Using Seedling and Pericarp Tissues to Determine Maternal Parentage of Dispersed Valley Oak Recruits

PETER E. SMOUSE, VICTORIA L. SORK, DOUGLAS G. SCOFIELD, AND DELPHINE GRIVET

From the Department of Ecology, Evolution and Natural Resources, School of Environmental and Biological Sciences, Rutgers University, New Brunswick, NJ 08901-8551 (Smouse); the Department of Ecology and Evolutionary Biology, University of California Los Angeles, Los Angeles, CA (Sork, Scofield, and Grivet); the Department of Plant Physiology, Umeå Plant Science Centre, Umeå University, Umeå, Sweden (Scofield); and the Department of Forest Ecology and Genetics, Center of Forest Research, CIFOR-National Institute for Agriculture and Food Research and Technology, Madrid, Spain (Grivet).

Address correspondence to Peter E. Smouse at the address above, or e-mail: smouse@aesop.rutgers.edu.

Abstract

The spatial pattern of established seedlings yields valuable information about variation in fecundity, dispersal, and spatial structure of distributed recruits, but separating maternal and paternal contributions in monoecious species has been hampered by the "2 parent" problem. It is now possible to determine the maternal parentage of established recruits with genetic assay of maternally derived tissues of the seed or fruit, but the DNA of weathered maternal tissues often yields unreliable genotypes, reducing the practical range of such assay. We develop a mixed assay of seedling and seed (pericarp) tissues and illustrate it with distributed recruits of California valley oak (*Quercus lobata* Née). Detailed analysis indicates correct maternal assignment rates of canopy patch recruits of 56% (seedling assay only) versus 94% (mixed assay). For open patch recruits, maternal assignment rates were less than 50% (seedling assay only) versus 91% (mixed assay). The strategy of choice is to use seedling genotypes to identify a small set of credible parental candidates and then deploy 3–4 well-chosen pericarp/endocarp loci to reduce that list to a single obvious maternal candidate. The increase in the number of recruits available for subsequent analysis is pronounced, increasing precision and statistical power for subsequent inference.

Key words: distributed recruits, maternal assay, microsatellites, Quercus

Seed dispersal is one of the most important demographic and genetic stages in the life of a plant because it is the only opportunity, barring vegetative spread, for colonization of new sites (Howe and Smallwood 1982; Howe and Miriti 2000). It also impacts the subsequent spread of genes across the landscape (c.f. Cain et al. 2000; Jordano and Schupp 2000; Dalling et al. 2002; Holbrook et al. 2002; Gómez 2003; Sork and Smouse 2006), affecting both local and regional patterns of genetic affinity. Notwithstanding its importance, precise delineation of the pattern of seed dispersal has been problematic because we have been severely restricted in our ability to designate the maternal source of dispersed seeds accurately, forcing us to study isolated trees (e.g., Augspurger and Kitajima 1992), or to track seeds by tagging them with materials, such as metal (Sork 1984), thread (Forget 1992), fluorescent dyes (Levey and Sargent 2000), or by radio tracking (Pons and Pausas

2007). Genetic assay of dispersed seedlings has provided limited parental resolution, particularly for monoecious organisms, where one must both identify and separate maternal and paternal parents (Meagher and Thompson 1987). For dioecious organisms, separating the pollen and seed parents of a 2-parent pair is straightforward. For monoecious organisms, the usual practice is to designate the closer candidate as the seed donor (Burczyk et al. 2006; González-Martínez et al. 2006; Hadfield et al. 2006). Although that choice may be attractive (on average), it can be seriously misleading for subsequent inference on both seed and pollen dispersal distributions (e.g., Aldrich and Hamrick 1998; Chapman et al. 2003; Jordano et al. 2007; Wang et al. 2007; Sezen et al. 2009; Moran and Clark 2011).

Recently, developed genetic assay methods for maternally derived tissues have enabled accurate maternal identification

of seeds from seed traps or storage granaries (Godoy and Jordano 2001; Schueler et al. 2003; Ziegenhagen et al. 2003; Grivet et al. 2005; Jones et al. 2005; Garcia et al. 2007; Hanson et al. 2007; Jordano et al. 2007; Scofield et al. 2010), effectively extending traditional paternity analysis (Devlin et al. 1988; Roeder et al. 1989) for pollen dispersal (Devlin and Ellstrand 1990; Adams and Birkes 1991; Adams et al. 1992; Smouse and Meagher 1994; Burczyk et al. 1996, 2002, 2004; Burczyk and Prat 1997; Smouse et al. 1999) to maternity analysis for seed dispersal (Grivet et al. 2005; Jones et al. 2005; Hadfield et al. 2006; Pairon et al. 2006; Robledo-Arnuncio and Garcia 2007; Jones and Muller-Landau 2008; Moran and Clark 2011; Scofield et al. 2011).

For assay of already germinated recruits, however, many field situations compromise the quality of maternal tissue DNA, making it difficult to obtain good quality genotypes, and yielding only partial genotypes for many seeds. Although improved error models may be constructed (Moran and Clark 2011), the accuracy of the resulting maternity analysis suffers in many cases, and we are often forced to exclude large numbers of natural recruits from analysis (e.g., Godoy and Jordano 2001; Grivet et al. 2005; Jones et al. 2005; Garcia et al. 2007). Meanwhile, seedling genotypes of those natural recruits often provide reliable maternal information, which can be used when the seedling and seed remain attached. We should be able to improve maternal inference considerably, if we were to combine seedling with maternal tissue information.

The object of this paper to present a novel method that combines genetic typing of seedling tissue with that of maternally inherited tissue (from the seed or fruit) for a joint maternal analysis of natural recruits. We test the analysis on natural recruits of valley oak (*Quercus lobata* Née), sampled from seed shadows directly beneath the canopies of single maternal adults (henceforth, "canopy patches"), heavily influenced by gravity but with some small-scale secondary (animal vectored) seed movements (Grivet et al. 2009). We then generalize seedling-based maternity analysis to the case where we have maternal tissue–based data from the pericarp and also to the case where we missing genotypes for the maternal candidates. We next conduct the analysis on a subset of these recruits for which maternity can be designated with considerable confidence, introducing random "missing loci" in their pericarps (in silico), to illustrate the impact of partial pericarp assay on maternal analysis. Finally, we extend the analysis to "open patch" recruits, found away from any reproductive adults (Grivet et al. 2009), where pericarp genotypes are more fragmentary. We show that adding even partial pericarp information to seedling information improves maternal inference greatly.

Materials and Methods

Naturally Distributed Recruits

The population of sampled recruits was located in the Figueroa Creek valley of Sedgwick Reserve, managed by the University of California at Santa Barbara, as part of the University of California Natural Reserve System, and located in the Santa Ynez Valley, Santa Barbara Co, California (34°40′52″N, 120°02′24″W). Sedgwick Reserve consists of oak savanna habitat, dominated by valley oak, *Q. lobata* Née, interspersed with *Q. agrifolia* Née and *Q. douglasii* Hook. & Arn. The study species is *Q. lobata* (valley oak), a California endemic distributed across savannah, woodland, and riparian forest ecosystems (Pavlik et al. 1991).

Canopy Patch Recruits

Valley oak acorns are often dispersed by gravity, resulting in establishment beneath the canopies of maternal adults. In January 2003, Grivet et al. (2009) sampled a series of 21 maternal canopy patches and assayed seedlings (and the pericarps of attached acorns) for 6 microsatellite markers, a total of 399 recruits. A total of 169 recruits provided both a 6-locus seedling and a 6-locus pericarp genotype, the latter identical with the maternal genotype, even when the adult above the recruit was not that parent (Grivet et al. 2009). We obviously know maternity for these 169 recruits, providing a "truth reference" benchmark, against which to compare maternal inference when there are missing genetic loci for the pericarp, which we can randomly insert in silico. The other 230 recruits had one or more pericarp genetic loci that were either missing or problematic. We have used these 399 recruits to answer 2 basic questions: How often can we correctly designate maternity with seedling assay alone? How does the performance of maternal designation improve as we increase the number of assayed pericarp loci? The results will show what can be done with seedling genotypes, with and without the availability of additional pericarp information from these same recruits.

Open Patch Recruits

Acorns are also dispersed away from any adult, being collected and scatter hoarded by western scrub jays (*Aphelocoma californica*) and western gray squirrels (*Sciurus griseus*) and buried in burrows by small rodents, such as the white-footed mouse (*Peromyscus fasciculata*). After searching the entire study site from March–June 2003, we located 5 patches of seedling recruits that had just germinated and matured leaves but that still had attached acorns (Grivet et al. 2009). From 4 of these open patches, we sampled 259 newly established seedlings with attached acorns (we excluded the southernmost patch of Grivet et al. [2009], as we had not exhaustively sampled all the adults in that area). Soil moisture in open patches was noticeably less than that for the canopy patches, the latter shaded by the leafy foliage of the adults above, and desiccation of the seeds is greater in the open. Moreover, the open patch recruits were sampled in the dry summer, whereas the canopy patch recruits were sampled in the moist winter. Probably as a consequence, pericarp assay was more challenging for the open patch recruits, but we used what pericarp information we could obtain to augment the corresponding seedling genotypes and evaluated our ability to designate maternity for these open patch recruits as well.

Microsatellite Analysis

Over a several year period, we have also obtained almost complete sampling and genotyping of valley oak adults for the Figueroa Creek site. The methods for both leaf and pericarp tissue DNA extraction and PCR analyses have been described by Grivet et al. (2005, 2009), and we used the same 6 microsatellite loci (msq4, qpzag1/5, qpzag9, qpZAG36, qrzag11, and qrzag20).

We provide allele frequencies for the 6-locus microsatellite battery for the 352 adults in our sample in Supplementary Appendix Table 1. The 6-locus adult match probability from this battery is 5.52×10^{-7}, the product of the single-locus values at the bottom of Supplementary Appendix Table 1 (data deposited in the Dryad repository: doi:10.5061/dryad.4bm3739j). There are a few missing genotypes for each locus, so the numbers of alleles that have yielded reliable assay are also indicated. We can deal productively with a few missing genotypes in the adult data set.

Seedling-Based Maternal Inference

The larger object of the Valley Oak Study is to elucidate differential parental fecundity and the seed dispersal distribution, as those contribute to the spatial pattern of genetic affinity, but the focus of this paper is maternity analysis per se. The mathematical argument (normally including both differential fecundity and a dispersal kernel) can be much simplified for our limited purposes here. Given a collection of K candidate maternal parents, the total probability of any particular seedling genotype (S_j) is

$$\Pr(S_j) = \sum_{k=1}^{k} \Pr(M_k) \cdot X_{jk}, \qquad (1)$$

where $\Pr(M_k)$ is the usual Hardy–Weinberg equilibrium frequency expected of the kth maternal genotype, and $X_{jk} = \Pr(S_j|M_k)$ is the probability that the kth maternal genotype would produce an offspring with seedling genotype S_j, computed for each seedling and maternal candidate, gene-by-gene ($g = 1, \ldots, G = 6$) as

$$\Pr(M_k) \cdot X_{jk} = \prod_{g=1}^{G} \Pr(M_{k,g}) \cdot \prod_{g=1}^{G=6} X_{jk,g}, \qquad (2)$$

where $\Pr(M_{k,g})$ is the Hardy–Weinberg frequency of the kth maternal candidate's gth genetic locus, and $X_{jk,g}$ is the Mendelian probability of that mother yielding the required seedling genotype for the gth genetic locus of seedling S_j. Reversing the argument, the likelihood that the jth seedling is attributable to the kth maternal candidate, is

$$L_{kj} = L(M_k|S_j) = \frac{\Pr(M_k) \cdot X_{jk}}{\Pr(S_j)} = \frac{\Pr(M_k) \cdot X_{jk}}{\sum_{j=1}^{k} \Pr(M_k) \cdot X_{jk}}. \qquad (3)$$

Given the high resolution of our 6-locus microsatellite battery, most maternal candidates (M_k) will have $X_{jk} = 0$ for most seedlings (S_j), and though few seedlings will achieve a categorically obvious maternal match ($L_{kj} = 1$) with a single maternal candidate, we can expect a small to modest number of credible maternal candidates with $0 < X_{jk} < 1$ for each seedling. We can utilize genetic information that is less than maternally categorical, but for which we are either confident of correct genotyping or can account for common forms of genotyping error with degraded tissue. The bookkeeping is elaborate but is a straightforward application of standard parentage analysis, as encoded in software, such as CERVUS (see Marshall et al. 1998), PATQUEST (Smouse et al. 1999), FAMOZ (Gerber et al. 2003), or NEIGHBOR (Burczyk et al. 2006).

Mixed Seedling–Pericarp Assay

If we sample the seedling recruits for *Quercus* species within 6–9 months of germination, we can usually find the buried acorn in the ground, still attached to the seedling, and that permits assay of maternal tissue. For oaks, the best tissue for that assay is the pericarp. The obvious attraction of adding pericarp analysis is that among those maternal candidates with $L_{jk} > 0$, from the seedling-based analysis, the kth maternal candidate either has a genotype that matches the pericarp ($X_{jk} = 1$) or it does not ($X_{jk} = 0$). With the addition of pericarp genotype, only the maternal candidates with $X_{jk} = 1$ survive in Equation 3, reducing the posterior likelihood of maternity to $L_{kj} = 1$ for the matching maternal candidate and $L_{kj} = 0$ for all other candidates. For the gth locus, the basic strategy is to replace the ($0 < X_{jk,g} < 1$)-value from the seedling with that from the pericarp (either $X_{jk,g} = 0$ or 1). With a complete 6-locus pericarp genotype, we will convert all seedling likelihood values ($0 < L_{kj} < 1$) to either ($L_{kj} = 0$, for non-maternal individuals) or ($L_{kj} = 1$, for the correct maternal candidate) and can reduce the list of maternal candidate from a small number of individuals to one.

The field reality, however, is that many of the pericarps are missing one or more loci. For some fraction of the recruits, we will achieve categorical designation of the maternal parent. For others, we are left with a reduced set of maternal candidates but without an obvious designee. At worst, however, we will have sharpened the posterior likelihood resolution on those remaining candidates, perhaps enough to justify a compelling likelihood-based assignment.

Null Alleles in the Pericarp

An additional issue that arises, as a result of degraded pericarp tissue, is the presence of null alleles (Dakon and Avise 2004; Scofield et al. 2010), wherein a single allele at a locus fails to amplify. With null alleles, the pericarp genotype appears to be homozygous but may actually be heterozygous. We follow the method described in Scofield et al. (2010) and assign this event a small probability $\Pr_{null}$, which in practice may differ among loci, depending on amplification conditions. Imagine that at the C-locus, the pericarp is homozygous ($P_j = C_h C_h$), whereas the maternal candidate is heterozygous ($M_k = C_h C_i$), with one maternal

allele matching the single pericarp allele. Denote the probability that the genotype observed in the pericarp is actually heterozygous by $\mu_N = \text{Pr}_{\text{null}}$; then the probability that the pericarp is both heterozygous and carrying the maternal C_i allele is $\mu_N \cdot q_i$. Without genotyping error, the forward probability of the pericarp–maternal candidate match is $X_{jk,C} = 0$; with genotyping error, we instead assign a (small) forward probability of a match, $X_{jk,C} = \mu_N \cdot q_i$. Alternatively, imagine that both the pericarp and the maternal candidate are homozygous and have matching genotypes at the C-locus ($P_j = C_h C_h = M_k$). There is a μ_N probability that the pericarp has a null allele and a $\mu_N \cdot (1 - q_h)$ probability that the missing allele is "not" the matching allele in the maternal candidate. With genotyping error, the probability of the pericarp–maternal match is $X_{jk,g} = 1 - \mu_N \cdot (1 - q_h)$.

The estimate used for μ_N may be obtained by regenotyping loci (Bonin et al. 2004; Scofield et al. 2010), by extension from a related study, or by establishing a baseline expectation, absent further data. The value we used here ($\mu_N = 0.02$) was estimated by regenotyping pericarps from *Q. agrifolia* at the same site (Scofield et al. 2010). By using a null allele error probability of $\mu_N = 0.02$, we penalized the Δ_{kj} likelihood values of candidates with an apparent null allele by ~1.7 (plus a factor due to frequency of the missing allele) for that particular (null) locus, ensuring that only strongly (or uniquely) supported maternal candidates would survive the adjustment. Some studies have found higher rates for other *Quercus* species (e.g., 5% by Moran and Clark [2011]), but the essential logic of the adjustment would be the same.

Missing Adult Genotypes

We also encounter cases where some of the sampled maternal candidates are missing particular marker locus information. To be definite, suppose that the kth maternal candidate is missing a D-locus genotype. If we have a D-locus genotype from the jth seedling, however, and if that maternal candidate is not excluded for that seedling on the basis of other loci, then summing over all possible maternal genotypes, we obtain:

$$X_{jk,D} = q_h^2, \quad \text{if } S_j = D_h D_h, \tag{4a}$$

$$X_{jk,D} = 2q_h q_i, \quad \text{if } S_j = D_h D_i. \tag{4b}$$

If the pericarp P_j is compatible with the maternal candidate for all other loci, then

$$X_{jk,D} = q_h^2, \quad \text{if } P_j = D_h D_h, \tag{5a}$$

$$X_{jk,D} = 2q_h q_i, \quad \text{if } P_j = D_h D_i. \tag{5b}$$

In short, we insert the appropriate Hardy–Weinberg frequencies for missing D-locus genotypes in maternal candidates, using either the seedling or pericarp genotypes from the recruits themselves. That allows us to add an occasional sampled adult for which we have one or more missing microsatellite loci.

Gaging Maternal Resolution

As we add either seedling or pericarp information to the assay battery for the recruit, 2 things should happen. First, more and more maternal candidates should be excluded ($X_{jk} = 0$) as we add genetic loci to the assay battery. Given enough seedling information, there would eventually be no more than 2 residual (maternal and paternal) candidates, though without pericarp genotype, we should allocate maternity equally to each of them for the monoecious case. A certain fraction of the paternal parents will be nonlocal (from off-site) and unsampled, of course, and for those cases, a single surviving local (maternal) parent will be present. Second, as we add genetic loci to the assay battery, the likelihood values of the surviving maternal candidates diverge, and even if the list is not reduced to 2 parental candidates, we can allocate maternity on the basis of relative likelihood values.

Our objective here is to determine how much more powerful is the mixed maternal assay than seedling assay only. We evaluate success both in terms of the numbers of maternal candidates that survive the winnowing process and in terms of their degree of likelihood divergence. We gage the latter by the LOD score ($\log_{10}$ odds ratio) of each candidate, when compared with the most likely candidate,

$$\Delta_{kj} = \log_{10}[L_{bj}(\text{best candidate} \mid R_j)] - \log_{10}[L_{kj}(k\text{th candidate} | R_j)], \tag{6}$$

computed both for seedling assay only ($R_j = S_j$) and for mixed assay ($R_j = P_j \cup S_j$). We tally all maternal candidates whose Δ_{kj} values < 2.0, the collection of candidates that are no more than 100 times less likely than the best candidate, a fairly generous inclusion criterion. We take the view that adults with $\Delta_{kj} > 2.0$ are not credible maternal candidates. An R program that implements this method (R Development Core Team 2011) is available from D.G.S. on request.

Results

The Truth Reference

We began the analysis with a collection of 169 truth reference canopy patch recruits, each of which had both complete 6-locus seedling and 6-locus pericarp genotypes, and for each of which maternity was categorically obvious. For the first trial, we used just the seedling genotypes to conduct a maternity analysis, attempting to assign each of these recruits to 1 of 352 sampled mothers. That assay yielded a collection of possible parents, but we almost never had enough seedling information to exclude all but 2 parental candidates. With seedling data alone, there were an average of 9.54 plausible maternal candidates per recruit (within a $\Delta < 2$ window for inclusion, but we were able to assign 94 of 169 of these maternally obvious canopy patch recruits (55.7%) to the correct maternal candidate (Figure 1), though generally by a small Δ_{kj} margin. In most cases, the 2 most likely candidates were undoubtedly the maternal and
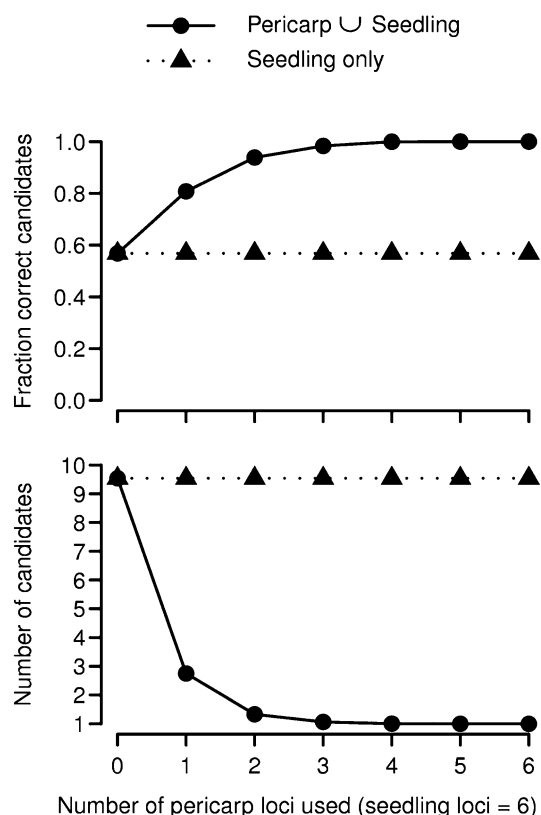
Figure 1. Maternal assignment success using the 169 recruits in the maternally obvious data set. In all cases, we use the full battery of 6 seedling loci. The panels indicate success criteria for all recruits versus the number of pericarp loci employed; the pericarp loci used for each recruit were chosen randomly in 100 independent trials. The top panel shows the mean fraction of recruits for which the correct (and obvious) candidate was chosen. The bottom panel shows the mean number of candidates with $\Delta(R_j) < 2 = [\log_{10}(100)]$, ranging from 10 (seedling only assay) down to 1 (a single high likelihood candidate).

paternal parents, but either could have had the higher likelihood, yielding 50% as the expected correct assignment rate. Earlier studies of pollen flow in this population had established that a small fraction of pollination could be expected from off-site (Sork et al. 2002; Austerlitz et al. 2007), so the correct paternal parent was sometimes not sampled and not listed among the collection of included candidates. That accounts for the fact that correct maternal assignment rate is slightly higher than 50%. The "seedling-only" success rate is the baseline, against which we compare all other assay combinations.

We then randomly added one of the 6 pericarp loci for each recruit and compared our maternal assignment success with that of the 6-locus seedling genotype only, using 100 random trials. We next added 2 random pericarp loci to the seedling genotype, a treatment again replicated 100 times. Subsequent treatments involved adding 3, 4, and 5 random pericarp loci to the 6-locus seedling genotype, each

replicated 100 times. For the final treatment, we added all 6 pericarp loci to all 6 seedling loci. The mixed assay results are shown in Figure 1, for contrast with the 6-locus seedling reference treatment, and indicate that maternal resolution rises steadily and quickly as we add pericarp loci to the seedling battery. As few as 3 pericarp loci added to the seedling battery provide almost 100% correct maternal designation and 4 loci is sufficient to provide categorical designation (Figure 1).

Not all the assignments were to the candidate immediately above the canopy patch, but all were assigned to adults quite local to the canopy patch, probably due to some secondary animal-vectored movement of acorns between neighboring canopy patches. Not surprisingly, maternal designation was virtually categorical where the pericarp data were complete and reliable because the Pr(multilocus random match) = 5.52×10^{-7}, the 6-locus product of the PI values on the bottom line of Supplementary Appendix Table 1. The more important observation, however, is that adding as few as 3–4 pericarp loci to the seedling genotype is sufficient to convert a modest likelihood preference for one or both parents into an unambiguous designation of maternity.

Seedling genotypes obviously provide valuable maternal information but are not maternally diagnostic. Among the 169 truth reference recruits, the average best maternal candidate exhibited a Δ(best–second best)-value of 1.09 (at least 12:1 odds in favor of the best) but was compatible with 9.54 credible maternal candidates within a $\Delta_{kj} < 2.0$ interval (less than 100:1 odds in favor of the best). As random pericarp loci were added progressively to the assay battery, however, the number of maternal candidates included in the $\Delta < 2$ window decreased steadily to 1.0, the single obvious candidate (Figure 1).

Canopy Patch Recruits with Incomplete Genotypes

On the strength of the analyses with the truth reference recruits (Figure 1), we can expect progressive improvement of seedling-based maternal assay as we include progressively more pericarp loci. The remaining (230) canopy patch recruits had missing seedling loci or pericarp loci, or both, and some recruits had single pericarp loci that were problematic, excluding all possible maternal candidates. This "field reality" is typical of many studies of this kind (c.f. Grivet et al. 2005, 2009; Jones et al. 2005; Bacles et al. 2006; Garcia et al. 2007). For such real world data sets, it is not uncommon to exclude a large fraction of the recruits from analysis on the basis of such "genetic assay problems," but if we could "rescue" these recruits with mixed tissue assay, we could increase our recruit sample sizes substantially.

To assess how much maternal resolution was available for such recruits, we evaluated mixed assay performance for these 230 partial data recruits, grouped by numbers of seedling loci available, and tallied the average number of pericarp loci available for assay. For both seedling-only (S) and mixed tissue ($P \cup S$) assay, we have indicated the average numbers of included maternal candidates ($\Delta_{kj} < 2$)

Table 1 Mixed maternal analysis for 399 Q*uercus lobata* canopy and 259 open patch recruits

Recruit tally	Number of loci		Average number included		Δ(best–second best)		Number assigned		Not assigned
	S	P	S-Only	P ∪ S	S-Only	P ∪ S	P ∪ S	One mismatch	
Maternally obvious recruits—complete (6-locus) seedling and pericarp genotypes									
169	6	6	9.54	1.00	1.09	7.89	169	0	0
Canopy patch recruits—incomplete seedling or pericarp genotypes									
117	6	4.85	19.21	0.39	1.00	5.80	44	59	14
79	5	4.89	16.47	0.73	0.69	6.50	58	15	6
16	4	4.69	35.31	0.88	0.26	5.67	13	2	1
18	1–3	5.00	67.22	0.83	0.28	6.50	11	4	3
Open patch recruits—very incomplete pericarp genotypes									
237	6	2.81	9.71	0.96	1.21	4.71	149	66	22
17	5	2.41	19.59	0.94	0.44	1.73	11	6	0
5	3–4	2.20	54.80	4.20	0.50	1.88	2	3	0

There were 169 canopy patch recruits with complete seedling (*S*) and pericarp (*P*) genotypes and a single maternal match, and 230 canopy patch recruits with missing loci or lacking a maternal match. The open patch recruits had similar characteristics. These data sets are grouped by numbers of seedling (*S*) and pericarp (*P*) loci. For seedling (*S*-only) and joint (*P* ∪ *S*) assay, the average numbers of maternal candidates included within a log-likelihood window, $\Delta(R_j) < 2 = [\log_{10}(100)]$, of the best candidate, and the average Δ(best vs. second best)-value provide measures of maternal likelihood resolution. The numbers of recruits with a specific maternal match are tallied for joint (*P* ∪ *S*) analysis and for a final treatment that removed single mismatched loci. We were able to designate maternity categorically for all but 24 canopy patch and all but 22 open patch recruits with seriously problematic genotypes.

per recruit, the average value of **Δ**(best–second best), as a measure of the confidence with which we can choose among the best candidates as well as the numbers of categorical maternal assignments from the joint assay.

The same trends that were evident for the truth reference recruits (Figure 1) were also evident for all the canopy patch recruits (Table 1). As expected, the overall performance of both seedling and joint assay improved as the numbers of loci increased. As the assay battery improved, the number of included ($\Delta_{kj} < 2$) maternal candidates declined and the value of **Δ**(best–second best) increased. In other words, as the assay battery improved, so did our ability to make a definitive maternal designation. As above, seedling-only assay (*S* columns) was not generally sufficient to render maternal analysis categorical, but when we augmented seedling genotypes with 3–4 pericarp loci, we usually rendered maternal designation all but categorical. Mixed assay often reduced the numbers of included ($\Delta_{jk} < 2$) maternal candidates to the best and second best (evidently the maternal and paternal parents), and more often than not, there was no candidate (other than the obvious maternal choice) within the ($\Delta(R_j) < 2$) window of inclusion (Table 1, columns labeled *P* ∪ *S*).

Single-Locus Mismatches

Even with mixed tissue assay, 104 canopy patch recruits did not have an assignable maternal candidate. A detailed comparison of these more difficult recruits with the genotypes of the 352 genotyped adults showed that in 78 of these 104 cases, the seedling was compatible with a single maternal candidate, but one of the pericarp loci did not match. If we removed that single pericarp locus (as for the truth reference) and used only the seedling genotype for that locus, the recruit was only compatible with that single maternal candidate, and we accepted that maternal designa-

tion. For 2 recruits, the pericarp loci we did have were compatible with more than one maternal candidate. Using the seedling loci we had for those same loci instead resulted in exclusion of all maternal candidates. By removing one of the mismatching seedling loci, we could identify a single compatible maternal candidate, and we accepted that choice. In the end, we were left with 24 recruits that required multiple removals to find a maternal candidate. These could either be from outside the area of maternal sampling or genetic assay unreliable, so rather than "pick and choose" among the available data, searching for a plausible maternal candidate, we chose to treat them as "not assignable." By allowing for single-locus mismatches of the pericarp or seedling, which we view as both compatible with the spirit of "pericarp rescue" and preferable to either removing them from consideration or allocating them arbitrarily to inflow from outside the study area, we were able to resolve maternity for all but 24 of 399 total progeny (6%) from the canopy patches. Our final measure of success was (169 pericarp-obvious + 126 mixed-assay compelling + 80 single-locus mismatches) equal to 375 with assigned mothers, of 399 recruits (94%). Though the dispersal kernel was not a consideration in these analyses, all these maternal assignments were to adults that were either just above the recruit or to a neighboring adult.

Open Patch Recruits

We also analyzed 259 natural recruits from 4 open patches from the same (2002) mast seeding episode in the Figueroa Creek study area. These large open areas were bordered by genotyped adults that represented the bulk of the more obvious maternal candidates. Pericarp assay in open patches was particularly challenging, however, because the buried acorn pericarp was decayed, and the DNA was somewhat degraded. This difference in DNA degradation in pericarps

from open sites could be due to different moisture regimes for the buried acorns or to the longer period of exposure of the open patch recruits, or both, but the frequency of the problem underscores the need to incorporate a provision for genotyping error into the assignment procedure. To illustrate, the *qrzag11* locus, though expressing well in DNA from seedling leaf tissue, could not always be reliably scored for pericarp tissue, particularly for open patch pericarps. We did include *qrzag11* as 1 of the 6 seedling markers to minimize the number of maternal candidates, but we did not use it for the pericarp DNA assays. We present the results of mixed assay for the open patch recruits in the lower panel of Table 1, for comparison with the canopy patch recruits. In general, there were fewer readable pericarp loci for open ($2.77/5 = 55.4\%$) than for canopy patches ($5.34/6 = 89.1\%$), but the numbers of readable seedling loci were roughly comparable for open ($5.88/6 = 98.0\%$) and canopy patches ($5.56/6 = 92.7\%$), respectively. A total of 97 of the 259 open patch recruits (37.5%) were challenging, even with mixed assay, a higher fraction than for the crown patches (104 of 399, 26.1%).

The same general patterns were evident for open patch as for canopy patch recruits. As we added genetic loci to the battery, the numbers of maternal candidates within the $\Delta < 2$ window decreased and the Δ value of the best candidate increased, relative to the second best. Removing single mismatching loci resolved maternity for 75 of the 97 difficult recruits (58 via removal of a single pericarp locus and 17 more via removal of a single seedling locus). Our final success rate was thus (126 pericarp-obvious + 36 mixed-assay compelling + 75 single-locus mismatch) equal to 237 of the 259 open patch recruits (91.5%). Open patch assay is more challenging than canopy patch assay, but mixed assay is consistently an effective rescue strategy.

Discussion

Optimizing the Mixed Assay Protocol

Before the development of maternal tissue assay of seeds and fruit, genetic maternity analysis of already-distributed recruits was based on seedling genotypes, and the best that could be done for the monoecious case was to choose the most likely pair of candidates and assume that the closer of the 2 was the mother (e.g., Bacles et al. 2006; Gonzalez-Martínez et al. 2006). Although that assumption may be reasonable for many cases, there are species and situations for which the inference on both the maternal and paternal dispersal kernels will be erroneous (e.g., Aldrich and Hamrick 1998; Chapman et al. 2003; Jordano et al. 2007; Wang et al. 2007; Sezen et al. 2009). With the advent of maternal tissue assay, routine practice has shifted to assaying those tissues instead, which has improved our ability to make maternal inference for at least a fraction of the new recruits, but as we have seen here, such assay often entails the elimination of a large fraction of recruits from analysis.

Meanwhile, seedling assay is generally more reliable than is seed assay under challenging field-sampling conditions.

With a seedling battery of no more than 6 microsatellites, we were able to reduce the list of included mothers substantially for any given recruit. We could probably reduce the list of included mothers a bit further by adding more seedling loci and could increase the Δ values of the better candidates, but with little improvement in the fraction of correctly designated maternal parents, inasmuch as the paternal parent is no less likely (genetically) than the maternal parent. The better strategy is to start with a reasonably effective seedling assay and augment with a small number (3 or 4) of reliably assayed maternal tissue loci, which should be sufficient to remove all but the correct maternal candidate from consideration under most circumstances.

Genetically Problematic Recruits

Even with the use of joint seedling and maternal tissue assay, we are typically presented with a nontrivial fraction of recruits for which no maternal candidate is evident, due either to genotyping errors or inflow from off site. Our treatment of null genotyping errors following Scofield et al. (2010) resulted in a notable increase in the number of assigned recruits (Table 1). The null allele error estimate used here ($\mu_N = 0.02$) was taken from regenotyping pericarps from acorns of *Q. agrifolia* from this same study site (Scofield et al. 2010). That rate estimate may not be generic for all oaks (Moran and Clark 2011), but we explored rates ranging from ($0.01 \leq \mu_N \leq 0.04$) and found that the value chosen had virtually no impact on our assignments (results not shown), whereas failure to allow for null allele errors substantially reduced our unambiguous assignment fractions (Table 1).

We have adopted the usual practice of removing single "problematic loci" to assign most of "problematic" recruits to a maternal candidate but that choice warrants a comment. The practical reality is that genetic assay of microsatellite markers, particularly for degraded seed tissues, entails a nontrivial fraction of assay errors, and such problems are not always obvious by inspection; they will (falsely) exclude the maternal parent from consideration (Bonin et al. 2004). Methods have been developed that recognize that erroneous microsatellite alleles may be observed that are similar in allele length to the true alleles and that one can account for such "misreads" when assigning parents (Hadfield et al. 2006; Moran and Clark 2011). Our approach, by contrast, takes advantage of the greater data available within a mixed tissue assay framework but is analytically more conservative. We discard genotype information for problematic pericarp loci altogether and replace them with the corresponding information for those loci available from the seedling. We showed that removal of a single problematic pericarp locus revealed an obvious maternal candidate in most cases, with which the seedling was compatible.

For a few "motherless recruits," a single seedling locus was incompatible with an otherwise (5-locus categorical) maternal candidate, with which the partial pericarp genotypes we did have were a perfect match. In those few

cases also, we chose to suppress that single seedling locus and assigned the recruit to that "obvious" maternal candidate. The maternal inclusion probability is $<10^{-6}$ for 6 loci and $<10^{-5}$ for 5 loci, so our results are analytically secure. We have refrained from removing both pericarp and seedling genotypes from the same locus and from suppressing 2 loci, though the additional data available from mixed assay provide some room for further maneuver. The analytic approach can be extended to allow for more complex error models, including multiple-locus errors, but even with simple error models, we achieve a substantial increase in maternal resolution.

Variations on a Theme

We have used seedling and maternal tissue genotypes for maternal assay here, but the utility of mixed assay for parentage analysis can be extended. Having used pericarp assay to determine the maternal parent, Grivet et al. (2009) combined pericarp-only maternal assay with seedling assay to infer the paternal gamete for pollen analysis. Gonzales et al. (2006) combined maternally derived diploid elaiosome tissue from seeds with triploid endosperm assay to infer maternity and deduce the paternal gamete for breeding system analysis. In angiosperms, one might combine diploid nuclear markers with haploid chloroplast markers in both adults and seedlings, rather than using pericarp/endocarp tissues, to accomplish the same objectives. That becomes more attractive as microsatellite markers become more available for chloroplast haplotype designation (cf. Ebert and Peakall 2009). In conifers, on the other hand, paternally inherited chloroplast markers can and have been used in conjunction with nuclear markers to infer paternity and pollen dispersal (e.g. Smouse and Robledo-Arnuncio 2005). The use of paternally or maternally inherited haploid genomes that are even modestly variable will provide valuable information that can only improve parental determination, when combined with multilocus nuclear genotypes. Mixed assay could also be used with embryonic leaf tissue and seed genotypes collected from seed traps or granaries, coupled with geolocation of maternal parents, we can improve our resolution on the purely spatial aspects of dispersal. The essential genetic and statistical principles underlying parental designation in all of these applications are simple extensions of classic parentage analysis.

Inferential Extension

There are very substantial challenges to seedling establishment (Tyler 2006) in valley oak, and evaluating seed dispersal after germination may involve a variety of differential genetic survival effects that can confound the measurement of dispersal per se (e.g., Janzen 1970; Connell 1971; Nathan and Cassagrandi 2004). Some of these effects last into adulthood, decades later (c.f. Dutech et al. 2005). For valley oak, we are most interested in describing the "effective dispersal" of successful recruits, plausible substrate for a series of consequential questions on the impact of variable maternal fecundity, translating into unequal reproductive contributions to the pool of new recruits, the impact of spatially asymmetric pollen and seed dispersal on the resulting pattern of male and female parental contributions to the genetic diversity among new recruits (Grivet et al. 2009; Scofield DG, unpublished data), as well as the relative impacts of seed and pollen dispersal on the spatial patterns of that genetic diversity, scattered patchily across micro- and mesoscale landscapes (Sork VL, unpublished data).

The larger payoff from improving our maternal inference will come when we translate these very substantial gains in available sample size for new recruits into inference about the processes determining the level and spatial patterns of recruitment. We have here almost doubled our sample sizes with mixed assay. That should not change the nature of broader inference on recruitment processes, but it has already increased both our precision and our statistical power on all the parameters of derivative interest. Many studies are beset with difficult genetic assay constraints under field conditions, and they should profit from similar analytical treatment.

Supplementary Material

Supplementary material can be found at http://www.jhered.oxfordjournals.org/.

References

Adams WT, Birkes DS. 1991. Estimating mating patterns in forest tree populations. In: Fineschi S, Malvolti M, Cannata F, Hattermer H, editors. Biochemical markers in the population genetics of forest trees. The Hague (the Netherlands): SPB Academic Publishing. p. 157–172.

Adams WT, Griffin AR, Moran GG. 1992. Using paternity analysis to measure effective pollen dispersal in plant populations. Am Nat. 140:762–780.

Aldrich PR, Hamrick JL. 1998. Reproductive dominance of pasture trees in a fragmented tropical forest mosaic. Science. 281:103–105.

Augspurger CK, Kitajima K. 1992. Experimental studies of seedling recruitment from contrasting seed distributions. Ecology. 73:1270–1284.

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.

Bonin A, Bellemain E, Broken Eidesen P, Pompanon F, Brochmann C, Taberlet P. 2004. How to track and assess genotyping errors in population genetic studies. Mol Ecol. 13:3261–3273.

Burczyk J, Adams WT, Birkes DS, Chybicki IJ. 2006. Using genetic markers to directly estimate gene flow and reproductive success parameters in plants based on naturally regenerated seedlings. Genetics. 193:363–372.

Burczyk J, Adams WT, Moran GF, Griffin AR. 2002. Complex patterns of mating revealed in a *Eucalyptus regnans* seed orchard using allozyme markers and the neighbourhood model. Mol Ecol. 11:2379–2391.

Burczyk J, Adams WT, Shimizu J. 1996. Mating patterns and pollen dispersal in a natural knobcone pine (*Pinus attenuata* Lemmon) stand. Heredity. 77:251–260.

Burczyk J, Lewandowski A, Chalupka W. 2004. Local pollen dispersal and distant gene flow in Norway spruce (*Picea abies* L. Karst.). For Ecol Manag. 197:39–48.

Burczyk J, Prat D. 1997. Male reproductive success in *Pseudotsuga menziesii* (Mirb.) Franco: the effect of spatial structure and flowering characteristics. Heredity. 79:638–647.

Cain ML, Milligan BG, Strand AE. 2000. Long-distance seed dispersal in plant populations. Am J Bot. 87:1217–1227.

Chapman CA, Chapman LJ, Vulinee K, Zanne A, Lawes MJ. 2003. Fragmentation and alteration of seed dispersal processes: an initial evaluation of dung beetles, seed fate, and seedling diversity. Biotropica. 35:382–393.

Connell JH. 1971. On the role of natural enemies in preventing competitive exclusion in some marine animals and in rain forest trees. In: den Boer PJ, Gradwell GR, editors. Dynamics of Populations: Proceedings of the Advanced Study Institute on Dynamics of Numbers in Populations, Ooseterbeek, 1970. Wageningen (the Netherlands): Centre for Agricultural Publishing and Documentation. p. 298–310.

Dakon EE, Avise JC. 2004. Microsatellite null alleles in parentage analysis. Heredity. 93:504–509.

Dalling JW, Muller-Landau HC, Wright SJ, Hubbell SP. 2002. Role of dispersal in the recruitment limitation of neotropical pioneer species. J Ecol. 90:714–727.

Devlin B, Ellstrand NC. 1990. The development and application of a refined method for estimating gene from angiosperm paternity analysis. Evolution. 44:248–259.

Devlin B, Roeder K, Ellstrand NC. 1988. Fractional paternity assignment: theoretical development and comparison to other methods. Theor Appl Genet. 76:369–380.

Dutech C, Sork VL, Smouse PE, Davis F. 2005. Relationships between fine-scale genetic structure and gene flow in a wind pollinated tree species, *Quercus lobata* Née. Am J Bot. 92:252–261.

Ebert D, Peakall R. 2009. Chloroplast simple sequence repeats (cpSSRs): technical resources and recommendations for expanding cpSSR discovery and applications to a wide array of plant species. Mol Ecol Resour. 9:673–690.

Forget PM. 1992. Seed removal and seed fate in *Gustavia superba* Lecythidaceae. Biotropica. 24:408–414.

Garcia C, Jordano P, Godoy JA. 2007. Contemporary pollen and seed dispersal in a *Prunus mahaleb* population: patterns in distance and direction. Mol Ecol. 16:1947–1955.

Gerber S, Chabrier P, Kremer A. 2003. FAMOZ: a software for parentage analysis, using dominant, codominant and uniparentally inherited markers. Mol Ecol Notes. 3:479–481.

Godoy JA, Jordano P. 2001. Seed dispersal by animals: exact identification of source trees with endocarp DNA microsatellites. Mol Ecol. 10:2275–2283.

Gómez JS. 2003. Spatial patterns in long-distance dispersal *of Quercus ilex* acorns by jays in a heterogeneous landscape. Ecography. 26:573–584.

Gonzales E, Hamrick JL, Smouse PE, Dyer RJ. 2006. Pollen mediated gene dispersal within continuous and fragmented populations of a forest understory species, *Trillium cuneatum*. Mol Ecol. 15:2047–2058.

González-Martínez SC, Burczyk J, Nathan R, Gil L, Alía R. 2006. Effective gene dispersal and mother-tree reproductive success in Mediterranean pine (*Pinus pinaster* Aiton). Mol Ecol. 15:4577–4588.

Grivet D, Robledo-Arnuncio JJ, Smouse PE, Sork VL. 2009. Relative contribution of contemporary pollen and seed dispersal to the neighborhood size of a seedling population of California valley oak (*Quercus lobata*, Née). Mol Ecol. 16:3967–3979.

Grivet D, Smouse PE, Sork VL. 2005. A novel approach to an old problem: tracking dispersed seeds. Mol Ecol. 14:3585–3595.

Hadfield JD, Richardson DS, Burke T. 2006. Towards unbiased parentage assignment: combining genetic, behavioural and spatial data in a Bayesian framework. Mol Ecol. 15:3715–3730.

Hanson T, Brunsfeld S, Finegan B, Waits L. 2007. Conventional and genetic measures of seed dispersal for *Dipteryx panamensis* (Fabaceae) in continuous and fragmented Costa Rican rain forest. J Trop Ecol. 23:635–642.

Holbrook KM, Smith TB, Hardesty BD. 2002. Implications of long-distance movements of frugivorous rain forest hornbills. Ecography. 25:745–749.

Howe HF, Miriti MN. 2000. No question: seed dispersal matters. Trends Ecol Evol. 15:434–436.

Howe HF, Smallwood J. 1982. Ecology of seed dispersal. Annu Rev Ecol Syst. 13:201–228.

Janzen DH. 1970. Herbivores and the number of tree species in tropical forests. Am Nat. 104:501–528.

Jones FA, Chen J, Weng G-J, Hubbell SP. 2005. A genetic evaluation of seed dispersal in the neotropical tree *Jacaranda copaia* (Bignoniaceae). Am Nat. 166:543–555.

Jones FA, Muller-Landau HC. 2008. Measuring long-distance seed dispersal in complex natural environments: an evaluation and integration of classical and genetic methods. J Ecol. 96:642–652.

Jordano P, Garcia C, Godoy JA, Garcia-Castano JL. 2007. Differential contribution of frugivores to complex seed dispersal patterns. Proc Natl Acad Sci U S A. 104:3278–3282.

Jordano P, Schupp EW. 2000. Seed disperser effectiveness: the quantity component and patterns of seed rain for *Prunus mahaleb*. Ecol Monogr. 70:591–615.

Levey DJ, Sargent S. 2000. A simple method for tracking vertebrate-dispersed seeds. Ecology. 81:267–274.

Marshall TC, Slate J, Kruuk L, Pemberton JM. 1998. Statistical confidence for likelihood-based paternity inference in natural populations. Mol Ecol. 7:639–655.

Meagher TR, Thompson EA. 1987. Analysis of parentage for naturally established seedlings of *Chamaelirium luteum* (Liliaceae). Ecology. 68:803–812.

Moran EV, Clark JS. 2011. Estimating seed and pollen movement in a monoecious plant: a hierarchical Bayesian approach integrating genetic and ecological data. Mol Ecol. 20:1248–1262.

Nathan R, Cassagrandi R. 2004. A simple mechanistic model of seed dispersal, predation and plant establishment: Janzen-Connell and beyond. J Ecology. 92:733–746.

Pairon M, Jonard M, Jacquemart A-L. 2006. Modeling seed dispersal of black cherry, an invasive forest tree: how microsatellites may help? Can J For Res. 36:1385–1394.

Pavlik B, Muick P, Johnson S, Popper M. 1991. Oaks of California. Los Olivos (CA): Cashuma Press & California Oak Found.

Pons J, Pausas JG. 2007. Acorn dispersal estimated by radio-tracking. Oecologia. 153:903–911.

R Development Core Team. 2011. R: a language and environment for statistical computing [Internet]. Vienna (Austria): R Foundation for Statistical Computing; [cited 2011 Dec 14]. Available from: URL http://www.R-project.org.

Robledo-Arnuncio JJ, Garcia CA. 2007. Estimation of the seed dispersal kernel from exact identification of source plants. Mol Ecol. 16:5098–5109.

Roeder K, Devlin B, Lindsay BG. 1989. Applications of maximum likelihood methods to population genetic data for the estimation of individual fertilities. Biometrics. 45:363–379.

Schueler S, Tusch A, Schuster M, Ziegenhagen B. 2003. Characterization of microsatellites in wild and sweet cherry (*Prunus avium* L.)—markers for individual identification and reproductive processes. Genome. 46:95–102.

Scofield DG, Alfaro VR, Sork VL, Grivet D, Martinez E, Papp J, Pluess AR, Koenig WD, Smouse PE. 2011. Foraging patterns of acorn woodpeckers of (*Melanerpes formicivorus*) on valley oak (*Quercus lobata* Née) in two California oak savannah-woodlands. Oecologia. 166:187–196.

Scofield DG, Sork VL, Smouse PE. 2010. Influence of acorn woodpecker social behaviour on transport of coast live oak (*Quercus agrifolia*) acorns in a southern California oak savanna. J Ecol. 98:561–572.

Sezen UU, Chazdon RL, Holsinger KE. 2009. Proximity is not a proxy for parentage in an animal-dispersed neotropical canopy palm. Proc R Soc B Biol Sci. 276:2037–2044.

Smouse PE, Meagher TR. 1994. Genetic analysis of male reproductive contributions in *Chamaelirium luteum* (L.) Gray (Liliaceae). Genetics. 136:313–322.

Smouse PE, Meagher TR, Kobak CJ. 1999. Parentage analysis in *Chamaelirium luteum* (L.): why do some males have disproportionate reproductive contributions? J Evol Biol. 12:1056–1068.

Smouse PE, Robledo-Arnuncio JJ. 2005. Measuring the genetic structure of the pollen pool as the probability of paternal identity. Heredity. 94:640–649.

Sork VL. 1984. Examination of seed dispersal and survival in red oak, *Quercus rubra* (Fagaceae), using metal-tagged acorns. Ecology. 65:1020–1022.

Sork VL, Davis F, Dyer RJ, Smouse PE. 2002. Mating patterns in a savanna population of valley oak (*Quercus lobata* Née). In: Standiford RB, McCreary D, Purcell KL, tech coordinators, editors. Proceedings of the Fifth Symposium on Oak Woodlands: Oaks in California's Changing Landscape. 2001 October 22–25, San Diego, CA. Gen. Tech. Rep. PSW-GTR-184. Albany (CA): Pacific Southwest Research Station, USFS, USDA. p. 427–439.

Sork VL, Smouse PE. 2006. Genetic analysis of landscape connectivity in tree populations. Landsc Ecol. 21:821–836.

Tyler CM. 2006. Demography and recruitment limitations of three oak species in California. Q Rev Biol. 81:127–152.

Wang BC, Sork VL, Leong MT, Smith TB. 2007. Hunting of mammals reduces seed removal and dispersal of the Afrotropical tree *Antocaryon klaineanum* (Anacardiaceae). Biotropica. 39:340–347.

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. 17:345–350.