Title: Reduced incompatibility in the production of second generation hybrids between two *Magnolia* species revealed by Bayesian gene dispersal modeling[1]

Authors: Ichiro Tamaki[2], Saori Tani[3], Suzuki Setsuko[4], Saneyoshi Ueno[4], Naotaka Wadasaki[3], and Nobuhiro Tomaru [3, 5]

[2] Gifu Academy of Forest Science and Culture, 88 Sodai, Mino, Gifu 501-3714, Japan

[3] Graduate School of Bioagricultural Sciences, Nagoya University, Furo-cho, Chikusa, Nagoya 464-8601, Japan

[4] Department of Forest Molecular Genetics and Biotechnology, Forestry and Forest Products Research Institute, Forest Research and Management Organization, 1 Matsunosato, Tsukuba, Ibaraki 305-8687, Japan

Short title: Reductions in incompatibility in producing hybrids

[1] Manuscript received _____; revision accepted _____.

[5] Author for correspondence (e-mail: tomaru@agr.nagoya-u.ac.jp)

Acknowledgements:

The authors thank the members of the Laboratory of Forest Ecology and Physiology of Nagoya University for assistance in the field, Aichi Kaisho Forest Center for permitting us to conduct the field survey and the anonymous reviewers for their helpful comments on an early version of our manuscript.

Abstract:

PREMISE OF THE STUDY: Hybrid zones are areas where gene flow between related species is currently occurring, so information on the compatibility between related species and their hybrids is essential for predicting the dynamics of such zones generated by introgressive hybridization. In this study, we quantified the compatibility among *Magnolia stellata*, *M. salicifolia* and their hybrids in a hybrid zone using gene dispersal modeling.

METHODS: After determining the genealogical classes of adult trees in the hybrid zone, a paternity analysis of 574 open pollinated seeds from 37 known maternal trees was performed with microsatellite markers. A neighborhood-based Bayesian gene dispersal model developed by us for estimating compatibility was then applied to the paternity data.

KEY RESULTS: When *M. stellata* or *M. salicifolia* were mothers, interspecific mating to produce F_1 hybrids yielded significant incompatibility but backcrossing with F_1 hybrids did not. Furthermore, when F_1 hybrids became mothers, no significant incompatibility resulted from backcrossing to parental species or intra-F_1 mating to produce F_2 hybrids. The estimated proportion of F_1 hybrids in the outcrossed seeds (1.7%) in the hybrid zone was much lower than that in the adult trees (14.0%).

CONCLUSIONS: While it is difficult to obtain F_1 hybrids, their low incompatibility makes it easy to produce advanced generation hybrids, once they have been successfully obtained. Although the production of F_1 seeds is rare, heterosis and/or weak selection pressure in an empty niche between the parental species' niches may have contributed to the increased proportion of adult F_1 hybrids in the hybrid zone.

KEY WORDS: Bayesian clustering; interspecific hybrids; interspecific mating; introgression; Magnoliaceae; neighborhood model; NewHybrids; pollen dispersal; simple sequence repeat (SSR); parentage analysis

Recent population genetic studies have reported that many plant species exhibit signatures of gene flow from their relatives after divergence to the present (Drummond and Hamilton, 2007; Scascitelli et al., 2010; Wachowiak et al., 2011; Tamaki and Okada, 2014; Wang et al., 2014; Filatov et al., 2016; Zhou et al., 2017). Although gene flow from distant lineages reduces the fitness of individuals in recipient species through outbreeding depression (Keller et al., 2000; Muhlfeld et al., 2009), it can introduce adaptive genes that increase the fitness of individuals in the species, i.e. adaptive introgression (Whitney et al., 2006; Norris et al., 2015; Whitney et al., 2015). Therefore, the formation of current species may have been partly driven by their hybridization history. Hybrid zones are areas where gene flow between related species is currently occurring, so information on the frequency and direction of interspecific gene flow in such zones could be useful in predicting the evolution of those species.

Interspecific gene flow in hybrid zones is generally somewhat more restricted than intraspecific gene dispersal. In general, the formation of a first generation hybrid (F_1) is very difficult, but once such hybrids have been formed, introgressive hybridization can advance, with the F_1 hybrids serving as a breakthrough point (Olrik and Kjaer, 2007; Abraham et al., 2011; Gailing and Curtu, 2014). It is therefore important to quantify the strength of barriers towards interspecific mating when attempting to predict the zone's dynamics. There are two types of barriers, pre- and post-mating barriers. For plant species, the former can be inferred by investigating the overlap of flowering phenology between species and the characteristics of their pollen vectors such as the commonality of pollinators and the wind direction. Although the latter can be inferred directly from artificial hand pollination experiments (Steinhoff, 1993), it is sometimes difficult to conduct such experiments in the field and to obtain adequate numbers of replicates, especially for larger trees. The magnitude of pre- and

post-mating barriers can be estimated indirectly by performing a paternity analysis of open pollinated seeds that were collected from the known maternal trees in the zone (Curtu et al., 2009; Abraham et al., 2011; Lepais and Gerber, 2011). Moreover, by incorporating interspecific compatibility parameters into a gene dispersal model, these effects can be quantified as parameter values. However, studies using such approaches are very scarce (Chybicki and Burczyk, 2013; Bialozyt et al., in press).

Natural hybridization studies using genetic markers are becoming increasingly common, but most of them are limited to major hybridization taxa such as *Quercus* (Lepais and Gerber, 2011; Gailing and Curtu, 2014; Tamaki and Okada, 2014), *Populus* (Lexer et al., 2005; Macaya-Sanz et al., 2011; Lindtke et al., 2014), *Picea* (Sun et al., 2014; Aizawa et al., 2016; Tsuda et al., 2016) and *Rhododendron* (Zha et al., 2010; Marczewski et al., 2015; Tamaki et al., 2017). Although *Magnolia* is one of the most famous cultivated tree genera and many cultivars are made by interspecific artificial crossings (Callaway, 1994), studies on its natural hybridization are very scarce (Muranishi et al., 2011, 2013). *Magnolia stellata* (Sieb. et Zucc.) Maxim. and *M. salicifolia* (Sieb. et Zucc.) Maxim., our study species, are trees belonging to the Magnoliaceae family that are known to form hybrid zones in areas where both species grow sympatrically or parapatrically (Muranishi et al., 2011, 2013). Our previous studies on adult trees of these species using nuclear and chloroplast microsatellite markers revealed pronounced differences between them in terms of the directions of hybridization and the subsequent introgression (Muranishi et al., 2011, 2013), and interspecific hand pollination experiments showed that fruit sets and seed sets were significantly lower for interspecific crosses where *M. stellata* is the maternal tree than for those in which *M. salicifolia* is the maternal tree (Tani et al., 2014). Although we have some understandings of the compatibility

between these two species, little is known about their compatibility with their hybrids, or that between their hybrids. To address this deficiency, in this work we determined the genealogical classes of adult trees in a hybrid zone between *M. stellata* and *M. salicifolia* (two parental species, F_1 and F_2 hybrids, and two backcross hybrids), identified the paternity of open pollinated seeds collected from known maternal trees, and quantified the compatibility between the genealogical classes using a gene dispersal model we developed.

MATERIALS AND METHODS

Study species—*Magnolia stellata* (also known as star magnolia) and *M. salicifolia* are tree species in the Magnoliaceae that are endemic to Japan. *M. stellata* is narrowly distributed in the Tokai region in central Honshu Island, while *M. salicifolia* is broadly distributed in Honshu, Shikoku and Kyushu Islands (see Fig. 1 in Muranishi *et al*., 2013). Both species have hermaphrodite flowers. The flowering seasons for both species and their hybrids run from late March to early April, and overlap extensively (Muranishi et al., 2011). They have common pollinators, notably small beetles, thrips and flies, and sometimes bumblebees and other bees (Yasukawa et al., 1992; Setsuko et al., 2012).

Study site and sampling—The study site is a warm temperate forest located in Aichi Prefecture in central Honshu Island, Japan (35°11′25″N, 137°06′55″E) that was also used as the study site in our previous investigation into these species (Muranishi et al., 2013). In this site, *M. stellata* and *M. salicifolia* grow at the bottom and on the middle slopes of the valley,

121 respectively, while their hybrids grow at intermediate locations (Muranishi et al., 2013). We
122 conducted a re-census for all individuals taller than 1.3 m in the study site in 2012 and found
123 new individuals that were recruited after the census conducted in 2009 for the previous study
124 (Muranishi et al., 2013) or were simply missed in the first census. Individuals taller than 1.3
125 m were considered to be adult trees with the potential for flowering. We resampled leaves for
126 DNA extraction from all 307 adult trees including the individuals used in the previous study
127 and recorded their DBH values and locations (Fig. 1). Sampled leaves were stored at –30°C
128 until DNA extraction. On one day in April 2013, we checked all adult trees to determine
129 whether they had flowered. If the tree had not flowered, we searched for flower buds and
130 assumed that trees with flower buds would flower in the season. In August 2013, a total of
131 574 seeds were sampled from 37 maternal trees.
132
133 ***DNA extraction and genotyping*—**Genomic DNA was extracted from leaves of adult trees or
134 embryos and endosperms of seeds by the hexadecyltrimethylammonium bromide method
135 (Murray and Thompson, 1980). For adult trees, 28 nuclear microsatellite loci [stm0002,
136 stm0114, stm0163, stm0184, stm0200, stm0214, stm0223, stm0225, stm0246, stm0251,
137 stm0353, stm0383, stm0415, stm0423 and stm0448 (Setsuko et al., 2005); M6D8 (Isagi et al.,
138 1999); and mag00402, mag01823, mag04151, mag04167, mag04769, mag05338, mag05534,
139 mag06266, mag08314, mag08400, mag10551 and mag14347 (Appendix S1; see the
140 Supplementary Data with this article), which were newly developed by Ueno *et al.*
141 (unpublished)] and 3 chloroplast microsatellite loci [*trn*L intron, *trn*H–*trn*K and *trn*G intron;
142 Ueno et al. (2005)] were amplified using the Multiplex PCR Kit (Qiagen, Venlo, Netherlands)
143 with the Gene Amp PCR System 9700 (Applied Biosystems, Waltham, Massachusetts, USA).

For seeds, PCR amplification was performed with only 16 of the 28 nuclear microsatellite loci (stm0002, stm0114, stm0163, stm0184, stm0200, stm0214, stm0223, stm0225, stm0246, stm0251, stm0353, stm0383, stm0415, stm0423, stm0448 and M6D8). PCR products were electrophoresed with a 3130-Avant Genetic Analyzer (Applied Biosystems). Genotypes were determined by GeneMapper version 4.0 (Applied Biosystems). Electrophoretic profiles of alleles at the locus stm0114 were not clear in some samples, so this locus was excluded from subsequent analyses. In a preliminary analysis, we analyzed genotype data at 27 loci for 307 adult trees using InStruct (Gao et al., 2007) and extracted 55 and 186 putative purebred trees that were assigned to genetic clusters of *M. stellata* and *M. salicifolia*, respectively, with probabilities above 0.95. Using these putative purebred trees, null allele frequencies at 27 loci in each species were estimated with INEst version 1.1 under the individual inbreeding model (Chybicki and Burczyk, 2009). Because the null allele frequencies at six loci in *M. stellata* (stm0223, stm0246, stm0423, mag04151, mag04769 and mag06266) and four loci in *M. salicifolia* (mag00402, mag04151, mag06266 and mag08400) were above 0.05, the eight corresponding unique loci (stm0223, stm0246, stm0423, mag00402, mag04151, mag04769, mag06266 and mag08400) were excluded from all subsequent analyses. These left 19 and 12 loci that were used for determining the genealogical classes of adult trees and the paternity of seeds, respectively (Table 1).

Identification of genealogical classes for adult trees—For each adult tree, the posterior probabilities of belonging to each genealogical classes were estimated using the developmental version 2.0 of NewHybrids (Anderson and Thompson, 2002). We conducted 10 independent runs with a burn-in period of 20 000 steps followed by 40 000 MCMC steps.

The threshold posterior probability was determined by the method of Muranishi et al. (2013). If the posterior probability for a given adult tree belonging to a particular genealogical class was above the threshold value, that tree was identified as a member of that class. The six genealogical classes of adult trees were: (1) MST (*M. stellata*), (2) MSA (*M. salicifolia*), (3) F_1 hybrids, (4) F_2 hybrids, (5) BxMST (backcrosses to *M. stellata*) and (6) BxMSA (backcrosses to *M. salicifolia*). There was also a seventh class (Unknown) for trees that could not be assigned to any other class, including unknown hybrids.

***Paternity assignment and estimation of selfing rates*—**To identify the pollen parents of seeds collected from the known mothers, a likelihood-based paternity analysis was conducted using CERVUS version 3.0 (Kalinowski et al., 2007). We identified 227 flowering trees among the 307 adult trees. However, because our investigation of flowering was not extensive, it is possible that we may have overlooked some flowering trees with flower buds. We therefore performed a preliminary analysis including all 307 adult trees as candidate fathers. This analysis identified 11 trees that were not apparently flowering as pollen parents at the 95% confidence level. These 11 trees were therefore added to the list of 227 trees for which flowering was observed, giving a total of 238 candidate father trees (Table 2). A threshold delta value for the difference in the LOD scores of the most likely father and the second most likely father was determined by performing 100 000 simulations with the following conditions. The number of candidate fathers was set to 350. The proportion of candidate fathers sampled was set to 80%. The proportion of typed loci was set to 98.84%, which was estimated by the observed data. The proportion of mistyped loci was set to 5%. Selfing was allowed at a rate of 20%. Using the threshold delta value from these simulations, the most

likely single father for each seed was determined at the 95% confidence level. To compare the mating systems of hybrids to those of purebreds, the selfing rate for each genealogical class and its variance components among genealogical classes and among maternal trees within genealogical class were estimated using the method of Tamaki et al. (2009).

Indirect quantification of inter-genealogical class compatibility using a gene dispersal model—To quantify compatibility between genealogical classes, we used a newly developed neighborhood model based approach. This analysis only considered outcrossed seeds whose father was identified. Selfed seeds and seeds whose father was not identified (in which case we assumed that the father was an individual outside the study site) were not used. Therefore, four maternal trees from which no outcrossed seeds were collected were removed as mothers and were included only as candidate fathers in this analysis. Because the sample sizes for the non-F_1 hybrids were small, we merged them into a single class, giving four genealogical classes in total: (1) MST, (2) MSA, (3) F_1 and (4) OTH (the other hybrids).

An exponential power function was used to model the pollen dispersal kernel (Clark, 1998; Austerlitz et al., 2004).

$$p = \frac{b_{[GC]}}{2\pi a^2 \Gamma(2/b_{[GC]})} \exp(-[D/a]^{b_{[GC]}}),$$

where Γ is a gamma function. The pollen density (p) at a given distance (D) from a pollen source is determined by the scale and shape parameters (a and b, respectively). The parameter b determines the shape of the kernel; $b < 1$, $b = 1$ and $b > 1$ correspond to fat-tailed, exponential and thin-tailed distributions, respectively. The parameter b was assumed to be different among four genealogical classes (GC). Using this dispersal kernel, the probability of the male reproductive success (q_{ij}) of the j-th father with the i-th mother can be expressed as

follows:

$$q_{ij} = \frac{c_{[\mathrm{GCM,\ GCF}]}\, f_j \exp(-[\mathrm{D}_{ij}/a]^{b_{[\mathrm{GC}]}})}{\sum_{k=1(k\neq i)}^{\mathrm{N}} c_{[\mathrm{GCM,\ GCF}]}\, f_k \exp(-[\mathrm{D}_{ik}/a]^{b_{[\mathrm{GC}]}})}$$

$$f_j = \mathrm{DBH}_j{}^{d}\, f\_error_j$$

Here, $c_{[\mathrm{GCM,\ GCF}]}$ is the relative compatibility between genealogical classes of mothers and fathers (GCM and GCF, respectively), and takes values greater than zero. The value of $c_{[\mathrm{MST,\ MST}]}$ was fixed at unity, and the other combinations of $c_{[\mathrm{GCM,\ GCF}]}$ were defined in relation to it. If the value of $c_{[\mathrm{GCM,\ GCF}]}$ for a given pairing is smaller (larger) than one, that pairing underperforms (overperforms) relative to intraspecific mating in *M. stellata*. The fecundity of the *j*-th father (f_j) is assumed to be determined by its diameter at breast height (DBH_j) and a random effect term (f_error_j) estimated for each father. The log value of f_error_j is assumed to be normally distributed with a mean of zero and a standard deviation of σ.

$$\log (f\_error_j) \sim \mathrm{Normal}\ (0, \sigma)$$

In the above paternity analysis, because we observed the number of outcrossed seeds of the *i*-th mother sired by the *j*-th father ($\mathrm{N\_seed}_{ij}$), the observed data can be assumed to follow the multinomial distribution with a probability of q_{ij}.

$$\mathrm{N\_seed}_{ij} \sim \mathrm{Multinomial}\ (q_{ij}, \textstyle\sum_{k=1} \mathrm{N\_seed}_{ik})$$

However, in insect-pollinated plants, dispersal of pollen grains may be correlated due to the behavior of pollinators. An insect carrying pollen grains from a source plant can pollinate multiple stigmas in single or multiple flowers of a recipient plant. This process can violate the assumption of independence between pollen gametes in the multinomial distribution with a probability of q_{ij} for the *i*-th mother. Thus, the multinomial distribution may be a crude approximation of the reality in insect-pollinated plants like the study species. This statistical model was implemented under the Bayesian scheme with the following sufficiently wide

uniform prior distributions.

$a \sim$ Uniform $(0, 10^4)$

$b_{[GC]} \sim$ Uniform $(0, 10)$

$c_{[GCM, GCF]} \sim$ Uniform $(0, 5)$

$d \sim$ Uniform $(-10, 10)$

$\sigma \sim$ Uniform $(0, 10)$

The R package rstan (version 2.15.1) was used to estimate the posterior distributions of parameters (Stan Development Team, 2016). Four independent chains consisting of 1 000 burn-in steps and then 3 000 MCMC steps with different initial values were run. The thinning period was set to 10, and 300 posteriors for each chain were obtained. Convergence was assumed if the R-hat statistic was < 1.01 and was visually confirmed by inspecting trace plots generated using R package ggmcmc (version 1.1) (Fernandez-i-Marin, 2016). Using a total of 1 200 posteriors, the posterior mode and 95% highest posterior density (HPD) were estimated with the R package coda (version 0.19.1) (Plummer et al., 2006). The significance of the differences between the values of $b_{[GC]}$ or $c_{[GCM, GCF]}$ was assessed by examining whether the 95% HPD of the differences between their posteriors contained zero.

RESULTS

The average numbers of alleles over the 19 loci in *M. stellata* and *M. salicifolia* were 8.16 and 9.84, respectively (Table 1). The average expected heterozygosity for *M. stellata* and *M. salicifolia* was 0.762 and 0.740, respectively.

The threshold value of the posterior probability for determining the genealogical class of adult trees was 0.59. There were 54 *M. stellata* (17.6%), 187 *M. salicifolia* (60.9%) and 66 hybrid (21.5%) trees at the study site (Table 2 and Fig. 2). The dominant hybrid type was F_1, which accounted for 65.1% of all hybrids. Five chloroplast haplotypes were detected among all adult trees (Table 3). Haplotypes were clearly separated between *M. stellata* (A, B and C) and *M. salicifolia* (D and E). Most of the hybrids had the haplotypes detected in *M. salicifolia*; only two individuals, both of which belonged to the *M. stellata* backcross genealogical class, had the haplotype detected in *M. stellata*.

The exclusion probability for the second parent calculated using 238 candidate fathers was 0.9999989. We could determine fathers for 509 of the 574 seeds that were analyzed (88.7%) with 95% confidence (Table 4 and Fig. 1). Eighty-five selfed seeds were detected and the selfing rates in genealogical classes estimated by the hierarchical Bayesian model were 1.7–6.3% (Table 5). The selfing rates were not significantly different between genealogical classes. The variance component among genealogical classes was lower than that among maternal trees within genealogical classes, but the difference was not significant with a posterior probability of 0.895. Among the 424 outcrossed seeds whose fathers were identified, the numbers of hybrid seeds produced by inter-purebred mating were 1 (0.5%) and 3 (2.7%) in the cases that *M. stellata* and *M. salicifolia* were mothers, respectively. On the other hand, the numbers of hybrid seeds produced by mating between purebreds and hybrids, i.e. backcrossing, were 20 (9.9%) and 27 (24.5%) in the cases that *M. stellata* and *M. salicifolia* were mothers, respectively, and were much larger than the numbers of hybrid seeds produced by inter-purebred mating. When F_1 hybrids were mothers, the numbers of outcrossed seeds sired by *M. stellata* and *M. salicifolia* fathers were 22 (23.7%) and 38 (40.9%), respectively,

and were almost the same as those resulting from intra-F_1 hybrid mating (29 seeds; 31.2%).

To estimate the actual proportions of seeds belonging to each genealogical class in all outcrossed seeds produced in the study site, we must assume that per-tree seed production does not differ among adult trees or genealogical classes, and that mating patterns are not different among adult trees within each genealogical class. Under these assumptions, the actual proportion of outcrossed *M. stellata* seeds was roughly estimated by multiplying the proportion of *M. stellata* adult trees in the study site (54 trees / 307 trees = 0.176) and the observed proportion of outcrossed *M. stellata* seeds produced from *M. stellata* mothers (182 seeds / 203 seeds = 0.897). This yields a value of 0.158. Similarly, the proportions of seeds for *M. salicifolia*, F_1 hybrids, and other hybrids were estimated to be: 0.609 × 0.727 = 0.443, 0.176 × 0.005 (for *M. stellata* mothers) + 0.609 × 0.027 (for *M. salicifolia* mothers) = 0.017, and 1 – 0.158 – 0.443 – 0.017 = 0.382, respectively.

The posterior mode (95% HPD) for the scale parameter (a) in the exponential power function was 43.3 (32.7–54.3) (Appendix S2). The posterior modes of shape parameters (b) ranged from 0.347–0.798 (Fig. 3 and Appendix S2). The values of $b_{[MSA]}$ and $b_{[OTH]}$ were significantly lower than 1.0. In the case that *M. stellata* became mothers, the value of the relative compatibility for interspecific mating ($c_{[MST, MSA]}$) was significantly lower than that for intraspecific mating (Fig. 4 and Appendix S2). However, the value for backcrossing ($c_{[MST, F1]}$) was not significantly different from that for intraspecific mating. In the case that *M. salicifolia* became mothers, the value for interspecific mating ($c_{[MSA, MST]}$) was significantly lower than that for intraspecific mating. However, the value for backcrossing ($c_{[MSA, F1]}$) was not significantly different from that for intraspecific mating. In the case that F_1 hybrids became mothers, all three values ($c_{[F1, MST]}$, $c_{[F1, MSA]}$ and $c_{[F1, OTH]}$) were not significantly

different from that for intra-F_1 mating ($c_{[F1, F1]}$). In the case of intra-genealogical class mating, the values for *M. stellata*, *M. salicifolia* and F_1 hybrids ($c_{[MST, MST]}$, $c_{[MSA, MSA]}$ and $c_{[F1, F1]}$, respectively) were not significantly different from each other. In the case of interspecific mating producing F_1 hybrids, although the value for the mating with *M. stellata* as the mother ($c_{[MST, MSA]}$) was lower than that for *M. salicifolia* as the mother ($c_{[MSA, MST]}$), the difference was not significant with a posterior probability of 0.938. In the case of backcrossing to F_1 hybrids, the value for mating with *M. stellata* as the mother ($c_{[MST, F1]}$) was significantly lower than that for the reverse mating ($c_{[F1, MST]}$), but that for mating with *M. salicifolia* as the mother ($c_{[MSA, F1]}$) was not significantly different from that for its reverse ($c_{[F1, MSA]}$). The posterior mode (95% HPD) of the parameter *d* was 0.680 (0.496–0.799), and was significantly greater than zero (Appendix S2). The posterior mode (95% HPD) of the standard deviation for random effects (σ) was 0.564 (0.460–0.719). The posterior distributions of the random effect term (*f_error*) on the log scale were overlapped around zero, with no candidate fathers that showed extreme deviations from zero (Appendix S3).

DISCUSSION

***Identification of genealogical classes for adult trees*—**By adding newly found individuals, increasing the number of markers and removing markers rich in null alleles, we identified a slightly different set of hybrids to that reported previously (Muranishi et al., 2013). Specifically, in this work we detected four hybrids produced by backcrossing to *M. stellata* that were not detected before. In the previous study, we concluded that introgression to *M.*

328 *stellata* did not occur because there were no individuals assigned to backcrosses to *M. stellata*.
329 However, this conclusion may have been inaccurate. The discovery of these backcrosses may
330 be partly due to the improvements listed above. Half of the chloroplast DNA haplotypes in the
331 backcrosses to *M. stellata* were identical to those detected in *M. stellata*. This suggests that
332 backcrossing can occur when both *M. stellata* and hybrids are mothers. This conclusion was
333 strengthened by the results of the seed paternity analysis because there were many seeds
334 produced by backcrossing to *M. stellata* in both directions.
335 The number of *M. stellata* individuals was quite similar to that reported previously
336 (Muranishi et al., 2013), but the numbers of *M. salicifolia* individuals and F_1 hybrids
337 increased while those of F_2 hybrids and backcrosses to *M. salicifolia* decreased. These
338 changes were attributed to a combination of newly discovered individuals and the use of a
339 larger number of loci after detailed checks for null alleles. The reduction in the number of
340 backcrosses to *M. salicifolia* was probably mainly due to the latter factor: the previous study
341 used loci rich in null alleles for the *M. stellata* population (stm0223, stm0246 and stm0423),
342 and individuals carrying null alleles inherited from *M. stellata* would have exhibited an
343 apparent excess of alleles inherited from *M. salicifolia*. Any genealogical class identification
344 based in part on such loci would inevitably be biased in favor of the backcross to *M.*
345 *salicifolia*. Therefore, our previous study may have overestimated the proportion of such
346 backcrosses. In this work, such individuals may have been classified as F_1 hybrids, explaining
347 the increase in this genealogical class.
348
349 ***Paternity*—**Because we could determine pollen parents for 88.7% of the analyzed seeds with
350 a very high exclusion probability for the second parent, it is assumed that the pollen parents

for the remaining 11.3% of seeds were trees outside the study site (for seeds from *M. stellata*, *M. salicifolia* and F_1 hybrid mothers, 8.0, 10.4 and 15.9%, respectively). Our study site (ca. 4 ha) was only a small fraction of a much larger continuous forest (> 3 000 ha), and is surrounded by several local populations of both parental species (Japan Association for Shidekobushi Conservation, 1996). Setsuko et al. (2013) investigated pollen flows among local populations in *M. stellata* over an area of ca. 100 ha, where *M. salicifolia* was not present, and reported that the average pollen immigration to local populations was 6.1%. The reported value is similar to that estimated for *M. stellata* in our study (8.0%). The selfing rates in each genealogical class were low, and did not differ significantly between genealogical classes. This indicates that hybrids have similar mating systems to their parental species.

Incompatibility between genealogical classes—The posterior modes of the parameter *b* for *M. salicifolia* and F_1 hybrids in the exponential power model were significantly lower than that for *M. stellata* (Fig. 3 and Appendix S2). Moreover, the value for *M. salicifolia* was significantly lower than 1.0, indicating that fat-tailed pollen dispersal occurs (Austerlitz et al., 2004), but the values for *M. stellata* and F_1 hybrids were not. However, the parameter *b* for *M. stellata* estimated in our study site was higher than that observed in an *M. stellata* metapopulation consisting of several local populations (where *M. salicifolia* was not present), for which the posterior median (95% credible interval) was 0.206 (0.182–0.257) (Setsuko et al., 2013). The difference in the dispersal kernel between *M. stellata* and *M. salicifolia* may be due to the difference in distribution patterns of individuals between the two species and inter-genealogical class mating. *M. stellata* individuals are aggregately distributed at the bottom of a valley, whereas *M. salicifolia* individuals are widely distributed along the slopes

on both sides of valleys. In general, density and spatial distribution of individuals condition effective pollen dispersal (Hardy, 2009). Therefore, it may be suggested that long-distance pollen dispersal is hard to occur for *M. stellata* compared to that for *M. salicifolia* particularly in hybrid zones because *M. stellata* individuals may be hybridized with their surrounding individuals of *M. salicifolia* and hybrids.

The parameter $c_{[GCM, GCF]}$ represents the relative compatibility for mating between maternal and paternal genealogical classes relative to the compatibility for intraspecific mating in *M. stellata* ($c_{[MST, MST]} = 1$). The relative compatibilities for intra-genealogical class mating ($c_{[MST, MST]}$, $c_{[MSA, MSA]}$ and $c_{[F1, F1]}$) were not significantly different. However, the relative compatibility for mating in non-F_1 hybrids ($c_{[OTH, OTH]}$) was significantly lower than 1. Because the other hybrids (OTH) group in this analysis consisted of only three mothers and 18 outcrossed seeds, the small sample size means there is considerable uncertainty associated with this estimated value, and so we do not discuss the incompatibility for this class further.

In the case of F_1 formation, the relative compatibility for interspecific mating was significantly lower than that for intraspecific mating, regardless of whether *M. stellata* or *M. salicifolia* was the mother. Therefore, there must be significant barriers that prevent mating between the two species. Although the relative compatibility for interspecific mating with *M. stellata* as the mother was lower than that for *M. salicifolia* as the mother, the difference was not significant. Tani et al. (2014) conducted reciprocal hand pollination experiments between *M. stellata* and *M. salicifolia*, and reported significant reductions in fruit and seed set rates for interspecific crossing relative to intraspecific crossing when *M. stellata* was the mother, but not when *M. salicifolia* was the mother. These differences in interspecific incompatibility may reflect post-mating barriers between *M. stellata* and *M. salicifolia*. However, our indirect

incompatibility estimates include the effects of both post- and pre-mating barriers. Although the flowering phenologies of *M. stellata*, *M. salicifolia* and their hybrids are well overlapped (Muranishi et al., 2011) and they have some common pollinators (Yasukawa et al., 1992; Setsuko et al., 2012), we cannot exclude the possibility of significant pre-mating barriers to F_1 formation.

For matings with F_1 hybrids, in both cases where *M. stellata* or *M. salicifolia* were mothers, the relative compatibility for backcrossing with F_1 hybrids was not significantly different from that for intraspecific mating. This indicates that there are no barriers in backcrossing with F_1 hybrids when *M. stellata* or *M. salicifolia* were mother. In the case that F_1 hybrids were mothers, the relative compatibilities for backcrossing to *M. stellata* and *M. salicifolia* and for F_2 formation were not significantly different from each other. This result also indicates that there were no barriers in backcrossing to *M. stellata* and *M. salicifolia* or in F_2 formation when the F_1 hybrids were mothers. However, the different tendency was observed when F_1 hybrids were fathers, i.e., the relative compatibilities for backcrossing to *M. stellata* and *M. salicifolia* were significantly different ($c_{[MST, F1]} < c_{[MSA, F1]}$). Moreover, the relative compatibility for backcrossing to *M. stellata* as a mother was significantly lower than that for backcrossing to *M. stellata* as a father, suggesting that seeds resulting from backcrossing to *M. stellata* are more likely to be produced with the combination of F_1 mothers and *M. stellata* fathers.

Overall, these findings indicate that although the formation of F_1 hybrids is difficult, especially with *M. stellata* as the maternal tree (Tani et al., 2014), advanced generation hybrids are relatively easily obtained once F_1 hybrids have been produced. The similar tendency has been observed in hybrid zones of oak species (Olrik and Kjaer, 2007; Abraham

et al., 2011; Lepais and Gerber, 2011; Gailing and Curtu, 2014).

The coefficient of DBH, *d*, was significantly higher than zero, indicating that DBH was positively related with siring success. It was previously found that *M. stellata* trees with larger DBH values have higher male reproductive success (Setsuko and Tomaru, 2011), and our results support this conclusion. No candidate fathers showed significant deviation from zero for the random effect term on the log scale. We also tested other models without the random effect term and consequently could not obtain good estimates of the parameters *a* and *b* (data not shown), indicating that the random effect term worked well to simultaneously estimate dispersal kernels, relative compatibilities and the effect of DBH. However, it may be noted that we could not specify the differences in individual fecundity between candidate fathers, because the number of seeds used in the modeling (424 seeds) was relatively small considering that of candidate fathers (238 individuals). To confirm the pattern detected in this study, more numbers of sites (hybrid zones) and seeds per site will be needed. Because our model cannot distinguish the effects of pre- and post-mating barriers, if one wants to know the strength of each barrier, one should also conduct hand pollination experiments. In our case, although intraspecific and interspecific hand pollination experiments have been already conducted (Tani et al., 2014), inter-hybrid ones have not been yet. To understand the detail of inter-hybrid matings, further studies with hand pollination experiments between hybrids will be needed.

Comparison of the composition of genealogical classes between adult trees and seeds—Because most selfed seeds are expected to die due to inbreeding depression as indicated by the very low values of the inbreeding coefficient in the adult trees (Table 1), the

composition of genealogical classes in the produced outcrossed seeds could help us to understand the dynamics in this hybrid zone. The estimated representation of genealogical classes among the outcrossed seeds was *M. stellata*, 15.8%; *M. salicifolia*, 44.3%; F_1 hybrids, 1.7%; and other hybrids, 38.2%. This distribution was significantly different from that for adult trees (*M. stellata*, 17.6%; *M. salicifolia*, 60.9%; F_1 hybrids, 14.0%; other hybrids, 7.5%; $P < 0.001$, the chi-square test), especially for F_1 hybrids and other hybrids. The inconsistency for F_1 hybrids may be explained by heterosis (hybrid vigor): fitness may be higher in F_1 hybrids than in their parental species, explaining the high proportion of adult F_1 hybrids in the hybrid zone. Another possible explanation is different strengths of natural selection in different locations on each genealogical class during germination or later life stages prior to adulthood (Milne et al., 2003; Fritz et al., 2006). Muranishi et al. (2013) reported that F_1 and F_2 hybrids tend to grow in locations intermediate between those preferred by *M. stellata* and *M. salicifolia*. *M. stellata* prefers wet environments at the bottom of a valley, whereas *M. salicifolia* prefers dry environments on the middle or upper slopes of valleys. Intermediate locations on the lower slopes may thus constitute an empty niche where competition between F_1 hybrids and parental species would be relatively weak. Although the production of F_1 seeds is rare, weakened selection pressure in the empty niche may have contributed to the high proportion of adult F_1 hybrids in the hybrid zone. The inconsistency for the other hybrids including F_2 hybrids and backcrosses in the second generation may be simply because there has not been sufficient time for development of many F_2 hybrids and backcrosses since the beginning of the hybridization process. Another possible explanation is hybrid breakdown (Breeuwer and Werren, 1995; Milne et al., 2003): hybrids in the second and more advanced generations may exhibit reduced fitness, explaining their comparatively low abundance.

Further studies on the dynamics of individuals including seedlings and saplings will be needed to test the above hypothesis.

ACKNOWLEDGEMENTS

The authors thank the members of the Laboratory of Forest Ecology and Physiology of Nagoya University for assistance in the field, Aichi Kaisho Forest Center for permitting us to conduct the field survey and the anonymous reviewers for their helpful comments on an early version of our manuscript.

LITERATURE CITED

Abraham, S. T., D. N. Zaya, W. D. Koenig, and M. V. Ashley. 2011. Interspecific and intraspecific pollination patterns of valley oak, *Quercus lobata*, in a mixed stand in coastal central California. *International Journal of Plant Science* 172: 691–699.

Aizawa, M., H. Yoshimaru, H. Ogawa, S. Goto, and M. Kaji. 2016. Natural hybridization of Yezo and Sakhalin spruce in central Hokkaido, revealed by DNA markers with contrasting modes of inheritance. *Plant Species Biology* 31: 188–195.

Anderson, E. C., and E. A. Thompson. 2002. A model-based for identifying species hybrids using multilocus genetic data. *Genetics* 160: 1217–1229.

Austerlitz, F., C. W. Dick, C. Dutech, E. K. Klein, S. Oddou-Muratorio, P. E. Smouse, and V. Sork. 2004. Using genetic markers to estimate the pollen dispersal curve. *Molecular Ecology* 13: 937–954.

Bialozyt, R., M. Niggemann, and B. Ziegenhagen. in press. Quantification of the zygotic barrier between interbreeding taxa using gene flow data. *Oikos*.

Breeuwer, J. a. J., and J. H. Werren. 1995. Hybrid breakdown between two halodiploid

494 species: the role of nuclear and cytoplasmic genes. *Evolution* 49: 705–717.
495 Callaway, D. J. 1994. The world of Magnolias. Timber Press, Oregon, USA.
496 Chybicki, I. J., and J. Burczyk. 2009. Simultaneous estimation of null alleles and inbreeding
497 coefficients. *Journal of Heredity* 100: 106–113.
498 Chybicki, I. J., and J. Burczyk. 2013. Seeing the forest through the trees: comprehensive
499 inference on individual mating patterns in a mixed stand of *Quercus robur* and *Q.*
500 *petraea*. *Annals of Botany* 112: 561–574.
501 Clark, J. S. 1998. Why trees migrate so fast: confronting theory with dispersal biology and the
502 paleorecord. *The American Naturalist* 152: 204–224.
503 Curtu, A. L., O. Gailing, and R. Finkeldey. 2009. Patterns of contemporary hybridization
504 inferred from paternity anslysis in a four-oak-species forest. *BMC Evolutionary*
505 *Biology* 9: 284.
506 Drummond, C. S., and M. B. Hamilton. 2007. Hierarchical components of genetic variation at
507 a species boundary: population structure in two sympatric varieties of *Lupinus*
508 *microcarpus* (Leguminosae). *Molecular Ecology* 16: 753–769.
509 Fernandez-I-Marin, X. 2016. ggmcmc: analysis of MCMC samples and Bayesian inference.
510 *Journal of Statistical Software* 70: 1–20.
511 Filatov, D. A., O. G. Osborne, and A. S. T. Papadopulos. 2016. Demographic history of
512 speciation in a *Senecio* altitudinal hybrid zone on Mt. Etna. *Molecular Ecology* 25:
513 2467–2481.
514 Fritz, R. S., C. G. Hochwender, B. R. Albrectsen, and M. E. Czesak. 2006. Fitness and genetic
515 architecture of parent and hybrid willows in common gardens. *Evolution* 60: 1215–
516 1227.
517 Gailing, O., and A. L. Curtu. 2014. Inter specific gene flow and maintenance of species
518 integrity in oaks. *Annals of Forest Research* 57: 5–18.
519 Gao, H., S. Williamson, and C. D. Bustamante. 2007. A Markov Chain Monte Carlo approach
520 for joint interface of population structure and inbreeding rates from multilocus
521 genotype data. *Genetics* 176: 1635–1651.
522 Hardy, O. J. 2009. How fat is the tail? *Heredity* 103: 437–438.
523 Isagi, Y., T. Kanazashi, W. Suzuki, H. Tanaka, and T. Abe. 1999. Polymorphic microsatellite
524 DNA markers for *Magnolia obovata* Thunb. and their utility in related species.
525 *Molecular Ecology* 8: 698–700.
526 Japan Association for Shidekobushi Conservation. 1996. Wild stand of Shidekobushi (star
527 magnolia) in Japan. Japan Association for Shidekobushi Conservation, Mizunami,
528 Gifu, Japan (in Japanese).

529 Kalinowski, S. T., M. L. Taper, and T. C. Marshall. 2007. Revising how the computer program
530 CERVUS accommodates genotyping error increases success in paternity assignment.
531 *Molecular Ecology* 16: 1099–1106.
532 Keller, M., J. Kollmann, and P. J. Edwards. 2000. Genetic introgression from distant
533 provenances reduces fitness in local weed population. *Journal of Applied Ecology* 37:
534 647–659.
535 Lepais, O., and S. Gerber. 2011. Reproductive patterns shape introgression dynamics and
536 species succession within the European white oak species complex. *Evolution* 65:
537 156–170.
538 Lexer, C., M. F. Fay, J. A. Joseph, M.-S. Nica, and B. Heinze. 2005. Barrier to gene flow
539 between two ecologically divergent *Populus* species, *P. alba* (white poplar) and *P.*
540 *tremula* (European aspen): the role of ecology and life history in gene introgression.
541 *Molecular Ecology* 14: 1045–1057.
542 Lindtke, D., Z. Gompert, C. Lexer, and C. A. Buerkle. 2014. Unexpected ancestry of Populus
543 seedlings from a hybrid zone implies a large role for postzygotic selection in the
544 maintenance of species. *Molecular Ecology* 23: 4316–4330.
545 Macaya-Sanz, D., L. Suter, J. Joseph, T. Barbara, N. Alba, S. C. Gonzalez-Gonzalez, A.
546 Widmer, and C. Lexer. 2011. Genetic analysis of post-mating reproductive barriers in
547 hybridizing Europian *Populus* species. *Heredity* 107: 478–486.
548 Marczewski, T., D. F. Chamberlain, and R. I. Milne. 2015. Hybridization in closely related
549 *Rhododendron* species: half of all species-differentiating markers experience serious
550 transmission ratio distortion. *Ecology and Evolution* 5: 3003–3022.
551 Milne, R. I., S. Terzioglu, and R. J. Abbott. 2003. A hybrid zone dominated by fertile F_1s:
552 maintenance of species barriers in *Rhododendron*. *Molecular Ecology* 12: 2719–2729.
553 Muhlfeld, C. C., S. T. Kalinowski, T. E. Mcmahon, M. L. Taper, S. Painter, R. F. Leary, and F.
554 W. Allendorf. 2009. Hybridization rapidly reduces fitness of a native trout in the wild.
555 *Biology Letters* 5: 328–331.
556 Muranishi, S., I. Tamaki, S. Setsuko, and N. Tomaru. 2011. Identification of natural hybrids
557 between *Magnolia stellata* and *M. salicifolia*. *Chubu Forestry Research* 59: 39–42 (in
558 Japanese).
559 Muranishi, S., I. Tamaki, S. Setsuko, and N. Tomaru. 2013. Asymmetric introgression
560 between *Magnolia stellata* and *M. salicifolia* at a site where the two species grow
561 sympatrically. *Tree Genetics & Genomes* 9: 1005–1015.
562 Murray, M. G., and W. F. Thompson. 1980. Rapid isolation of high molecular weight plant
563 DNA. *Nucleic Acids Research* 8: 4321–4326.

564 Norris, L. C., B. J. Main, Y. Lee, T. C. Collier, A. Fofana, A. J. Cornel, and G. C. Lanzaro.
565 2015. Adaptive introgression in an African malaria mosquito coincident with the
566 increased usage of insecticide-treated bed nets. *Proceedings of the National Academy*
567 *of Sciences* 112: 815–820.
568 Olrik, D. C., and E. D. Kjaer. 2007. The reproductive success of a *Quercus petraea* × *Q. robur*
569 F1-hybrid in back-crossing situations. *Annals of Forest Science* 64: 37–45.
570 Plummer, M., N. Best, K. Cowles, and K. Vines. 2006. CODA: Convergence Diagnosis and
571 Output Analysis for MCMC. *R News* 6: 7–11.
572 Scascitelli, M., K. D. Whitney, R. A. Randell, M. King, C. A. Buerkle, and L. H. Rieseberg.
573 2010. Genome scan of hybridizing sunflowers from Texas (*Helianthus annuus* and *H.*
574 *debilis*) reveals asymmetric patterns of introgression and small islands of genomic
575 differentiation. *Molecular Ecology* 19: 521–541.
576 Setsuko, S., and N. Tomaru. 2011. The effects of plant size and light availability on male and
577 female reproductive success and functional gender in a hermaphrodite tree species,
578 *Magnolia stellata*. *Botany* 89: 593–604.
579 Setsuko, S., T. Nagamitsu, and N. Tomaru. 2013. Pollen flow and effects of population
580 structure on selfing rates and female and male reproductive success in fragmented
581 *Magnolia stellata* populations. *BMC Ecology* 13: 10.
582 Setsuko, S., S. Ueno, Y. Tsumura, and N. Tomaru. 2005. Development of microsatellite
583 markers in *Magnolia stellata* (Magnoliaceae), a threatened Japanese tree.
584 *Conservation Genetics* 6: 317–320.
585 Setsuko, S., T. Nagamitsu, K. Ishida, and N. Tomaru. 2012. Relationships between
586 flower-visiting insects and female reproductive success in *Magnolia stellata*. *Chubu*
587 *Forestry Research* 60: 37–42.
588 Stan Development Team. 2016. RStan: the R interface to Stan. R package version 2.14.1.
589 Steinhoff, S. 1993. Results of species hybridizaition with *Quercus robur* L and *Quercus*
590 *petraea* (Matt) Liebl. *Annals of Forest Science* 50: 137s–143s.
591 Sun, Y., R. J. Abbott, L. Li, L. Li, J. Zou, and J. Liu. 2014. Evolutionary history of purple
592 cone spruce (*Picea purpurea*) in the Qinghai-Tibet Plateau: homoploid hybrid origin
593 and Pleistocene expansion. *Molecular Ecology* 23: 343–359.
594 Tamaki, I., and M. Okada. 2014. Genetic admixing of two evergreen oaks, *Quercus acuta* and
595 *Q. sessilifolia* (subgenus *Cyclobalanopsis*), is the result of interspecific introgressive
596 hybridization. *Tree Genetics & Genomes* 10: 989–999.
597 Tamaki, I., S. Setsuko, and N. Tomaru. 2009. Estimation of outcrossing rates at hierarchical
598 levels of fruits, individuals, populations and species in *Magnolia stellata*. *Heredity*

102: 381–388.

Tamaki, I., W. Yoichi, Y. Matsuki, Y. Suyama, and M. Mizuno. 2017. Inconsistency between morphological traits and ancestry of individuals in the hybrid zone between two *Rhododendron japonoheptamerum* varieties revealed by a genotyping-by-sequencing approach. *Tree Genetics & Genomes* 13: 4.

Tani, S., I. Tamaki, S. Setsuko, and N. Tomaru. 2014. Differences in seed formation and germination rates between reciprocal interspecific crosses in *Magnolia stellata* and *M. salicifolia*. *Journal of Japanese Forest Society* 96: 200–205 (in Japanese with English abstract).

Tsuda, Y., J. Chen, M. Stocks, T. Kallman, J. H. Sonstebo, L. Parducci, V. Semerikov, et al. 2016. The extent and meaning of hybridization and introgression between Siberian spruce (*Picea obovata*) and Norway spruce (*Picea abies*): cryptic refugia as stepping stones to the west? *Molecular Ecology* 25: 2773–2789.

Ueno, S., S. Setsuko, T. Kawahara, and H. Yoshimaru. 2005. Genetic diversity and differentiation of the endangered Japanese endemic tree *Magnolia stellata* using nuclear and chloroplast microsatellite markers. *Conservation Genetics* 6: 563–574.

Wachowiak, W., A. E. Palme, and O. Savolainen. 2011. Speciation history of three closely related pines *Pinus mugo* (T.), *P. uliginosa* (N.) and *P. sylvestris* (L.). *Molecular Ecology* 20: 1729–1743.

Wang, J., T. Kallman, J. Liu, Q. Guo, Y. Wu, K. Lin, and M. Lascoux. 2014. Speciation of two desert poplar species triggered by Pleistocene climatic oscillations. *Heredity* 112: 156–164.

Whitney, K. D., R. A. Randell, and L. H. Rieseberg. 2006. Adaptive introgression of herbivore resistance traits in the weedy sunflower *Helianthus annuus*. *The American Naturalist* 167: 794–807.

Whitney, K. D., K. W. Broman, N. C. Kane, S. M. Hovick, R. A. Randell, and L. H. Rieseberg. 2015. QTL mapping identifies candidate alleles involved in adaptive introgression and range expansion in a wild sunflower. *Molecular Ecology* 24: 2194–2211.

Yasukawa, H., H. Kato, R. Yamaoka, H. Tanaka, H. Arai, and S. Kawano. 1992. Reproductive and pollination biology of *Magnolia* and its allied genera (Magnoliaceae) I. Floral volatiles of several *Magnolia* and *Michelia* species and their roles in attracting insects. *Plant Species Biology* 7: 121–140.

Zha, H.-G., R. I. Milne, and H. Sun. 2010. Asymmetric hybridization in *Rhododendron agastum*: a hybrid taxon comprising mainly F_1s in Yunnan, China. *Annals of Botany* 105: 89-100.

Zhou, Y., L. Duvaux, G. Ren, L. Zhang, O. Savolainen, and J. Liu. 2017. Importance of incomplete lineage sorting and introgression in the origin of shared genetic variation between two closely related pines with overlapping distributions. *Heredity* 118: 211–220.

DATA ACCESIBILITY

Genotype data and program codes have been submitted to Dryad: doi:… (If the manuscript will be accepted, we will submit the data sets).

TABLE 1. Population genetic statistics for 27 microsatellite loci in *Magnolia stellata* and *M. salicifolia* adult trees, which were determined by InStruct analysis with probability > 0.95.

Species	Locus	*A*	H_O	H_E	*F*	*Null*
M. stellata	stm0002[a]	8	0.764	0.746	-0.024	0.020
(*N* = 55)	stm0163[a]	7	0.691	0.732	0.057	0.032
	stm0184[a]	7	0.655	0.726	0.098	0.043
	stm0200[a]	7	0.691	0.655	-0.055	0.024
	stm0214[a]	6	0.618	0.661	0.064	0.039
	stm0223	6	0.309	0.689	0.551	0.208
	stm0225[a]	10	0.818	0.864	0.053	0.024
	stm0246	13	0.691	0.872	0.208	0.098
	stm0251[a]	9	0.855	0.813	-0.051	0.016
	stm0353[a]	9	0.818	0.811	-0.009	0.026
	stm0383[a]	15	0.800	0.789	-0.014	0.015
	stm0415[a]	6	0.746	0.709	-0.051	0.023
	stm0423	14	0.455	0.840	0.459	0.210
	stm0448[a]	6	0.655	0.640	-0.022	0.026
	M6D8[a]	7	0.836	0.791	-0.057	0.018
	mag00402	17	0.927	0.879	-0.055	0.019
	mag01823[b]	7	0.909	0.804	-0.131	0.013
	mag04151	5	0.582	0.607	0.041	0.061
	mag04167[b]	7	0.818	0.758	-0.079	0.019
	mag04769	5	0.618	0.692	0.106	0.052
	mag05338[b]	10	0.746	0.848	0.121	0.048
	mag05534[b]	9	0.909	0.787	-0.155	0.012
	mag06266	7	0.473	0.718	0.342	0.136
	mag08314[b]	4	0.600	0.654	0.083	0.041
	mag08400	10	0.946	0.864	-0.094	0.011
	mag10551[b]	9	0.964	0.873	-0.104	0.010
	mag14347[b]	12	0.909	0.822	-0.106	0.012
	Average (27 loci)	8.59	0.733	0.765	0.043	0.046
	Average (19 loci)	8.16	0.779	0.762	-0.020	0.024
	Average (12 loci)	8.08	0.745	0.745	-0.001	0.026

Notes: A, number of alleles; H_O, observed heterozygosity; H_E, expected heterozygosity; *F*, inbreeding coefficient; *Null*, null allele frequency.

[a] These 12 loci were used for the identification of genealogical classes and paternity analysis.

[b] These seven loci were only used for the identification of genealogical classes. In total 19 loci were used in the analysis.

650 **TABLE 1,** *continued*

Species	Locus	*A*	H_O	H_E	*F*	*Null*
M. salicifolia	stm0002[a]	7	0.640	0.702	0.088	0.026
(*N* = 186)	stm0163[a]	11	0.710	0.859	0.173	0.032
	stm0184[a]	8	0.640	0.674	0.051	0.014
	stm0200[a]	14	0.812	0.865	0.061	0.020
	stm0214[a]	6	0.710	0.713	0.004	0.011
	stm0223	9	0.731	0.723	-0.011	0.007
	stm0225[a]	16	0.785	0.908	0.135	0.010
	stm0246	13	0.871	0.853	-0.021	0.008
	stm0251[a]	9	0.753	0.780	0.035	0.011
	stm0353[a]	7	0.366	0.380	0.038	0.017
	stm0383[a]	13	0.866	0.868	0.002	0.008
	stm0415[a]	9	0.672	0.736	0.087	0.040
	stm0423	11	0.790	0.838	0.056	0.020
	stm0448[a]	9	0.753	0.756	0.005	0.010
	M6D8[a]	7	0.720	0.783	0.080	0.028
	mag00402	13	0.091	0.641	0.857	0.454
	mag01823[b]	5	0.280	0.333	0.161	0.036
	mag04151	3	0.323	0.716	0.550	0.087
	mag04167[b]	10	0.742	0.786	0.055	0.021
	mag04769	8	0.532	0.607	0.123	0.044
	mag05338[b]	8	0.731	0.678	-0.078	0.013
	mag05534[b]	14	0.828	0.859	0.036	0.012
	mag06266	3	0.231	0.266	0.131	0.054
	mag08314[b]	9	0.823	0.855	0.038	0.013
	mag08400	8	0.468	0.786	0.405	0.192
	mag10551[b]	11	0.758	0.699	-0.084	0.008
	mag14347[b]	14	0.823	0.820	-0.004	0.009
	Average (27 loci)	9.44	0.646	0.722	0.110	0.045
	Average (19 loci)	9.84	0.706	0.740	0.047	0.018
	Average (12 loci)	9.67	0.702	0.752	0.063	0.019

651 *Notes: A*, number of alleles; H_O, observed heterozygosity; H_E, expected heterozygosity; *F*,

652 inbreeding coefficient; *Null*, null allele frequency.

653 [a] These 12 loci were used for the identification of genealogical classes and paternity analysis.

654 [b] These seven loci were only used for the identification of genealogical classes. In total 19 loci

655 were used in the analysis.

TABLE 2. Numbers of adult trees and candidate fathers in each genealogical class.

		Number of candidate fathers		
Genealogical class	Number of adult trees	Flowering was observed	Flowering was not observed	Total
MST	54	38	4	42
MSA	187	137	3	140
F_1	43	34	2	36
F_2	5	5	0	5
BxMST	4	4	0	4
BxMSA	13	9	2	11
Unknown	1	0	0	0
Total	307	227	11	238

Note: MST, MSA, F_1, F_2, BxMST, BxMSA and Unknown indicate *M. stellata*, *M. salicifolia*, F_1 hybrids, F_2 hybrids, backcrosses to *M. stellata*, backcrosses to *M. salicifolia* and unknown hybrids, respectively.

TABLE 3. Numbers of adult trees with each chloroplast haplotype in each genealogical class.

	Chloroplast DNA haplotype					
Genealogical class	A	B	C	D	E	Total
MST	49	1	4	0	0	54
MSA	0	0	0	186	1	187
F_1	0	0	0	43	0	43
F_2	0	0	0	5	0	5
BxMST	2	0	0	2	0	4
BxMSA	0	0	0	12	1	13
Unknown	0	0	0	1	0	1
Total	51	1	4	249	2	307

Note: MST, MSA, F_1, F_2, BxMST, BxMSA and Unknown indicate *M. stellata*, *M. salicifolia*, F_1 hybrids, F_2 hybrids, backcrosses to *M. stellata*, backcrosses to *M. salicifolia* and unknown hybrids, respectively.

TABLE 4. Paternity of seeds collected from maternal trees in each genealogical class.

Genealogical class	Mother ID	Number of analyzed seeds	Number of seeds whose father was identified	Number of selfed seeds	Number of outcrossed seeds by each genealogical class of its father							
					Total	MST	MSA	F_1	F_2	BxMST	BxMSA	Unknown
MST	OG59	44	42	1	41	41	0	0	0	0	0	0
(*N* = 11)	PH318	8	8	5	3	2	0	1	0	0	0	0
	PH503	10	10	2	8	7	0	1	0	0	0	0
	PH504	2	1	0	1	1	0	0	0	0	0	0
	PH547	53	52	14	38	37	0	0	0	1	0	0
	PH569	25	24	0	24	14	0	10	0	0	0	0
	PH608	24	24	0	24	22	0	1	1	0	0	0
	PH613	2	2	2	0	0	0	0	0	0	0	0
	PH622	75	59	0	59	55	0	4	0	0	0	0
	PH624	12	12	11	1	0	0	1	0	0	0	0
	YC152	6	6	2	4	3	1	0	0	0	0	0
	Total	261	240	37	203	182	1	18	1	1	0	0
MSA	PH319	23	23	14	9	0	4	4	1	0	0	0
(*N* = 11)	PH347	6	4	1	3	0	2	1	0	0	0	0
	PH401	2	2	2	0	0	0	0	0	0	0	0
	PH402	67	56	0	56	3	47	2	0	0	4	0
	PH412	11	9	0	9	0	9	0	0	0	0	0
	PH443	1	1	0	1	0	1	0	0	0	0	0
	PH469	6	5	0	5	0	2	1	0	0	2	0
	PH506	4	4	0	4	0	0	4	0	0	0	0
	PH510	23	23	6	17	0	13	4	0	0	0	0
	PH526	3	3	2	1	0	0	0	1	0	0	0
	YC179	8	8	3	5	0	2	0	0	0	3	0
	Total	154	138	28	110	3	80	16	2	0	9	0
F_1	N2	14	12	5	7	0	4	2	0	0	1	0
(*N* = 11)	PH320	13	11	2	9	0	7	2	0	0	0	0
	PH450	11	6	1	5	0	4	1	0	0	0	0
	PH451	22	21	0	21	10	10	1	0	0	0	0
	PH497	12	11	1	10	8	1	0	1	0	0	0
	PH513	2	2	0	2	1	0	0	0	0	1	0
	PH591	38	29	3	26	1	9	15	0	0	1	0
	PH595	9	8	0	8	1	0	7	0	0	0	0
	PH604	4	4	4	0	0	0	0	0	0	0	0
	PH606	1	1	0	1	1	0	0	0	0	0	0
	PH641	6	6	2	4	0	3	1	0	0	0	0
	Total	132	111	18	93	22	38	29	1	0	3	0
F_2	YC220	7	7	2	5	3	2	0	0	0	0	0
BxMST	PH575	7	5	0	5	3	0	2	0	0	0	0
BxMSA	PH322	5	0	0	0	0	0	0	0	0	0	0
(*N* = 2)	YC190	8	8	0	8	6	0	2	0	0	0	0
	Total	13	8	0	8	6	0	2	0	0	0	0
All		574	509	85	424	219	121	67	4	1	12	0

Note: MST, MSA, F_1, F_2, BxMST, BxMSA and Unknown indicate *M. stellata*, *M. salicifolia*, F_1 hybrids, F_2 hybrids, backcrosses to *M. stellata*, backcrosses to *M. salicifolia* and unknown hybrids, respectively.

TABLE 5. Posterior mode and 95% highest posterior density (HPD) for selfing rates and variance components.

Parameter		Mode (95% HPD)
Selfing rate	MST	0.061 (0.001 – 0.284)
	MSA	0.063 (0.004 – 0.266)
	F_1	0.053 (0.003 – 0.242)
	OTH	0.017 (0.000 – 0.209)
Variance component	among genealogical classes	0.30 (0.00 – 16.95)
	among maternal trees within genealogical classes	7.89 (3.18 – 21.89)

Note: MST, MSA, F_1 and OTH indicate *M. stellata*, *M. salicifolia*, F_1 hybrids and the other hybrids, respectively.

672 FIGURE LEGENDS

673

674 **FIGURE 1** Locations of adult trees and directions of pollen movements. MST, MSA, F_1, F_2,
675 BxMST, BxMSA and Unknown indicate *M. stellata*, *M. salicifolia*, F_1 hybrids, F_2 hybrids,
676 backcrosses to *M. stellata*, backcrosses to *M. salicifolia* and unknown hybrids, respectively.
677 Large marks indicate maternal trees whose seeds were collected for the paternity analysis.
678 Observed pollen movements are shown by arrows (from fathers to mothers). Curved lines
679 indicate 5-m contours.

680

681 **FIGURE 2** Posterior distributions of individuals belonging to each genealogical class
682 estimated by NewHybrids. MST, MSA, F_1, F_2, BxMST, BxMSA and Unknown indicate *M.*
683 *stellata*, *M. salicifolia*, F_1 hybrids, F_2 hybrids, backcrosses to *M. stellata*, backcrosses to *M.*
684 *salicifolia* and unknown hybrids, respectively. Genealogical classes shown under the bar plot
685 indicate assigned classes using the threshold value of 0.59.

686

687 **FIGURE 3** Posterior distributions of shape parameter (*b*) in the exponential power function.
688 Dots and bars indicate posterior modes and 95% highest posterior densities, respectively.
689 Different letters above each bar indicate significant difference in posterior distribution with
690 95% confidence. MST, MSA, F_1 and OTH indicate *M. stellata*, *M. salicifolia*, F_1 hybrids and
691 the other hybrids, respectively.

692

693 **FIGURE 4** Posterior distribution of parameters for relative compatibility against intraspecific
694 mating of *M. stellata* (i.e., $c_{[MST, MST]}$ was fixed to 1). Dots and bars indicate posterior modes

695 and 95% highest posterior densities, respectively. Different letters above each bar indicate
696 significant difference in posterior distribution with 95% confidence. MST, MSA, F_1 and OTH
697 indicate *M. stellata*, *M. salicifolia*, F_1 hybrids and the other hybrids, respectively.

FIGURE 1

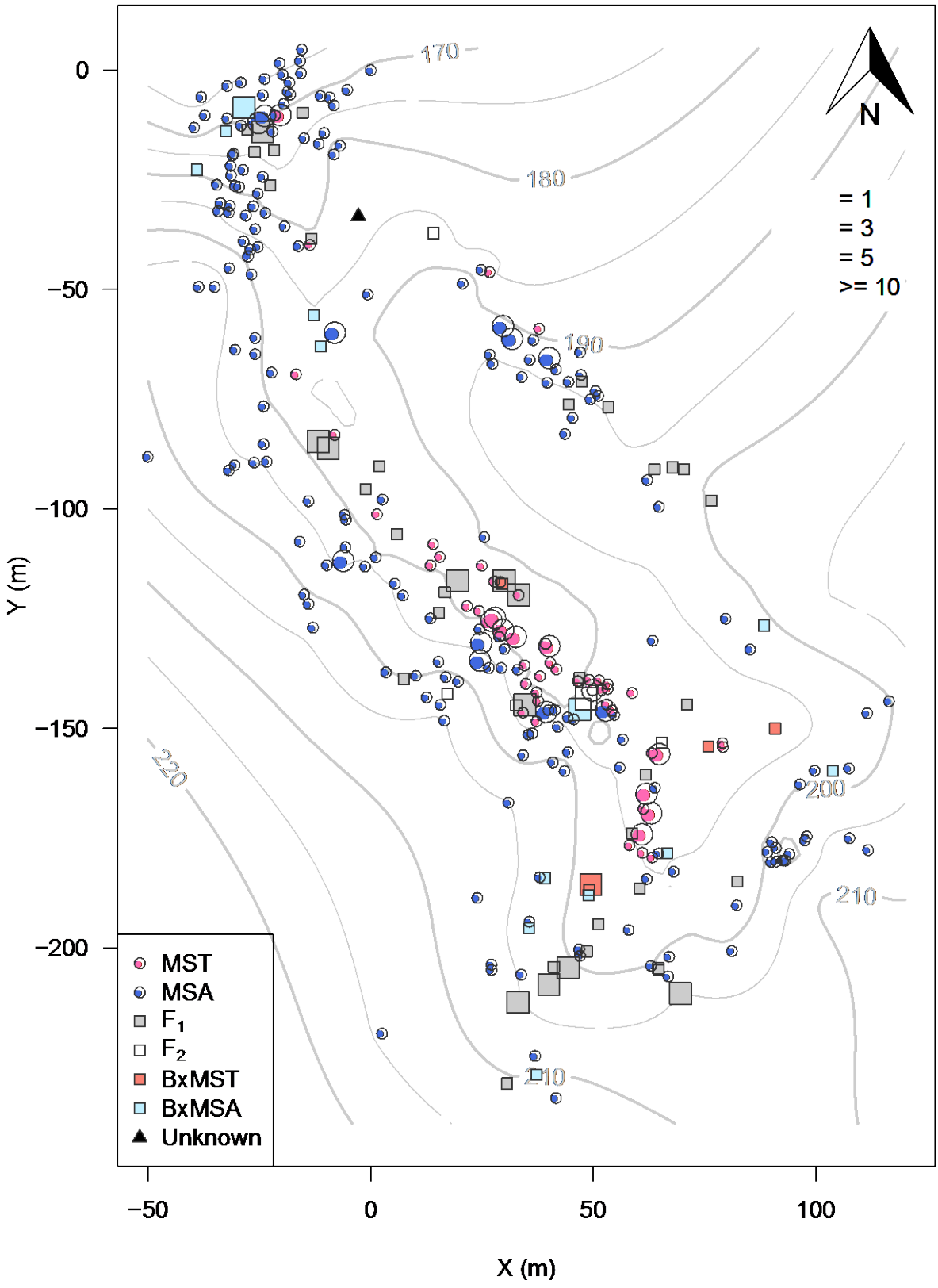

0
−50
−100
−150
−200
Y (m)
−50
0
50
100
X (m)
170
180
190
200
210
220
210
N
= 1
= 3
= 5
>= 10
MST
MSA
F1
F2
BxMST
BxMSA
Unknown

FIGURE 2

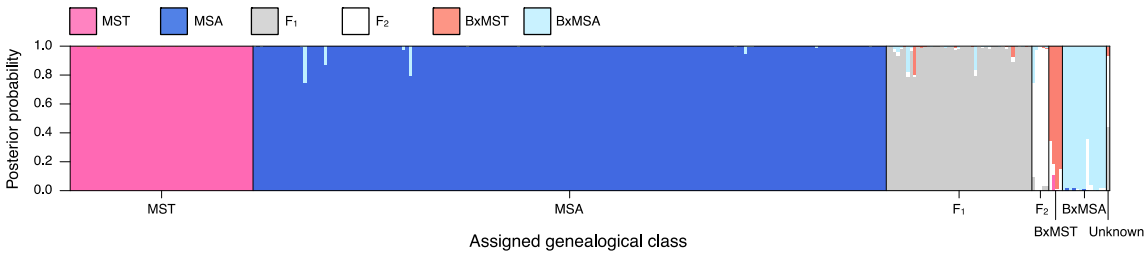

MST
MSA
F1
F2
BxMST
BxMSA
Posterior probability
1.0
0.8
0.6
0.4
0.2
0.0
MST
MSA
F1
F2
BxMST
BxMSA
Unknown
Assigned genealogical class

FIGURE 3

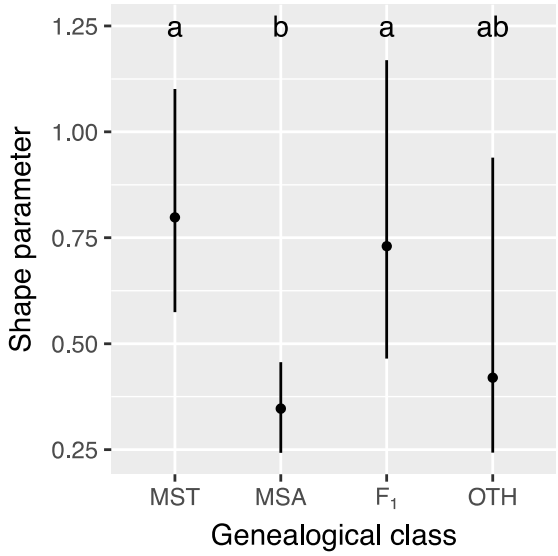
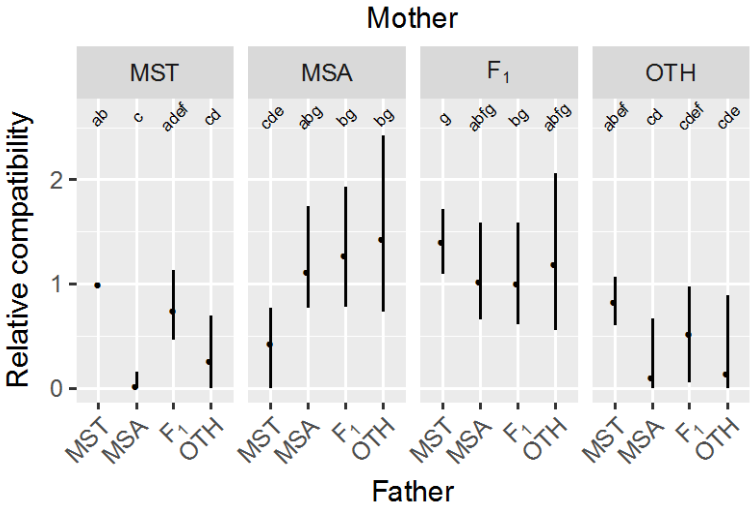

a
b
a
ab
1.25
1.00
0.75
0.50
0.25
Shape parameter
MST
MSA
F1
OTH
Genealogical class

FIGURE 4


Mother
MST
MSA
F1
OTH
ab
c
adef
cd
cde
abg
bg
bg
g
abfg
bg
abfg
abef
cd
cdef
cde
Relative compatibility
0
1
2
MST
MSA
F1
OTH
Father

Appendix S1. Characteristics of 12 newly developed microsatellite markers for *Magnolia stellata*, *M. salicifolia* and *M. kobus* that were used in this study.

Locus		Primer sequences (5′–3′)	GenBank accession no.	Repeat motif	Ta (C°)	Allele size range
mag00402	F:	CCAGCTTATCATCCTCCGAAACC	LC213632	$(TC)_{15}$	57	246–308
	R:	TTACTTCGACCAGTTGGAAAGGC	LC213633			
mag01823	F:	GAATAATCTACGTTGGCGGGTACG	LC213638	$(GA)_{10}$	57	218–236
	R:	ATTAACGGGCACGATTCCTTCTC	LC213639			
mag04151	F:	GGCATTCAAGAAGAAGCCTTTGG	LC213660	$(GA)_{11}$	57	331–351
	R:	TTTCCACAACATTGCCATCTCC	LC213661			
mag04167	F:	TCGGGAGGAGGAGAGCGATTAG	LC213662	$(TC)_{10}$	57	216–248
	R:	GTCCGTTAGGATGTCGTTGATGC	LC213663			
mag04769	F:	AAGATTCCTGCGATGATTCCGAC	LC213670	$(AG)_{12}$	57	116–138
	R:	GCCAACGAGAGAGAAATCAAACG	LC213671			
mag05338	F:	TTGCTGCATCTGCTCATCCTC	LC213674	$(GA)_{16}$	57	184–243
	R:	TTCTTCAAGATGAAGCTGGCACC	LC213675			
mag05534	F:	CGTTAATCTCGTTACTTCCCGCC	LC213678	$(CT)_{12}$	57	277–325
	R:	TCTTATCCTCCGCACCCTCTCTC	LC213679			
mag06266	F:	GATAAGATACCGGAGCAAACGGG	LC213684	$(AG)_{12}$	57	123–146
	R:	ACGAGTCACCGAATCCAGACATC	LC213685			
mag08314	F:	CCAATTTGGAAGAGAATGCCCTC	LC213700	$(TC)_{13}$	57	346–378
	R:	TAATCTGATGCGGGAGTTGGAAG	LC213701			
mag08400	F:	CACAGACTTAGCAAAGATGCCCG	LC213706	$(GA)_{11}$	57	325–350
	R:	CGCTCGCATACAAGCTAAATTGG	LC213707			
mag10551	F:	GCACCAACAACCTACCTGCAATC	LC213712	$(CT)_{16}$	57	161–208
	R:	AATACAAGAGGCCCACTGTCACG	LC213713			
mag14347	F:	GAGGAGGATGAGAGCTTTCCGAG	LC213722	$(CT)_{10}$	57	111–184
	R:	TCGAGAGAAAGGGAAGAGAAGCC	LC213723			

Note: Sequence data have been submitted to GenBank: accession numbers LC213632–LV213725.

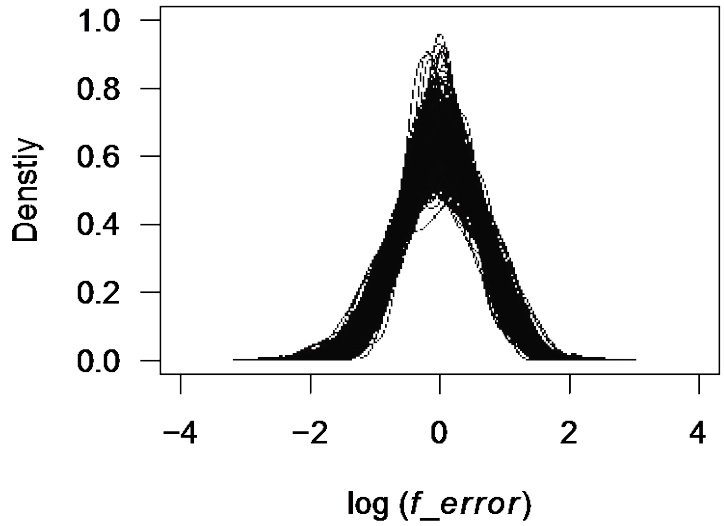
Appendix S2. Posterior mode and 95% highest posterior density (HPD) for all parameters in the gene dispersal model.

Parameter	Mode (95% HPD)
a	43.3 (32.7 – 54.3)
$b_{[MST]}$	0.798 (0.574 – 1.101)
$b_{[MSA]}$	0.347 (0.242 – 0.456)
$b_{[F1]}$	0.730 (0.465 – 1.169)
$b_{[OTH]}$	0.420 (0.243 – 0.939)
$c_{[MST, MSA]}$	0.022 (0.000 – 0.164)
$c_{[MST, F1]}$	0.748 (0.470 – 1.132)
$c_{[MST, OTH]}$	0.264 (0.000 – 0.700)
$c_{[MSA, MST]}$	0.426 (0.000 – 0.773)
$c_{[MSA, MSA]}$	1.113 (0.775 – 1.750)
$c_{[MSA, F1]}$	1.273 (0.788 – 1.939)
$c_{[MSA, OTH]}$	1.432 (0.733 – 2.427)
$c_{[F1, MST]}$	1.404 (1.099 – 1.718)
$c_{[F1, MSA]}$	1.025 (0.667 – 1.588)
$c_{[F1, F1]}$	1.002 (0.621 – 1.595)
$c_{[F1, OTH]}$	1.190 (0.559 – 2.060)
$c_{[OTH, MST]}$	0.824 (0.611 – 1.072)
$c_{[OTH, MSA]}$	0.106 (0.000 – 0.675)
$c_{[OTH, F1]}$	0.524 (0.061 – 0.982)
$c_{[OTH, OTH]}$	0.143 (0.000 – 0.893)
d	0.685 (0.496 – 0.799)
σ	0.564 (0.460 – 0.719)

Notes: a and $b_{[GC]}$ are the scale and shape parameters of pollen dispersal kernels, respectively; the parameter b was estimated for each genealogical class (GC). $c_{[GCM, GCF]}$ indicates the relative compatibility between genealogical classes of mothers and fathers (GCM and GCF, respectively). The value of $c_{[MST, MST]}$ was fixed to 1.0 and the other combinations of $c_{[GCM, GCF]}$ were defined relative to it. The significance of the differences between the values of $b_{[GC]}$ or $c_{[GCM, GCF]}$ was assessed by examining whether the 95% HPD of the differences between

their posteriors contained zero (see Figures 3 and 4). MST, MSA, F_1 and OTH indicate *M. stellata*, *M. salicifolia*, F_1 hybrids and the other hybrids, respectively.

In the model, the fecundity of a father was assumed to be determined by its diameter at breast height (DBH) and a random effect estimated for each father. *d* indicates the effect of father's DBH on siring success; σ indicates the standard deviation of the parameter (*f_error*) for random effects on the log scale (see also Appendix S3). The mode of the parameter *d* is positive and its 95% HPD does not contain zero, indicating that DBH is positively related with siring success.



Appendix S3. Posterior distribution of the parameter for random effects (*f_error*) on the fecundity of 238 candidate fathers. Rugs under the curves indicate the location of their modes.