MOLECULAR ECOLOGY

Molecular Ecology (2009) **18**, 3967–3979 doi: 10.1111/j.1365-294X.2009.04326.x

Relative contribution of contemporary pollen and seed dispersal to the effective parental size of seedling population of California valley oak (*Quercus lobata*, Née)

DELPHINE GRIVET,*†[1] JUAN J. ROBLEDO-ARNUNCIO,†[1] PETER E. SMOUSE‡ and VICTORIA L. SORK*

**Department of Ecology and Evolutionary Biology, University of California Los Angeles, Box 951606, Los Angeles, CA 90095-1606, USA, †Department of Forest Systems and Resources, Forest Research Institute, CIFOR-INIA, Carretera de la Coruña km 7.5, 28040 Madrid, Spain, ‡Department of Ecology, Evolution and Natural Resources, Rutgers University, New Brunswick, NJ 08901, USA*

Abstract

For plant populations, gene movement through pollen and seed dispersal governs the size of local genetic neighbourhoods and shapes the opportunities for natural selection and genetic drift. A critical question is how together these two processes influence the evolutionary dynamics of local populations. To assess the respective contributions of pollen and seed flow, we propose a novel indirect assessment of the separate male and female gametic contributions to total effective parental size (N_e), based on parental correlations estimated via kinship coefficients, that can be applied to data sets that include unambiguous genotypes for male and female gametic contributions. Using the endemic Californian valley oak (*Quercus lobata*) as our study species, we apply this method to a set of microsatellite genotypes for two distinct ecological sets of naturally recruiting seedlings with acorns attached. We found that the effective numbers of contributing male parents (N_{ep}) exceed effective numbers of female parents (N_{em}) for seedlings established beneath adult trees (N_{ep} = 8.1 and N_{em} = 1.1), as well as for seedlings established away from adult trees (N_{ep} = 15.4 and N_{em} = 2.7), illustrating that seed dispersal enhances pollen dispersal and increases the effective number of seed sources in open seedling patches. The resulting effective parental size of seedling populations translates into smaller effective numbers of parents for undispersed vs. dispersed seedlings (N_e = 3.6 and N_e = 6.7, respectively). This study introduces a novel statistic method and provides important new evidence that, on a short-term temporal scale, seed dispersal shapes the local neighbourhood size of new recruits.

Keywords: effective parental size, neighbourhood, pollen, *Quercus lobata*, seed, seedling

Received 16 March 2009; revision received 3 July 2009; accepted 4 July 2009

Introduction

The potential for rapid evolution of populations facing dramatic environmental changes, due to landscape alteration and climate change, has focused attention on contemporary evolution (Stockwell *et al.* 2003; Smith & Bernatchez 2008). For plant populations, the opportunity for gene movement comes through pollen or seed and, once a seedling becomes established, the outcome of that movement determines the fine scale genetic structure of the population within which genetic drift and selection will take place. Given human-induced landscape changes, many biologists are concerned that the populations are jeopardized through loss of genetic diversity when pollen or seed dispersal is disrupted (Ledig 1992; Ellstrand & Elam 1993; Young *et al.* 1996), although many tree populations may have sufficient local genetic variation to mitigate this problem, at least

Correspondence: Delphine Grivet, Fax: +34 913572293; E-mail: dgrivet@inia.es

[1]These authors contributed equally to this study.

for the near term future (Hamrick 2004). Nonetheless, extant populations will be challenged by increased climatic variability, temperature rise, and other environmental changes and the ability to respond to these changes will depend on the amount of local genetic variation in locally subdivided populations.

Sewall Wright (1943, 1946) proposed the concept of neighbourhood size to characterize the extent of population subdivision and to gauge the combined effects of restricted gene flow and genetic drift on spatial genetic subdivision of continuous, uniform populations (isolation-by-distance model; Wright 1969). The neighbourhood size is defined as $N_b = 4\pi\sigma^2 d$, where σ^2 is the parent–offspring axial dispersal variance and d is the effective population density. In plant populations, σ^2 is a function of the variance of both haploid (pollen) and diploid (seed) dispersal (Crawford 1984a, b), and their respective impacts on neighbourhood size will depend on their respective variances. If both pollen and seed dispersal are restricted, it is clear that the population should show substantial genetic subdivision. If pollen flow is limited but seed dispersal widespread, one would expect that seed dispersal would dilute the genetic structure produced by restricted pollen flow. Conversely, if seed dispersal is restricted, extensive pollen flow might mitigate the impact of seed dispersal on genetic structure. While pollen represents the first phase of gametic dispersal and is responsible for most long-distance gene movement in numerous tree species (Ennos 1994; Petit & Hampe 2006; Dick 2008), dispersal of seed may have a larger impact on the genetic neighbourhood size because this process moves both maternal and paternal gametic genomes (Crawford 1984a, b; Hamilton 1999). In addition, seed dispersal determines the final location of genotypes, which contributes to fine scale genetic structure. However, the extent of its impact is shaped by the interaction of both seed and pollen movements. Therefore, it would be useful to measure their separate and combined influence on population subdivision at the landscape level and to assess their respective impacts on effective population size of naturally dispersed seedling recruits.

Several recent studies have analysed the relative contributions of pollen and seed movements on naturally recruiting seedling and sapling populations. Pollen flow generally exceeds seed flow, as shown in studies using direct (Burczyk *et al.* 2006; Goto *et al.* 2006; Bittencourt & Sebbenn 2007; Isagi *et al.* 2007; Oddou-Muratorio & Klein 2008; Nakanishi *et al.* 2009) and indirect approaches (Tero *et al.* 2005; Oddou-Muratorio & Klein 2008). Occasionally, however, seed dispersal is more effective than pollen dispersal (Bacles *et al.* 2006) or the two processes are comparable, as shown in one tropical tree species (Hardesty *et al.* 2006). In the studies so far, the one factor that has not been explored is how pollen and seed dispersal interact with each other. Thus, the next issue to explore is the synergy between pollen and seed movement that structures seedling populations.

The overall objective of this study is to quantify the relative contributions of pollen and seed dispersal to established seedling populations, incorporating the interaction of these two processes. To do so, we introduce a novel approach to demonstrate the respective contributions of pollen and seed to the total effective parental size of the same population of natural recruits. We take advantage of a particularly informative data set, consisting of seedlings (with an attached acorn), to estimate, via parental correlations computed from kinship coefficients, the effective number of mothers (N_{em}) and fathers (N_{ep}) contributing to overall effective parental numbers for natural seedling patches. Our study species is valley oak (*Quercus lobata* Née), an endemic California tree species that occurs in oak savannah, oak woodland and riparian oak habitats (Pavlik *et al.* 1991). As in other plant species where the seed remains attached to the seedling, the acorn coexists with the seedling long after germination. Thus, after acorns are dispersed and germinate we can identify the genotype of the maternal seed sources from the maternally inherited pericarp tissue, the genotype of the seedlings from the leaf tissue, and by standard paternal inference, the haploid paternal (pollen) contribution. We compare the maternal and paternal contributions to the effective parental size of discrete demes (patches) of naturally dispersed seedlings, one set established beneath the canopies of adult trees and one set located away from adult trees. These two ecological settings are likely to exhibit different patterns of seedling genetic structure. Most seedlings beneath the adult canopy come from the seed tree above, pollinated by nearby adults; while seedlings in open patches probably result from animal dispersal from multiple seed parents, each pollinated by somewhat distinct sets of paternal parents. We thus expect these two ecological settings to display different values of N_{ep}, N_{em}, as well as their combination N_e.

This study uses an indirect approach (i.e. parental correlations, as introduced by Robledo-Arnuncio *et al.* 2006) to estimate the respective pollen and seed contributions to effective parentage of a set of naturally dispersed seedlings and to illustrate how the two components determine overall effective parental population size, represented by newly established recruits. Specifically, we have four objectives: (i) to estimate parental correlations within and among seedling patches in order to infer the effective number of parents

of individual seedling patches, using kinship coefficients; (ii) to analyse the respective contribution of effective number of pollen (N_{ep}) and seed (N_{em}) donors to the total parental size (N_e) for individual seedling patch; (iii) to test whether N_{ep}, N_{em} and N_e computed from parental correlations lead to similar estimates to those indirectly estimated from genetic differentiation among seedling patches; and (iv) to compute N_{ep}, N_{em} and N_e in two distinct ecological settings: seedlings found in patches beneath seed trees and patches away from potential parent plants.

Methods

Study site and study species

Our study site is located at the UC Santa Barbara Sedgwick Reserve in the Santa Ynez Valley (34°42′N, 120°02′W) of Santa Barbara County (California, USA), which is managed as part of the University of California Natural Reserve System. The study population of valley oak (*Quercus lobata*) co-occurs with coast live oak (*Quercus agrifolia* Née) and blue oak (*Quercus douglasii* Hook and Arn.) within the oak savannah habitat and is situated in a valley along the Figueroa Creek in an area that encompasses about 180 ha (~900 × 2000 m). This low-density population (1–6 trees per hectare) is composed of centenary adult trees with scattered saplings and juveniles. We have elsewhere reported on pollen dispersal (Sork *et al.* 2002; Austerlitz *et al.* 2004, 2007; Pluess *et al.* 2009), acorn movement by acorn woodpeckers (Grivet *et al.* 2005), and fine scale genetic structure of a small number of adults (Dutech *et al.* 2005); all these studies were done in various subareas of the same Figueroa Valley studied here.

Valley oak is wind pollinated with estimates of effective mean pollen dispersal distances ranging from 60–350 m (Sork *et al.* 2002; Austerlitz *et al.* 2004, 2007; Pluess *et al.* 2009). Adult trees produce a massive quantity of acorns every 5–6 years, which was the case in the year of the study (2002). From 2001 to 2008, acorn production was poor (2001, 2003, 2005, 2006) to moderate (2004 and 2008), and seedlings were found solely beneath adult trees, with very few of those surviving. By contrast, during 2002, the year of high acorn production, seedlings were found both beneath adult trees and in areas away from any adults.

Acorns mature in autumn, drop between September and November (VLS personal observation), and germinate without any dormancy period starting in the late fall. A variety of animals disperse the acorns, including mice (*Peromyscus* spp.), California ground squirrels (*Spermophilus beecheyi*), scrub jays (*Aphelocoma coerulescens*) and acorn woodpeckers (*Melanerpes formicivorus*).

Sampling

Following the acorn season of 2002, we collected seedlings with their attached acorns from throughout the Figueroa Creek valley (Fig. 1). A first set was sampled in January 2003, composed of 400 progeny collected from patches directly beneath the canopies of 21 valley oak adults, henceforth designated as 'canopy patches'. In addition, we waited a few months to allow for germination and leaf emergence, after the conditions in the field were optimal to spot the emerging seedlings among the grasses. Then, in March though June 2003, we performed a systematic search of the Figueroa Creek valley to locate dispersed seedlings and found 341 seedlings with their attached acorns, aggregated in five open patches, henceforth designed as 'open patches'.

DNA extraction and amplification

DNA from the maternally inherited pericarp tissue (Ziegenhagen *et al.* 2003) and the corresponding

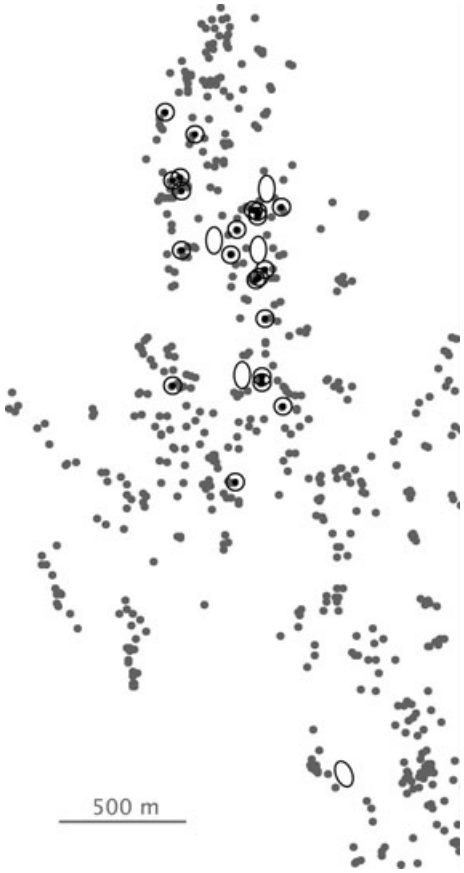

Fig. 1 Localization of the 21 seedling canopy patches (circles with central black dots) and the five seedling patches in open area (ovals), along with the *Quercus lobata* adult trees (small grey dots) in Figueroa Valley of the UC Santa Barbara Sedgwick Reserve, Santa Ynez, California.

biparentally inherited leaf tissue from seedlings were extracted using the DNeasy Plant Mini Kit (Qiagen). For leaf tissue, we started from 100 mg of frozen material and followed the supplier recommendations and the DNA was eluted in 100 μL of water before polymerase chain reaction (PCR) reactions. For the pericarp tissue, we proceeded as described in Grivet *et al.* (2005), starting from 50 mg of material and eluting the DNA in 50 μL of water before PCR reactions. We amplified both sets of DNA with seven microsatellite pairs: MSQ4 (Dow *et al.* 1995), QpZAG1/5, QpZAG9, QpZAG36, QpZAG110 (Steinkellner *et al.* 1997), QrZAG11 and QrZAG20 (Kampfer *et al.* 1998). Amplifications were conducted with the Qiagen Multiplex PCR kit (Qiagen) in 10 μL reaction mixture containing: 2.5 μL of diluted DNA, 5 μL of Multiplex PCR Master Mix, 0.4 μL of bovine serum albumin 10x, 1 μL of each primer pair at 2 μM, 1.1 μL distilled water qsp. PCR reactions were carried out following the conditions reported in Grivet *et al.* (2005).

Genotyping

PCR products were run on an ABI 3700 automatic sequencer, and genotyped based on the statistical packages of the UCLA Sequencing and Genotyping Core Facility (for details see, http://www.genetics.ucla.edu/sequencing/index.php). Each reaction was repeated twice for leaf tissue, and four times for pericarp tissue, to check for genotypic consistency. Due to the low yield of PCR amplification for the pericarp tissue collected from dispersed progeny in June 2003, our samples yielded missing data and genotyping errors as indicated by genetic incompatibilities between pericarp and seedling genotypes.

Missing data can bias the outputs of the genetic structure analyses, so in creating our dataset we used a conservative approach to reduce substantially the number of missing data, with the drawback of reducing our sample size greatly: (i) we only analysed the four best loci (i.e. displaying less missing data) of the pericarp in open patches, and analysed them across all our data sets, namely MSQ4, QpZAG1/5, QpZAG36, QrZAG20 and (ii) we checked the genetic incompatibilities between pericarps and leaf tissue of the same seedlings and zeroed-out the incompatible genotypes (i.e. incompatible alleles were coded as zeros). Taking into account (i) and (ii) our final sample size was 384 seedlings collected beneath 21 adult trees (Table 1, Fig. 1) and 148 seedlings collected in five open patches, away from any adult trees (Table 1, Fig. 1). Male gametic contribution to each seedling was computed using our own programs (available upon request to JJRA). When pericarps and seedlings presented the same heterozygote genotypes, paternal contribution was inferred by fractionally assigning paternity to each of the two possible alleles according to their posterior likelihood value, given the pollen pool frequencies estimated from the unambiguous cases (as done in Robledo-Arnuncio *et al.* 2007). A minimum likelihood of 0.60 was required to accept an allele as the paternal candidate. Multilocus number of allele and unbiased gene diversity (Nei 1987) were computed for the different data sets (i.e. ecological settings and tissue type) using FSTAT version 2.9.3.2 (Goudet 2001) (Table 2). Additional information for the seven loci analysed in the adult population in the Figueroa Creek valley is provided in Pluess *et al.* (2009).

Table 1 Sampling summary of acorns and seedlings along with the percentage of missing data per marker, tissue type and ecological setting

	Canopy patch	Open patch
Total number of individuals	384	148
Number of patches	21	5
Average individual per patch (range)	18.3 (11–25)	29.6 (20–57)
Missing pericarp data (%)	4	16
MSQ4	7	25
QpZAG1/5	8	15
QpZAG36	0	21
QrZAG20	1	2
Missing pollen data (%)	10	17
MSQ4	11	26
QpZAG1/5	14	16
QpZAG36	8	22
QrZAG20	5	5
Missing seedling data (%)	5	1
MSQ4	5	1
QpZAG1/5	5	3
QpZAG36	6	1
QrZAG20	4	1

For each data set, maximized for sample size and minimized for missing data, we estimated per cent missing data by counting the number of missing alleles (canopy patches: 122 for pericarp, 148 for pollen, 154 for progeny; open patches: 186 for pericarp, 101 for pollen, 16 for progeny) and dividing by total alleles [canopy patch: 3072 for pericarp and progeny (4 loci•2 alleles•384 individuals) and 1536 for pollen (4 loci•1 allele•384 individuals); open patches: 1184 for pericarp and progeny (4 loci•2 alleles•148 individuals) and 592 for pollen (4 loci•1 allele•148 individuals)] (Table 1). To test the impact of missing data on our analyses, we produced some data sets with additional random missing data (10% and 20% more missing data than our best data set) for the two ecological settings (canopy and open patches) and the three parental contributions (pericarp, pollen and progeny). Simulated data sets

Table 2 Multilocus genetic diversity for the 384 seedlings in canopy patches and the 148 seedlings in open patches

		N_a			Gene diversity		
Population	N	Pericarp	Pollen	Seedling	Pericarp	Pollen	Seedling
Canopy patch							
1	16	8	21	24	0.500	0.705	0.725
2	19	7	16	18	0.375	0.592	0.565
3	18	7	27	30	0.375	0.665	0.615
4	15	6	16	18	0.250	0.683	0.558
5	14	12	15	20	0.584	0.651	0.668
6	25	7	24	26	0.375	0.711	0.632
7	18	6	17	22	0.250	0.701	0.503
8	18	8	17	21	0.500	0.513	0.621
9	19	7	26	27	0.375	0.725	0.629
10	11	8	16	20	0.500	0.605	0.683
11	13	7	15	17	0.375	0.631	0.588
12	16	6	19	22	0.250	0.630	0.538
13	16	8	20	22	0.388	0.698	0.607
14	20	9	16	19	0.495	0.627	0.627
15	16	6	13	14	0.250	0.391	0.326
16	19	8	21	26	0.499	0.657	0.752
17	18	8	19	22	0.500	0.616	0.640
18	23	6	24	26	0.022	0.606	0.440
19	20	7	19	19	0.375	0.626	0.547
20	19	6	23	24	0.250	0.659	0.527
21	13	5	19	20	0.125	0.632	0.476
Average	17	7	19	22	0.362	0.634	0.584
Open patch							
1	27	20	27	29	0.696	0.752	0.739
2	22	12	13	15	0.449	0.424	0.468
3	57	22	34	35	0.705	0.712	0.775
4	21	12	20	21	0.574	0.703	0.654
5	20	11	18	20	0.501	0.584	0.581
Average	29	15	22	24	0.585	0.635	0.643

N, number of samples; N_a, number of alleles; unbiased gene diversity (Nei 1987).

with additional missing data yielded parental correlation estimates (see definitions below) close to those corresponding to our best data set (relative difference between 0.1% and 5% in most cases; maximum relative difference of 10%). In addition, none of the analyses resulted in different inferences than the findings reported here. Thus, the level of missing data reported here does no bias our results.

Statistical analyses

We used an indirect approach, based on kinship coefficients, to estimate the effective parental size of seedling populations. Our procedure is indirect in that it does not employ genetic parentage assignments. Although kinship analyses may give less precise estimates than parentage analyses when the number of loci is low, they do not require exhaustively genotyping the parental population, representing an useful alternative when the spatial scale of analysis is large or when adult sampling is difficult.

Definitions of parentage correlations and effective parental numbers

Effective population size (N_e) is an important parameter in evolutionary genetics because it influences the rate of inbreeding and loss of genetic variation, as well as the efficiency of natural selection in adaptively shaping the local gene pool. Estimating the current effective population size in natural populations with genetic data is not without its difficulties, and we propose here an original approach to evaluate the number of effective mother-trees (N_{em}) and father-trees (N_{ep}) that contribute to naturally established seedling patches. This approach characterizes the parental structure of seedling patches in terms of average probabilities of paternal and maternal identity for pairs of seedlings. Specifically, we define the following parentage correlations:

- Q_w^m (correlation of maternity within a patch): probability that two randomly drawn seedlings from a patch have been dispersed from the same mother.
- Q_w^p (correlation of paternity within a patch): probability that two randomly drawn seedlings from a patch have been sired by the same father.
- Q_b^m (correlation of maternity among patches): probability that two randomly drawn seedlings from two different patches have been dispersed from the same mother.
- Q_b^p (correlation of paternity among patches): probability that two randomly drawn seedlings from two different patches have been dispersed from the same father.
- Q_w^{mp} (cross-parental correlation within a patch): probability that two randomly drawn seedlings from a patch show a cross-parental match, i.e. the mother of the first is the father of the second, or vice versa.
- Q_b^{mp} (cross-parental correlation among patches): probability that two randomly drawn seedlings from two different patches show a cross-parental match, i.e. the mother of the first is the father of the second, or vice versa.

Using the first two quantities above, we additionally defined the effective number of mothers per seedling patch (N_{em}) as the inverse of the probability of maternal identity within a patch, $N_{em} = 1/Q_w^m$, and the effective number of fathers per seedling patch (N_{ep}) as the inverse of the probability of paternal identity within a patch: $N_{ep} = 1/Q_w^p$.

Effective seedling patch size as a function of effective parental numbers

In order to quantify the relative contribution of pollen and seed to the effective parental size of seedling patches, we derived a formal relationship between the total effective patch size, N_e, and the parental correlations within patches (Q_w^m, Q_w^p and Q_w^{mp}). We define N_e as the reciprocal of q, the probability of coalescence in the previous generation of two randomly sampled genes within a patch: $N_e = 1/2q$. This can be regarded as an instantaneous inbreeding effective patch size, with the factor of two yielding an effective number of diploid individuals, instead of an effective number of genes. It can also be interpreted as the effective number of parents contributing to the seedling patch, i.e. the idealized number of individuals mating and dispersing seeds at random, the descendants of which would exhibit an associated q probability equal to that observed.

To calculate the expectation of q, as a function of parental correlations, we compute the coalescence probabilities of two randomly sampled genes within a patch, depending on whether they came via two female gametes, two male gametes, or one of each. The probability of randomly sampling two maternally inherited genes from a set of diploid individuals is ¼; the probability that they come from the same mother is Q_w^m; the probability that they coalesce, given that they came from the same mother is ½; and the probability that they coalesce, given that they come from different mothers is zero. The joint probability that two randomly sampled genes within a patch are both maternally inherited and that they coalesce in the previous generation is $[Q_w^m/2 + (1 - Q_w^m) \cdot 0]/4 = Q_w^m/8$. Similarly, the joint probability that two randomly sampled genes within a patch are both paternally inherited and that they coalesce in the previous generation is $Q_w^p/8$. Finally, the probability that two randomly sampled genes from a diploid population are one maternally and one paternally inherited is ½ and the probability that they come from the same individual is Q_w^{mp}; using similar logic to that above, the joint probability that two randomly sampled genes within a patch are one paternally and one maternally inherited and that they coalesce in the previous generation is $Q_w^{mp}/4$. Adding the three possible ways of inheritance of two genes that coalesce in the previous generation, we have

$$q = \frac{Q_w^m}{8} + \frac{Q_w^p}{8} + \frac{Q_w^{mp}}{4} = \frac{Q_w^m + Q_w^p + 2Q_w^{mp}}{8}, \qquad (eqn\ 1)$$

and thus

$$N_e \equiv \frac{1}{2q} = \frac{4}{Q_w^m + Q_w^p + 2Q_w^{mp}}, \qquad (eqn\ 2)$$

which represents the desired relationship between effective patch size (or effective number of parents) and maternal and paternal correlations.

In the particular case that cross-parental correlations within patches are negligible ($Q_w^{mp} \approx 0$), we would have

$$N_e \approx \frac{4}{Q_w^m + Q_w^p} = \frac{4N_{em}N_{ep}}{N_{em} + N_{ep}}. \qquad (eqn\ 3)$$

This formula is, without cross-parental correlations, our effective parental number N_e and it exhibits the same relationship with N_{em} and N_{ep} as the classical relationship between effective population size and effective numbers of the two sexes (Wright 1931). Note that estimates of N_e obtained using eqn 3 (i.e. when $Q_w^{mp} \approx 0$) may (in principle) differ substantially from those expected from eqn 2. As an example, if $Q_w^m = 0.5$, $Q_w^p = 0.1$ and $Q_w^{mp} = 0.05$, eqn 2 would predict N_e = 5.7, while eqn 3 would overestimate this number by 17% (with N_e = 6.7). The overestimation of N_e would reach 33% for $Q_w^{mp} = 0.1$ and the same values of Q_w^m and Q_w^p.

Estimation of parentage correlations from genetic data

Given the combined pericarp-leaf assay, we used the maternally inherited diploid pericarp genotypes and extracted the paternal haplotype of each seedling, which also allowed us to decompose the diploid leaf genotype of each seedling into its two gametic phases, the corresponding maternal and paternal haplotypes received from its parents, which we can use to estimate the cross-parental correlations Q_w^{mp} and Q_b^{mp}. Following logic analogous to that in Hardy *et al.* (2004), we estimated parental correlations by averaging kinship coefficients (F_{ij}) among particular pairs (i, j) of seedling haplotypes or genotypes. As in Hardy *et al.* (2004), we used the kinship coefficient given by Loiselle *et al.* (1995), using the target sample (i.e. canopy and open patches) as the reference population to estimate allelic frequencies. Note that this method assumes that parents are unrelated. If parents were related, kinship coefficients would tend to overestimate parental correlations. In particular, we estimated:

- Q_w^m: as twice the average F_{ij} over all the within-patch maternal pericarp genotype pairs.
- Q_w^p: as twice the average F_{ij} over all the within-patch paternal haplotype pairs.

- Q_b^m: as twice the average cross-patch F_{ij} between pairs of maternal pericarp genotypes (one from each patch).
- Q_b^p: as twice the average cross-patch F_{ij} between pairs of paternal haplotypes (one from each patch).
- Q_w^{mp}: as twice the average cross-parent F_{ij} between maternal–paternal gametic phase pairs of seedling within patches. That is, since we need to estimate the average probability that the father of one seedling is the mother of a second seedling, we compute twice the average F_{ij} between the maternal gametic phase of a seedling and the paternal gametic phase of another (and vice versa).
- Q_b^{mp}: as twice the average cross-patch, cross-parent F_{ij}-between maternal–paternal gametic phase pairs (one from each patch). That is, we compute twice the average F_{ij} between the paternal gametic phase of a seedling (from one patch) and the maternal gametic phase of another (from another patch).

To estimate N_e(patch), we simply inserted the genetic estimates of Q_w^m, Q_w^p, Q_w^{mp} into eqn 2.

Relationship between F_{ST}*-statistics and effective parental numbers*

An alternative procedure to estimate N_{em} and N_{ep} from the combined pericarp-leaf assay would be, following Austerlitz & Smouse (2001), to perform an analysis of molecular variance (Excoffier *et al.* 1992) to compute F_{ST}(maternal) on maternal pericarp genotypes and F_{ST}(paternal) on the extracted paternal haplotypes. Assuming noninbred and unrelated parents, as we also did for the kinship-coefficient approach, we have the following expectations (Austerlitz & Smouse 2001):

$$F_{ST}(\text{maternal}) = \frac{Q_w^m - Q_b^m}{2 - Q_b^m} \qquad (eqn\ 4)$$

$$F_{ST}(\text{paternal}) = \frac{Q_w^p - Q_b^p}{2 - Q_b^p}. \qquad (eqn\ 5)$$

If parental correlations among patches were negligible ($Q_b^m \approx Q_b^p \approx 0$), we could then obtain $N_{em} = 1/Q_w^m \approx 1/2F_{ST}$ (maternal) and $N_{ep} = 1/Q_w^p \approx 1/2F_{ST}$ (paternal). Note that this alternative approach, as opposed to that based on kinship coefficients, relies on ignoring parental correlations between patches; moreover, it does not allow estimating either the latter nor cross-parental correlations within and among patches.

Finally, measuring genetic differentiation among seedling patches, using diploid progeny genotypes derived from seedling leaf tissue, F_{ST}(progeny), might be useful to estimate the effective patch size N_e when lacking a combined pericarp-leaf genetic assay. The expected differentiation among patches can be written (Slatkin 1991):

$$F_{ST}(\text{progeny}) = \frac{f_s - f_t}{1 - f_t} = 1 - \frac{h_s}{h_t}, \qquad (eqn\ 6)$$

where f_s and f_t are the probabilities of identity in state (IIS) of two genes within and among patches, respectively, and where $h_s = 1 - f_s$ and $h_t = 1 - f_t$ are the gene diversities within and among patches, respectively. We will compute here the probabilities of non-IIS of two randomly sampled genes within a patch as a function of parental correlations and, as we did before for coalescence probabilities, we will consider separately whether they came via two female gametes, two male gametes, or one of each. The probability of randomly sampling two maternally inherited genes from a set of diploid individuals is ¼; the probability that they come from the same mother is Q_w^m; the probability that they are different given that they come from the same mother is $h_i/2$, with $h_i = 1 - f_i$, where f_i is the probability of IIS of two genes within an individual; and the probability that they are different given that they come from different mothers is $h_p = 1 - f_p$, with f_p being the probability of IIS of two genes from the whole population. We discover that the joint probability that two randomly sampled genes within a patch are both maternally inherited and that they are not IIS is $[Q_w^m h_i/2 + (1 - Q_w^m)h_p]/4$. Analogously, the joint probability that two randomly sampled genes within a patch are both paternally inherited and that they are not IIS is $[Q_w^p h_i/2 + (1 + Q_w^p)h_p]/4$, and the joint probability that two randomly sampled genes within a patch are one paternally and one maternally inherited and that they are not IIS is $[Q_w^{mp} h_i/2 + (1 - Q_w^{mp})h_p]/2$. The gene diversity within a patch is thus

$$h_s = \frac{1}{4}\left[\frac{Q_w^m}{2}h_i + (1 - Q_w^m)h_p\right] + \frac{1}{4}\left[\frac{Q_w^p}{2}h_i + (1 - Q_w^p)h_p\right] + \frac{1}{2}\left[\frac{Q_w^{mp}}{2}h_i + (1 - Q_w^{mp})h_p\right]. \qquad (eqn\ 7)$$

If we assume that adult individuals are not inbred, we have $h_i = h_p$ and eqn 7 simplifies to

$$h_s = h_p\left[\frac{1}{4}\left(1 - \frac{Q_w^m}{2}\right) + \frac{1}{4}\left(1 - \frac{Q_w^p}{2}\right) + \frac{1}{2}\left(1 - \frac{Q_w^{mp}}{2}\right)\right]. \qquad (eqn\ 8)$$

Using similar reasoning for h_s, we obtain an expression for gene diversity among patches:

$$h_t = h_p\left[\frac{1}{4}\left(1-\frac{Q_b^m}{2}\right)+\frac{1}{4}\left(1-\frac{Q_b^p}{2}\right)+\frac{1}{2}\left(1-\frac{Q_b^{mp}}{2}\right)\right] \quad (eqn\ 9)$$

Replacing eqns 8 and 9 in eqn 6 and doing some algebra yields:

$$F_{ST}(\text{progeny}) = \frac{Q_w^m + Q_w^p - Q_b^m - Q_b^p + 2Q_w^{mp} - 2Q_b^{mp}}{8 - Q_b^m - Q_b^p - 2Q_b^{mp}}. \quad (eqn\ 10)$$

In the special case that cross-parental and parental correlations among patches are negligible ($Q_b^m \approx Q_b^p \approx Q_b^{mp} \approx 0$), we would have

$$\frac{1}{2F_{ST}(\text{progeny})} \approx \frac{4}{Q_w^m + Q_w^p + 2Q_w^{mp}} = N_e. \quad (eqn\ 11)$$

That is, F_{ST}(progeny) would approximate the probability of coalescence in the previous generation of two genes within a patch under the assumptions of: (i) no adult inbreeding and (ii) negligible parental and cross-parental correlations among patches. Note once more, however, that if we were in the more favourable position of being able to estimate parental correlations within patches (Q_w^m, Q_w^p and Q_w^{mp}) with kinship coefficients, as we will be if a combined pericarp-leaf assay is available, we could actually estimate N_e using eqn 2, without relying on assumption (ii) above. Positive parametric values of parental and/or cross-parental correlations among patches will result in overestimates of N_e using eqn 11.

Computer programs to calculate the parental correlations described in this paper are available from the authors upon request to JJRA. F_{ST}-statistics were computed as weighted averages across loci, using Arlequin 3.1 (Schneider *et al.* 2000).

Results

We found an effective number of males (N_{ep}) that was obviously larger than the effective number of females (N_{em}) contributing to any single patch of dispersed oak seedlings, for both the canopy and open patches (Table 3). At the within-patch level, the estimated maternal correlation (Q_w^m) was 0.879 for canopy patches, indicating that almost 90% of seedling pairs below tree crowns were maternal half-sibs, while the estimated maternal correlation (Q_w^m) for open patch recruits was only 0.365. The estimated paternal correlations Q_w^p (0.123 for canopy patches and 0.065 for open patches, respectively) are substantially lower than the corresponding maternal Q_w^m estimates; recruits were produced by more fathers than mothers for either type of patch. The Q_w^m and Q_w^p estimates translate into N_{em} = 1.137 seed parents and N_{ep} = 8.099 pollen parents for the average canopy patch and into N_{em} = 2.742 and N_{ep} = 15.409 for recruits sampled in the average open patch. Estimates of Q_w^{mp} (0.052 for canopy patches and 0.085 for open patches, respectively) reflect some degree of cross-parental correlation (i.e. the mother of one seedling is the father of a second), within a single patch, interpretable as the result of the physical proximity of the maternal and paternal parents for a given patch. Because the effective numbers of both maternal and paternal parents were higher for open patches, the effective patch sizes (effective numbers of parents contributing to the patches) were higher for open patches (N_e = 6.667; eqn 2) than for canopy patches (N_e = 3.616; eqn 2). The two-sex value of N_e is, as expected, intermediate between those of maternal and paternal contributions, reflecting the impact of both seed and pollen flow on effective parental size of seedling patches.

At the among-patch level, all the estimates (Q_b^m, Q_b^p and Q_b^{mp}) are slightly below zero for the two seedling data sets, which translates into no maternal

Table 3 Summary of the statistics computed to estimate the effective parental number (N_e), and the effective number of mothers (N_{em}) and fathers (N_{ep}) for the two seedling data sets (canopy patch vs. open patch) and for two statistical approaches (kinship coefficient vs. F_{ST})

	Canopy patch	Open patch
Kinship coefficient		
Q_w^m	0.879	0.365
Q_w^p	0.123	0.065
Q_b^m	−0.041	−0.059
Q_b^p	−0.007	−0.014
Q_w^{mp}	0.052	0.085
Q_b^{mp}	−0.026	−0.052
$N_{em} = 1/Q_w^m$	1.137	2.742
$N_{ep} = 1/Q_w^p$	8.099	15.409
$N_e = \frac{4}{Q_w^m+Q_w^p+2Q_w^{mp}}$	3.616	6.667
F_{ST}-statistics		
F_{ST} maternal (pericarp)	0.439	0.197
F_{ST} paternal (pollen)	0.060	0.024
F_{ST} progeny (seedling)	0.137	0.087
$Q_w^m \approx 2F_{ST}$	0.878	0.394
$Q_w^p \approx 2F_{ST}$	0.121	0.047
$N_{em} \approx 1/2F_{ST}$	1.139	2.541
$N_{ep} \approx 1/2F_{ST}$	8.289	21.177
$N_e \approx \frac{4N_{em}N_{ep}}{N_{em}+N_{ep}}$	4.006	9.076

Q, parentage correlations; w, within patch; b, between patch; m, maternal; p, paternal; mp, cross-parental.

or paternal sharing across patches, as well as no cross-parental correlation. This finding means that we are not seeing any sharing of parents between patches, regardless of their degree of seed dispersal. Thus, seedlings from different patches were respectively dispersed and sired by distinct maternal and paternal sources. This finding suggests that the average inter-patch distances for both data sets (mean inter-canopy patch distance $\approx$ 520 m; mean inter-open patch distance $\approx$ 1085 m) are larger than the scale of dispersal. We cannot discount the remote possibility, however, that estimation errors associated with kinship coefficients might have precluded detecting small positive parametric values of Q_b^m, Q_b^p or Q_b^{mp}. Indeed, since kinship coefficients were computed using the sample allele frequencies as the reference population, kinship estimates between seedling pairs that are less related than average may exhibit negative biases (Robledo-Arnuncio *et al.* 2006), which might result in underestimates of Q_b^m, Q_b^p or Q_b^{mp}, were their parametric value very small.

It is useful to quantify spatial genetic structure and to check the theoretical relationship between parentage correlations and gene diversities within and among patches (eqn 10). The computation of F_{ST}(progeny), based on the diploid genotypes of established seedlings, revealed a higher genetic structure for seedling canopy patches [$\hat{F}_{ST}(\text{progeny}) = 0.137$] than for seedlings in open patches [$\hat{F}_{ST}(\text{progeny}) = 0.087$] (Table 3), confirming the lower level of gene flow into seedling patches located beneath tree canopies suggested by parentage correlations. These observed values matched quite closely the theoretical expectations, as a function of paternal and maternal correlations (eqn 10), which predict F_{ST}-values of 0.146 and 0.095 for canopy and open patches, respectively. Additionally we found higher F_{ST} for seed than for seedling than for pollen, which translates into $N_{em} < N_e < N_{ep}$ for both seedling data sets, the same ranking that was revealed by estimates based on kinship coefficients (Table 3). However, although estimates of effective parental numbers derived from F_{ST} statistics (eqn 4) were quite close to those obtained from kinship coefficients in the case of N_{em}, they were about 40% larger for N_{ep} and N_e in open patches (Table 3).

Overall our results reflect more extensive gene flow by pollen than by seeds into a given patch, and larger numbers of both paternal and maternal sources contributing to open than to canopy patches. In valley oak, despite the fact that pollen movement is relatively extensive, seed dispersal is sufficiently restricted to reduce total effective parental numbers substantially, thus tending to enforce a considerable level of local genetic structure.

Discussion

The findings of this study illustrate why limited seed dispersal has such a strong impact on local genetic structure, despite the fact that valley oak, like many tree species, has the potential for long distance pollen flow. The evidence is compelling because we decompose the genetic structure of a newly recruited seedling population into male and female gametic contributions, using the exact same set of individuals. Given our previous studies of this oak population on pollen flow (Sork *et al.* 2002; Austerlitz *et al.* 2004, 2007; Pluess *et al.* 2009) and seed dispersal (Grivet *et al.* 2005), it is not unexpected to find that our patches of naturally regenerated valley oak seedlings had higher paternal (N_{ep}) than maternal (N_{em}) effective numbers of parents. In fact, the effective number of pollen sources is 6–7 times greater than seed sources for both types of seedling patches. Nonetheless, the low numbers of maternal parents resulted in a relatively low effective parental size of seedling populations, inducing a great deal of genetic structure (as shown by the inter-patch F_{ST}-values). The tradeoff between restricted seed and more extensive pollen dispersal becomes clear from a comparison of eqns 2 and 10, predicting that large proportions of maternal half-sibs will substantially reduce the total effective parental size and increase the spatial genetic structure among seedling patches. Because seed dispersal moves genes from both maternal and paternal sources, the number of maternal parents of a given set of seedlings will necessarily limit the number of paternal parents siring those same seedlings (see below).

Our study compares patches of seedlings found beneath the canopies of trees, representing largely non-dispersed individuals, with those in open patches, composed of dispersed seedlings. Most of the recruits beneath adult trees represent gravity-mediated dispersal, but with a small fraction of seed movement from neighbouring seed parents, via short-distance animal movements. By contrast, recruits in open patches are almost inevitably the result of animal dispersal from diverse seed parents. The important issue is not the difference in the extent of seed dispersal into the two types of patches, but rather the interaction with pollen flow that this comparison allows. In the canopy patches, we find lower values of N_{ep} than in open patches, reflecting the fact that recruits are primarily those from the tree just above, pollinated predominantly by the pool of neighbouring adults (Sork *et al.* 2002; Pluess *et al.* 2009). In contrast, open patch recruits come from multiple maternal seed parents, each pollinated by a relatively distinct (and spatially restricted) set of neighbouring adults, the net result of which is a more

heterogeneous pollen cloud for open patch recruits, and thus a higher N_{ep}.

The effective parental size (N_e) is obviously determined by both male and female gametic dispersal, but seed dispersal also has an impact on male gametic flow, and when seed flow is more restricted, both N_{em} and N_{ep} are reduced. The contribution of male gametic flow to seedling genetic structure is shaped by the extent of seed dispersal and by the number of maternal trees contributing to the seedling cohort. Crawford (1984a, b) emphasizes the distinct contributions of pollen and seed movements, and here we show how seed contribution has a strong impact on the genetic structure and effective parental size of valley oak naturally dispersed recruits.

Ecological implications

Studying naturally dispersed seedlings gives access to the realized gene flow, a measure that may differ from that estimates at the pollen and seed stages because of the impact of various factors influencing seedling establishment (see Burczyk *et al.* 2006). One of the main challenges of Californian valley oak is the lack of regeneration of seedlings and saplings, making it difficult for the populations to maintain their current population sizes (Tyler *et al.* 2006). After germination and during their establishment phase, these seedlings will have to face various threats before reaching the sapling stage and contributing to the next generation of reproductive adult trees: low rainfall, soil compaction, diseases, competition with annuals and grasses, herbivory by mammals (Tyler *et al.* 2006). Although seedlings from canopy patches benefit from both high acorn density and favourable conditions for germination—high humidity being critical for early seedling persistence (Tyler *et al.* 2006)—they are probably not the ones that will survive in the long run, because there are more seedlings in that location than can develop into canopy adults. On the other hand, recruits in open patches experience less favourable conditions for germination, but once established they may more likely contribute to the next generation.

Because long term gene flow may be closer to that observed for open than that for canopy patches, it is worthwhile to focus on these postdispersal seedling recruits. In our study, we found a relatively high level of spatial genetic divergence at the seedling stage. That pattern should evolve by the time individuals reach reproductive maturity, but we should still expect some residual signature of this early input (see Dutech *et al.* 2005). The effective parental size (N_e) determines the local rate of genetic drift: as N_e decreases, local genetic differentiation among patches increases as does the opportunity for local drift. It also determines the magnitude of local inbreeding and/or the potential for kin competition within a patch. In the presence of genetic incompatibilities, N_e should also influence mate availability within patches once seedlings become adults. For valley oak, the N_e of patches is small and future studies might explore the fitness consequences of this pattern.

Kinship- vs. F-statistics

Our approach based on kinship coefficients potentially allows estimating all considered parentage correlations, i.e. parental correlations within and between patches, as well as cross-parental correlations within and between patches, making possible an accurate estimation of the effective number of parental sources (N_e, N_{em}, N_{ep}). However, measuring genetic differentiation (F_{ST}) among seedling patches using diploid progeny genotypes can be a useful approach to estimate the effective patch size, in the absence of a combined pericarp-seedling genetic assay, though only if it is safe to assume that there are no parental correlations between patches. Individual estimates of effective parental numbers obtained with both methods (kinship coefficients and F_{ST}-statistics) were quite similar for canopy patches but not for open patches, a discrepancy that could be due to additional simplifying assumptions when estimating effective parental numbers from the F_{ST}-based approach (among-patch parental correlations are ignored, from which we expect positive biases in the estimation of N_e) and/or to potentially different effects on each method of the higher rate of missing data and the smaller sample size for open patches. Note that slightly negative estimates of among-patch parental correlations do not necessarily imply that their true value is zero, because of potential estimation errors via kinship coefficients. Finally, it could simply be that the observed discrepancy is not statistically significant. We note that future studies should address the nontrivial problem of obtaining reliable uncertainty measures (i.e. accurate estimates of the variance and confidence intervals) for parental correlations and effective parental numbers obtained from kinship coefficients and F_{ST}-statistics, respectively.

Wright's neighbourhood size (N_b) vs. 'effective parental size' (N_e)

The novel approach we propose here computes the probabilities of parental identity and effective parental size (N_e) among a limited number of individuals from small discrete patches and over a single generation. By contrast, Wright's neighbourhood size (N_b) is defined as a function of 'the probability of identity of two uniting gametes' in an infinite, continuous and uniform

population, the inverse of which takes the form $N_b = 4\pi\sigma^2 d$ (with σ^2: parent–offspring axial dispersal variance and d: population density) under Gaussian dispersal only (Wright 1946). Note that under non-Gaussian dispersal, N_b does not measure any probability of identity, though it determines the increase of genetic differentiation with distance for any dispersal function (Rousset 2004). The latter formal relationship has been used to estimate N_b and the dispersal variances indirectly from spatial genetic structure data by regressing the relatedness coefficient between pairs of individuals on the logarithm of geographic distance (Rousset 2000; Vekemans & Hardy 2004). For an isolation-by-distance model in a two-dimensional landscape, we expect a log-linear regression, the slope of which provides an estimate of $1/4\pi\sigma^2 d$ (Rousset 2000). This indirect approach assumes population equilibrium and yields historical N_b estimates, with the estimated dispersal variance being the asymptotic mean square displacement between gene lineages, which will not generally be the same as the dispersal variance measured over one generation through parentage analysis, especially in populations with demographic heterogeneities (J. J. Robledo-Arnuncio & F. Rousset, unpublished data). On the other hand, Crawford (1984a) proposed a formula that decomposes the averaged squared axial parent–offspring distance into pollen and seed dispersal variances, as $\sigma^2_{axial} = 1/2\sigma^2_p + \sigma^2_s$. Note that these dispersal variance components can be estimated directly by means of paternity and maternity analysis over one generation, but they would translate into contemporaneous estimates of dispersal that should not generally be used directly for Wright's neighbourhood size equation, in estimating the probability of identity of two uniting gametes or the neighbourhood area, since this would only be meaningful under Gaussian dispersal. Thus, while Wright's neighbourhood model characterizes spatial genetic structuring for a continuous, uniform and infinite population in equilibrium under Gaussian dispersal, the approach we present here establishes a theoretical framework to characterize the contemporary parental structure of finite plant demes in terms of probabilities of parental identity and effective parental size, without assumptions on the shape of the dispersal distribution or on the spatio-temporal demographic structure of the population. Consequently, our estimate of effective parental size N_e is not equivalent to the Wright's neighbourhood size N_b.

Conclusions

An important issue for evolutionary biology, conservation biology, and landscape genetics is the relative impact of pollen and seed dispersal on the effective population size and spatial pattern of genetic variation among new recruits. Here, the decomposition of N_e into N_{em} and N_{ep}, using parental correlations computed from kinship coefficients, allows the characterization of naturally dispersed recruits from different perspectives: (i) the effective number of parental sources contributing to a seed patch (N_e), (ii) the relative contributions of maternal and paternal effective gametic flow into a patch (N_{em}, N_{ep}), and (iii) the comparison of kinship vs. genetic structure (F_{ST}) of seedling patches. The original approach we present here not only applies to the specific case of pericarp attached to seedlings, as illustrated with valley oak, but can be extended to any system where maternally inherited tissue coexists with diploid embryo tissue, such as in gymnosperms, for which genotypes of the megagametophyte and embryo allow the estimation of paternal and maternal contributions.

Our findings point out to the importance of the cross-parental correlations: even in cases where these correlations between maternal and paternal contributions cannot be estimated, this factor should be taken into account when interpreting the findings, because neglecting it may overestimate effective size of the progeny population. In the specific case of valley oak, the finding of a very high N_{ep} in the progeny population, especially for the dispersed seedlings (N_{ep}=15.4), suggests that pollen flow is extensive and therefore an important mechanism for genetic connectivity. However, the low values of N_{em} for dispersed seedlings (N_{em} = 2.7) demonstrates that seed dispersal can create genetic bottlenecks because the maternal sources represent a smaller subsample of the adult population. This point is shown by the small N_e for the dispersed seedling population (N_e = 6.7), illustrating the critical role of seed dispersal in shaping genetic structure. If seed dispersal is restricted, more extensive pollen flow may not mitigate against the strong impact of seed flow on the local genetic structure of new recruits, at least over the short term.

Acknowledgments

The authors thank Jeanette Papp (from the UCLA Sequencing and Genotyping Core Facility) for handling the genotypic data, Rachel Buchwalter, Joe Epperson and Kurt Merg for field assistance, staff from the UC Santa Barbara Sedgwick Reserve for technical support, Frederic Austerlitz for early comments on the analytical section, and Douglas Scotfield for providing input files with random missing alleles to test the effect of missing data on the output. The authors also thank Kenichiro Shimatani and one anonymous referee for constructive comments on the manuscript. Special thanks to Etienne Klein for insightful comments that improved the clarity of the analysis substantially. V.L.S. and D.G. were supported by NSF-DEB-0089445 and UCLA. J.J.R was supported by a Ramón y Cajal research fellowship from the Spanish 'Ministerio de Ciencia e

Innovación'. P.E.S. was supported by the US Department of Agriculture and the New Jersey Agricultural Experiment Station (USDA/NJAES-17111) and by NSF-DEB-0514956.

References

Austerlitz F, Smouse PE (2001) Two-generation analysis of pollen flow across a landscape. II. Relation between $\varphi_{\varphi\tau}$, pollen dispersal and interfemale distance. *Genetics*, **157**, 851–857.

Austerlitz F, Dick CW, Dutech C *et al.* (2004) Using genetic markers to estimate the pollen dispersal curve. *Molecular Ecology*, **13**, 937–954.

Austerlitz F, Dutech C, Smouse PE, Davis F, Sork VL (2007) Estimating anisotropic pollen dispersal: a case study in *Quercus lobata*. *Heredity*, **99**, 193–204.

Bacles CFE, Lowe AJ, Ennos RA (2006) Effective seed dispersal across a fragmented landscape. *Science*, **311**, 628.

Bittencourt JVM, Sebbenn AM (2007) Patterns of pollen and seed dispersal in a small, fragmented population of the wind-pollinated tree *Araucaria angustifolia* in southern Brazil. *Heredity*, **99**, 580–591.

Burczyk J, Adams WT, Birkes DS, Chybicki IJ (2006) Using genetic markers to directly estimate gene flow and reproductive success parameters in plants on the basis of naturally regenerated seedlings. *Genetics*, **173**, 363–372.

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

Crawford T (1984b) What is a population? In: *Evolutionary Ecology* (ed. Shorrocks B), pp. 135–173. Blackwell Scientific Publications, Oxford.

Dick CW (2008) New interpretations of fine-scale spatial genetic structure. *Molecular Ecology*, **17**, 1873–1874.

Dow BD, Ashley MV, Howe HF (1995) Characterization of highly variable (GA/CT)(N) microsatellites in the bur oak, *Quercus macrocarpa*. *Theoretical and Applied Genetics*, **91**, 137–141.

Dutech C, Sork VL, Irwin AJ, Smouse PE, Davis FW (2005) Gene flow and fine-scale genetic structure in a wind-pollinated tree species *Quercus lobata* (Fagaceaee). *American Journal of Botany*, **92**, 252–261.

Ellstrand NC, Elam DR (1993) Population genetic consequences of small population size—implications for plant conservation. *Annual Review of Ecology and Systematics*, **24**, 217–242.

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

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

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.

Goudet J (2001) FSTAT: *A program to estimate and test gene diversities and fixation indices (version 2.9.3)*. Available from http://www2.unil.ch/popgen/softwares/fstat.htm.

Grivet D, Smouse PE, Sork VL (2005) A novel approach to an old problem: tracking dispersed seeds. *Molecular Ecology*, **14**, 3585–3595.

Hamilton MB (1999) Tropical tree gene flow and seed dispersal. *Nature*, **401**, 129–130.

Hamrick JL (2004) Response of forest trees to global environmental changes. *Forest Ecology and Management*, **197**, 323–335.

Hardesty BD, Hubbell SP, Bermingham E (2006) Genetic evidence of frequent long-distance recruitment in a vertebrate-dispersed tree. *Ecology Letters*, **9**, 516–525.

Hardy O, González-Martínez SC, Colas B *et al.* (2004) Fine-scale genetic structure and gene dispersal in *Centaurea corymbosa* (Asteraceae). II. Correlated paternity within and among sibships. *Genetics*, **168**, 1601–1614.

Isagi Y, Saito D, Kawaguchi H, Tateno R, Watanabe S (2007) Effective pollen dispersal is enhanced by the genetic structure of an *Aesculus turbinata* population. *Journal of Ecology*, **95**, 983–990.

Kampfer S, Lexer C, Glossl J, Steinkellner H (1998) Characterization of (GA)(n) microsatellite loci from *Quercus robur*. *Hereditas*, **129**, 183–186.

Ledig FT (1992) Human impacts on genetic diversity in forest ecosystems. *Oikos*, **63**, 87–108.

Loiselle BA, Sork VL, Nason J, Graham C (1995) Spatial genetic structure of a tropical understory shrub, *Psychotria officinalis* (Rubiaceae). *American Journal of Botany*, **82**, 1420–1425.

Nakanishi A, Tomaru N, Yoshimaru H, Manabe T, Yamamoto S (2009) Effects of seed- and pollen-mediated gene dispersal on genetic structure among *Quercus salicina* seedlings. *Heredity*, **102**, 182–189.

Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Oddou-Muratorio S, Klein EK (2008) Comparing direct vs. indirect estimates of gene flow within a population of a scattered tree species. *Molecular Ecology*, **17**, 2743–2754.

Pavlik BM, Muick PC, Johnson SG, Popper M (1991) *Oaks of California*. California Press and the California Oak Foundation, Los Olivos.

Petit RJ, Hampe A (2006) Some evolutionary consequences of being a tree. *Annual Review of Ecology, Evolution, and Systematics*, **37**, 187–214.

Pluess AR, Sork VL, Dolan B *et al.* (2009) Short distance pollen movement in a wind-pollinated tree, *Quercus lobata* (Fagaceae). *Forest Ecology and Management*, **258**, 735–744.

Robledo-Arnuncio JJ, Austerlitz F, Smouse PE (2006) A new method of estimating the pollen dispersal curve independently of effective density. *Genetics*, **173**, 1033–1045.

Robledo-Arnuncio JJ, Austerlitz F, Smouse PE (2007) POLDIST: a software package for indirect estimation of contemporary pollen dispersal. *Molecular Ecology Notes*, **7**, 763–766.

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

Rousset F (2004) *Genetic Structure and Selection in Subdivided Populations*. Princeton University Press, Princeton.

Schneider S, Roessli D, Excoffier L (2000) ARLEQUIN: *A software for population genetics data analysis, Ver 2.000*. Genetics and Biometry Lab, Department of Anthropology, University of Geneva, Geneva.

Slatkin M (1991) Inbreeding coefficients and coalescence times. *Genetical Research*, **58**, 167–175.

Smith TB, Bernatchez L (eds) (2008) Evolutionary change in human-altered environments. *Molecular Ecology*, **17**, 1–500.

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

Steinkellner H, Fluch S, Turetschek E *et al.* (1997) Identification and characterization of (GA/CT)(n)-microsatellite loci from *Quercus petraea*. *Plant Molecular Biology*, **33**, 1093–1096.

Stockwell CA, Hendry AP, Kinnison MT (2003) Contemporary evolution meets conservation biology. *Trends in Ecology & Evolution*, **18**, 94–101.

Tero N, Aspi J, Siikamaki P, Jakalaniemi A (2005) Local genetic population structure in an endangered plant species, *Silene tatarica* (Caryophyllaceae). *Heredity*, **94**, 478–487.

Tyler CM, Kuhn B, Davis FW (2006) Demography and recruitment limitations of three oak species in California. *Quarterly Review of Biology*, **81**, 127–152.

Vekemans X, Hardy OJ (2004) New insights from fine-scale spatial genetic structure analyses in plant populations. *Molecular Ecology*, **13**, 921–935.

Wright S (1931) Evolution in Mendelian populations. *Genetics*, **16**, 97–159.

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

Wright S (1946) Isolation by distance under diverse systems of mating. *Genetics*, **31**, 39–59.

Wright S (1969) *Evolution and the Genetics of Populations, Vol. 2. The Theory of Gene Frequencies*. University of Chicago Press, Chicago, IL, USA.

Young A, Boyle T, Brown T (1996) The population genetic consequences of habitat fragmentation for plants. *Trends in Ecology & Evolution*, **11**, 413–418.

Ziegenhagen B, Liepelt S, Kuhlenkamp V, Fladung M (2003) Molecular identification of individual oak and fir trees from maternal tissues of their fruits or seeds. *Trees, Structure and Function*, **17**, 345–350.

D.G. is interested in population genetics and genomics, gene flow and conservation genetics of temperate forest trees. J.J.R.A. is a population geneticist with special interest in dispersal estimation methods and the evolutionary consequences of gene flow. P.E.S. has interests in population genetics and biostatistical modelling. V.L.S. studies the ecology, evolution and the conservation of temperate and tropical trees, especially oaks.
