ELSEVIER

Review

Measuring pollen flow in forest trees: an exposition of alternative approaches

Peter E. Smouse[a,*], Victoria L. Sork[b]

[a]Department of Ecology, Evolution & Natural Resources, Cook College, Rutgers University,
14 College Farm Road, New Brunswick, NJ 08901-8551, USA
[b]Department of Organismic Biology, Ecology & Evolution, and Institute of the Environment,
University of California, P.O. Box 951786, Los Angeles, CA 90095-1786, USA

Abstract

Steady reduction of once-extensive forested habitats into isolated fragments will have unintended consequences for genetically disrupted tree populations. Gene flow in forest trees involves both pollen and seed flow, and here we describe two alternative analytical models for pollen flow analysis, PARENTAGE and TWOGENER. With PARENTAGE models, we can assess the numbers and frequency spectra of pollinating males for a single female, and the spatial distribution of those pollinating males. Parentage analysis establishes the male parentage of offspring, and uses the inter-parent distances to establish the spatial distribution of pollination. With TWOGENER analysis, we gauge the degree of pollen pool non-overlap of widely spaced females, to derive estimates of the effective number of fathers per mother and the average distance of pollination. Parentage analysis suggests large numbers of fathers per mother while TWOGENER analysis suggests smaller numbers of 'effective pollinators'. Because of TWOGENER's emphasis on "effective pollen flow", it tends to yield shorter average pollination distances. The two approaches are complementary, but suited to different sampling designs and questions. Here, we attempt to reconcile sometimes-divergent findings while also discussing emerging findings on pollen movement across complex landscapes.
© 2004 Elsevier B.V. All rights reserved.

Keywords: Gene flow; Parentage analysis; Pollen dispersal; Mating system

1. Introduction

With our ever-larger human footprint on natural landscapes, our forested habitats have become increasingly fragmented (Ledig, 1992). Abundant theory tells us that we should expect a reduction in gene flow among populations of many species, with all the maladaptive consequences that can follow from genetic isolation (Templeton et al., 1990; Ellstrand and Elam, 1993), but the real question of how much

genetic isolation we can expect remains unanswered (Ehrlich, 1996). For those tree populations that remain, many are managed in some fashion that potentially alters future genetic structure, either because individuals are being removed or because pollen and seed flow are being modified (Sork et al., 1999). For both of these scenarios, assessments of the extent of gene flow and risk of genetic isolation are timely.

Traditionally, gene flow has been estimated indirectly, using genetic structure statistics (Neigel, 1997; Sork et al., 1999). Standard theory tells us that at evolutionary equilibrium, there should be a balance between the genetic population radiation caused by

* Corresponding author. Tel.: +1 732 932 1124;
fax: +1 732 932 1126.
E-mail address: smouse@aesop.rutgers.edu (P.E. Smouse).

genetic drift and the convergence caused by gene flow, expressed as a relationship between $\theta = 4N_e m$ (measuring the effective number of immigrants received by a subpopulation, each generation) and F_{st} (a measure of genetic divergence among subpopulations), usually expressed as $F_{st} \approx [\theta + 1]^{-1}$ (Cockerham and Weir, 1993), or some analogue such as G_{st} (Nei, 1972), for which estimates are standard in population genetic survey of forest tree species (e.g., Loveless and Hamrick, 1984).

We can assess whether subpopulations show 'isolation by distance' with an analysis of the pair-wise decay of the per generation rate of migration among subpopulations, m, with geographic separation (Slatkin, 1993). Elaborate analytical methods have been developed to describe paired population differences, given a multi-population migration matrix and information on local population sizes (reviewed in Smouse and Long, 1992). These methods have continued to be developed, allowing us to estimate the scale of the neighbor area (Rousset, 1997). All of these approaches assume equilibrium conditions, but in view of the level of disruption recently experienced by most populations, it would be unreasonable to expect historical processes to continue at their pre-disturbance rates. Traditional 'genetic survey' estimates of θ will be useful in setting a 'population structure' baseline, but will be of limited utility for predicting the future course of the genetic structure of fragmented and remnant populations, particularly where pollination dynamics are changing, due to anthropogenic disturbance. We need to be able to assess 'real time' gene flow in forest species, and we need that information for various species and breeding systems, under a variety of ecological conditions and forest management practices likely to be encountered in routine practice.

Both seed and pollen dispersal aspects are of critical interest, but for the immediate future, the colonization implications of seed dispersal are a different issue from the genetic connections provided by pollen dispersal, and we will confine ourselves here to the latter. The need to understand contemporary processes requires that we examine current mating episodes. For any particular species of interest, there are two primary questions needing attention: (1) What are the numbers and frequency spectrum of pollinating males for a single female? (2) What is the spatial distribution of those pollinating males? These questions are fundamental to evolutionary biology, because they concern the effective size of the population that experiences evolutionary processes and the area of connectedness among individuals and populations. Our object here is to concentrate on a pair of complementary approaches used to address these questions (PARENTAGE analysis and TWOGENER analysis), and to learn what we may about the pattern of pollen flow in forest tree species.

2. Parentage methods

The most direct approach for examining pollen flow is provided by paternity analysis. Its basic thrust is to use genetic markers to: (1) determine the precise fathers for each of numerous offspring from each of a small number of focal mothers, (2) measure the maternal–paternal distances, (3) ascertain the number of males contributing pollen to a focal female, and (4) then summarize the frequency and spatial spectrum of those contributors (cf., Devlin et al., 1988; Adams et al., 1992; Chase et al., 1996b; Dow and Ashley, 1998; Smouse et al., 1999; Streiff et al., 1999; Burczyk et al., 2002). Because this method requires an exhaustive and thorough map of all the reproductive adults within some circumscribed area, the sample size is constrained by practical limitations. For most forest tree studies, a few focal females are chosen, centrally located within the area, and mature seed are collected from those mothers. All of the potentially contributing adults are genotyped, as are a large sample of offspring from the focal females.

Consider Fig. 1, with mother M_i, receiving gametes from a potentially large array of fathers ($F_k: k = 1, \ldots, K$). For each offspring, we compute the Mendelian probability of obtaining its genotype, given its known mother and each of the potential fathers (from within the area), using standard paternity procedures (Essen-Moller, 1938). With enough genetic markers, only one male will be possible for each of her offspring, all the others being excluded by paternity analysis (cf., Meagher and Thompson, 1986; Roeder et al., 1989), but as a practical matter, our allozyme batteries have had limited exclusion probability, and until the advent of microsatellite markers, the challenge of excluding all but one male was elusive (Chakraborty et al., 1988). Moreover, in most practical situations, the array of
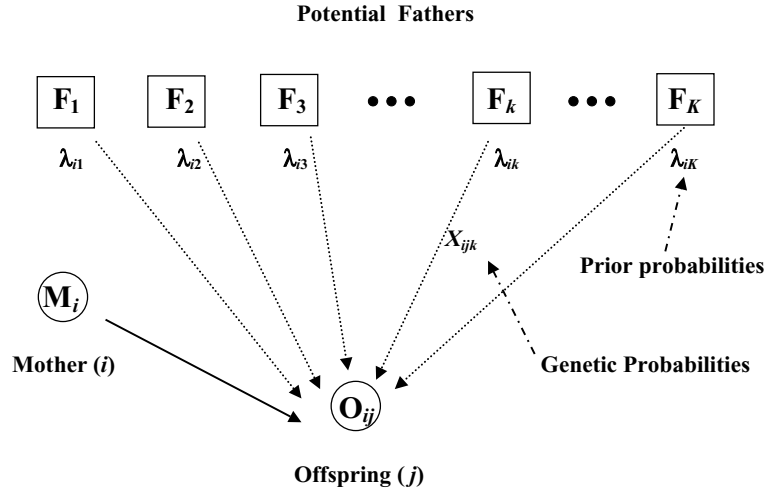
Potential Fathers

Fig. 1. Likelihood paternity analysis, from an array of potential fathers (F$_k$: $k = 1, \ldots, K$), for the jth offspring (O$_{ij}$) of the ith mother (M$_i$), where λ_{ik} ($k = 1, \ldots, K$) are the prior probabilities that each of the potential fathers produced the jth pollen grain sampled by the ith mother, where X_{ijk} are the genetic probabilities that potential father (F$_k$) and mother (M$_i$) could have given the genotype of offspring (O$_{ij}$)

potential males is so large that it is beyond our ability to characterize exhaustively. The best approach is to allocate paternity proportionately, using likelihood methods (Devlin et al., 1988; Smouse and Meagher, 1994) and also allow for the possibility of unsampled males (cf., Adams and Birkes, 1991; Burczyk et al., 1996). If we can assume, for Fig. 1, that we have sampled the candidate males exhaustively, we can compute the posterior likelihood (L_{ijk}) that candidate male F$_k$ is the father, by denoting the mendelian probability of the genotype of the jth offspring of the ith mother, O$_{ij}$, given the genotypes of F$_k$ and M$_i$, as X_{ijk}, where $0 \leq X_{ijk} \leq 1$, and multiplying that probability by λ_{ik}, his a priori probability of fertilizing the ith female. In the absence of any other information, the posterior likelihood that the kth male has fathered the jth offspring of the ith mother is

$$L_{ijk} = \frac{\lambda_{ik}X_{ijk}}{\sum_{k=1}^{K} \lambda_{ik}X_{ijk}}, \qquad (1)$$

where these posterior likelihoods, L_{ijk}, sum to unity over the collection of all possible males.

2.1. Non-informative priors and auxiliary information

Recourse to likelihood estimation of reproductive parameters frees us from the necessity of categorical determination of the precise fathers, and shifts attention to the λ-parameters themselves. The usual (minimalist) starting point is a non-informative prior, with $\lambda_{ik} = 1/K$, proportional contributions of all males to all females. The larger point of the exercise is that the true λ (to be estimated) are anything but equal, and more elaborate specifications allow us to incorporate auxiliary information that might impact on a candidate father's performance. We can model λ_{ik} on features of the male candidates themselves, say on their size or reproductive dominance or pollen volumes (Smouse et al., 1999),

$$\lambda_{ik} \propto \exp\{\alpha_0 + \alpha_1 z_{1k} + \alpha_2 z_{2k} + \cdots + \alpha_H z_{Hk}\}, \qquad (2)$$

where the α are regression coefficients to be estimated, and the z_{hk} are the potentially predictive measures of the male phenotype for the kth male. Morgan and Connor (2001) have extended this treatment to include both linear and quadratic terms, allowing us to model both selection gradients and adaptive peaks. It is also possible to model λ_{ik} on features of male–female pairs (cf., Adams et al., 1992; Burczyk and Prat, 1997; Smouse et al., 1999), such as the distance between them (d_{1ik}) or their phenological overlap (d_{2ik}), but one could easily incorporate any set of interesting pair-wise features, say ($d_{3ik} \cdots d_{Hik}$)

$$\lambda_{ik} \propto \exp\{\beta_0 + \beta_1 d_{1ik} + \beta_2 d_{1ik} + \cdots + \beta_H d_{Hik}\}, \qquad (3)$$

where the β are the regression coefficients that must be estimated. The difference between Eqs. (2) and (3) is that the z_{hk}-variables characterize features of the males themselves, whereas the d_{hik}-variables reference pair-wise features of the ith female and the kth male.

2.2. Flow from outside

It is often possible to genotype only the paternal candidates close to the mother, say within a circle of arbitrary (but convenient) radius, say 25 or 50 or 100 m, and then treat the collective pollen coming in from unsampled males beyond the circle as 'inflow'. It is usual to model the λ from inside the circle as exponentially declining functions of inter-mate distance, while treating the incoming pollen as a global pollen cloud, whose allele frequencies are either sampled from the surrounding population or estimated from the data. One can even add a provision for selfing (cf., Adams et al., 1992; Burczyk et al., 2002). The likelihood model for the jth seedling for

the ith mother takes the form of a NEIGHBORHOOD model,

$$L_{ij} \propto \left[sX_{iji} + (1 - m - s)\sum_{k=1}^{K}\lambda_{ik}X_{ijk} + mX_{ijo} \right], \quad (4)$$

where s is the selfing rate, m the proportion of incoming pollen (from outside the circular 'neighborhood'), X_{iji} the probability of the ith mother yielding both gametes for the jth offspring, and X_{ijo} the probability of the ith female yielding the jth offspring, given the allele frequencies from the pollen pool outside the neighborhood. The usual finding is that the λ within the circular neighborhood decay rapidly, as a function of inter-mate distance, but that a considerable fraction of the pollen comes from outside the neighborhood (Table 1). An added plus for the NEIGHBORHOOD treatment is that it can also be used to estimate the rate of selfing, as well as selection gradients for local populations, where inter-mate distance may be only one of the several interesting determinants of mating pattern.

Table 1
Eleven studies using paternity analysis to estimate gene flow in tree species, with inflow rate equal to the fraction of pollen coming in from outside the arbitrary circle defining the 'neighborhood' and with δ equal to the average pollination distance

Name of species	Type of landscape	Selfing rate (s)	Inflow rate	Distance (δ)	# Seed	# Mothers	# Sites	Area of Site (ha)	Literature citation
(a) Wind-pollinated species									
Pinus attenuata	Mixed forest	0.03	0.56	15	880	4	4	0.04	Burczyk et al. (1996)
Pinus flexilis	Coniferous forest	na	0.06	140	518	71	5	na	Schuster and Mitton (2000)
Quercus humboldtii	Forest fragment	∼0.06	0.32	na	406	3	5	na	Fernández-Manjarrés (2002)
Quercus macrocarpa	Savanna fragment	0.00	0.57	76.9	282	3	1	5	Dow and Ashley (1998)
Quercus petrae & Q. robur	Deciduous forest	∼0.02	0.67	43.6[a]	984	13	1	5.8	Streiff et al. (1999)
(b) Animal-pollinated species									
Calophyllum longifoloium	Tropical forest	0.00	0.06	na	∼400	5	1	84	Stacy et al. (1996)
Eucalyptus regnans	Seed orchard	0.13	0.50	21.4	∼1800	30	9	1/2	Burczyk et al. (2002)
Neobalanocarpus heimee	Dipterocarp forest	0.03	0.32	191.2[a]	238[b]	5	1	42	Konuma et al. (2000)
Pithecellobium elegans	Tropical fragment	0.13	0.29	142	167	6	1	∼80	Chase et al. (1996b)
Spondias mombum	Tropical forest	0.00	0.45	na	na	8–10	2	∼3	Nason and Hamrick (1997)
Spondias mombum	Tropical fragment	0.00	0.90	na	na	1–22	7	0.5–4	Nason and Hamrick (1997)
Turpinia occidentalis	Tropical forest	0.00	0.01	na	115	3	1	5.3	Stacy et al. (1996)

Note that the studies vary in type of marker used, level of genetic resolution, sample sizes, and sample area: (a) wind-vectored pollination, and (b) animal-vectored pollination.

[a] Within-stand estimate; beyond-stand extrapolation yields 287 m for *Q. petraea*, 333 m for *Q. robur*, and 542 m for *N. heimee*.

[b] Determined at sapling stage.

We can elaborate the treatment to include multiple external sources, but that extension is only useful if we can also provide estimates of the allele frequencies from those external sources. For *Cecropia obtusifolia*, where this has been done (Kaufman et al., 1998), the results show that the probability of paternity decays outward for kilometers. Similar results are available from *Swietenia humilis* (White et al., 2002). Evidently, the homogeneous 'global pollen pool' is an expedient statistical myth, but while the 'global pollen pool' may not actually exist, the essential message is robust; many fathers contribute from outside the neighborhood.

2.3. Limitations of paternity analysis

With the advent of microsatellite markers, the hope has been expressed that if we can improve our genetic battery sufficiently, we can designate virtually all of the fathers precisely. Resolution improves with exclusion probability, but to exclude $(K - 1)$ other males, where K is a large number, is an extremely daunting challenge (Chakraborty et al., 1988). In the absence of categorical genetic delineation of all paternal parents, the estimates of their relative contributions are biased, upward for the (mostly unobserved) rare contributors and downward for the apparent major contributors. As it is, one must iterate to a solution to minimize the bias (cf., Smouse et al., 1999). We can use iterative proportional allocation procedures, but that still leaves one critical issue unresolved. To measure the distance to any particular male, we have to be able to locate and sample it; without having the father's location, we lose much of the value of being able to identify the genotype exactly. Establishing paternity is only a means to an end here. Parentage precision is no substitute for sampling the relevant males. To work out the pollination distance distribution, we need to know where they are, not just who they are.

As a statistical consequence of the fact that much of the pollen is coming in from outside the circle, we are in a position to model λ as a declining function of inter-mate distance for the males inside the circle, but must extrapolate outside the circle. If, as seems to be the case in many studies (see Table 1), a large fraction of the pollen is coming in from outside the circle, our statistical delineation of the tail of the distribution from the distribution inside the circle amounts to little

more than informed conjecture. As we shall see later, the shape of the decay curve is not all that well established, and we are not currently reaching out far enough to establish the shape of the tail of that distribution with much credibility. The use of a 'global pollen pool' to represent the incoming male gametes implies that homogeneous frequencies apply regionally. While that may be an expedient assumption, the few data that do exist on this point suggest that the assumption is not warranted (Kaufman et al., 1998). Burczyk et al. (2002) have shown that we can relax that assumption, and can even proceed without additional samples from that pollen pool, but the matter needs a great deal more attention in the near future.

The use of paternity analysis requires consideration of whether it is preferable to use a paternal assignment approach or a fractional paternity allocation approach. Some ML-based software programs, in the presence of paternal uncertainty, assign offspring to the most likely of the sampled male parents. A less biased procedure is to allocate paternity proportionately, while allowing for the possibility of inflow from outside the sampled area. Jones and Ardren (2003) review the software packages available and their optimal use. On balance, the paternity approach provides valuable precision for determining where pollen is coming from and how much comes from outside the area, but since the external pollen sources are imprecisely identified, the estimated rates of inflow are subject to non-trivial sampling error (see discussion of sampling and other issues related to paternity analysis in Sork et al. (1998).

3. TWOGENER method

The TWOGENER model is an indirect method of estimating pollen movement that is a hybrid of traditional paternity and genetic structure analyses (Austerlitz and Smouse, 2001a; Smouse et al., 2001). The approach is based on the idea that mothers, spaced across the landscape, sample partially or completely non-overlapping sets of pollen donors. Typically, we sample many mothers, with few offspring each, so we can investigate a wide variety of situations. The idea is to space the mothers out widely, but to have some clusters that are relatively close together, so that we

can examine the effect of inter-maternal distance on the degree of overlap in pollen pools (Smouse et al., 2001). We genotype both mothers and offspring, but ignore the potential fathers.

3.1. Genetic assay

We determine the male gametes by 'subtraction of the female genotype' from the genotypes of each of her offspring, in the same fashion that is done for paternity analysis. In conifers, we can genotype haploid megagametophytes, revealing the female gametic contribution and rendering the male contribution obvious by inspection. For angiosperms, there is some gametic ambiguity, but we can allocate the paternal contribution among the possible male gametes in the same fashion we use for paternity analysis (Smouse et al., 2001). In either case, given a battery of N male gametic characterizations, we compute a pair-wise genetic distance between each pair of male gametes. We pack the pair-wise distances into an $N \times N$ matrix, and extract an analysis of molecular variance (AMOVA: Excoffier et al., 1992) from that distance matrix, using the sampled mothers as strata and their male gametes as the within-stratum replicates (Table 2). The intra-class correlation of male gametes within a female (Φ_{ft}) measures the divergence of maternal pollen pools, measuring non-overlap of the males they have sampled.

3.2. Effective number of pollinators

It is possible to convert our measure of pollen structure (Φ_{ft}) into other measures of interest, the first of which is the 'effective number of fathers', $N_{ep} = [2\Phi_{ft}]^{-1}$ (Austerlitz and Smouse, 2001a). The basic question is: "Having drawn a random male gamete from a particular female, what is the probability that the next male gamete, drawn randomly and independently from the same female, will have come from the same male?" If we had N_{ep} 'ideal males', all contributing with the same probability, that probability would be simply one over the number of those ideal males, $[N_{ep}]^{-1}$.

Real males are neither ideal nor do they contribute with equal probability, but whatever the array of contributors and however uneven their contributions to a given female, there is some calculable probability that two pollen grains, drawn from that same female, are from the same male, and that probability is denoted $[N_{ep}]^{-1}$, where N_{ep} is the 'effective number' of fathers, i.e., the number of idealized males that would give the same result.

The early empirical returns from TWOGENER analysis show that the effective number of fathers is sometimes quite small; for *Quercus alba*, $N_{ep} \sim 8.22$ (Smouse et al., 2001); for *Q. lobata*, $N_{ep} \sim 3.43$ (Sork et al., 2002); for *Q. velutina*, $N_{ep} \sim 6.25$ (Fernández-Manjarrés and Sork, unpublished); for *Pinus echinata*, $N_{ep} \sim 3.4$–5.5 (Dyer, 2002); for *Albizia julibrissin*, $N_{ep} \sim 2.87$ (Irwin et al., 2003); for *Cornus florida*, $N_{ep} = 2.9$–5.6 (Sork et al., unpublished), smaller than we might have imagined, given what we have learned about pollen flow from the parentage studies. For some species, on the other hand, N_{ep} can be quite large; in *Pinus sylvestris*, $N_{ep} > 70$, and dependent on silvicultural regime (Robledo-Arnuncio et al., 2004). As is the case with any effective number, N_{ep} is sensitive to the unevenness of paternal contribution, so N_{ep} is often considerably smaller than the number of potential males, i.e., $N_{ep} \ll N$.

3.3. Average pollination distance

Austerlitz and Smouse (2001a) show that Φ_{ft} can also be related to the average distance ($\bar{z}$) between sampled mothers, because

$$\Phi_{ft} = \frac{Q_0 - Q(\bar{z})}{2 - Q(\bar{z})}, \qquad (5)$$

Table 2
Molecular analysis of variance for pollen variation within and among sampled females, with M mothers, J progeny within each mother, and with Φ_{ft} defined as the intra-class correlation of male gametes

Source of variation	Degrees of freedom	Sums of squares	Mean squares	Expected mean squares	Variance estimates	Intra-class correlation (Φ_{ft})
Among mothers	$(M-1)$	SS(a)	$MS(a) = SS(a)/(M-1)$	$V(w) + JV(a)$	$\hat{V}(a) = (MS(a) - MS(w))/J$	$\Phi_{ft} = \hat{V}(a)/(\hat{V}(w) + \hat{V}(a))$
Within mothers	$M(J-1)$	SS(w)	$MS(w) = SS(w)/M(J-1)$	$V(w)$	$\hat{V}(w) = MS(w)$	

where Q_0 is the probability that two pollen grains, drawn at random from the same mother, came from the same father (i.e., $Q_0 = 1/N_{ep}$), and $Q(\bar{z})$ is the comparable probability for two pollen grains, drawn from two different mothers, an average distance ($\bar{z}$) apart. If we can assume the distribution of pollen flight distances, we can also translate Φ_{ft} into the average distance of pollination (Austerlitz and Smouse, 2001a). We have modeled bivariate normal and bivariate exponential models, both of which lead to translations of Φ_{ft} (or equivalently, N_{ep}) into an estimate of the average distance of pollination, δ. Defining Δ as the adult stem density per hectare, we have

$$N_{ep} = 8\delta^2\Delta \text{ for the bivariate normal,} \qquad (6a)$$

and

$$N_{ep} = 2\pi\delta^2\Delta \text{ for the bivariate negative exponential.} \qquad (6b)$$

More recent work has considered more general pollen flight distributions, of which the normal and exponential are special limiting cases, and the matter is still under active investigation (Dick et al., 2003). The essential assertion is that the pollen distribution has a 'fatter' tail than either the normal or exponential. As a purely empiric matter, that is a difficult issue to settle, given our penchant for restricting attention to a circle of ≤100 m in radius for most such work.

Distributional details aside, it would appear that while pollen can (and does) move long distances, much of it is not moving very far, with closer males contributing a disproportionate fraction of the pollen. Using TWOGENER analysis, we estimate that the average pollination distance is $\delta \sim 17$ m in *Q. alba* (Smouse et al., 2001), $\delta \sim 65$ m in *Q. lobata* (Sork et al., 2002), $\delta \sim 22$ m in *Q. velutina* (Fernández-Manjarrés and Sork, unpublished), $\delta \sim 17$–21 m in *P. echinata* (Dyer, 2002). The accumulated experience from parentage studies (Table 1) indicates that the average distance of pollination is in excess of 20 m and sometimes much greater.

3.4. Population pattern among the adults

The adult stand often exhibits some form of 'genetic structure', representing the residual historical signature of past selection, gene flow and genetic drift, and that signature can inflate our estimates of Φ_{ft}. As an initial attempt to deal with 'genetic structure' among the adults, Dyer et al. (2004) have described the inflation of Φ_{ft} by subtle allele-frequency gradients across the study area, and have developed methods for measuring and correcting for such gradients. Imagine that there are trends in the allele frequencies among the adults, spread across an extended landscape, and that the usual TWOGENER estimate of Φ_{ft}, in addition to including the usual 'among-mothers' variation, is inflated by the effects of the gradient. The essence of the solution is to embed the AMOVA used by TWOGENER within a more general linear model framework, and then to extract the variation attributable to the gradient, before testing the 'among-mothers' component of variation. We have explored this extension with the ST AMOVA analysis of *Q. alba* (Dyer et al., 2004), and show that the intra-class correlation decreases from $\Phi_{ft} = 0.063$ (ignoring the adult gradient) to $\Phi_{ft} = 0.056$ (correcting for that gradient), a modest but significant difference. That change increases the estimate of N_{ep} from 7.9 to 8.9 and that of δ from 16.3 m to 17.4 m. In *Q. alba*, the bias from ignoring the gradient is modest, but in other organisms, it might not be. Whether the bias from ignoring the gradient is large or small, it can be estimated and extracted (Dyer et al., 2004).

For many forest species, consanguinity and local genetic structure among the adults also complicates our assessment of pollen structure, since the historical signature of population structure tends to inflate our estimates of Φ_{ft} (Austerlitz and Smouse, 2001b). Consanguinity among the adults, representing historical population structure, yields

$$\Phi = \frac{[Q_0 - Q(\bar{z})](1 + F)}{2 - Q(\bar{z})(1 + F)}, \qquad (7)$$

where F is the consanguinity coefficient among adults. For *Quercus* and *Pinus*, sampled in the Missouri Ozark Forest Ecosystem Project (MOFEP), which evaluated three forest treatments—clearcut, selective cuts, and uncut controlled stands (Dyer, 2002), $F \approx 0$, but departures from panmixia are evident in *C. florida*, ranging from $-0.105 \leq F \leq +0.024$ (Sork, unpublished). Allowance for that adult structure alters our estimates of Φ_{ft} from 0.174 to 0.170 for uncut control stands, from 0.125 to 0.140 for selectively cut stands, and from 0.090 to 0.091 for clearcut stands, all of which lead to small derivative changes in N_{ep} and the

average distance of pollination, δ. Adult structure is a complication, because it is confounded with pollen structure, but with independent information on adult structure, we can separate their effects.

We should distinguish between adult consanguinity and selfing, the probability that the female is sampling her own pollen. Current versions of TWOGENER do not deal with selfing. Since a great many species that are facultative selfers will soon come under examination, and since an elevated selfing rate will inflate our estimate of Φ_{ft}, elaboration of the standard TWOGENER estimation scheme (to include selfing) would be timely. Pending such developments, analysis of inbred species may require the NEIGHBORHOOD model (cf., Burczyk et al., 2002).

3.5. Temporal variation

Studies of pollen movement are generally restricted to a single reproductive season, by virtue of the field and laboratory effort required. Questions remain about the adequacy of single-season measures for long-lived, iteroparous forest species. Drawing from a multi-year study of *A. julibrissin*, Irwin et al. (2003) were able to estimate the difference between single- and multi-year sampling, relative to estimation of N_{ep}. They demonstrated less among-year variation within mothers than variation among-mothers, but showed that $N_{ep} \sim 2.0$, when estimated from the traditional single-season study, but increased to $N_{ep} \sim 2.9$ with allowance for year-to-year variation in pollen profile for a single mother, a 40% increase. Clearly, we must devote some serious attention to the question of year-to-year variation in future studies of pollen flow.

3.6. Limitations of TWOGENER analysis

For TWOGENER, we have given up pursuit of the fathers, focusing on what we can learn from mothers and offspring. That choice has moved us along rapidly, and our comparative work is accelerating. We are now beginning to answer some larger questions, and larger patterns are beginning to emerge, but all of this comes at a price. There are two major limitations of TWOGENER. First, N_{ep}—while it does constitute a convenient and generic gauge of pollen structure—conceals more than it reveals. We

expect that N_{ep} should be smaller than N, but how much smaller? We know that Φ_{ft} can be upwardly biased by either consanguinity or the historical remnants of adult genetic structure (Austerlitz and Smouse, 2001b; Dyer et al., 2004), and with independent assessments of both factors, we can (in principle) correct for these potential confounding effects, but how much attention are we going to need to devote to sampling and analysis of those factors in the adult population? Our deeper interest is in 'pollen flow', but for populations with suspected high inbreeding or substantial local genetic structure, it is desirable to pay some serious attention to 'adult structure', and that will be costly. So far, biases in $\hat{\Phi}_{ft}$ from these sources have not been great, but future work will be needed to demonstrate whether adult structure is generally a serious confounding factor or a only a minor aggravation.

A second limitation is the estimation of the distance of pollen movement, a problem that plagues paternity analysis as well. Our estimation of the average distance of pollination from TWOGENER is indirect, and is based on particular assumptions about the shape of the pollen dispersal curve. We initially modeled bivariate normal and exponential functions (Austerlitz and Smouse, 2001a), but empiric data now suggest a distribution that is steeper near the origin, with a fatter tail than either the normal or exponential (Dick et al., 2003), better modeled with the exponential power distribution,

$$p(a, b; x, y) = \frac{b}{2\pi a^2 \Gamma(b)} \exp\left(-\left(\frac{\sqrt{x^2 + y^2}}{a}\right)^b\right),$$

(8)

where $\Gamma(b)$ is the classical gamma function (Abramowitz and Stegun, 1964). The parameter b is the shape parameter, determining the form of the dispersal function, and a is the scale parameter for distance. When ($b < 1$—subexponential), the distribution is both steeper and fatter-tailed than the exponential, and δ is greater (Fig. 2a). For $b = 1$, Eq. (8) becomes the bivariate exponential distribution (with $a = \gamma^{-1}$), where γ is the classic scale parameter of the exponential function (Fig. 2b). When ($1 < b < 2$—subnormal), the distribution (and average distance, δ) are intermediate between the normal and the exponential (Fig. 2c). For $b = 2$, the distribution becomes a
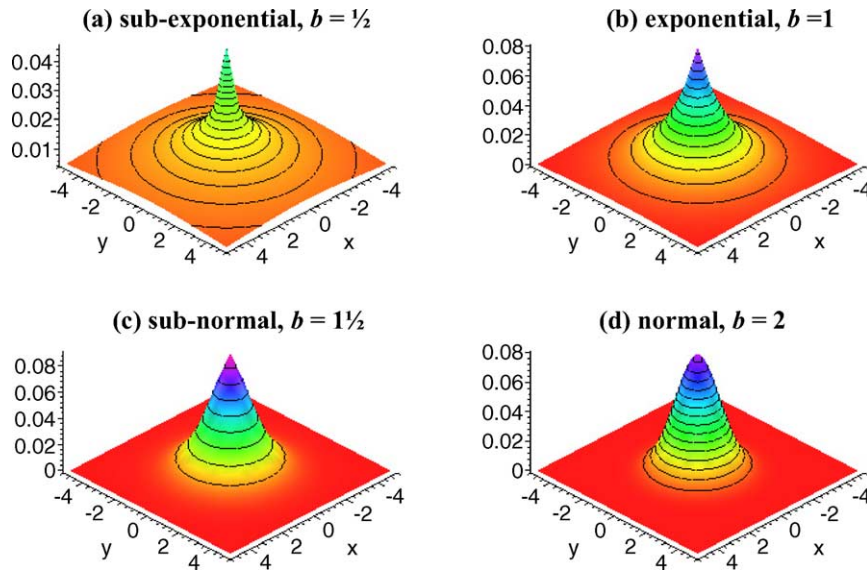
Fig. 2. Illustrative pollen distributions under four different scenarios, described in terms of the a- and b-parameters of Eq. (8): (a) bivariate sub-exponential ($a = \sqrt{2}$, $b = 1/2$); (b) bivariate exponential ($a = \sqrt{2}$, $b = 1$); (c) bivariate sub-normal ($a = \sqrt{2}$, $b = 1.5$); and (d) bivariate normal distribution ($a = \sqrt{2}$, $b = 2$); the tail probability increases as b declines.

bivariate normal, with $a = (\sigma\sqrt{2})$ as the scale parameter (Fig. 2d). Using non-linear estimation techniques, Austerlitz and Smouse (2002) modeled different possibilities and estimated a and b from the data, rather than assuming them a priori. Dick et al. (2003) found that $b \sim 0.9$ for *Dinizia excelsa* and noted that other tested species also showed $b < 1$ (subexponential). Most paternity work fits the data to an exponential model ($b = 1$), by virtue of the construction of Eq. (3). There is nothing, in principle, to prevent fitting Eq. (8) as $d_{ij} =$ inter-mate distance in Eq. (3), but that has not entered standard practice to date, and the limitations of using the exponential model plague current parentage analysis as well. More recently, Austerlitz et al. (2004) have explored several competing models with fat tails, and it now seems clear that whether we assume the shape of the distribution beforehand or attempt to estimate it from the data, the estimates are somewhat model-dependent.

As is also evident from Eq. (6a) and (b), our translation of Φ_{ft} (or equivalently, N_{ep}) into an estimate of the average distance of pollen movement, δ, comes in the form of the product ($\delta^2 \Delta$). We have commented elsewhere (Austerlitz and Smouse, 2002) that the adult stem density is almost surely an overestimate of

'effective adult density'. Some pollen donors produce less pollen; others are out of synchrony phenologically; others are (potentially) incompatible mates with the mother in question. Accumulated experience with paternity methods, where such factors have been explicitly modeled (Adams et al., 1992; Burczyk et al., 1996, 2002; Smouse et al., 1999; Morgan and Connor, 2001) inspires limited confidence that we can describe the nuances of variable male reproductive availability with any precision, but the effects are surely there, and they contribute to the unevenness of male parentage contributions. To the extent that stem count per hectare (Δ) is an overestimate of Δ_e, then our estimate of δ is an underestimate of its analogue, 'effective distance' δ_e. Rousset (1997) uses F_{st} to estimate the product ($\delta^2 \Delta$) among adults, effectively scaling distance inversely with density, and Smouse et al. (2001) use Φ_{ft} to do the same thing for male gametes. In either case, estimates of δ and Δ are highly confounded. We can obtain a partial separation of δ_e and Δ_e (Austerlitz et al., 2004), but it has to be said that those estimates are negatively (and strongly) correlated. A better joint-estimation scheme has yet to emerge, and the matter requires some additional work.

4. Number of fathers versus N_{ep}

Parentage and TWOGENER studies would seem to be telling us two different stories. The parentage studies yield a general impression of widespread pollen flow, with male contributions declining with distance, at least within (say) 100 m of the focal mother, but with many males contributing from outside the circle. Distance is ignored from those males, because the incoming pollen is usually (though not always, see Kaufman et al., 1998) viewed as a homogeneous 'global pollen pool'. TWOGENER studies, on the other hand, yield the impression of a small number of pollen donors (small N_{ep}) and a short average distance of pollination (δ), though there is circumstantial evidence for a pollen distribution with a longer and fatter tail, miming the 'global pollen pool' of the parentage studies. To resolve the apparent inconsistencies compellingly, we will need some direct studies that facilitate both approaches on the same material. The difficulty is that parentage studies work best with small numbers of sampled mothers, a restricted set of genetically delineated competing fathers, and large numbers of offspring from each mother (Roeder et al., 1989; Devlin and Ellstrand, 1990), not a particularly convenient design for forest species with widely-dispersed pollen and large numbers of (unsampled) candidate fathers. Conversely, TWOGENER works best with large numbers of mothers and few offspring each (Smouse et al., 2001; Irwin et al., 2003), which would be too few offspring to characterize the paternal array for any one mother with any accuracy or precision. We are going to need large numbers of mothers and large numbers of offspring per mother, and we will have to locate and genotype a large fraction of the potential fathers over an extended area, if we are to mount a compelling 'head to head' comparison. The challenge is not out of reach, but it is formidable. We will have to choose our species and ecological setting carefully, if we are to mount a compelling 'head to head' comparison.

In the meantime, we are inclined to ask whether the discrepancy itself is real, or whether it results from divergent, albeit reconcilable perspectives. For example, even though the paternity studies reveal a large number of fathers, they also show that these males contribute unevenly to the seed pool and that nearby males father a disproportionate share of the progeny (e.g, Burczyk et al., 1996; Kaufman et al., 1998; Streiff et al., 1999). So, it is quite possible to have a large number of actual fathers but a low number of effective fathers, in the same way that the census population size does not equal the effective population size. To illustrate what we mean by that, consider the ith mother, who draws pollen from a potentially infinite array of male contributors. Just to simplify the mathematical illustration, imagine a geometric distribution, describing the unevenness of the relative probabilities of those males contributing to the ith mother, where the λ-value for the kth most prolific male, contributing to ith mother, is defined as

$$\lambda_{ik} = \alpha(1-\alpha)^{k-1}, \tag{9}$$

where α is the 'dominance' parameter, the expected relative contribution of the most prominent pollen donor. The α-value might vary among females, and the ordering of male prominence would certainly vary among females, but the important point here is that as α decreases from 1 to 0, the evenness of male contributions increases. For example, if $\alpha = 0.90$, then the series of λ_{ik}-values is $\lambda_{i1} = 0.9$, $\lambda_{i2} = 0.09$, $\lambda_{i3} = 0.009$, $\lambda_{i4} = 0.0009$, etc. However, if $\alpha = 0.30$, then the series is flatter, $\lambda_{i1} = 0.30$, $\lambda_{i2} = 0.21$, $\lambda_{i3} = 0.147$, $\lambda_{i4} = 0.1029$, etc., and if α is as low as 0.10, then the series becomes very flat, $\lambda_{i1} = 0.10$, $\lambda_{i2} = 0.09$, $\lambda_{i3} = 0.081$, $\lambda_{i4} = 0.0729$, etc.

All three examples are drawn from an infinite array of potential fathers. The number of different fathers, $E(\hat{K})$, expected from a sample of J offspring will increase with sample size, but it will also depend on the value of α, via standard rarifaction logic (Chakraborty et al., 1988)

$$E(\hat{K}) = \sum_{k=1}^{\infty}[1 - (1 - \alpha(1-\alpha)^{k-1})^J], \tag{10}$$

plotted in Fig. 3a, as a function of sample size, J. For $\alpha = 0.9$, a sample of $J = 50$ offspring (per mother) could be expected to yield a majority of offspring from the dominant male, a few from the first subdominant male, and a smattering of essentially unpredictable singletons from all the other males; the expected (average) number of fathers sampled per mother would be only $E(\hat{K}) \approx 2.40$. For the case of $\alpha = 0.30$, that same sample of $J = 50$ offspring would yield a more balanced array of paternity, with no one

(a) Expected number of different fathers, $E(\hat{K})$
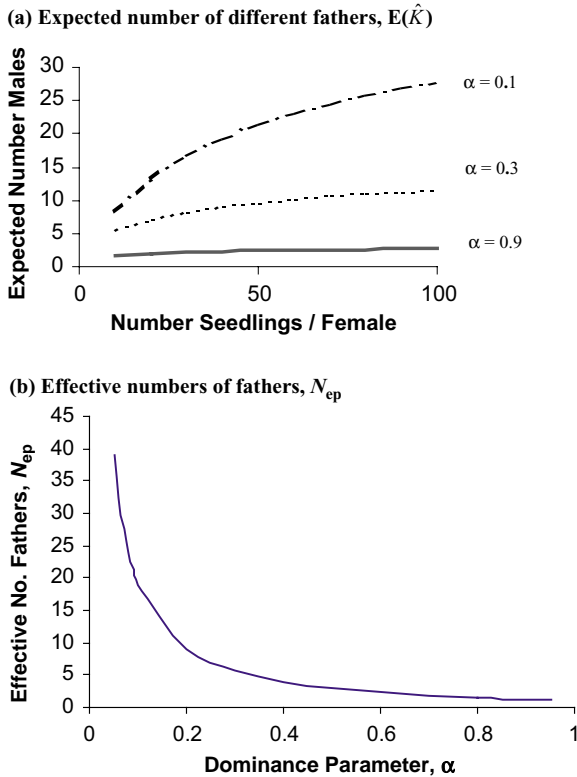

(b) Effective numbers of fathers, N_{ep}

Fig. 3. Male parentage distribution predictions of the geometric distribution: (a) expected numbers of male parents for different numbers of offspring per mother and for dominance parameters, $\alpha = 0.1$, 0.3 and 0.9; (b) effective numbers of pollinators, as a function of α.

male represented quite so heavily, and more males represented; the average number of fathers sampled per mother would be $E(\hat{K}) \approx 9.74$. For the case of $\alpha = 0.10$, the frequency distribution of male parentage would be quite flat, and the expected number of fathers per mother would be $E(\hat{K}) \approx 21.4$. While the number of different fathers recovered is expected to rise with sample size, J, for an uneven distribution of λ_{ik}-values (larger α), the expected number recovered will asymptote more quickly.

In contrast to the number of expected fathers, the number of 'effective fathers' is not very sensitive to sample size, because small λ_{ik}-values contribute almost nothing to N_{ep}. From a given array of J offspring, drawn from the ith mother, we can compute the probability that two pollen grains were drawn from the same male, which is the definition of $(1/N_{ep})$,

computed as

$$\frac{1}{N_{ep}} = \sum_{k=1}^{\infty} \lambda_{ik}^2 = \sum_{k=1}^{\infty} \alpha^2 (1-\alpha)^{2(k-1)} = \cdots$$
$$= \frac{\alpha}{2-\alpha}, \qquad (10\prime)$$

so evidently, $N_{ep} = (2 - \alpha)/\alpha$, which decreases with increasing unevenness of the male parentage distribution in a predictable way (Fig. 3b), even though the array of potential males is infinite. For $\alpha = 0.9$, $N_{ep} = 1.22$; for $\alpha = 0.3$, $N_{ep} = 5.67$; for $\alpha = 0.1$, $N_{ep} = 19.00$.

Reworking Eq. (6a) and (b), we discover that $(N_{ep}/8$—bivariate normal$) < \delta^2 \Delta < (N_{ep}/2\pi$—negative exponential$)$, so the estimate of average pollination distance (δ) is meant to represent N_{ep} effective (idealized) individuals, all contributing equally and without regard to their location within the population. By computation, N_{ep} is closer to a harmonic than to an arithmetic mean, so δ will be smaller than the arithmetic mean, not measured directly with TWOGENER methods, but separately estimable (at least for the fraction of fathers within the 'neighborhood') from NEIGHBORHOOD methods (see Section 2.1).

Now, the geometric distribution is too simple to represent the real pollen distribution, but the essential point survives changes in the specification of that distribution. The effective number of males N_{ep} increases with the evenness of male parentage distribution, and small N_{ep} represents a very uneven distribution of relative male contributions, with a few repeat contributors and many minor contributors, represented once each or not at all. Burczyk et al. (1996) have used a different treatment but the same basic premise (unequal contributions from a potentially infinite pool) to extract estimates of N_{ep} from the NEIGHBORHOOD model. The essential result is the same; whatever the detailed male parentage distribution may be, the evidence supports the assertion that the local males are more frequent contributors.

5. Discussion

We began this review with questions about the expectation of reduced pollen flow across disturbed and fragmented landscapes. A major motivation for measuring gene flow is that we need to assess the

impact of various contextual disturbances (human-generated or natural) that fragment the population, reduce conspecific density, or otherwise alter the reproductive landscape. The paternity allocation approach has generated valuable insights concerning the extent of immigration into a local stand (reviewed in Adams et al., 1992; Sork et al., 1999; Burczyk et al., 2002; see Table 1). Many studies of fragmentation have used this approach because identification of parents within the sample also informs us about how many progeny were sired from outside the sample area. In contrast, studies of the impact of forest management practices that tend to alter conspecific density and canopy closure have tended to use the TWOGENER approach, which can be mounted under a wider variety of circumstances. Both approaches have yielded valuable lessons about gene flow under different circumstances.

5.1. Fragmentation

The paternity approach is particularly useful for isolated populations in a delimited area, because every adult within the patch can be genotyped, and the progeny that must have resulted from incoming pollen become categorically obvious. Chase et al. (1996b), examining *Pithecellobium elegans*, a self-incompatible, hawkmoth-pollinated species, showed that 28% of the progeny found within a fragment containing 28 adults were pollinated from outside (and distant) pollen sources. From identifiable male parents, inside the fragment, the average distance of pollen movement was 142 m. Pollen coming in from outside the fragment could only make that average larger. Nason and Hamrick (1997) report that 90–100% of *Spondias mombin* progeny produced in small fragments were the product of gene flow events from forest stands located 80–1000 m away (see Table 1).

The evidence suggests that fragmentation can sometimes lead to increased pollen flow. Dick (2001) reports that African honeybees, which have replaced native insects in isolated pastures, facilitate extensive pollen movement across the landscape. Mean pollinator distance for pasture and gallery forest trees was 1288 m, while that into a 10 ha fragment was only 417 m. It is not always clear whether pollen flow into fragments is greater because, with fewer 'inside males', a greater fraction of successful pollen must

have come from 'outside males', or whether pollen simply moves farther across open landscapes. How the increased pollen dispersal that results from opening the forest canopy translates into changes in N_{ep} will depend on the local conspecific density, before and after disturbance. Sometimes, increased pollen flow will more than make up for the reduction in local mate availability; but sometimes, it will not. The tradeoffs for pollen-dispersal between conspecific density and canopy closure are complex.

5.2. The impact of forest management

A major effect of forest management is the alteration of conspecific density and/or general canopy closure. Early TWOGENER studies of *Q. alba*, in a closed-canopy, mixed conifer-deciduous forest, with conspecific stand density on the order of 93 trees per hectare, yielded $N_{ep} \sim 8.22$ and $\delta \sim 17$ m (Smouse et al., 2001), whereas a similar study of *Q. lobata*, growing under savanna conditions, with conspecific density less than two trees per hectare, yielded $N_{ep} \sim 3.43$ and $\delta \sim 65$ m (Sork et al., 2002). At face value, these studies suggest that pollen flow is wider under open stand conditions, but in this case greater pollen movement translates into a decrease in N_{ep}. We caution that this comparison involves two different species, growing in two different regions, with conspecific density and canopy closure confounded.

An opportunity to examine canopy closure alone was provided by a silvicultural study of *P. echinata*, through the Missouri Ozark Forest Ecosystem Project (Mofep), which evaluated three forest treatments—clearcut, selective cuts, and uncut control stands (Dyer, 2002). The cutting treatments removed all tree species except *P. echinata*, which altered canopy closure with minimal consequences for the density of *P. echinata* itself. The parameter estimates were $N_{ep} \sim 5.5$ and $\delta \sim 22$ m for clearcut, $N_{ep} \sim 5.2$ and $\delta \sim 17$ m for selectively cut, and $N_{ep} \sim 3.4$ and $\delta \sim 17$ m for uncut controls, respectively. Opening the canopy both enhances the average distance of pollen movement and increases N_{ep}, with no change in the conspecific density of *P. echinata*. In the same MOFEP landscape as for the *P. echinata* study, Sork et al. (unpublished) found for *C. florida*, an insect-pollinated understory tree, that N_{ep} and δ were both highest in clearcut areas and lowest in uncut control areas. For

C. florida, however, conspecific density varied slightly with treatment (Apsit et al., 2002), remaining confounded with canopy closure.

A direct comparison of silvicultural practices that changed conspecific density is provided by studies of *P. sylvestris* (Robledo-Arnuncio et al., 2004), growing in monospecific stands in Spain. Both shelterwood and group selection cuts were compared with uncut (but otherwise matched) controls in mature stands. The shelterwood cut yielded $\Delta \sim 183$ trees ha^{-1}, $N_{ep} \sim 83$ and $\delta \sim 24$ m, compared with control values of $\Delta \sim 315$ trees ha^{-1}, $N_{ep} \sim 71$ and $\delta \sim 17$ m. The group selection system yielded $\Delta \sim 80$ trees ha^{-1}, $N_{ep} \sim \infty$ and $\delta \sim \infty$, while its uncut control yielded $\Delta \sim 183$ trees ha^{-1}, $N_{ep} \sim 125$ and $\delta \sim 29$ m. Here, both δ and N_{ep} values increase with stand thinning, but the pollen distance results reflect the fact that uncut canopies were densely packed. For the most open stand (group selection), pollen was moving freely across the area.

These three case studies illustrate the potential impact of forest management on pollen movement in both wind and insect-pollinated species. The evidence raises the question of whether the impact is due primarily to the modified forest canopy structure or reductions in conspecific density, and future studies that clearly separate the two effects are needed.

5.3. The importance and utility of N_{ep}

The TWOGENER work draws attention naturally to the effective number of fathers, N_{ep}, which provides a natural link to theoretical discussions of effective population size among the standing adults (e.g., Chesser et al., 1993; Lande and Barrowclough, 1987). It develops that Ritland's MLTR analysis (Ritland, 1989, 1990) can also be used to estimate N_{ep}, and Burczyk et al. (2002) show that parentage analysis can provide an estimate as well. Greef et al. (2001) discuss and illustrate the concept of N_{ep} for hermaphroditic plant and animal populations, with an interesting application to Ficus. Statistical nuances of different estimation schemes aside, however, the point here is that N_{ep} is a parameter of general interest and utility.

As a summary parameter, N_{ep} allows ready comparison among species, among habitats, and among

circumstances. A question of emerging interest is whether the type of pollination system influences the effective number of fathers, which would have different implications for genetic drift across extended landscapes for species that were wind pollinated, versus those that were animal-pollinated. Drawing from a set of available studies reporting such data for woody species, and from which we can extract an estimate (Table 3), we discern three trends.

First, we note that some wind-pollinated species show much higher values of N_{ep} than do animal-pollinated species, suggesting that N_{ep} is inherently smaller in animal-pollinated than in wind-pollinated species, in spite of the fact that the pollinators can be shown to move considerable distances. While wind-dispersed pollen can move great distances, given open-stand aerodynamic conditions, N_{ep} tends to be smaller under closed canopy conditions, everything else being equal. In general, N_{ep} is inversely correlated with the level of pollen flow across the local landscape, with all of the consequences that usually follow from genetic drift, and the data suggest that pollen flow is more affected by the landscape context than by the pollination system.

Second, N_{ep} increases with conspecific density, everything else being equal. With fewer near neighbors, the number of effective contributors is higher, notwithstanding the fact that the most frequent contributors remain the near neighbors (Stacy et al., 1996). White et al. (2002) have shown, with *S. humilis*, that the fraction of incoming pollen increases as fragment population size (stem density) decreases. Third, thinning the canopy (of other species) seems to increase the distance of pollination (Dyer, 2002; Sork et al., unpublished) and in some cases increases the effective number of fathers.

5.4. Seed flow

We have arbitrarily concentrated this paper on pollen flow, indicating at the outset that estimation of seed flow has its own set of issues and problems. Several workers are beginning to report studies of contemporary seed flow (e.g., Aldrich and Hamrick, 1998; Godoy and Jordano, 2001; Nathan et al., 2002). Colonization considerations aside, seed flow into fragments may be common enough that it contributes to real-time gene flow. In instances where seed flow

Table 3

A selection of N_{ep}-values, estimated via MLTR analysis (Ritland, 1989; Ritland, 1990), TWOGENER analysis (Smouse et al., 2001), or direct assessment of the paternity spectrum (Burczyk et al., 2002), for tree species sampled from different landscape situations: (a) species that have wind-vectored, and (b) animal-vectored pollination

Species under study	Landscape context	Effective number of males		Literature citations
		N_{ep}	Method	
(a) Wind-vectored pollination				
Larix occidentalis	High density	50–100	MLTR	El-Kassaby and Jaquish (1996)
Larix occidentalis	Low density	10–16	MLTR	El-Kassaby and Jaquish (1996)
Picea abies	Dense forest	33–46	TWOGENER	Finkeldey (1995)
Picea glauca	Seed orchard	1–2	Paternity	Schoen and Stewart (1986)
Picea mariana	Boreal forest	6–9	MLTR	Perry and Bousquet (2001)
Pinus echinata	Closed forest	3–6	TWOGENER	Dyer and Sork (2001)
Pinus sylvestris	Dense forest	>70	TWOGENER	Robledo-Arnuncio et al. (2004)
Pinus washoensis	Coniferous forest	14	MLTR	Mitton et al. (1997)
Quercus alba	Closed forest	8	TWOGENER	Smouse et al. (2001)
Quercus humboldtii	Forest fragments	1–2	MLTR	Fernández-Manjarrés (2002)
Quercus lobata	Savanna	3–4	TWOGENER	Sork et al. (2002)
Quercus velutina	Montane forest	6	TWOGENER	Fernández-Manjarrés and Sork (unpublished)
(b) Animal-vectored pollination				
Acacia melaxylon	Eucalypt forest	16–33	Paternity	Muona et al. (1991)
Albizia julibrissin	Scattered clumps	3	TWOGENER	Irwin et al. (2003)
Carapa pocera	Tropical forest	12	MLTR	Doligez and Joly (1997)
*Caryocar brazielensi**	Cerrado fragments	5–12	MLTR	Collevatti et al. (2001)
Cornus florida	Northern aspect	11	MLTR	Bailey (2002)
Cornus florida	Southern aspect	5	MLTR	Bailey (2002)
Enterolobium cyclocarpum	Closed forest	2–4	MLTR	Rocha and Aguilar (2001a,b)
Enterolobium cyclocarpum	Open pasture	10	MLTR	Rocha and Aguilar (2001a,b)
Ficus (4 species)	Scattered trees	3–11	Paternity	Greef et al. (2001)
Jacaranda copaia	Scattered dryforest	4–5	TWOGENER	James et al. (1998)

exceeds pollen flow, estimation of the rate will be imperative.

Genetic studies of seed flow are more challenging than pollen flow studies, because once the seed is separated from its maternal tree, we have entered the inferential realm of the 'two-parent' problem (Meagher and Thompson, 1986), where neither parent is known a priori. Most workers have been forced to assume that seed collected from the 'seed shadows' of different mothers are progeny of those same mothers, and while that is (probably) a reasonable first approximation under field conditions, it necessarily confines the approach to low-density situations, with non-overlapping seed shadows. Aldrich and Hamrick (1998), in studying gene flow in *Symphonia globulifera*, a hummingbird-pollinated and bat-dispersed tropical canopy tree, discovered that a majority of seedling recruitment in remnant forest was derived from pasture trees. Apparently, high fruit production of these pasture trees attracted foraging bats, while the dense foliage of the remnant forest provided more suitable feeding and roosting sites, resulting in the movement of seed from pasture trees into the fragment. The authors were unable to separate pollen-versus seed-mediated gene flow events definitively, but they could estimate that total gene flow into fragmented forest ($m = 0.16$) was less than that into continuous forest ($m = 0.56$), but higher than the rate of pollen flow alone. For this species and setting at least, seed constituted a major component of propagule flow.

More recently, genetic assay methods have been developed for diploid seed coats (diploid maternal tissue), which can be used for powerful 'maternity analysis', uncoupled from male parentage (Godoy and

Jordano, 2001). Armed with well-identified maternal genotypes, we can (at least in principle) extend standard parentage models to separate established progeny into: (a) those with both parents from within the fragment, (b) those with an internal mother and external father, (c) those with an internal father and external mother, and even (d) those for whom both parents are external. It is becoming obvious that we are going to have to separate gene flow into pollen- and seed-mediated components, for which we will need to follow the movements of both paternal and maternal gametes. Where the parentage approach is used, we will have to map and assay exhaustively over a considerable area. With a generalized population structure-like approach, we will have to develop cross-classified TWOGENER analogues that use both the females and males as strata. For conifers, with separately inherited maternal (mtDNA), paternal (cpDNA), and biparental (nuclear) genomes, the analytical way forward is clear, but for angiosperms, without a paternally inherited genome, the nature of that analysis is not yet clear. In any case, all of these challenges, representing unfinished business, lie ahead of us.

5.5. Conclusions

Analytical challenges aside, the accumulating evidence from the field suggests that forest tree populations occupying human-altered landscapes will sometimes experience reduced gene flow and sometimes enhanced gene flow, depending on the species and circumstances. The challenge is detecting this impact, when it occurs. In this manuscript, we reviewed two major approaches one can take and their findings thus far. Each has its own virtues and vices. The parentage approach provides a description of who came from where, the total number of pollen donors that came from within a circumscribed area, and the number of progeny that represent pollen-mediated immigration. If one can identify potential fathers at great enough distances, this approach should also help us to document the shape of the dispersal curve and the length of the tail. Its requisites are the ability to identify every potential father within a specified area, high genetic resolution (ideally >99%), and larger numbers of progeny than potential fathers, preferably for each seed tree. The need to identify all potential pollen donors will be challenging for tree species that occur at low densities, with specialized long-distance pollinators.

TWOGENER methods estimate the effective distance of pollen movement and the effective number of pollen donors per maternal tree. This method is ideal for studies where the geographical scale is sufficiently large that identification of all pollen donors is beyond reach. In addition, this model can be deployed even when the genetic resolution for the battery of available markers is less than 95%. TWOGENER estimates are influenced by assumptions of the distribution of the pollen dispersal curve, but not to such an extent that it obviates hypothesis tests about differences among sites experiencing different conditions (e.g., fragmented versus continuous forest, or clear-cut versus selective cut forest management). Average pollen movement distances may be underestimated, due to an emphasis on effective pollen movement. Pollination distances can also be underestimated if the density of reproductive adults is substantially lower than the nominal stem count. As we continue studying the impact of habitat disturbance on forest species, parentage and TWOGENER analyses will provide complementary tools that can be expected to elucidate the factors that impact gene flow across the landscape.

References

Abramowitz, M., Stegun, I.A., 1964. Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. US Government Printing Office, Washington.

Adams, W.T., Birkes, D.S., 1991. Estimating mating patterns in forest tree populations. In: Fineschi, S., Malvolti, M.E., Cannata, F., Hattermer, H.H. (Eds.), Biochemical Markers in the Population Genetics of Forest Trees. SPB Publishing, The Hague, pp. 157–172.

Adams, W.T., Griffin, A.R., Moran, G.F., 1992. Using paternity analysis to measure effective pollen dispersal in plant populations. Am. Nat. 140, 762–780.

Aldrich, P.R., Hamrick, J.L., 1998. Reproductive dominance of pasture trees in fragmented tropical forest mosaic. Science 281, 103–105.

Apsit, V.J., Sork, V.L., Dyer, R.J., 2002. Patterns of mating in an insect-pollinated tree species in the Missouri Ozark Forest Ecosystem Project (MOFEP). In: Shifley, S.R., Kabrick, J.M. (Eds.), Proceedings of the Second Missouri Ozark Forest Ecosystem Symposium: Post-Treatment Results of the Landscape Experiment, 17–18 October 2000, St. Louis, MO. Gen. Tech. Rep. NC-227. US Department of Agriculture, Forest Service, North Central Forest Experiment Station, St. Paul, MN, pp. 213–227.

Austerlitz, F., Smouse, P.E., 2001a. Two-generation analysis of pollen flow across a landscape. II. Relation between ϕ_{ft}, pollen dispersal, and inter-female distance. Genetics 157, 851–857.

Austerlitz, F., Smouse, P.E., 2001b. Two-generation analysis of pollen flow across a landscape. III. Impact of within-population structure. Genet. Res. 78, 271–280.

Austerlitz, F., Smouse, P.E., 2002. Two-generation analysis of pollen flow across a landscape. IV. Estimating the dispersal parameter. Genetics 161, 355–363.

Austerlitz, F., Dick, C.W., Dutech, C., Klein, E., Oddou-Muratorio, S., Smouse, P.E., Sork, V.L., 2004. Using genetic markers for the estimation of the pollen dispersal curve. Mol. Ecol. 13, 937–954.

Bailey, K.J., 2002. Influence of environmental factors on the mating system of *Cornus florida*. Unpublished Ph.D. Dissertation, Department of Biology, University of Missouri, St. Louis.

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.

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

Burczyk, J., Adams, W.T., Moran, G.F., Griffin, A.R., 2002. Complex patterns of mating revealed in a *Eucalyptus regnans* seed orchard using allozyme markers and the neighborhood model. Mol. Ecol. 11, 2379–2391.

Chakraborty, R., Meagher, T., Smouse, P.E., 1988. Parentage analysis with genetic markers in natural populations. I. The expected proportion of offspring with unambiguous paternity. Genetics 118, 527–536.

Chase, M., Moller, R.C., Kesseli, R., Bawa, K.S., 1996b. Distant gene flow in tropical trees. Nature 383, 398–399.

Chesser, R.K., Rhodes Jr., O.E., Sugg, D.S., Schnabel, A., 1993. Effective sizes for subdivided populations. Genetics 135, 1221–1232.

Cockerham, C.C., Weir, B.S., 1993. Estimation of gene flow from *F*-statistics. Evolution 47, 855–863.

Collevatti, R.G., Grattapaglia, D., Hay, J.D., 2001. High resolution microsatellite based analysis of the mating system allows the detection of significant biparental inbreeding in *Caryocar brasiliense*, an endangered tropical tree species. Heredity 86, 60–67.

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

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

Dick, C.W., 2001. Genetic rescue of remnant tropical trees by an alien pollinator. Proc. R. Soc. Lond. Ser. B 268, 2391–2396.

Dick, C.W., Etchelecu, G., Austerlitz, F., 2003. Pollen dispersal of Neotropical trees (*Dinizia excelsa*: Fabaceae) by native insects and African honeybees in pristine and fragmented Amazonian rainforest. Mol. Ecol. 12, 753–764.

Doligez, A., Joly, H.I., 1997. Genetic diversity and spatial structure within a natural stand of a tropical forest tree species, *Carapa procera* (Meliaceae), in French Guiana. Heredity 79, 72–82.

Dow, B.D., Ashley, M.V., 1998. High levels of gene flow in bur oak revealed by paternity analysis using microsatellites. Heredity 89, 62–70.

Dyer, R.J., 2002. Contemporary pollen movement in shortleaf pine, *Pinus echinata* Mill. Unpublished Ph.D. Dissertation, Department of Biology, University of Missouri, St. Louis, pp. 140.

Dyer, R.J., Sork, V.L., 2001. Pollen pool heterogeneity in shortleaf pine, *Pinus echinata* Mill. Mol. Ecol. 10, 859–866.

Dyer, R.J., Westfall, R.D., Sork, V.L., Smouse, P.E., 2004. Two-generation analysis of pollen flow across a landscape. V. A stepwise approach for extracting factors contributing to pollen structure. Heredity 92, 204–211.

Ehrlich, P.R., 1996. Conservation in temperate forests: what do we need to know and do? For. Ecol. Manage. 85, 9–19.

El-Kassaby, Y.A., Jaquish, B., 1996. Population density and mating pattern in western larch. J. Hered. 87, 438–443.

Ellstrand, N.C., Elam, D.R., 1993. Population genetic consequences of small population size: implications for plant conservation. Ann. Rev. Ecol. Syst. 23, 217–242.

Essen-Moller, E., 1938. Die Beweiskraft der Ähnlichkeit im Vatershaftsnachweis. Theoretische Grundlagen. Mitt. Anthropol. Ges. Wien 68, 9–53.

Excoffier, L., Smouse, P.E., Quattro, J.M., 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131, 479–491.

Fernández-Manjarrés, J.F., 2002. Genetic diversity and mating system in a fragmented population of the Andean oak *Quercus humboldtii* Bonpl. (Fagaceae). Unpublished Ph.D. Dissertation, University of Missouri, St. Louis.

Finkeldey, R., 1995. Homogeneity of pollen allele frequencies of single seed trees in *Picea abies* (L.) Karst. plantations. Heredity 74, 451–463.

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

Greef, J.M., Nason, J.D., Compton, S.G., 2001. Skewed paternity and sex allocation in hermaphroditic plants and animals. Proc. R. Soc. B 268, 2143–2147.

Irwin, A.J., Hamrick, J.L., Godt, M.J.W., Smouse, P.E., 2003. A multi-year estimate of the effective pollen donor pool for *Albizia julibrissin*. Heredity 90, 187–194.

James, T., Vege, S., Aldrich, P., Hamrick, J.L., 1998. Mating systems of three tropical dry forest tree species. Biotropica 30, 587–594.

Jones, A.G., Ardren, W.R., 2003. Methods of parentage analysis in natural populations. Mol. Ecol. 12, 2511–2523.

Kaufman, S.R., Smouse, P.E., Alvarez-Buylla, E.R., 1998. Pollen mediated gene flow and differential male reproductive success in a tropical pioneer tree, *Cecropia obtusifolia* Bertol. (Moraceae.). Heredity 81, 164–173.

Konuma, A., Tsumura, Y., Lee, C.T., Lee, S.L., Okuda, T., 2000. Estimation of gene flow in the tropical-rainforest tree *Neobalanocarpus heimii* (Dipterocarpaceae), inferred from paternity analysis. Mol. Ecol. 9, 1843–1852.

Lande, R., Barrowclough, G.F., 1987. Effective population size, genetic variation, and their use in population management. In: Soulé, M.E. (Ed.), Viable Populations for Conservation. Cambridge University Press, Cambridge, pp. 87–123.

Ledig, F.T., 1992. Human impacts on genetic diversity in forest ecosystems. Oikos 63, 87–108.

Loveless, M.D., Hamrick, J.L., 1984. Ecological determinants of genetic structure in plant populations. Ann. Rev. Ecol. Syst. 15, 65–95.

Meagher, T.R., Thompson, E.A., 1986. The relationship between single parent and parent pair genetic likelihoods in genealogy reconstruction. Theor. Popul. Biol. 29, 87–106.

Mitton, J.B., Latta, R.G., Rehfeldt, G.E., 1997. The pattern of inbreeding in Washoe pine and survival of inbred progeny under optimal environmental conditions. Silvae Genet. 46, 215–219.

Morgan, M.T., Connor, J.K., 2001. Using genetic markers to directly estimate male selection gradients. Evolution 55, 272–281.

Muona, O., Moran, G.F., Bell, J.C., 1991. Hierarchical patterns of correlated mating in *Acacia melanoxylon*. Genetics 127, 619–626.

Nason, J.D., Hamrick, J.L., 1997. Reproductive and genetic consequences of forest fragmentation: two case studies of neotropical canopy trees. J. Hered. 88, 264–276.

Nathan, R., Katul, G., Horn, H.S., Thomas, S.M., Oren, R., Avissar, R., Pacala, S.W., Levin, S.A., 2002. Mechanisms of long-distance dispersal of seeds by wind. Nature 418, 409–413.

Nei, M., 1972. Genetic distance between populations. Am. Nat. 106, 283–292.

Neigel, J.E., 1997. A comparison of alternative strategies for estimating gene flow from genetic markers. Ann. Rev. Ecol. Syst. 28, 105–128.

Perry, D.J., Bousquet, J., 2001. Genetic diversity and mating system of post-fire and post-harvest black spruce: an investiga-tion using codominant sequence-tagged-site (STS) markers. Can. J. For. Res. 31, 32–40.

Ritland, K., 1989. Correlated matings in the partial selfer *Mimulus guttatus*. Evolution 43, 848–859.

Ritland, K., 1990. A series of FORTRAN computer programs for estimating plant mating systems. J. Hered. 81, 235–237.

Robledo-Arnuncio, J.J., Smouse, P.E., Gil, L., Alia, R., 2004. Pollen movement under alternative silvicultural practices in native populations of Scots pine (*Pinus sylvestris* L.) in central Spain. For. Ecol. Manage. 197, 243–253.

Rocha, O.J., Aguilar, G., 2001a. Variation in the breeding behavior of the dry forest tree *Enterolobium cyclocarpum* (Guanacaste) in Costa Rica. Am. J. Bot. 88, 1600–1606.

Rocha, O.J., Aguilar, G., 2001b. Reproductive biology of the dry forest tree *Enterolobium cyclocarpum* (Guanacaste) in Costa Rica: a comparison between trees left in pastures and trees in continuous forest. Am. J. Bot. 88, 1607–1614.

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

Rousset, F., 1997. Genetic differentiation and estimation of gene flow from *F*-statistics under isolation by distance. Genetics 145, 1219–1228.

Schuster, W.S.F., Mitton, J.B., 2000. Paternity and gene dispersal in limber pine (*Pinus flexilis* James). Heredity 84, 348–361.

Schoen, D.J., Stewart, S.C., 1986. Variation in male reproductive investment and male reproductive success in white spruce. Evolution 40, 1109–1120.

Slatkin, M., 1993. Isolation by distance in equilibrium and nonequilibrium populations. Evolution 47, 264–279.

Smouse, P.E., Long, J.C., 1992. Matrix correlation analysis in anthropology and genetics. Yearbk. Phys. Anthropol. 35, 187–213.

Smouse, P.E., Meagher, T.R., 1994. Genetic analysis of male reproductive contributions in *Chamaelirium luteum* (L.) Gray (Liliaceae). Genetics 136, 313–322.

Smouse, P.E., Meagher, T.R., Kobak, C.J., 1999. Parentage analysis in *Chamaelirium luteum* (L.): why do some males have disproportionate reproductive contributions? J. Evol. Biol. 12, 1069–1077.

Smouse, P.E., Dyer, R.J., Westfall, R.D., Sork, V.L., 2001. Two-generation analysis of pollen flow across a landscape. I. Male gamete heterogeneity among females. Evolution 55, 260–271.

Sork, V.L., Campbell, D., Dyer, R., Fernández-Manjarrés, J.F., Nason, J., Petit, R., Smouse, P., Steinberg, E., 1998. In: Proceedings from a Workshop on Gene Flow in Fragmented, Managed, and Continuous Populations. National Center for Ecological Analysis and Synthesis, Santa Barbara, CA. http://www.nceas.ucsb.edu/papers/geneflow.

Sork, V.L., Nason, J., Campbell, D.R., Fernández-Manjarrés, J.F., 1999. Landscape approaches to the study of gene flow in plants. Trends Ecol. E 142, 219–224.

Sork, V.L., Davis, F., Smouse, P.E., Apsit, V., Dyer, R., Fernández-Manjarrés, J.F., Kuhn, B., 2002. Pollen movement in declining populations of California valley oak, *Quercus lobata*: where have all the fathers gone? Mol. Ecol. 11, 1657–1668.

Stacy, E.A., Hamrick, J.L., Nason, J.D., Hubbell, S.P., Foster, R.B., Condit, R., 1996. Pollen dispersal in low-density populations of three neotropical tree species. Am. Nat. 148, 275–298.

Streiff, R., Ducousso, A., Lexer, C., Steinkellner, H., Gloessl, J., Kremmer, A., 1999. Pollen dispersal inferred from paternity analysis in a mixed oak stand of *Quercus robur* L. and *Q. petraea* (Matt.) Liebl. Mol. Ecol. 8, 831–841.

Templeton, A.R., Shaw, K., Routman, E., Davis, S.K., 1990. The genetic consequences of habitat fragmentation. Ann. Mol. Bot. Gard. 77, 13–27.

White, G.M., Boshier, D.H., Powell, W., 2002. Increased pollen flow counteracts fragmentation in a tropical dry forest: an example from *Swietenia humilis* Zuccarini. Proc. Natl. Acad. Sci. U.S.A. 19, 19 2038–2042.