$\pi_{jk} = \frac{F_k p_{jk}}{\sum_{l:father} F_l p_{jl}}, \tag{eqn SM2.2}$$

where  $(x_j, y_j)$  and  $(x_k, y_k)$  are the coordinates of mother  $j$  and father  $k$  and  $p_{jk} = p(a,b; x_j - x_k, y_j - y_k)$ .

Following the classical mating models, we assume that an ovule of a mother  $j$  has the probabilities  $s$ ,  $m$  and  $1-m-s$  to be fecundated by a pollen grain from respectively (i) the same plant (selfing), (ii) outside of the site (immigration) and (iii) a known father inside the site. The likelihood of the set of genotypes  $\mathbf{g} = \{g_o\}_{o \in \text{offspring}}$  of the sampled seeds is thus given by

$$L(\mathbf{g}|\mathbf{F}, \delta, b, s, m) = \prod_{o:offspring} \left[ sT(g_o|g_{j_o}, g_{j_o}) + (1-s-m) \sum_{k:father} \pi_{j_o k} T(g_o|g_{j_o}, g_k) + mT(g_o|g_{j_o}, AF) \right] \tag{eqn SM2.3}$$

where  $\pi_{jk}$  are the pollen pool compositions defined by equation (3.2).  $T(g_o|g_{j_o}, X)$  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  $X$ , where  $X$  corresponds to (i) the genotype of the mother in case of selfing, (ii) the genotype of the considered male ( $g_k$ ) in case of outcrossing with a sampled male  $k$ , or (iii) the allelic frequencies in the pollen pool external to the neighbourhood ( $AF$ ) in case of outcrossing with a non sampled 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).

*MCMC algorithm*

We used the following prior distributions for the parameters:  $PDF(\delta) \propto \delta^{-1}$  for  $\delta$  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  $(0, 0.1)$ ,  $PDF(\alpha_l) \propto \alpha_l^{-1}$ ,  $PDF(\beta_l) \propto \beta_l^{-1}$ ,  $PDF(\varphi_l) \propto \varphi_l^{-1}$  for all  $\alpha$ 's,  $\beta$ 's and  $\varphi$ 's.

We first started by choosing a set of realistic values for all the parameters,  $(\delta_0, b_0, m_0, s_0) = (5000, 1, 0.5, 0.1)$  and all parameters  $(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \varphi_1, \varphi_2, \varphi_3, \varphi_4)_0$  equal to 1. Then, at each iteration  $t$  of the

MCMC we achieved the following steps:

Step 1— For each of the parameters  $\alpha$ ,  $\beta$ , and  $\varphi$  successively, we generated a candidate value by drawing in the proposal distribution (multiplying the present value by a random number drawn in a  $LN(0,0.1)$  distribution). Compute the individual fecundities with the proposed parameters  $\mathbf{F}_t^*$  and with the present parameters  $\mathbf{F}_t$

Using equation (3.3), we computed the ratio

$$r_F = \frac{L(\mathbf{g}|\mathbf{F}_t^*, \delta_t, b_t, m_t, s_t)}{L(\mathbf{g}|\mathbf{F}_t, \delta_t, b_t, m_t, s_t)}$$

With probability  $\min(1, r_F)$ , the candidate value was accepted, otherwise it was rejected .

Step 2— For each parameter  $\delta$ ,  $b$ ,  $m$ ,  $s$  successively, we generated a candidate value by drawing in the proposal distribution (multiplying the present value by a random number drawn in a  $LN(0,0.1)$  distribution)

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

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

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

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

We used  $(\delta_0, b_0, m_0, s_0) = (5000, 1, 0.5, 0.1)$  and all  $\alpha$ 's,  $\beta$ 's and  $\varphi$ 's equal to 1 as starting values for the MCMC. We simulated 110000 iterations and deleted the first 10000. The posterior distribution for each estimated parameters was computed from the values of the 100000 remaining iterations and we provided the mean and median of the posterior distribution. At each iteration  $t$ , the effective density was computed through the ratio

$$\frac{d}{d_{em,t}} = \frac{\text{Var}(F_{k,t})}{(\bar{F}_{k,t})^2} + 1,$$

where  $\text{Var}()$  stands for the empirical variance and  $\bar{F}$  for the mean. We computed the posterior distribution for this ratio as for the other parameters.

### Results.

The MCMC remained within a range of parameters values that provide high values for likelihood: the conditional log-likelihood was mainly between -13960 and -13970 when the maximum attainable value that is -13955 reached by the maximum likelihood approach

(MLA) in Oddou-Muratorio et al. (2005) (Fig. SM1 A). The maximum value reached over the 100000 iterations was -13956.1 and the chain spent 11260 steps over 100000 at values higher than -13960.

For the mean dispersal distance  $\delta$ , we found a 95%-credible set (542.6, 1227.4), with a posterior mean of 788.5, a posterior median of 757.3 (Fig. SM1 B). These values were close to those obtained from the maximum likelihood approach ( $\delta=743$  with confidence interval (540, 1160)). For the shape parameter  $b$ , we found a credible set of (0.26, 0.40) and mean and median were equal to 0.33 (Fig. SM1 C). Maximum likelihood provided  $b = 0.33$  with confidence interval (0.27, 0.40). A good agreement was also obtained for the migration rate (credible set=(0.41, 0.47); mean = median = 0.44; MLA provided 0.44 with CI= (0.41, 0.47)) and for selfing rate (credible set = (0.001, 0.01); mean = median = 0.005; MLA provided 0.005 with CI= (0.0 , 0.1)) (Fig. SM1 D-E).

The Bayesian approach provided values almost equal to that obtained with the maximum likelihood approach for the effects of flowering intensity and neighbourhood density (Table SM1 A and C, Figure SM2 A, C and D). The values obtained for the effect of tree diameter was slightly different in absolute value but provided the same qualitative result (fecundity markedly increasing with diameter). The differences between both approaches observed for all diameter classes and the large values for SE may be due to the low power for estimating the effect for class <10cm, which is used as reference level (fecundity fixed to 1.0).

A.

Year	Flowering intensity	$f(2SE)$
2000	Anecdotal	1.0
	Abundant	2.4 (0.9)
	Massive	4.5 (1.9)
	Unknown	2.1 (1.2)

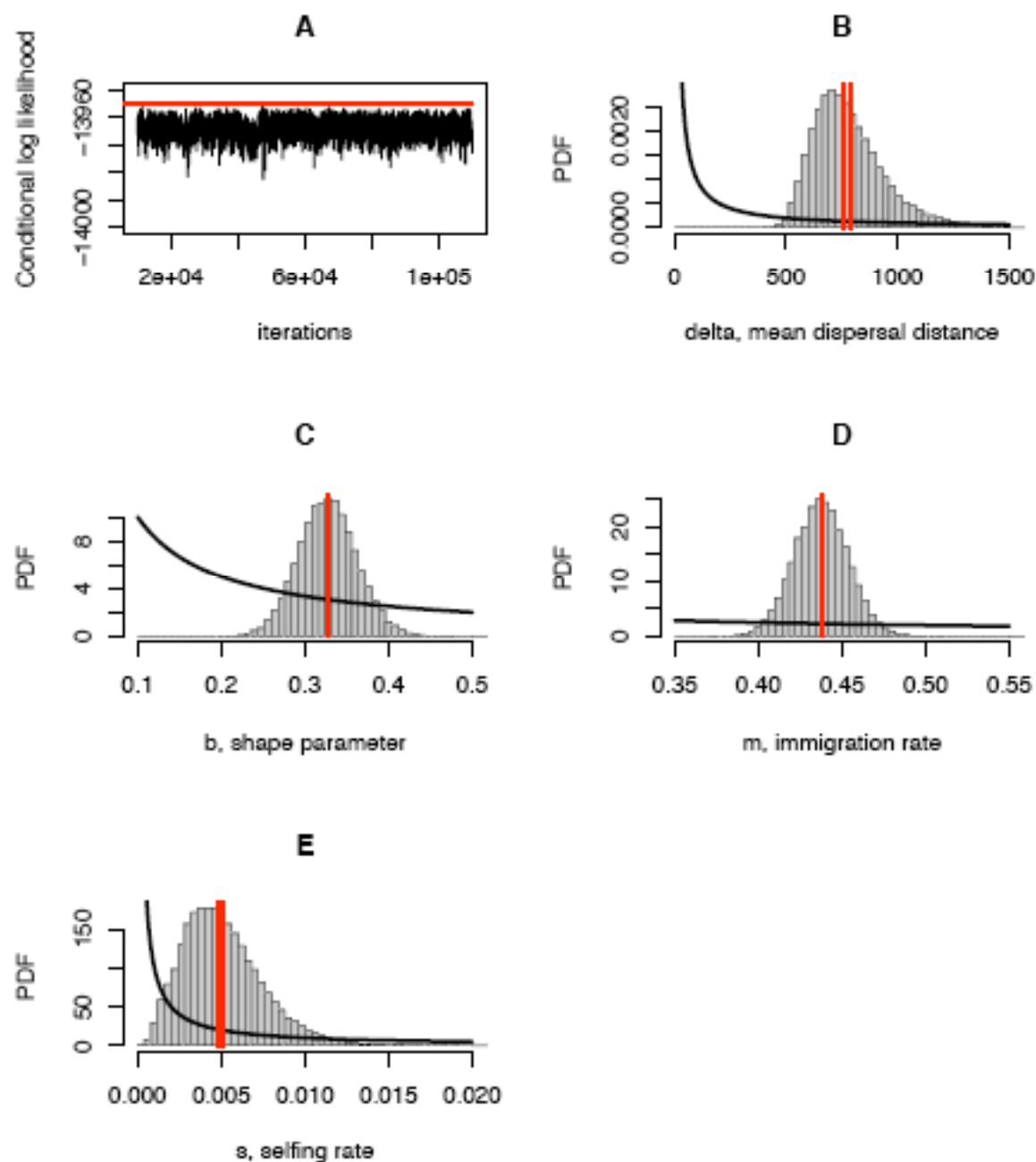
B.

Year	Tree diameter	$f(2SE)$
2000	< 10 cm	1.0
	< 20 cm	2.8 (4.2)
	< 30 cm	7.4 (10.8)
	< 40 cm	11.3 (15.8)
	< 50 cm	14.7 (21.2)
	$\geq$ 50 cm	21.3 (30.6)

C.

Year	Neighbourhood density	$f(2SE)$
2000	Suppressed stand tree	1.0
	Codominant stand tree	1.3 (0.4)
	Dominant stand tree	1.5 (0.5)
	Edge tree	2.5 (1.2)
	Isolated tree	0.9 (0.3)

**Table SM1** Effects of the three covariates on individual male fecundity estimated from the Bayesian approach. For each level of each covariate (A. Flowering intensity; B. Tree diameter; C. Neighbourhood density) we computed the posterior mean from 100000 iterations of the MCMC together with the posterior standard-deviation provided here as 2SE.



**Figure SM1** Estimations of the parameters with the Bayesian approach. (A) Conditional log-likelihood along the 100 000 iterations of the MCMC, together with the log-likelihood previously reached in the fixed effects model (red horizontal line). (B-F) Prior distribution (black line) and posterior distribution (histogram) for the mean dispersal distance  $\delta$ , the shape parameter  $b$ , the immigration rate  $m$  and the selfing 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) red vertical lines.

