Evolution, 55(2), 2001, pp. 260–271

TWO-GENERATION ANALYSIS OF POLLEN FLOW ACROSS A LANDSCAPE. I. MALE GAMETE HETEROGENEITY AMONG FEMALES

Peter E. Smouse,[1] Rodney J. Dyer,[2] Robert D. Westfall,[3] and Victoria L. Sork[2]
[1]*Department of Ecology, Evolution and Natural Resources, Cook College, Rutgers University, New Brunswick, New Jersey 08901-8551 E-mail: Smouse@AESOP.Rutgers.Edu*
[2]*Department of Biology, University of Missouri, St. Louis, Missouri 63121-4499*
[3]*Institute of Forest Genetics, United States Department of Agriculture Forest Service, Pacific Southwest Research Station, P.O. Box 245, Berkeley, California 94701*

Abstract.—Gene flow is a key factor in the spatial genetic structure in spatially distributed species. Evolutionary biologists interested in microevolutionary processess and conservation biologists interested in the impact of landscape change require a method that measures the real time process of gene movement. We present a novel two-generation (parent-offspring) approach to the study of genetic structure (TwoGener) that allows us to quantify heterogeneity among the male gamete pools sampled by maternal trees scattered across the landscape and to estimate mean pollination distance and effective neighborhood size. First, we describe the model's elements: genetic distance matrices to estimate intergametic distances, molecular analysis of variance to determine whether pollen profiles differ among mothers, and optimal sampling considerations. Second, we evaluate the model's effectiveness by simulating spatially distributed populations. Spatial heterogeneity in male gametes can be estimated by Φ_{FT}, a male gametic analogue of Wright's F_{ST} and an inverse function of mean pollination distance. We illustrate TwoGener in cases where the male gamete can be categorically or ambiguously determined. This approach does not require the high level of genetic resolution needed by parentage analysis, but the ambiguous case is vulnerable to bias in the absence of adequate genetic resolution. Finally, we apply TwoGener to an empirical study of *Quercus alba* in Missouri Ozark forests. We find that $\Phi_{FT} = 0.06$, translating into about eight effective pollen donors per female and an effective pollination neighborhood as a circle of radius about 17 m. Effective pollen movement in *Q. alba* is more restricted than previously realized, even though pollen is capable of moving large distances. This case study illustrates that, with a modest investment in field survey and laboratory analysis, the TwoGener approach permits inferences about landscape-level gene movements.

Key words.—AMOVA, gene flow, genetic structure, parentage analysis, Φ-statistics, pollen movement, *Quercus alba*, TwoGener.

Received January 4, 2000. Accepted September 1, 2000.

The questions that drive our studies of gene flow are changing. As evolutionary biologists explore the mechanisms creating genetic structure, they need a direct means of assessing ongoing gene movement. Moreover, instead of assuming that propagule movement is independent of environmental context, evolutionary biologists are increasingly interested in how landscape features influence the paths of gene flow (e.g., Husband and Barrett 1996; Kudoh and Whigham 1997). Simultaneously, conservation biologists have become concerned that disruption of gene flow processes by anthropogenic landscape change might isolate populations and make them vulnerable to loss of genetic variation, leading to a loss of fitness through genetic drift and inbreeding (Ledig 1992; Ellstrand and Elam 1993). In managed forests, it is necessary to assess how forest fragmentation influences propagule flow, neighborhood size, inbreeding, and pollen contamination of seed orchards (Adams and Birkes 1991; Adams et al. 1992a,b; Ledig 1992). For all these questions, we need to know how far propagules are moving now and how changing population spatial arrays and landscape features impact those movement patterns. We also need to measure contemporary gene movement, rather than its long-term average.

Historically, research on gene flow has emphasized the evolutionary consequences of gene flow for population differentiation and species cohesion (Neigel 1997). Traditional *F*-statistic models (Wright 1969) and recently developed coalescent models (Slatkin 1989; Hudson 1991) provide indirect estimates of the average *effective* number of migrants exchanged per generation among a set of populations. Both treatments require assumptions of evolutionary (migration-drift) equilibrium and selective neutrality that limit their utility for the study of contemporary gene flow (Bossart and Prowell 1998) because contemporary genetic affinities among populations are confounded by processes other than gene flow and widespread anthropogenic landscape alteration may change gene flow from its historical pattern, thus confounding any attempt to estimate contemporary gene flow patterns (Sork et al. 1998, 1999). Genetic structure analysis provides an estimate of long-term effective gene flow, but it fails to reveal the contribution of (sometimes changing) demographic processes to that gene flow.

Many researchers have switched to a more direct approach, using parentage analysis, which provides a means of quantifying localized gene movements. Using genetic markers, statistical treatments are now available to gauge pollen-mediated gene movement within and among populations (Devlin and Ellstrand 1990; Adams and Birkes 1991; Smouse et al. 1999). The model can be used to estimate pollen immigration into a reference site, but in most cases, it is not possible to identify where the pollen came from (but see Kaufman et al. 1998). In spite of valuable detail on local gene movement (e.g., Chase et al. 1996; Dow and Ashley 1996; Nason and Hamrick 1997; Streiff et al. 1998), this approach will always be constrained by the need to characterize a large number of local pollen donors and a large number of progeny per female, requiring very large sample sizes.

For most plant species, pollen is most responsible for gene movement (Ennos 1994). The first step toward understanding the process of pollen flow is to increase the number of sites and the spatial scale of assessment. In doing so, we can begin to address questions about the pattern of pollen flow, relative to the demographic and environmental factors that emerge from the current ecological context or recent landscape history. For these questions, the exact parentage of any given seedling is no more than a means to an end. Instead of identifying the father, we need to answer several questions. What is the real time pattern of pollen flow? What can we infer about effective pollination distance and effective neighborhood size? What is the impact of differential/changing ecological context on pollen flow and its population structure consequences? We need an analysis that permits sampling a wider array of situations, one that can be mounted within the scope of feasible sampling and laboratory efforts.

We develop a two-generation procedure, dubbed TwoGener, which combines the survey simplicity of the population structure approach with the parent-offspring-deductive aspects of the parentage approach. In the first section of this paper, we develop the estimation strategy that underlies TwoGener and, using information on the genetic distances between male gametes, we describe how to answer the initial question of whether different females, spread across the landscape, are sampling from heterogeneous male gamete pools. We then explore the sampling aspects of pollen movement, characterized by isolation by distance, and indicate how to allocate the sampling effort within and among females.

In the second section of this paper, we simulate a spatially distributed population, and explore three specific objectives that clarify the application of our model to real data: (1) we describe and contrast two cases, that of categorical gametic assay, possible with conifers, and that of ambiguous gametic assay, typical of angiosperms; (2) we explore the strength of isolation by (pollination) distance, as a function of the decay parameter of our pollen distribution; and (3) we evaluate the statistical resolution provided by genetic loci of different parentage exclusionary power.

In the third section, we apply TwoGener to data from a natural population of *Quercus alba*, which was collected as part of the Missouri Ozark Forest Ecosystem Project, designed to address gene flow patterns on a landscape scale.

Intergametic Distance Analysis

Sampling Strategy

The TwoGener model is based on an analysis of genetic distances among male gametes. The first step is to determine the impact of spatial separation among females on the heterogeneity of pollen donors they sample. In essence, we use the females as spatially distributed ''pollen traps'' (as in Sorensen 1972). The sampling design requires some pairs in close proximity (within a single patch), some at intermediate distances (e.g., single trees in a dispersed population), and some (from end to end of the study) that will surely sample pollen pools that are effectively nonoverlapping. We sample J females, and from the jth female we extract n_j seed. Because the total sample size (N) is commonly constrained by limited field and laboratory resources, we set $N = \Sigma_{i=1}^{J} n_i = 400$ progeny here, a manageable sample for most studies, particularly if we must evaluate multiple populations. We are faced with a trade-off between replication for each female and the number of females sampled. For most of what follows, we use a sample design of $K = n_j = 20$ for each of $J = 20$ females. From each seedling, we infer both paternal and maternal gametic contributions and use the paternal gametes to gauge the heterogeneity of the pollen pools sampled by different females.

Gametic Assay

It has been shown that genetic inference improves with a multilocus treatment (Smouse et al. 1982). Therefore, consider a set of L codominant loci reasonably presumed to be unlinked and segregating independently, for which we can characterize each sampled female. Consider the genotype of the jth female, for example, A_1A_3, B_2B_2, C_3C_4,. . . , L_4L_4. This female is assumed to show meiotic segregation for the heterozygous A- and C-loci, presumably in Mendelian proportions ½:½, but does not segregate for the homozygous B- and L-loci. We next examine the genotypes of n_j seedlings, derived from this same mother. We imagine two cases: (1) both maternal and paternal gametes are categorically obvious by inspection; and (2) some ambiguity exists. The first (categorical) case would apply to conifers, for which a separate assay of the megagametophyte associated with each seed would indicate the maternal contribution categorically; the male contribution could then be obtained by subtraction. The second (ambiguous) case would apply to angiosperms, where some mother-offpring pairs are of genotypes G_iG_k and G_iG_k, for which both paternal and maternal gametic contributions are ambiguous.

For our jth female, Table 1a provides the gametic inference available from a collection of four seedlings and the corresponding megagametophyte (for a conifer), and Table 1b provides the comparable inference for the angiosperm case. Under the conifer scenario, we have categorical assay for both maternal and paternal gametic contributions and the post hoc likelihoods of those gametes are all one, given the observed maternal, seedling, and megagametophytic genotypes. In the angiosperm case, where direct gametic assay is not available, we have some ambiguity. The first two seedlings both yield obvious maternal and paternal gametic genotypes, but the third seedling is ambiguous for the C-locus. The maternal probabilities are the Mendelian proportions ½:½, but the paternal proportions are determined by the frequencies of C_3 and C_4 in the pollen pool, q_3 and q_4, respectively. The posterior likelihoods of paternal-maternal gametic combinations, given the maternal and seedling genotypes, are $\gamma_{34} = ½q_3/(½q_3 + ½q_4)$ for the C_3-C_4 pair and $\gamma_{43} = ½q_4/(½q_3 + ½q_4) = (1 - \gamma_{34})$, for the C_4-C_3 pair, respectively. For the fourth seedling, both the A- and C-loci are ambiguous, yielding four possible paternal-maternal gametic combinations. The four maternal gametes have expected proportions of $(½)^2 = ¼$ each, whereas the expected male proportions are determined by the allele frequencies at both the C-locus (given above) and the A-locus (p_1 and p_3). The posterior paternal-maternal gametic likelihoods, given the maternal and seedling genotypes, are determined by γ_{34} and γ_{43} (as defined above) and $\alpha_{13} = ½p_1/(½p_1 + ½p_3)$ and $\alpha_{31} = ½p_3/(½p_1 + ½p_3)$, as shown in Table 1b.

TABLE 1. Inference on gametic genotypes of a A_1A_3, B_2B_2, C_3C_4, . . . , L_4L_4 mother and various fathers: (a) for categorical assay (as with a conifer, having separate megagametophytic assay); and (b) for ambiguous assay (as with an angiosperm, having some ambiguity).

(a) Categorical assay, with known male and female gametes

Diploid offspring genotype	Megagamete (maternal) contribution		Microgramete (paternal) contribution	Posterior gametic likelihood
A_1A_2, B_2B_2, C_2C_4, . . . , L_2L_4	A_1 B_2 C_4 . . . L_4	+	A_2 B_2 C_2 . . . L_2	1
A_3A_3, B_2B_2, C_1C_3, . . . , L_1L_4	A_3 B_2 C_3 . . . L_4	+	A_3 B_2 C_1 . . . L_1	1
A_2A_3, B_2B_2, C_3C_4, . . . , L_2L_4	A_3 B_2 C_4 . . . L_4	+	A_2 B_2 C_3 . . . L_2	1
A_1A_3, B_2B_2, C_3C_4, . . . , L_3L_4	A_1 B_2 C_3 . . . L_4	+	A_3 B_2 C_4 . . . L_3	1

(b) Ambiguous assay, with some ambiguous male and female gametes

Diploid offspring genotype	Maternal gametic contribution		Paternal gametic contribution	Posterior gametic likelihood
A_1A_2, B_2B_2, C_2C_4, . . . , L_2L_4	A_1 B_2 C_4 . . . L_4	+	A_2 B_2 C_2 . . . L_2	1
A_3A_3, B_2B_2, C_1C_3, . . . , L_1L_4	A_3 B_2 C_3 . . . L_4	+	A_3 B_2 C_1 . . . L_1	1
A_2A_3, B_2B_2, C_3C_4, . . . , L_2L_4	A_3 B_2 C_4 . . . L_4	+	A_2 B_2 C_3 . . . L_2	γ_{34}
		or		
	A_3 B_2 C_3 . . . L_4	+	A_2 B_2 C_4 . . . L_2	γ_{43}
A_1A_3, B_2B_2, C_3C_4, . . . , L_3L_4	A_1 B_2 C_3 . . . L_4	+	A_3 B_2 C_4 . . . L_3	$\alpha_{31}\gamma_{43}$
		or		
	A_1 B_2 C_4 . . . L_4	+	A_3 B_2 C_3 . . . L_3	$\alpha_{31}\gamma_{34}$
		or		
	A_3 B_2 C_3 . . . L_4	+	A_1 B_2 C_4 . . . L_3	$\alpha_{13}\gamma_{43}$
		or		
	A_3 B_2 C_4 . . . L_4	+	A_1 B_2 C_3 . . . L_3	$\alpha_{13}\gamma_{34}$

The only maternal-offspring combinations that yield ambiguity are those of the type G_iG_k-G_iG_k for a particular locus, those where mother and offspring share the same pair of alleles in heterozygous form. Clearly, categorical assay is better, but it is possible to use the ambiguous treatment. If we are forced to make allele frequency assumptions, it seems best to use the average allele frequency estimates for the entire male pollen profile of the study, rather than using a separate set of estimates from the pollen profile of each female. That will provide common (and well-estimated) values for the whole study, but to the extent that different females are drawing from pollen pools with different allele frequencies, it will favor the null hypothesis of no pollen profile divergence among those females and, thus, the degree of pollen structure among them. We are trading robust estimation (of the average) against statistical power (to detect small differences), but over the spatial scale of interest here, that trade-off seems reasonable.

Distance Metrics

The intent of this component of the model is to determine whether the average pollen profiles for different females are different. We gauge divergence among male gametes by means of pairwise genetic distances. We then use these pairwise distances, from the same and from different females, to gauge the degree of pollen structure among those females. To measure genetic distance, we prefer multiple-allelic loci, because paternal gametic resolution improves with the level of polymorphism (Chakravarti and Li 1983; Jamieson 1994). The four-allele case is sufficient to describe the scoring convention. We can describe the distances between different alleles with an equilateral tetrahedron (a perfect pyramid), with each vertex representing an allele and each edge the distance between a pair of alleles (Fig. 1a). We set each edge length to unity (one), because the alleles are assumed to be equally divergent.

We can also obtain the distance in more formal fashion. We use the vector representation in Table 2a and define the *squared* genetic distance between the ith and kth gametes as

$$d_{ik}^2 = \frac{1}{2}\sum_{a=1}^{H}(Y_{ia} - Y_{ka})^2 = \frac{1}{2}[\mathbf{Y}_i - \mathbf{Y}_k]^T[\mathbf{Y}_i - \mathbf{Y}_k], \quad (1)$$

where $\mathbf{Y}_i$ and $\mathbf{Y}_k$ are the vectors for the ith and kth gametes and H is the total number of different alleles at this locus. Equation (1) is defined in such a fashion as to provide the same values as the tetrahedron (Fig. 1a). The multilocus squared distance is simply the sum of the squared distances for the separate loci, a tally of the allelic differences at the L loci

$$d_{ik}^2(L \text{ loci}) = \sum_{l=1}^{L} d_{ik}^2(l\text{th locus}). \quad (2)$$

For the categorical case, this is all we need.

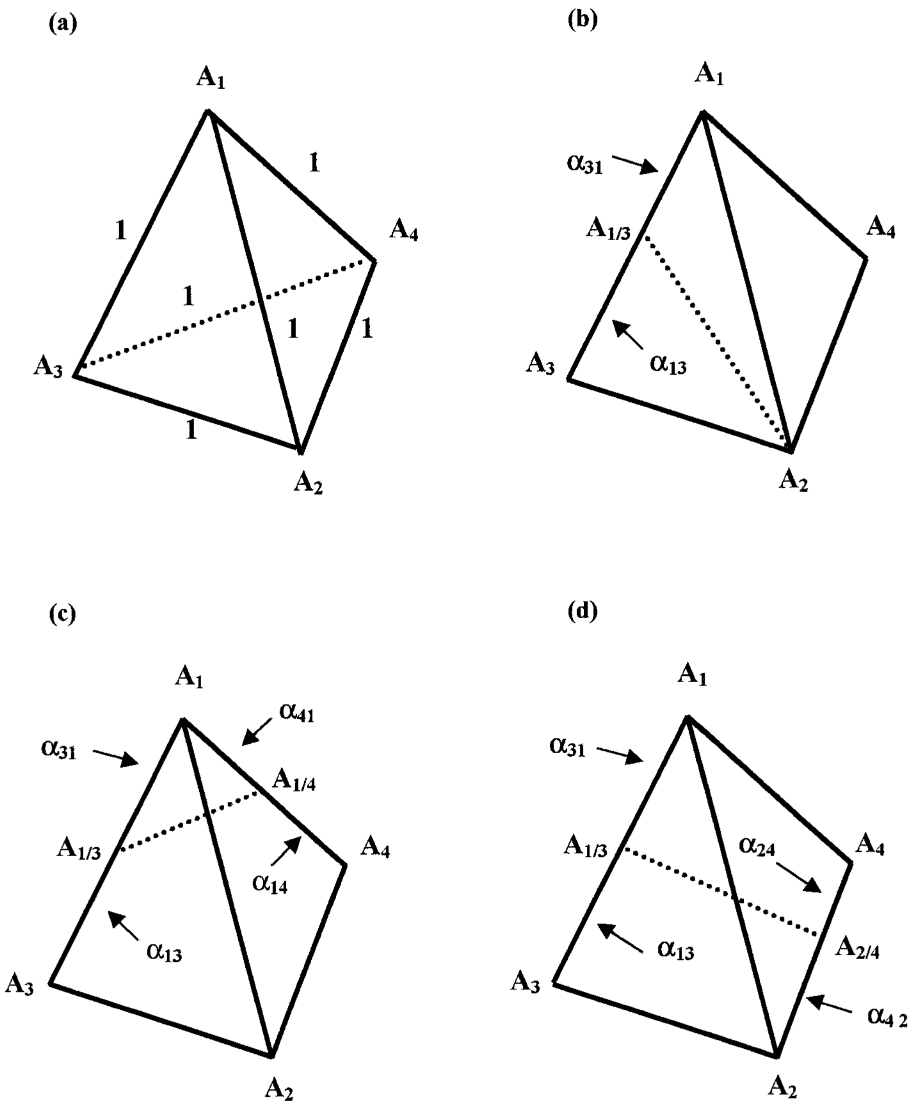

FIG. 1. Intergametic distance tetrahedra: (a) categorical assay of paternal alleles, comparing all pairs of alleles; (b) ambiguous assay of paternal alleles, for the case of mother-offspring pairs sharing the same heterozygous diploid genotype, comparing $A_{1/3}$ with A_2; (c) ambiguous case, comparing $A_{1/3}$ with $A_{1/4}$; (d) ambiguous case, comparing $A_{1/3}$ with $A_{2/4}$.

For the ambiguous case, the treatment is a bit more elaborate and is illustrated in Figure 1b and in Table 2b. Consider the fourth seedling, ambiguous for the parental alleles A_1 and A_3 and for the parental alleles C_3 and C_4. Consider the A-locus alleles; the paternal allele is A_1 (and maternal allele A_3) with likelihood $\alpha_{13,}$ and the paternal allele is A_3 (and maternal allele A_1) with likelihood α_{31}. The appropriate vector representation (Table 2) of the paternal gamete is

$$\mathbf{Y}_{1/3} = \alpha_{13}\begin{vmatrix}1\\0\\0\\0\end{vmatrix} + \alpha_{31}\begin{vmatrix}0\\0\\1\\0\end{vmatrix} = \begin{vmatrix}\alpha_{13}\\0\\\alpha_{31}\\0\end{vmatrix}, \qquad (3)$$

with the positions of α_{13} and α_{31} reversed in the maternal ($A_{3/1}$) gamete. Concentrating on the paternal gamete, we have the representation in Figure 1b, with $A_{1/3}$ gamete located along the side of the tetrahedron, at distance α_{13} from the A_3 vertex and distance α_{31} from the A_1 vertex. (The ambiguous gamete is closer to the allelic vertex with the higher frequency in the total population of paternal gametes.) The $A_{1/3}$ gamete's vector difference from a categorical A_1 gamete is

$$[\mathbf{Y}_{1/3} - \mathbf{Y}_1] = \begin{vmatrix}\alpha_{13}\\0\\\alpha_{31}\\0\end{vmatrix} - \begin{vmatrix}1\\0\\0\\0\end{vmatrix} = \begin{vmatrix}-\alpha_{31}\\0\\\alpha_{31}\\0\end{vmatrix}, \qquad (4)$$

and its squared distance to the A_1 vertex is

$$d_{ik}^2(\mathrm{A}_{1/3} \text{ vs. } \mathrm{A}_1) = \frac{1}{2}[\mathbf{Y}_{1/3} - \mathbf{Y}_1]^T[\mathbf{Y}_{1/3} - \mathbf{Y}_1] = \alpha_{31}^2. \qquad (5)$$

Similar arguments, using either the geometry of the tetrahedron (Fig. 1b–d) or vector algebra (Table 2b), yield

TABLE 2. Four-allele scoring of paternal (p) and maternal (m) gametes for (a) categorical and (b) ambiguous gametic analysis illustrated for the A_1A_3, B_2B_2, C_3C_4, . . . , L_4L_4–A_1A_3, B_2B_2, C_3C_4, . . . , L_3L_4 mother-offspring pair of Table 1; each **Y**-variable corresponds to one allele of a four-allele locus.

(a) Four-allele, categorical scoring for the fourth seedling in Table 1a

Gametic vector		A-locus		B-locus		C-locus		L-locus	
		p	m	p	m	p	m	p	m
$\mathbf{Y}_1$	=	0	1	0	0	0	0	0	0
$\mathbf{Y}_2$		0	0	1	1	0	0	0	0
$\mathbf{Y}_3$		1	0	0	0	0	1	1	0
$\mathbf{Y}_4$		0	0	0	0	1	0	0	1

(b) Ambiguous scoring for the fourth seedling in Table 1b, with posterior likelihood weights

Gametic vector		A-locus		B-locus		C-locus		L-locus	
		p	m	p	m	p	m	p	m
$\mathbf{Y}_1$	=	α_{13}	α_{31}	0	0	0	0	0	0
$\mathbf{Y}_2$		0	0	1	1	0	0	0	0
$\mathbf{Y}_3$		α_{31}	α_{13}	0	0	γ_{34}	γ_{43}	1	0
$\mathbf{Y}_4$		0	0	0	0	γ_{43}	γ_{34}	0	1

$\alpha_{13} = \mathrm{fr}(A_1)/[\mathrm{fr}(A_1) + \mathrm{fr}(A_3)]$ $\alpha_{31} = \mathrm{fr}(A_3)/[\mathrm{fr}(A_1) + \mathrm{fr}(A_3)]$

$\gamma_{34} = \mathrm{fr}(C_3)/[\mathrm{fr}(C_3) + \mathrm{fr}(C_4)]$ $\gamma_{43} = \mathrm{fr}(C_4)/[\mathrm{fr}(C_3) + \mathrm{fr}(C_4)]$

$$d^2_{ik}(A_{1/3} \text{ vs. } A_3) = \alpha^2_{13}, \tag{6a}$$

$$d^2_{ik}(A_{1/3} \text{ vs. } A_2) = 1 - \alpha_{13}\alpha_{31}, \tag{6b}$$

$$d^2_{ik}(A_{1/3} \text{ vs. } A_{2/4}) = 1 - \alpha_{13}\alpha_{31} - \alpha_{24}\alpha_{42}, \tag{6c}$$

$$d^2_{ik}(A_{1/3} \text{ vs. } A_{1/4}) = 1 - \alpha_{13}\alpha_{31} - \alpha_{14}\alpha_{41} - \alpha_{13}\alpha_{14}, \tag{6d}$$

and

$$d^2_{ik}(A_{1/3} \text{ vs. } A_{1/3}) = 0, \tag{6e}$$

with additional definitions for $A_{2/4}$ and $A_{1/4}$ heterozygous mother-offspring pairs,

$$\alpha_{24} = \frac{1}{2}p_2 \bigg/ \left(\frac{1}{2}p_2 + \frac{1}{2}p_4\right), \tag{7a}$$

$$\alpha_{42} = \frac{1}{2}p_4 \bigg/ \left(\frac{1}{2}p_2 + \frac{1}{2}p_4\right), \tag{7b}$$

$$\alpha_{14} = \frac{1}{2}p_1 \bigg/ \left(\frac{1}{2}p_1 + \frac{1}{2}p_4\right), \quad \text{and} \tag{7c}$$

$$\alpha_{41} = \frac{1}{2}p_4 \bigg/ \left(\frac{1}{2}p_1 + \frac{1}{2}p_4\right). \tag{7d}$$

The same sort of treatment applies to the ambiguous C-locus or any other ambiguous locus. In any case, we can compute squared distances between all pairs of alleles, regardless whether there is gametic ambiguity. For the purposes to follow, squared distances are sufficient, albeit sometimes algebraically cumbersome. To obtain the multiple-locus distance, we simply add across loci, as in equation (2).

The Intergametic Distance Matrix

Given N seedlings, we have N paternal gametes. However, all the information is contained within the full set of squared intergametic distances, and further analysis amounts to manipulation of those squared distances. It is convenient to pack the pairwise distances into an $N \times N$ squared distance matrix

$$\mathbf{D} = \begin{vmatrix} \mathbf{D}_{11} & \mathbf{D}_{12} & \mathbf{D}_{13} & \cdots & \mathbf{D}_{1J} \\ \mathbf{D}_{21} & \mathbf{D}_{22} & \mathbf{D}_{23} & \cdots & \mathbf{D}_{2J} \\ \mathbf{D}_{31} & \mathbf{D}_{32} & \mathbf{D}_{33} & \cdots & \mathbf{D}_{3J} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \mathbf{D}_{J1} & \mathbf{D}_{J2} & \mathbf{D}_{J3} & \cdots & \mathbf{D}_{JJ} \end{vmatrix}, \tag{8}$$

where $\mathbf{D}_{gg}$ is the matrix of pairwise distances among the paternal gametes of the gth female, and $\mathbf{D}_{gh}$ is the matrix of pairwise distances between paternal gametes of the gth and hth females.

Analysis of Molecular Variance

Our analytical task is to determine whether the pollen profiles for different females were random samples from the same pollen pool or whether different females have sampled different collections of males by virtue of their own physical isolation and the tendency for pollen to fall close to its donor. We hope to be able to show that females have sampled from different pollen pools, but the null hypothesis is that they have not. We answer this question with an analysis of molecular variance (AMOVA; Excoffier et al. 1992), using females as the strata and paternal gametes within them to establish replication error. AMOVA is essentially a multivariate analysis of variance (MANOVA), but it uses the pairwise distance matrix as input, rather than the raw **Y**-vectors themselves. If we were dealing with diploid genotypes, rather than gametes, we would have results that were equivalent to the standard Weir and Cockerham (1984) multilocus F-statistics extraction, as we have described elsewhere (Peakall et al. 1995). The other novelty is that instead of using Normal theory test procedures for multinomial variables, we use permutational (data reuse) procedures. Briefly, we need a classic one-level nest, separating variation into within-female and among-female components of pollen variation. AMOVA computes these components from the distance matrix, **D,** obtained from equation (8), among paternal gametes. The usual molecular sums of squares, estimated mean squares, and estimated variance components are shown in Table 3. Instead of using Fisher's F-ratio, it is customary to use the intraclass correlation, Φ_{FT}, which is estimated as the fraction of the total variance that is accounted for by interfemale differences, analogous to Wright's (1969) F_{ST}-statistic, but with females replacing populations as strata and male gametes replacing diploid individuals as replicates within strata,

$$\hat{\Phi}_{\mathrm{FT}} = s^2_{\mathrm{A}}/(s^2_{\mathrm{A}} + s^2_{\mathrm{W}}). \tag{9}$$

Significance testing is conducted by randomly shuffling the male gametes among females, on the null premise that if there are no differences among the pollen pools, it will not matter which collection of male gametes is associated with which female. That shuffling is done by permuting rows (and corresponding columns) of **D** (see Excoffier et al. 1992).

TABLE 3. Analysis of molecular variance (AMOVA), adapted to determine heterogeneity of pollen gene pools among individual females; J, number of females; K, number of progeny for each female; SS_A, sum of squares among females; SS_W, sum of squares within females; MS_A, mean square among females; MS_W, mean square within females; s^2_A, estimated variance among females; σ^2_A, parametric variance among females; s^2_W, estimated variance within females; σ^2_W, parametric variance within females; Φ_{FT}, heterogeneity estimate.

Source of variation	Degrees of freedom	Sums of squares	Estimated mean squares	Expected mean squares	Variance and heterogeneity estimates
Among female sibships	$(J - 1)$	SS_A	$MS_A = \frac{SS_A}{(J - 1)}$	$\sigma^2_W + K\sigma^2_A$	$s^2_A = \frac{MS_A - MS_W}{K}$
Within female sibships	$J(K - 1)$	SS_W	$MS_W = \frac{SS_W}{J(K - 1)}$	σ^2_W	$s^2_W = MS_W$
					$\Phi_{FT} = \frac{s^2_A}{s^2_W + s^2_A}$

Allocation of Sampling Effort

The total size of a study is constrained by limited resources, both in terms of field sampling and laboratory assay. It is cheaper to obtain more seed from a given tree than to sample another tree. However, for a fixed total sample size, the more maternal parents we sample, the better our coverage of the landscape and associated auxiliary variables, but the fewer male gametes we will have to characterize the pollen pool of each mother's pollen draw.

For the analysis at hand, we were interested in estimating Φ_{FT}. To do that with any precision, we needed a balance between the number of females (J) and the number of progeny (K) per female. The best choice of J and K, for fixed $N = JK$, depends on the unknown value of Φ_{FT} (the intraclass correlation), because the variance of Φ_{FT} is (Falconer 1981)

$$s^2_\Phi = 2\frac{[1 + (K - 1)\Phi_{FT}]^2(1 - \Phi_{FT})^2}{K(K - 1)(J - 1)}. \qquad (10)$$

By examining the variance of the intraclass correlation, equation (10), over a range of J, K, and Φ_{FT} values, we discover that the minimum variance is achieved when $K = (\Phi_{FT})^{-1}$. To increase overall precision, we must increase total sample size, N. For a fixed total sample size of $N \approx 400$, for example, and a value of $\Phi_{FT} \approx 1/32$, sample allocations of $J = 12$, $K = 33$ or $J = 13$, $K = 31$ would be best, but for a value of $\Phi_{FT} \approx 1/8$, a sample allocation of $J = 50$, $K = 8$ would be optimal. To add an element of realism, we note that for *Quercus alba* $\Phi_{FT} \approx 1/16$ (see below), for which we should have $J = 25$ and $K = 16$. Detailed analysis, however, shows that the standard error of Φ_{FT} is relatively insensitive to the J:K ratio over this range of Φ_{FT} values. Any allocation between ($16{:}25 \leq J{:}K \leq 25{:}16$) should be adequate; for an intermediate value of $\Phi_{FT} = 1/16$, any allocation between 10:40 and 40:10 would make very little difference. On the strength of these theoretical considerations, we have chosen to evaluate an intermediate combination of 400 total progeny, involving $J = 20$ families and $K = 20$ progeny per family. To spread the females across our simulated landscape, 100 units on a side (see below), we have chosen the spacing scheme in Figure 2.

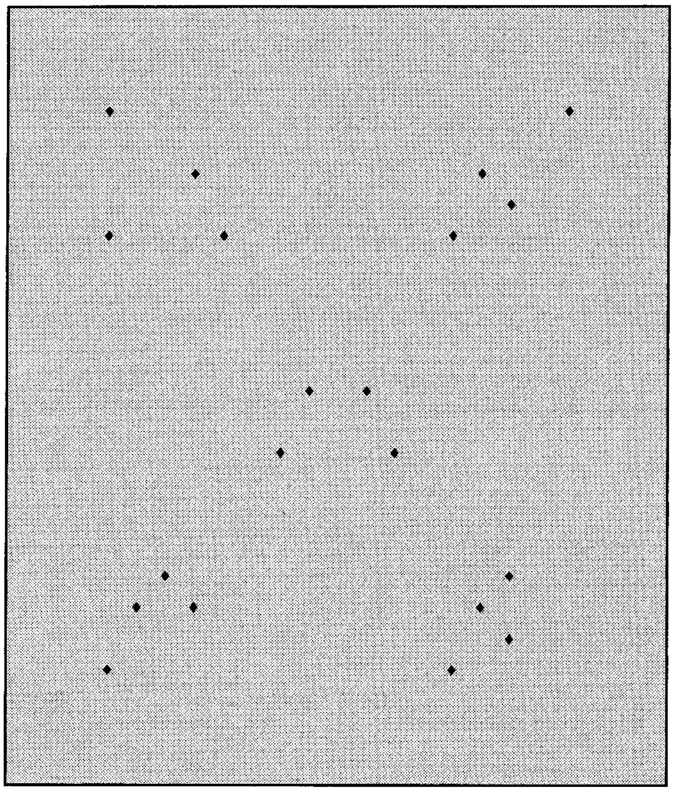

FIG. 2. Physical layout 10,000 simulated individuals, within a panmictic population, used to evaluate the power of the analysis under different conditions. The positions of the 20 sampled females are highlighted with diamonds.

SIMULATION ANALYSIS OF A UNIFORMLY DISTRIBUTED POPULATION

Methods

Reference population.—We benchmarked the TwoGener model with simulated reference populations consisting of 10,000 mature, hermaphroditic, diploid, eight-locus individuals, uniformly distributed across a 100 × 100 unit landscape (density, $d = 1$ per unit^2), represented by the speckling in Figure 2. We assigned genotypes to each individual, drawn from a panmictic gene pool array, corresponding to the allele frequencies of *Q. alba*, which we use later for illustration of this model (Table 4). For example, a two-allele locus in Hardy-Weinberg equilibrium will have genotype frequencies $Pr(A_1A_1) = p^2$, $Pr(A_2A_2) = q^2$, and $Pr(A_1A_2) = 2pq$. A cumulative frequency vector was created from these frequencies $[p^2, 2pq, q^2]$, and the genotype of the individual was determined by drawing a random number from the U(0,1) distri-

TABLE 4. Allele frequencies used for the simulations, drawn from the *Quercus alba* example. Most polymorphic loci are in the top panel, least polymorphic loci in the bottom panel; expected exclusion probabilities (E) are given for each locus. *Per-2,* peroxidase; *Tpi,* triosephosphate isomerase; *Adh,* alcohol dehydrogenase; *Fe-1* and *Fe-3,* fluorescent esterases 1 and 3; *Pgi-2,* phosphoglucoisomerase; *Pgm,* phosphoglucomutase; *MNR,* menadione reductase.

Per-2		*Tpi*		*Adh*		*Fe-1*	
Allele	Freq.	Allele	Freq.	Allele	Freq.	Allele	Freq.
1	0.000	1	0.004	1	0.001	1	0.001
2	0.000	2	0.000	2	0.000	2	0.000
3	0.189	3	0.469	3	0.363	3	0.622
4	0.231	4	0.000	4	0.000	4	0.000
5	0.580	5	0.527	5	0.631	5	0.376
7	0.000	7	0.000	7	0.005	7	0.001
E	0.3095		0.1923		0.1857		0.1822

Pgi-2		*Pgm*		*MNR*		*Fe-3*	
Allele	Freq.	Allele	Freq.	Allele	Freq.	Allele	Freq.
1	0.000	1	0.000	1	0.000	1	0.000
2	0.000	2	0.000	2	0.034	2	0.000
3	0.112	3	0.032	3	0.965	3	0.002
4	0.001	4	0.000	4	0.000	4	0.000
5	0.863	5	0.954	5	0.001	5	0.980
7	0.024	7	0.014	7	0.000	7	0.018
E	0.1177		0.0444		0.0329		0.0194

bution. If the random number x was $\leq p^2$, we assigned the genotype A_1A_1 to that individual. If $p^2 < x \leq (p^2 + q^2)$, then the individual was assigned the genotype A_2A_2. If $(p^2 + q^2) < x \leq 1$, then the genotype was A_1A_2. We repeated this process until all 10,000 individuals had eight-locus genotypes.

To illustrate the impact of differing amounts of polymorphic variation within the assay battery on the estimated values of Φ_{FT}, we modeled three different situations. Model 1 uses the four most polymorphic loci (top panel of Table 4), twice each. Model 2 uses the eight loci presented, typical of the genetic resolution found in a routine allozyme survey of forest trees. Model 3 uses the four least polymorphic loci (bottom panel of Table 4), twice each. Polymorphism is limited in this last case; the battery would usually be considered inadequate for productive parentage analysis and, as we shall see, it limits the utility of TwoGener as well.

The mating array.—The major feature of interest here is the isolation by distance that follows from limited pollen dispersal; the operative variable becomes the effective size of the pollen pool drawn by a particular female. For a two-dimensional landscape, with isotropic (nondirectional) pollen flow, we modeled the probability that a particular pollen grain was drawn from a male at distance z from the female (in any direction) as a negative exponential distribution of the form

$$f(z) = \lambda \exp\{-\lambda z\}, \tag{11}$$

where the expected (average) distance from which a pollen grain is drawn is $\gamma = \lambda^{-1}$. We used seven settings for γ to seed the pollen flow dynamics of the model, γ = 2.5, 5.0, 7.5, 10.0, 12.5, 15.0, and 50.0, the last approaching broadcast (panmictic) pollination on a landscape of this size.

We expect the estimated value of Φ_{FT} to increase as the average pollen dispersal distance decreases. That is, we expect Φ_{FT} to increase as γ decreases (or to increase as λ increases). For each sampled female, the locations of which are shown in Figure 2, we randomly selected a single allele from each locus. That provided the eight-locus maternal gamete. To determine the male who was mating with that female, we randomly selected a distance (z) measured from that female based on equation (11). We found the male that was closest to this random distance (z) from that female, irrespective of direction (isotropic pollen flow). We randomly selected corresponding alleles from each locus from that male to construct an eight-locus paternal gamete. We repeated the sampling process until all maternal trees had complete progeny arrays, each offspring represented by different maternal and paternal (eight-locus) gametes, a total of 400 gametic pairs for the study. For each simulated dataset, we extracted the matrix **D** (eq. 8) of distances among the inferred male gametes, conducted the AMOVA described in Table 3, and then tallied Φ_{FT}. For each set of initial parameters, we repeated this entire process 1000 times, allowing the construction of an empiric frequency distribution for Φ_{FT}. Simulation programs for population creation and TwoGener are available from R. J. Dyer upon request.

Results

The relationship between γ and Φ_{FT}.—Using the categorical treatment, we evaluated Φ_{FT} (denoted Φ_{FT}^{C}) for values of γ representing a range of isolation by distance from very proximal pollination to broadcast (virtually panmictic) pollination. For any given value of γ or λ, there was variation from run to run, but on average, Φ_{FT}^{C} was inversely proportional to γ (Fig. 3a). Thus, Φ_{FT}^{C} is positively proportional to (and linear in) $\lambda = \gamma^{-1}$ (Fig. 3b). The pattern and the conclusion were precisely the same for the ambiguous treatment (Φ_{FT}^{A}, results not shown). This relationship is the central result of the enterprise, and it represents what makes TwoGener work. Widely spaced females draw from different pollen pools, and the resulting differences in their male gamete arrays reflect the average distance of pollen flow.

Polymorphic resolution.—The ability to detect spatial structure of the pollen pool depends on the level of polymorphism of the genetic loci used. Traditional parentage resolution is dependent on the exclusion probability (E_L), the probability of being able to reject paternity for a random nonfather for the average mother-offspring pair with the genetic data available. The value of E_L has been worked out for the multiple-allelic, ambiguous case and is a standard result (Selvin 1980; Chakravarti and Li 1983; Smouse and Chakraborty 1986; Jamieson 1994). If E_l is the exclusion probability for the lth locus, then that for the complete L-locus battery is given by

$$E_L = 1 - \prod_{l=1}^{L} (1 - E_l). \tag{12}$$

Although we are not assigning paternity here, because we have no particular interest in the precise father, resolution does depend on the level of polymorphism. We have modeled three situations to illustrate the impact of differing amounts of polymorphic variation on the average value and variance

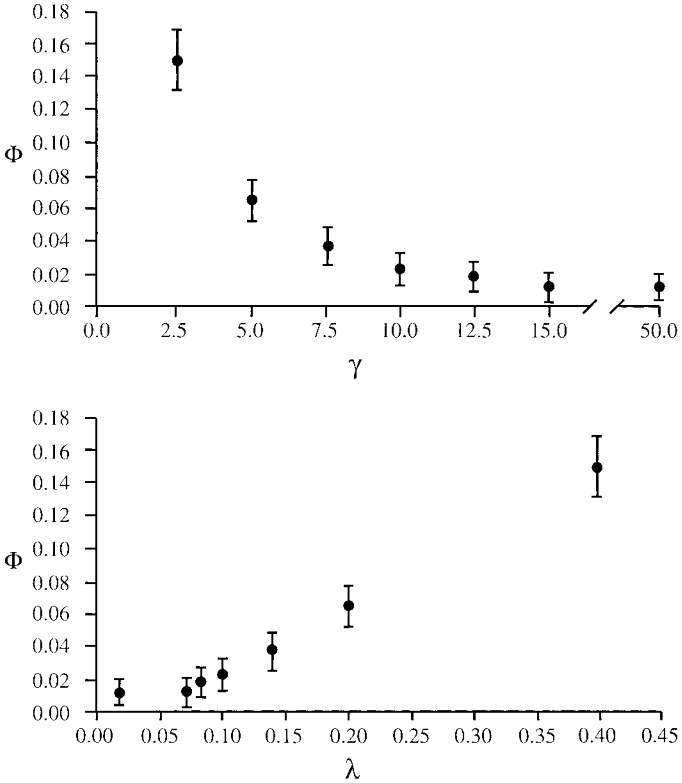

FIG. 3. Distributions of Φ^{C}_{FT}; 1000 simulated runs with actual *Quercus alba* allozyme data (eight loci), 20 mothers with 20 progeny each, for several different values of γ (top panel) and $\lambda = \gamma^{-1}$ (bottom panel).

of Φ_{FT}: (1) the four most polymorphic loci (top panel of Table 4), twice each, yielding E_8(best loci) = 0.862; (2) the eight loci presented, describing the genetic resolution encountered in a typical allozyme survey of forest trees, yielding E_8(actual loci) = 0.707; and (3) the four least polymorphic loci (bottom panel of Table 4), twice each, yielding E_8(worst loci) = 0.361. The expectation is that analytical resolution will increase with the value of E_8.

Results indicate that the impact of genetic polymorphism on Φ_{FT} differs for the categorical and ambiguous cases (Fig. 4). For the categorical case, the estimated average pollen structure does not differ significantly across the three levels of genetic polymorphism. In contrast, for the ambiguous case, the analysis significantly underestimates the amount of pollen structure among females for the low-polymorphism case (paired $t = 22.00$, $P < 0.001$). The ambiguous treatment is especially vulnerable to limited genetic polymorphism, probably reflecting the conservative (null-hypothesis favoring) consequences of using global allele frequencies to construct male gametic vectors for ambiguous mother-offspring (G_iG_k-G_iG_k) pairs. For either treatment, limited polymorphism yields larger variances in the Φ_{FT} estimates, thus lower statistical power. For empiric work, it will obviously be important to deploy a highly polymorphic genetic assay battery. A genetic assay battery with $E_8 = 0.361$ is simply inadequate to the task, and it should be augmented. For the *Q. alba* illustration (see below), we have used the allele frequencies (Table 4) corresponding to the middle treatment ($E_8 = 0.707$), for which the average values of Φ_{FT} are not statistically different for the categorical and ambiguous assay. That level of polymorphism is certainly adequate to support the enterprise, but the variances would be reduced with an even more polymorphic battery.

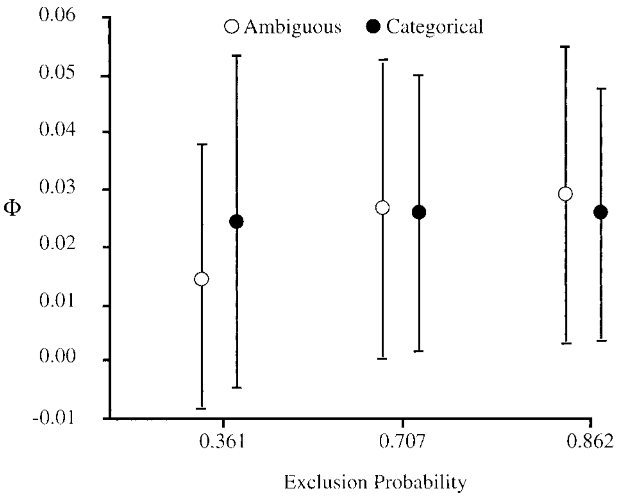

FIG. 4. Distributions of estimated Φ^{C}_{FT} and Φ^{A}_{FT}, for different choices of eight allozyme loci, least polymorphic (left), actual *Quercus alba* loci (center), and most polymorphic (right), 1000 replicates of 20 mothers and 20 progeny each; vertical lines represent actual ranges of 90% of the individual estimates.

QUERCUS ALBA IN MISSOURI OZARK FORESTS

We recently conducted a study of *Q. alba* (Fagaceae) to evaluate variation in mating system among Missouri Ozark forest stands (V. L. Sork, V. Apsit, and J. Raveill, unpubl. ms. a). *Quercus alba* is a common, continuously distributed tree species in the Missouri Ozarks, one of several focal tree species in a study of the impact of forest management on genetic diversity in the Missouri Ozark Forest Ecosystem Project, MOFEP (Koop 1996; Gram and Sork 1999, in press; Sork et al. 1997; V. L. Sork, A. R. Templeton, M. A. de la Fuente, P. Foster, A. L. Koop, and R. D. Westfall, unpubl. ms. b). For more details about the MOFEP, see Brookshire et al. (1997). Previous analysis of the adult genetic structure of *Q. alba* based on *F*-statistic models (Weir and Cockerham 1984) showed no evidence of genetic structure ($F_{ST} \approx 0.00$, $P \leq 0.98$; Koop 1996) among 36 adult subpopulations distributed across nine forest compartments (~250–500 ha each, within a region 20 km in diameter), which is compatible with long-term spatial homogenization over tens of kilometers.

Our interest here is to apply TwoGener to the *Q. alba* progeny data to understand the genetic structure within a single bout of pollination. We used the allele frequencies from *Q. alba* (Table 4) in our earlier simulations, and those were derived from 1586 progeny of these 54 maternal trees, distributed over the same 36 stands in nine compartments that we sampled for the earlier study of adult structure. Within a compartment, distances between pairs of maternal trees ranged from 50 m to 500 m. Methods for sampling, germination, and electrophoresis are described elsewhere (Sork et al. 1997, unpubl. ms. b). To maintain comparability with the

TABLE 5. Pooled within-compartment analysis of molecular variance (AMOVA) for *Quercus alba* using a set of 54 adults and 1586 progeny; the estimated value of Φ_{FT} is a measure of pollen pool heterogeneity among females.

Source of variation	Degrees of freedom	Sum of squares	Estimated mean squares	Variance estimate extracted	Heterogeneity (Φ_{FT}) estimate
Among females	45	125.15	2.7811	0.0638	$\Phi_{FT} = 0.061$
Within females	1531	1502.74	0.9815	0.9815	$P = 0.001$

mating system study, we chose to illustrate the TwoGener analysis with a pooled within-compartment analysis. We scored the gametes as in Table 2b, extracted a pairwise distance matrix for the male gametes within each compartment, conducted an AMOVA for each compartment, and tallied the sums of squares and degrees of freedom. By pooling the corresponding sums of squares and degrees of freedom across compartments, we obtained the results in Table 5, from which we obtained a pooled estimate of $\Phi_{FT} = 0.061$ ($P < 0.001$).

DISCUSSION

We have introduced a novel gene-flow analysis, based on a two-generation (parent-offspring) genetic structure analysis that identifies the scale of male gametic heterogeneity among females, in a manner that allows translation into the average distance of pollen movement. We show that pollen pool structure sampled by these females, as measured by Φ_{FT}, is inversely proportional to the average pollen dispersal distance, suggesting that even proximal females are sampling different portions of the global pollen pool. Our results show that this local pollen pool structure is informative about the pollen dispersal distribution. It is particularly helpful that we can estimate such structure with modest levels of genetic resolution (preferably, $E_L > 0.80$), because we can achieve this degree of resolution for many species with only modest numbers of progeny per female, which allows us to sample more maternal plants, spread over the larger spatial distances needed for landscape-scale studies. Our ability to apply this approach to *Q. alba* demonstrates empirically our ability to detect pollen pool heterogeneity among females. Our findings, based both on simulations and on natural population data, illustrate the promise of this new method for assessment of real-time pollen flow. We will turn now to some extensions of our simulations and empirical findings, elaborate on design issues and their implications, and close with some future applications.

Effective Pollination Neighborhood

An important feature of this model is its relationship to practical outcomes of gene flow. It would be especially profitable to convert our estimate of Φ_{FT} for *Q. alba* into a statement about the effective number of males and to provide some sense of the spatial extent of the mating neighborhood (the effective pollination neighborhood) for the Missouri study. Those are the questions to which we now turn our attention. Using a bivariate normal distribution of pollen dispersion, Wright (1946) defined the genetic neighborhood size as $N_e = 4\pi\sigma^2 d$, where σ^2 is the variance of parent-offspring natal distances and d is the population density. We are concerned here with pollen flow only, not parent-offspring natal distance, and Crawford (1984) has shown that Wright's formula needs adjustment for that case. Moreover, we have modeled negative exponential (rather than bivariate normal) pollen dispersal.

As was the case with Wright's model, we assume adaptively neutral marker loci. Unlike Wright, we assume that there is no local genetic structure among the adults themselves—caused either by previous drift and propagule flow or by selectively driven allele frequency heterogeneity—over the spatial scale of interest. If there is divergence among the pollen pools of the sampled mothers, we interpret it as a statement about limited pollen flow among spatially randomized adult male genotypes, rather than as evidence of adult genetic structure that is spatially organized. Any violation of that assumption will lead to an inflation of our estimate of Φ_{FT}. Our previous analysis of F_{ST} among *Q. alba* adults suggests no significant adult structure (Koop 1996), but a multivariate analysis of that genetic structure shows substantial heterogeneity (Gram and Sork, in press; Sork et al., unpubl. ms. b). We need to pursue the matter further, both as it affects the impact of adult structure (F_{ST}) and selective gradients (see Tonsor et al. 1993) on our inference concerning pollen structure (Φ_{FT}); we will leave those issues for later reports (F. Austerlitz and P. E. Smouse, unpubl. ms.; R. J. Dyer, R. D. Westfall, V. L. Sork, and P. E. Smouse, unpubl. ms.).

Our estimate of Φ_{FT} can be shown (Austerlitz and Smouse 2001) to be related to the probability of identity by descent (IBD) for two paternal alleles, drawn from among the progeny of the same and different females. If we assume idealized adults (equally fecund and phenologically synchronized) and that the females are spaced sufficiently far apart across the landscape, Φ_{FT} is (to a first approximation) a measure of the average per locus probability of IBD: $\Phi_{FT} \approx [2N_{ep}]^{-1}$. From our estimate of $\Phi_{FT} = 0.061$, we infer that a set of $N_{ep} \approx 8.2$ *effective* (idealized) males, providing all the pollen for a given female and all with equal likelihood, could be expected to yield the observed Φ_{FT} value. As an independent check on our results, we have used Ritland's (1990) mating system program to analyze the mating patterns within *Q. alba*; we found significant biparental inbreeding of about 6%, as well as four to nine effective pollen donors per mother (Sork et al., unpubl. ms. b). The analyses have different bases, but they both suggest a small effective number of pollen donors for the average female in a single bout of reproduction.

Our low estimates of the effective number of males sampled by individual mothers contrast sharply with the estimates of total number of pollen donors in oak paternity studies (e.g., Dow and Ashley 1996; Streiff et al. 1998). The two approaches seemingly give different results, but only because they measure different parameters. Because paternity analysis concentrates on identifying who the fathers are, it can provide valuable detail on the absolute numbers of fathers and their locations (see Nason et al. 1996; Sork et al. 1998). In contrast, TwoGener examines the impact of those fathers on pollen pool structure. For example, Streiff et al. (1998) have estimated that about 64% of progeny found in a 5.76-ha site were pollinated from external fathers (census number of pollen donors not reported). From these same data, TwoGener (based on the equations for F. Austerlitz and P. E. Smouse, unpubl. ms.) yields an effective number of (idealized) pollen donors of $N_{ep} \approx 18$–20. The total number of contributors may well be large, but most of them contribute very little; N_{ep} accounts for the unevenness of their relative contributions. TwoGener sacrifices paternal designation and enumeration for *effective* genetic impact.

One advantage of our parameterization is the potential to translate our results into an effective pollination area. For our simulations, the total area sampled was 100 units on a side, so that, for example, $\gamma = 5$ would represent five units. With 10,000 simulated individuals, population density was set to $d = 1$ idealized adults per square unit, Φ_{FT} was inversely proportional to the implicit product (γd). For the actual data on *Q. alba*, $\Phi_{FT} = 0.061$ translates into $N_{ep} \approx 8.2$, but how large an area is that on the ground? Our population density measurements for *Q. alba* in the nine compartments yield an average value of 92.8 trees ($\geq$11.4 cm dbh) per hectare. Adults are neither equally fecund nor phenologically synchronous, so the effective density is less than 92.8 trees per hectare. However, lacking more elaborate data on the variance and timing of relative pollen output from adults, we will treat the raw stem tallies as though they represented idealized adults.

Given that crude device, we estimate that the effective pollination area, A_{ep}, is 8.2/92.8 = 0.0884 ha (in extent), a circle of radius 16.77 m centered on the female in question. That is, if pollen contributions were even for all males and pollination were at random and synchronized, 8.2 idealized males—from within a circle of radius 16.77 m around the female—would yield the observed value of Φ_{FT}. For real adults, the circle containing 8.2 effective pollen donors will obviously have a larger radius and represent a larger area. Future work will need to address the impact of fecundity and phenological variance on A_{ep}, but with due allowance for the unknowns in this situation, it is clear that effective pollination is quite localized for *Q. alba* in this setting.

Several alternative pollen distributions might apply to *Q. alba*, and each would have its own translation into ground area and the effective number of equally contributing males. We will deal with those other distributions elsewhere (Austerlitz and Smouse 2001) but the point here is that Φ_{FT} is estimable, irrespective of the precise assumptions about pollen distribution. In addition, Φ_{FT} is large enough to indicate that effective pollination is quite localized, in spite of the fact that pollen can move large distances.

What We Have Wrought

TwoGener uses a combination of old tricks to do something new and different. With *Q. alba*, the F_{ST} analysis of adult genotypes showed no evidence of genetic structure. One might argue about the arbitrary definition of population strata for a continuously distributed species like *Q. alba*, but the absence of spatial structure among adults was striking. However, by concentrating on a two-generation comparison of parents (maternal trees) and their own progeny (male gametes drawn by those females), we have been able to estimate the pollination neighborhood for a single mother tree during a single bout of reproduction. For *Q. alba*, we estimate that pollination neighborhood to be about eight males, representing an area of less than 0.1 ha. That pattern has obviously not contributed much to genetic fragmentation for continuously distributed *Q. alba*, but for any species that is sparsely or patchily distributed, either naturally or by virtue of recent disruption, the genetic consequences would be considerable genetic structure among subsequent adults.

For a fixed total sample size, it seems best to obtain enough seed per female (replication) to ensure that $K \approx \Phi_{FT}^{-1}$, although the analytical scheme is reasonably robust over a broad range of sample allocations. There are two other ways to improve precision. The first is to increase total sample size, N, while holding $K \approx \Phi_{FT}^{-1}$ by increasing the number of females, J. Inspection of equation (10) shows that $s_\Phi \propto J^{-1/2}$, which implies $s_\Phi \propto N^{-1/2}$ for constant K. To reduce the standard deviation by half, we must quadruple N and thus the number of females. The second is to improve the genetic battery. We are best served by highly polymorphic, codominant loci. Genetic markers that are minimally polymorphic should be replaced by others with more information content. Multiple-allelic loci with relatively balanced allele frequencies are best for either classic paternity analysis (Chakraborty et al. 1988) or population survey analysis (Smouse and Chevillon 1998). Increasing the multilocus exclusion probability, E_L (eq. 12), can be done either by choosing more effective loci (those with higher E_l values) or by adding more loci (Smouse and Chakraborty 1986; Chakraborty et al. 1988).

The proper spatial scale for this sort of survey work needs further development. We anticipate that the spatial array of individuals and the heterogeneity of the landscape will determine the sampling design, but some generalizations would be helpful. It seems clear that by spacing our sampled females out over a substantially larger scale than has been possible for parentage studies, we will be able to deal with a variety of spatial distributions (e.g., high vs. low density, continuous vs. patchy distributions). A larger spatial array should aid in comparative work, without having to be concerned about whether the arbitrary sampling locations are panmictically cohesive. Given our concern with spatial processes, however, a number of questions concerning optimal spacing design need additional exploration.

Future Work

We can now estimate the pollination neighborhood for a single bout of reproduction. The scope of work required is manageable, so we can mount replicate studies over several seasons to evaluate the genetic consequences of multiple bouts of pollination for long-lived organisms. Alternatively, instead of increasing N for a single location and investing all of our effort in extracting a single estimate of Φ_{FT}, we can consider investing modest amounts of effort in each of several locations, attempting to compare and contrast pollination dynamics under different circumstances. With *Q. alba*, the interesting contrasts are those involving alternative cutting practices, the prime focus of the MOFEP, but other ecological factors might also impinge on pollen flow, not the least of which are the effective stem density of adults and the degree of habitat fragmentation for the species in question. We will explore some of those questions in later papers. Here, we have conveniently assumed that male and female gametes are unrelated. Given that most pollen is drawn from the local vicinity of a given female and that some of the contributing males may be related, particularly if there is preexisting genetic structure among the adults themselves, uniting gametes that are related need attention. We are pursuing that matter elsewhere (F. Austerlitz and P. E. Smouse, unpubl. ms.).

Acknowledgments

The TwoGener collaborative project was conducted as part of the Gene Flow Working Group, supported by the National Center for Ecological Analysis and Synthesis (NCEAS), a center funded by National Science Foundation (DEB-94–21535), the University of California–Santa Barbara, the California Resources Agency, and the California Environmental Protection Agency. PES was supported by the New Jersey Agricultural Experiment Station (NJAES/USDA-32106 and McIntire-Stennis-17309). VLS was supported by a sabbatical fellowship from NCEAS and research grants from the Missouri Department of Conservation for participation in the Missouri Ozark Forest Ecosystem Project. The authors thank V. Apsit, F. Austerlitz, K. Bailey, C. Kelly, J. Nason, and J. Wall for many helpful suggestions on the manuscript.

Literature Cited

Adams, W. T., and D. S. Birkes. 1991. Estimating mating patterns in forest tree populations. Pp.157–172 *in* S. Fineschi, M. E. Malvolti, F. Cannata, and H. H. Hattemer, eds. Biochemical markers in the population genetics of forest trees. SBP Academic, The Hague.

Adams, W. T., D. S. Birkes, and V. J. Erickson. 1992a. Using genetic markers to measure gene flow and pollen dispersal in forest tree seed orchards. Pp. 37–61 *in* R. Wyatt, ed. Ecology and evolution of plant reproduction. Chapman and Hall, New York.

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

Austerlitz, F., and P. E. Smouse. 2001. Two generation analysis of pollen flow across a landscape. II. Alternative dispersal functions and sampling strategies. Genetics: *In press*.

Bossart, J. L., and D. P. Prowell. 1998. Genetic estimates of population structure and gene flow: limitations, lessons, and new directions. Trends Ecol. Evol. 13:202–206.

Brookshire, B. L., R. Jensen, and D. C. Dey. 1997. The Missouri Ozark Forest Ecosystem Project: past, present and future. Pp.1–25 *in* B. L. Brookshire and S. R. Shifley, eds. Proceedings of the Missouri Ozark Forest Ecosystem Project symposium: an experimental approach to landscape research; 3–5 June 1997. St. Louis, MO. Gen. Tech. Rep. NC-193. USDA For. Serv., North Centr. For. Exp. Sta. St. Paul, MN.

Chakravarti, A., and C. C. Li. 1983. The effect of linkage on paternity calculations. Pp. 411–422 *in* R. H. Walker, ed. Inclusion probabilities in parentage testing. American Association of Blood Banks, Arlington, VA.

Chakraborty, R., T. R. Meagher, and P. E. Smouse. 1988. Parentage analysis with genetic markers in natural populations. 1. Paternity exclusion and expected proportions of offspring with unambiguous paternity. Genetics 118:527–536.

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

Crawford, T. J. 1984. The estimation of neighborhood parameters for plant populations. Heredity 52:273–283.

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

Dow, B. D., and M. V. Ashley. 1996. Microsatellite analysis of seed dispersal and parentage of saplings in bur oak, *Quercus macrocarpa*. Molec. Ecol. 5:615–627.

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

Ennos, R. A. 1994. Estimating the relative rates of pollen and seed migration among plant populations. Heredity 72:250–259.

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

Falconer, D. R. 1981. Introduction to quantitative genetics. 2d ed. Longman, London.

Gram, W. K. and V. L. Sork. 1999. Does population density reflect genetic diversity? Conservation Biology 13:1079–1087.

———. Association between environmental and genetic heterogeneity in forest tree populations. Ecology *In press*.

Hudson, R. 1991. Gene genealogies and the coalescent process. Pp. 1–44. *in* P. Harvey and L. Partridge, eds. Oxford Surveys in Evolutionary Biology. Oxford Univ. Press, New York.

Husband, B. C., and S. C. H. Barrett. 1996. A metapopulation perspective in plant population biology. J. Ecol. 84:461–469.

Jamieson, A. 1994. The effectiveness of using co-dominant polymorphic allelic series for (1) checking pedigrees and (2) distinguishing full-sib pair members. Anim. Genet. 25:37–44.

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

Koop, A. L. 1996. Genetic variation and structure in *Quercus alba* L. in a Missouri Ozark landscape. Dept. of Biology, Univ. of Missouri, St. Louis.

Kudoh, H., and D. F. Whigham. 1997. Microgeographic genetic structure and gene flow in *Hibiscus moscheutos* (Malvaceae) populations. Am. J. Bot. 84:1285–1293.

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

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

Nason, J. D., E. A. Herre, and J. L. Hamrick. 1996. Paternity analysis of the breeding structure of strangler fig populations: evidence for substantial long-distance wasp dispersal. J. Biogeogr. 23:501–512.

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

Peakall, R., P. E. Smouse, and D. R. Huff. 1995. Evolutionary implications of allozyme and RAPD variation in diploid populations of dioecious buffalograss (*Buchloë dactyloides* (Nutt.) Engelm.). Molec. Ecol. 4:135–147.

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

Selvin, S. 1980. Probability of nonpaternity determined by multiple allele codominant systems. Am. J. Hum. Genet. 32:276–278.

Slatkin, M. 1989. Detecting small amounts of gene flow from phylogenies of alleles. Genetics 121:609–612.

Smouse, P. E., and R. Chakraborty. 1986. The use of restriction fragment length polymorphisms in paternity analysis. Am. J. Hum. Genet. 38:918–939.

Smouse, P. E., and C. Chevillon. 1998. Analytical aspects of population-specific DNA-fingerprinting for individuals. Heredity 89:143–150.

Smouse, P. E., R. S. Spielman, and M.-H. Park. 1982. Multiple-locus allocation of individuals to groups as a function of the genetic variation within and differences among human populations. Am. Natural. 119:445–463.

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

Sorensen, F. C. 1972. The seed orchard tree as a pollen sampler: a model and example. USDA For. Serv. Res. Note, PNW-175 175:1–11.

Sork, V. L., A. L. Koop, M. A. de la Fuente, P. Foster, and J. Raveill. 1997. Patterns of genetic variation in woody plant species in the Missouri Ozark Forest Ecosystem Project (MOFEP). Pp. 233–249 *in* B. L. Brookshire and S. R. Shifley, eds. Proceedings of the Missouri Ozark Forest Ecosystem Project symposium: an experimental approach to landscape research; 3–5 June 1997; St. Louis, MO. Gen. Tech. Rep. NC-193. USDA For. Serv., North Centr. For. Exp. Sta., St. Paul, MN.

Sork, V. L., D. Campbell, R. Dyer, J. Fernandez, J. Nason, R. Petit, P. Smouse, and E. Steinberg. 1998. Proceedings from a workshop on gene flow in fragmented, managed, and continuous populations. Research Paper no. 3. National Center for Ecological Analysis and Synthesis, Santa Barbara, CA. *http://www.nceas.ucsb.edu/nceas-web/projects/2057/nceas-paper3/*.

Sork, V. L., J. Nason, D. R. Campbell, and J. F. Fernandez. 1999. Landscape approaches to the study of gene flow in plants. Trends Ecol. Evol. 142:219–224.

Streiff, R., T. Labbe, R. Bacilieri, H. Steinkellner, J. Glossl, and A. Kremer. 1998. Within population genetic structure in *Quercus robur* L. and *Quercus petraea* (Matt.) Liebl. assessed with isozymes and microsatellites. Molec. Ecol. 7:317–328.

Tonsor, S. J., S. Kalisz, J. Fisher, and T. P. Holtsford. 1993. A life-history based study of population genetic structure: seed bank to adults in *Plantago lanceolata*. Evolution 47:833–843.

Weir, B. S., and C. C. Cockerham. 1984. Estimating *F*-statistics for the analysis of population structure. Evolution 38:1358–1370.

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

———. 1969. Evolution and the genetics of populations. Vol. 2. The theory of gene frequencies. Univ. Chicago Press, Chicago.

Corresponding Editor: S. Tonsor