Theor Appl Genet (1988) 76:369–380

Fractional paternity assignment: theoretical development and comparison to other methods

B. Devlin [1], K. Roeder [2] and N. C. Ellstrand [1]

[1] Department of Botany and Plant Sciences and Program in Genetics, University of California, Riverside, CA 92521-0124, USA
[2] Department of Statistics, Pennsylvania State University, University Park, PA 16802, USA

Received October 10, 1987; Accepted March 10, 1988
Communicated by PMA Tigerstedt

Summary. There has recently been a burgeoning interest in the analysis of paternity patterns for natural populations because of its relevance to population genetic phenomena such as the distance between successful mates, relative male reproductive success and gene flow. In this paper we develop a method of analyzing populational patterns of paternity, the fractional paternity method, and compare its performance to two other commonly used methods of paternity analysis (simple exclusion and the most-likely methods). We show that the fractional method is the most accurate method for determining populational patterns of paternity because it assigns paternity to all progeny examined, and because it avoids biases inherent in the other paternity analysis methods when model assumptions are met. In particular, it avoids a systematic bias of the most-likely paternity assignment method, which has a tendency to over-assign paternity of progeny to certain male parents with a greater than average number of homozygous marker loci. We also demonstrate the effect of linkage of some of the marker loci on paternity assignment, showing how the knowledge of the linkage phase of male and female parents in the population can significantly improve the accuracy of the estimates of populational patterns of paternity. Knowledge of the linkage phase of individuals in a population is usually unknown and difficult to assess without progeny testing, which involves considerable labor. However, we show how the linkage phase of hermaphroditic individuals in a population can be obtained in conjunction with the paternity analysis if progeny can be obtained from each hermaphroditic individual in the population, thereby avoiding the problem of traditional progeny testing. Applications of the fractional paternity approach developed herein should contribute significantly to our understanding of the mating patterns in, and hence the evolution of, natural populations.

Key words: Paternity assignment – Likelihood – Mating patterns – Exclusion probability – Linkage phase

Offprint requests to N. C. Ellstrand

Introduction

The most critical evolutionary event in a population is the successful transmission of genes from one generation to the next. In many species, the identification of the successful female parent is straightforward because progeny are held on that individual for a period of time (e.g., seeds). Identification of the successful male parent is considerably more difficult. Recently, a number of studies have used multilocus data to document mating patterns within populations. Specifically, such studies have sought to measure the number of sires contributing to a sibship (e.g., Hanken and Sherman 1981; Ellstrand and Marshall 1986; Brown et al. 1986), the distances among successful mates within a population (e.g., Neale 1983; Hamrick and Schnabel 1985; Meagher 1986), the relative male fitnesses and/or functional gender of bisexual individuals (e.g., Muller-Starck and Ziehe 1984; Schoen and Stewart 1986; Cheliak et al. 1987; Ennos and Dodson 1987), the role of phenology on mating success (Ennos and Dodson 1987) and the rate of interpopulational gene flow by pollen (e.g., Ellstrand and Marshall 1985; Smith and Adams 1983; Friedman and Adams 1985).

In any paternity analysis, it is important to distinguish between the problem of paternity assignment where interest is focussed on the likelihood of a single triplet of progeny, mother and putative male parent (e.g., human paternity analysis) and that of paternity assignment for

populations, where the objective is to discern the populational patterns of paternity (or mating patterns). A number of methods have been used to identify populational patterns of paternity when the identity of the mother is known. Simple exclusion techniques (Ellstrand 1984; Hamrick and Schnabel 1985) compare the multilocus genotype of the progeny with that of its female parent, subtract the maternal contribution, and compare the remaining paternal gametic contribution with all possible local fathers' genotypes. Those individuals who could not have produced the appropriate multilocus gamete are excluded, and one or more individuals are assigned as possible parents.

Another method (Meagher 1986; Meagher and Thompson 1986, 1987), which we call the "most-likely" method, is a direct extension of the theory developed for inference of human paternity (e.g., Smouse and Chakraborty 1986 and references therein). In this method, the likelihood of paternity for each potential male parent, given the female parent and her progeny, is calculated based on segregation probabilities – paternity is assigned to the male parent with the highest likelihood value. If there is no most-likely male parent (ties), no father is assigned. This method should provide more information than the simple exclusion method because paternity can generally be assigned for more progeny.

A third method, which we term the "fractional method" (Brown et al. 1985) and develop later, calculates paternal likelihoods in the same way as the most-likely method. However, the fractional method assigns paternity of a progeny to one or more non-excluded male parents, the fraction assigned to any particular male parent being proportional to its likelihood of paternity relative to all other male parent likelihoods. Consequently, in contrast to the most-likely method, paternity will be assigned for all progeny, although some progeny will not be assigned a single father.

An alternative approach is to not assign progeny to fathers in any manner (fractional or otherwise), but to model the probability structure of the entire sample of offspring simultaneously, determining the most likely fertility parameters that would have generated such a sample. Similar genetic problems have been modeled in this fashion (e.g., Elandt-Johnson 1971). Schoen and Stewart (1986) apply this approach to analyze fertilities of gymnosperm populations, modeling the fertilities as a series of linear equations. Roeder et al. (1988) generalize this approach of modeling fertilities using the set of progeny genotypes, presenting exact likelihood models for three cases: the estimation of male fertilities when the maternal parent is known and the fertility parameters depend upon or are independent of the maternal parent and the estimation of fertilities when neither parent is known.

In this study, we develop the fractional method for two situations: (1) where there is no linkage among the marker loci; and (2) where some of the marker loci are linked. In the latter case, we show how the linkage phase of hermaphroditic individuals can be determined by progeny testing and how these data can be used to refine the paternity assignment. We compare the performance of the fractional method to the most-likely and simple exclusion methods through simulation analysis. We demonstrate that the efficacy of all methods improves with the exclusion probability of a given population, that the fractional method is the most appropriate of the three methods for determining mating pattern parameters in populations, and that knowledge of linkage phase further improves assignment.

General background

The exclusion probability and its significance

The power of any paternity exclusion method depends on the exclusion probability of a given population and its loci under study. Consider a female parent of genotype aa, one of her progeny of genotype Aa, and a set of M potential male parents from the population (sexually mature individuals that are either functionally male or hermaphroditic). If the allele frequencies of A and a are p and q for a population in Hardy-Weinberg equilibrium, then the paternal genotypes AA, Aa and aa are expected to occur in M^*p^2, M^*2pq and M^*q^2 potential male parents. Given this scenario, it is clear that on the average M^*q^2 potential male parents (genotype aa) can be excluded as fathers for this particular female parent and progeny pair. This leads to the definition of a population statistic, the exclusion probability for a population for the locus or loci under observation, defined as the expected value of an indicator of exclusion, I, where

$$I = \begin{cases} 1 \text{ if a randomly selected male parent is excluded} \\ \text{from a randomly selected female parent and progeny pair, 0 otherwise.} \end{cases}$$

Since

$$P(I=1) = \sum_{\text{all pairs}} P(I=1 \,\&\, \delta_1, \delta_2)$$

where (δ_1, δ_2) denote the genotypes of a mother and progeny pair respectively. If follows that the expected value of exclusion is calculated as

$$= \sum_{\text{all pairs}} P(I=1 \mid \delta_1, \delta_2)\, P(\delta_2 \mid \delta_1)\, P(\delta_1) \qquad (1)$$

(Neel and Schull 1954; Selvin 1980). For the single locus, 2 allele case, the exclusion probability is given by

$$\varepsilon_i = pq(1-pq) \qquad (2)$$

and the exclusion probability for a set of n loci by

$$\varepsilon = 1 - \prod_{i=1}^{n} (1-\varepsilon_i) \qquad (3)$$

(Ryman and Chakraborty 1982). An equivalent empirical calculation consists of summing the number of male parents in the population that can be excluded for each observed female parent and progeny pair and dividing by the total number of observed pairs.

The exclusion probability increases with the number of loci that can be used as genetic markers, with the number of alleles at each locus and with the evenness of the allele frequency at each locus (Chakraborty et al. 1974; Selvin 1980; Ryman and Chakraborty 1982; Smouse and Chakraborty 1986). However, the relationship exhibits a diminishing marginal return.

The importance of exclusion probability to paternity assignment is that an increase in the exclusion probability increases the probability of paternity among the set of non-excluded male parents (Smouse and Chakraborty 1986). This fact can be illustrated by example. Consider 3 populations, each containing 40 potential male parents but having different exclusion probabilities, say 0.5, 0.7 and 0.9. Then, on average, the number of non-excluded male parents for a progeny will be 20, 12 and 4, respectively, for the 3 populations. Clearly, the likelihood of choosing the correct non-excluded male parent of the progeny increases when the number of non-excluded male parents decreases. In the extreme case, as the exclusion probability approaches 1, most progeny can be assigned exclusively to a single male parent in the population.

The simple exclusion procedure of paternity analysis considers the paternity of progeny assignable only in cases with a single non-excluded male parent. This class of progeny, however, will be relatively infrequent. For the example above with a population size of 40 and a relatively high exclusion probability of 90%, approximately 30% of the progeny are assignable. We will demonstrate later that this level of assignment is not sufficient to derive accurate population statistics on mating patterns. Consequently, it is usually necessary to use other methods to determine the likelihood of paternity.

Human paternity analysis

Most of the theory of paternity assignment based on likelihood has been developed for application to human populations where a particular male is accused of paternity (Thompson 1986; Smouse and Chakraborty 1986 and references therein). We will briefly review the theory for human paternity analysis herein because it is pertinent to the theoretical development of methods to determine the patterns of paternity for natural populations.

Consider a triplet of female parent i (FP_i), putative male parent j (MP_j), and offspring k (O_k) with vectors α_i, β_i, and γ_k denoting their respective multilocus genotypes. Let i and j be indices that range over the individual parents, but let k index the distinctive offspring genotypes. The likelihood of putative male parent j being the actual father is generally formulated as the posterior odds of paternity versus non-paternity given the available genetic information, call it λ_j.

$$\lambda_j = \frac{P(\alpha_i, \beta_j, \gamma_k \mid \text{paternity})}{P(\alpha_i, \beta_j, \gamma_k \mid \text{non-paternity})}. \tag{4}$$

Assuming random mating and independent loci, the conditional probabilities can be rewritten as

$$\lambda_j = \frac{P(\beta_j)\, P(\alpha_i)\, T(\gamma_k \mid \alpha_i, \beta_j)}{P(\beta_j)\, P(\alpha_i)\, T(\gamma_k \mid \alpha_i)}, \tag{5}$$

where T denotes the transition (Mendelian) probabilities of the child's multilocus genotype given either the mother and putative father's multilocus genotype (numerator) or that of the mother and a random draw of complementary alleles from the population (denominator), and $P(\cdot)$ denotes the probability of that multilocus genotype for the population. The method of calculating these transition probabilities is well known, particularly when loci assort independently. With simplification,

$$\lambda_j = \frac{T(\gamma_k \mid \alpha_i, \beta_j)}{T(\gamma_k \mid \alpha_i)}. \tag{6}$$

The decision on paternity (versus non-paternity) is then usually based on some arbitrary threshold value for the likelihood ratio (Valentin 1980).

Modification to populational patterns of paternity

Theory

To analyze the patterns of paternity of a natural population, we assume that the genotypes of all parents are known. Further, we assume that a set of progeny have been collected from either a sub sample of mothers or the entire set of mothers in the population, and that the multilocus genotype of each progeny has been determined. Ideally, we would like to know the number of offspring fathered by MP_j on FP_i, denote this number F_{ij}. The information we have for estimating this is the number of offspring of each genotype (γ_k) from this female parent, call it X_{ik}.

We argue, developing the fractional paternity approach suggested by Brown et al. (1985), that F_{ij} should be estimated by

$$\hat{F}_{ij} = \sum_k X_{ik}\, P(MP = j \mid FP = i,\, 0 = k). \tag{7}$$

Equation (7) can be generalized to matrix form. Let $\hat{F}_{i\cdot} = (\hat{F}_{i1}, \hat{F}_{i2}, \ldots \hat{F}_{im})$ denote the number of offspring from FP_i fathered by each of the male parents. Suppose we have M male parents and b_i distinct offspring genotypes.

Let $\mathbf{P}_i$ equal an M by b_i matrix containing elements $P(MP=j|FP=i, 0=k)$, where each column is this probability for each MP for a fixed γ_k, and each row this probability for each γ_k for a fixed MP_j. In addition, let $\mathbf{X}_i'=(X_{i1}, X_{i2}, \ldots, X_{ib_i})$. Then

$$\hat{\mathbf{F}}_i = \mathbf{P}_i \mathbf{X}_i, \tag{8}$$

$$E[\hat{\mathbf{F}}_i] = \mathbf{P}_i E[\mathbf{X}_i], \text{ and} \tag{9}$$

$$\text{cov}(\mathbf{F}_i) = \mathbf{P}_i' \text{cov}(\mathbf{X}_i) \mathbf{P}_i. \tag{10}$$

Since $\mathbf{X}_i$ has a multinomial distribution, we can estimate the covariance matrix in the usual way.

It remains to estimate the conditional probability that the putative male parent is the actual male parent, given the female parent and the multilocus genotype of the progeny. Using Bayes Theorem, we get

$$P(MP=j^*|FP=i, 0=k) = \frac{P(0=k|FP=i, MP=j^*)\, P(MP=j^*|FP=i)}{\sum_j P(0=k|FP=i, MP=j)\, P(MP=j|FP=i)}. \tag{11}$$

We can rewrite the term $P(0=k|FP=i, MP=j)$ as $T(\gamma_k|\alpha_i, \beta_j)$, the transition probability of the offspring genotype given the multilocus genotype of the female parent and putative male parent.

$$P(MP=j^*|FP=i, 0=k) = \frac{T(\gamma_k|\alpha_i, \beta_{j^*})\, P(MP=j^*|FP=i)}{\sum_j T(\gamma_k|\alpha_i, \beta_j)\, P(MP=j|FP=i)}. \tag{12}$$

The term $P(MP=j|FP=i)$, however, is not easily dismissed. Let us assume that $P(MP=j|FP=i)$ is constant for all j (we will justify this shortly). We can then show that (11) is identical to the usual likelihood ratio of paternity versus non-paternity for male parent j divided by the sum all potential paternal likelihood ratios.

$$\frac{\lambda_{j^*}}{\sum \lambda_j} = \left[\frac{T(\gamma_k|\alpha_i, \beta_{j^*})}{T(\gamma_k|\alpha_i)}\right] \Big/ \sum_j \frac{T(\gamma_k|\alpha_i, \beta_j)}{T(\gamma_k|\alpha_i)} = \frac{T(\gamma_k|\alpha_i, \beta_{j^*})}{\sum_j T(\gamma_k|\alpha_i, \beta_j)} = P(MP=j^*|FP=i, 0=k). \tag{13}$$

The term $P(MP=j|FP=i)$ is not usually included or discussed in human paternity analysis or in studies of paternity patterns in natural populations, although an understanding of this term is important for the latter case. This term can be thought of as the prior probability of paternity, encompassing all of the ecological/genetic parameters of the population that make certain individuals more or less likely to be the male parent of a particular FP's progeny. In plant populations, for instance, it is generally assumed that near neighbors of FP_i have a much greater probability of siring seed than distant individuals of the same population (the result of leptokurtic pollen movement; Levin and Kerster 1974). Consequently, if the assumption of limited pollen dispersal is correct, then a priori, male parents closest to FP_i are more likely to father FP_i's seed than distant male parents; this information would be accounted for in the term $P(MP=j|FP=i)$. Likewise, differential attractiveness among males (Trivers 1985 and references therein), differential pollen fertility (Devlin and Stephenson 1987) and the like would affect this prior probability. Hence, although most studies to date assume that the population is a random mating population in which the paternal probabilities are equal, as we will, the assumption is clearly not accurate.

We justify the assumption that $P(MP=j|FP=i)$ is constant for all j with two arguments. The first and most obvious argument is that there are little to no data on these prior probabilities; consequently, we do not know what probability density function maps the paternity probabilities to the pertinent ecological/genetic variables nor, for that matter, what the pertinent variables are. The second argument is that a major goal of analyzing patterns of paternity in natural populations is to identify the pertinent variables that affect this prior probability. Hence, it would be circular to assume a certain functional relationship for an ecological parameter like inter-parent distance and then perform the paternity analysis with the goal of ascertaining the effects of inter-parent distance.

The distinction between the fractional paternity approach we have just developed and the most-likely paternity approach developed by Meagher (1986) and Meagher and Thompson (1986) is straightforward. In our method, some proportion of the seed of genotype γ_k from FP_i are assigned to each potential male parent; only the set of non-excluded male parents receive a fraction greater than 0. For the set of non-excluded male parents, the fraction of X_{ik} assigned to MP_j is proportional to its likelihood of paternity relative to the sum of non-excluded male parents' likelihoods. In the most-likely method, however, all seeds of γ_k from FP_i (i.e., X_{ik}) are assigned to the MP_j with the greatest likelihood of paternity; specifically, assign X_{ik} to MP_{j^*} if and only if $T(\gamma_k|\alpha_i, \beta_{j^*}) > T(\gamma_k|\alpha_i, \beta_j)$, $j^* \neq j$, over the set of all potential male parents. Moreover, in the case where there is no "most-likely" potential male parent, X_{ik} is not assigned while all seed are assigned in the fractional method.

We argue that the most-likely method, as formulated above, embodies a bias in that the most-likely paternal parent will always be that individual in the population that has the highest number of loci homozygous for the necessary paternal gamete contribution that complements the maternal contribution for a particular γ_k. We can illustrate this fact with a simple example. Consider a random mating population of 4 individuals, 3 male and

1 female. The 3 males have genotypes AA, Aa and aa. The female of genotype AA will produce progeny of genotypes AA and Aa in equal proportions. In the most-likely method, all AA progeny are assigned to the male with genotype AA although, all other things being equal, the male of genotype Aa would have fathered 0.33 of these progeny. Additionally, all progeny of genotype Aa are assigned to the male of genotype aa, although again the male of genotype Aa would have fathered 0.33 of the progeny. Hence, the most-likely method is biased even when the model assumptions are met. While this problem of bias is ameliorated with multilocus paternity estimation, we demonstrate from simulation results (below) that the estimator remains biased in the multilocus case.

In Appendix 1, we demonstrate that the fractional method is unbiased only if the model assumptions are met, and how a bias is introduced when these assumptions are violated. This bias is minimized with greater genetic information and with larger samples of progeny.

Simulations

The relative abilities of the simple exclusion, most-likely and fractional methods to determine the patterns of paternity of a population were evaluated with a simulation analysis. In this analysis, 48 hermaphroditic individuals functioned as both male and female parents but individuals were self-incompatible. Eleven independent loci were used as genetic markers, with 2–4 alleles per locus; this variation is similar to natural systems in which paternity has been assessed (e.g., Meagher 1986). Each individual produced 60 progeny as a female parent. Paternity of each progeny was assigned randomly to one of the 47 potential paternal parents, with all potential male parents having 1/47 probability of being chosen. After assignment, the multilocus genotype of this progeny was produced by randomly choosing 1 of the 2 alleles from the maternal and paternal parent for each of the 11 loci; each allele at each locus had a 0.5 probability of being chosen. The program kept track of each value of F_{ij}, i.e., the number of progeny of maternal parent i fathered by paternal parent j.

With the progeny and parental data sets, $\hat{F}_{ij}$ was determined for each of the three paternity assignment methods. From these data, two mean square error statistics (MSE) were calculated. The first MSE statistic used the squared difference between F_{ij} and $\hat{F}_{ij}$, specifically

$$MSE_{GM} = \left[\sum_i \sum_j \left(\frac{2880}{NA}\hat{F}_{ij} - F_{ij}\right)^2\right] \Big/ M(M-1). \quad (14)$$

Note that $\hat{F}_{ij}$ was adjusted by a scaling parameter (2880/NA), where NA equals the number of progeny assigned by a particular method. Recall that both the simple exclusion and the most-likely methods do not generally assign all progeny. This scaling parameter made the three methods comparable and, if assignment for either method was incomplete but proportionately correct, then $MSE_{GM} = 0$. The second MSE statistic used the squared difference between the estimated total progeny fathered by each individual and the actual number, specifically

$$MSE_{RS} = \left[\sum_j \left(\sum_i \frac{2880}{NA}\hat{F}_{ij} - \sum_i F_{ij}\right)^2\right] \Big/ M. \quad (15)$$

The subscripts of the mean square statistics are acronyms chosen to represent possible applications of the particular estimator. In plant population genetics, values for $\hat{F}_{ij}$ would generally be used to evaluate realized gene movement (GM) by pollen and $\sum_i \hat{F}_{ij}$, paternal reproductive success (RS). In essence, these MSE statistics tell us the relative abilities of the three methods to track the values of these variables for a population.

We replicated the above procedure 20 times. For 5 of the replications, parental gene frequencies were adjusted to achieve an expected exclusion fraction of 0.7, and likewise 5 replications each for expected exclusion fractions of 0.8, 0.9 and 0.96. Because parental alleles were assigned randomly for each replication (with the restriction placed on gene frequency), the actual exclusion fraction varied slightly from their expected values.

In every replication, the minimum MSE_{GM} and the minimum MSE_{RS} were produced by the fractional paternity method (Figs. 1 and 2). Consequently, we concluded that the fractional approach is the most appropriate of the three methods when all of the assumptions are met. We noted that the square root of the MSE's overestimates the mean difference between F_{ij} and $\hat{F}_{ij}$. For instance, the average difference between $\sum_i F_{ij}$ and $\sum_i \hat{F}_{ij}$ over all replications with an expected exclusion fraction of 0.96 was 3.06, 10.22 and 21.17 for the fractional, most-likely and simple exclusion methods, respectively.

We performed a second set of simulations in which we violated the assumption of equal male parent fertility (holding female fertility constant). In this simulation, 10 individuals out of the 48 were chosen to have twice the male fertility; that is, they were twice as likely to father progeny than the remaining 38 hermaphroditic individuals. In addition, we restricted the range of expected exclusion fractions to 0.96 and 0.90, because we felt that the genetic information for a population of 48 individuals with exclusion fractions of 0.8 and 0.7 would not be sufficient to track the population patterns of paternity. Except for these changes, the structure of the simulations remained as described above.

In 9 of 10 replications, the minimum MSE_{GM} was produced by the fractional method, while in 1 replication, the fractional and most-likely methods produced equal values (Fig. 3). In every replication, the minimum MSE_{RS} was produced by the fractional paternity method (Fig. 4).

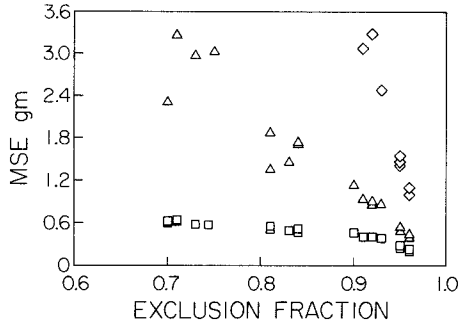

Fig. 1. The relative performance of the fractional (□), most-likely (△) and simple exclusion (◇) methods of paternity assignment in relation to the population's exclusion fraction. All loci assort independently and model assumptions are satisfied. Smaller values for this MSE statistic (GM) indicate greater accuracy in the estimation of the number of progeny of each female parent fathered by each male parent in the population. Each point represents a separate simulation run. Note that most points for the simple exclusion method are out of range on the *Y-axis*

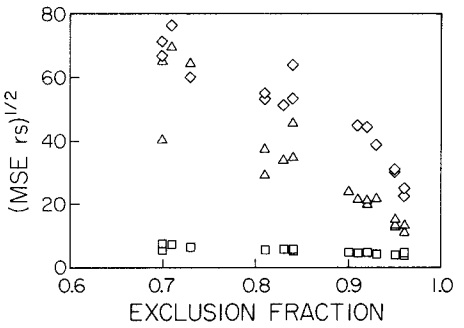

Fig. 2. The relative performance of the fractional (□), most-likely (△) and simple exclusion (◇) methods of paternity assignment in relation to the population's exclusion fraction. All loci assort independently and model assumptions are satisfied. Smaller values for this MSE statistic (RS) indicate greater accuracy in the estimation of the each male parent's reproductive success. Each point represents a separate simulation run. Note that a few points for the simple exclusion method are out of range on the *Y-axis* and that some points of each method overlap

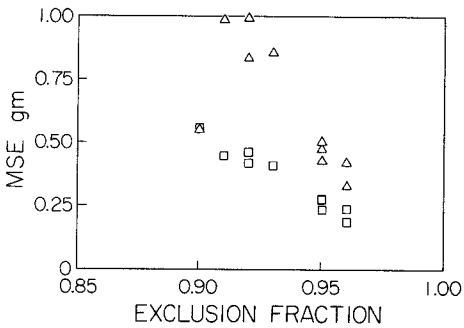

Fig. 3. The relative performance of the fractional (□) and most-likely (△) methods of paternity assignment in relation to the population's exclusion fraction. All loci assort independently but the model assumptions are violated, that is, 10 male parents have twice the fertility of the remaining 38 male parents. See caption of Fig. 1 for the definition of MSE (GM). Each point represents a separate simulation run

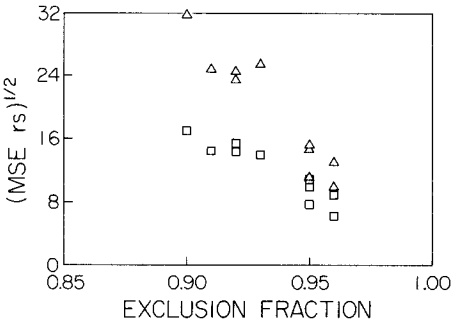

Fig. 4. The relative performance of the fractional (□) and most-likely (△) mathods of paternity assignment in relation to the population's exclusion fraction. All loci assort independently but the model assumptions are violated, that is, 10 male parents have twice the fertility of the remaining 38 male parents. See caption of Fig. 2 for the definition of MSE (RS). Each point represents a separate simulation run

Moreover, the fractional paternity method tracked the patterns of paternity quite well; for instance, the average difference between $\sum_i F_{ij}$ and $\sum_i \hat{F}_{ij}$ was 8.5. The other methods produced considerably greater values. While the fractional paternity method revealed a reasonably accurate pattern of paternity, the violation of the model assumption of equal paternal fertilities increased the MSE substantially (Figs. 1 and 2 versus Figs. 3 and 4).

Summarizing this section, we illustrated a method of analyzing populational patterns of paternity, namely the fractional paternity method. We showed that the performance of this measure is superior to other suggested methods of paternity analysis when the objective is to determine populational patterns of paternity. The weak performance of the simple exclusion method is the result of limited assignment of paternity; those assignments that occur do not accurately reflect populational patterns of paternity (Figs. 1–4), but rather are biased to identify potential male parents with unusual alleles and unusual combinations of maternal/paternal gametes. While the foregoing is less true for the most-likely method, that method also suffers from the additional bias that highly homozygous potential male parents are overrepresented as the chosen most-likely male parent. This assertion is supported by the simulation results. Using a nested analysis of variance with dependent variable, MSE_{RS}, and independent variables, "exclusion fraction and the num-

ber of homozygous loci per MP_j nested in the exclusion fraction", the later variable accounted for 12.7% of the variance ($F = 10.93$; $p < 0.0001$).

While fractional assignment minimizes the difference between the estimated and actual paternity values relative to the most-likely and the simple exclusion methods, the technique is only as powerful as the genetic information in the population, specifically the exclusion fraction (Figs. 1–4). Moreover, we emphasize that the genetic information must also be considered in light of the number of potential male parents. In addition, while the fractional assignment is unlike the simple exclusion and most-likely methods in that its estimates are unbiased when assumptions of the model are met, its estimates are also biased when the assumptions are violated. In this case, the fractional method consistently underestimates F_{ij} for high fertility males. For hypothesis testing, this bias could be argued to be conservative since the null hypothesis would generally be that there are no differences among individuals in male fertility. Consequently, it is more difficult to reject the null hypothesis when the estimated values are lower than the true value for individuals who are truly more male fertile. There is, however, another and more troublesome issue. The variance estimates [Eq. (10)] will now also underestimate the true variance. This aspect of the problem and the general experimental design of paternity studies are a focus of our current research.

The effects of linkage on paternity assignment

Theory

To appreciate the effect of linkage on paternity analysis, recall that the key element in the estimating equation is the transition probability. When loci are linked, alleles at the linked loci do not segregate independently, and hence linkage has an effect on the transition probabilities for each pair of parents in the population (Chakraborty and Hedrick 1983). The transition probabilities for the linked loci will depend on which alleles are associated on the parental chromosomes, termed the linkage phase. To calculate the transition probabilities, we treat the linked loci as a pseudolocus. Given the pseudolocus construct, two linked loci, each with 2 alleles, would be considered a single pseudolocus with 4 possible alleles, say AB, aB, Ab and ab. Call the probability of recombination for this locus pair r. For a heterozygous individual with linkage phase Ab/aB, the gametes Ab, aB, AB and ab are produced by this individual with probabilities $0.5(1-r)$, $0.5(1-r)$, $0.5r$ and $0.5r$.

In general, if we have l loci involved in the pseudolocus construction there are 2^l possible gametes, but if the parent is homozygous at any locus then some of the gametes are not distinct. Let $\mathfrak{F}(\alpha)$ and $\mathfrak{M}(\beta)$ denote the set of possible gametes formed by the female and male parents of genotypes α and β respectively (note that in this section α and β represent not only the genotypes, but also the linkage phase of the parents.) Let ϕ and μ represent particular elements in $\mathfrak{F}$ and $\mathfrak{M}$ and let $P_f(\phi)$ and $P_m(\mu)$ denote the probability densities for the elements in $\mathfrak{F}$ and $\mathfrak{M}$, respectively. The transition probability for a triplet $(\gamma_k, \alpha_i, \beta_j)$ can be calculated as

$$T(\gamma_k|\alpha_i, \beta_j) = \sum_{\phi \in \mathfrak{F}(\alpha)} \sum_{\mu \in \mathfrak{M}(\beta)} P_f(\phi)\, P_m(\mu)\, I(\gamma) \tag{16}$$

where $I(\gamma) = 1$ if the two gametes ϕ and μ can combine to form an offspring of genotype γ_k, and 0 otherwise.

As an example, consider 2 linked marker loci with $r = 0.1$, a female parent of genotype AABB, her progeny of genotype AaBb and two potential male parents: MP_1 with genotype Ab/aB and MP_2 with genotype ab/AB. The female parent contributes gamete AB with probability 1. Thus $\mathfrak{F} = \{AB\}$ and $P_f(AB) = 1$. Both potential male parents produce gamete sets $\mathfrak{M} = \{AB, Ab, aB \text{ and } ab\}$, but the probabilities of producing these gametes are reversed. For MP_1 the respective probabilities $P(\cdot)$ are 0.05, 0.45, 0.45 and 0.05, while the respective probabilities for MP_2 are 0.45, 0.05, 0.05 and 0.45. $T(AaBb|AABB, Ab/aB) = 0.05$ and $T(AaBb|AABB, ab/AB) = 0.45$. It is apparent in this example that linkage can have a significant effect on the transition probabilities and hence on paternity assignment. Finally, the calculation of transition probabilities for mixed linked and unlinked loci is a simple extension of the normal method; calculate the transition probabilities at each pseudolocus, then take the product of the independent groups as usual.

Simulations

The effect of linkage in the example above is apparent but its effects on the determination of populational patterns of paternity are probably less apparent. In order to illustrate these effects, we performed a simulation analysis where 4 of the 11 marker loci were linked, with $r = 0.05$ for adjacent pairs of loci. Furthermore, genotypes were assigned such that the population was in linkage equilibrium. All other aspects of the simulation remained as reported in the previous section. In these simulations, two scenarios concerning knowledge of the assortment of the marker loci were developed: (1) that 4 of the loci were linked, that we had perfect knowledge concerning this linkage and about the linkage phase of all individuals in the population; and (2) that 4 of the loci were linked but that we had no information that this was the case.

Because the simple exclusion technique is unaffected by linkage (unless there is no recombination), we ignored this method in the analysis. Moreover, because the fractional method is the most appropriate of the three methods for the determination of populational patterns of

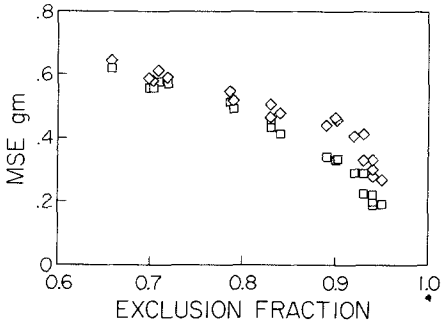

Fig. 5. The relative performance of the fractional method of paternity assignment in relation to the population's exclusion fraction – 4 of the 11 marker loci are linked in the case. □ indicate values obtained when the method is modified to take this linkage into account and ◇ indicate values obtained when this information is either ignored or unknown. See caption of Fig. 1 for the definition of MSE (GM). Each point represents a separate simulation run

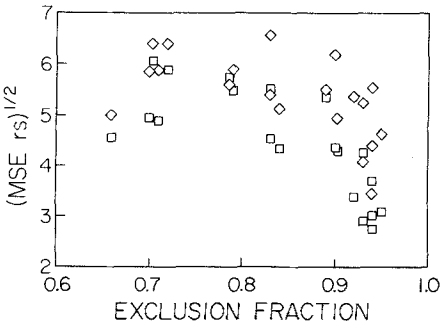

Fig. 6. The relative performance of the fractional method of paternity assignment in relation to the population's exclusion fraction – 4 of the 11 marker loci are linked in the case. □ indicate values obtained when the method is modified to take this linkage into account and ◇ indicate values obtained when this information is either ignored or unknown. See caption of Fig. 2 for the definition of MSE (RS). Each point represents a separate simulation run

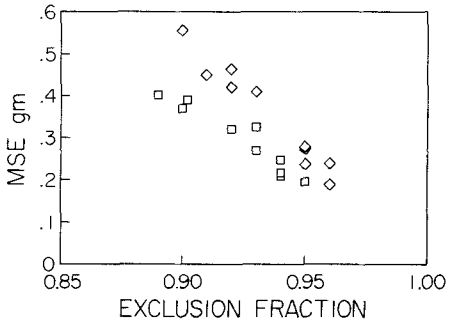

Fig. 7. The relative performance of the fractional method of paternity assignment in relation to the population's exclusion fraction – 4 of the 11 marker loci are linked and male parent fertilities vary as discussed in Fig. 3's caption. □ indicate values obtained when the method is modified to take this linkage into account and ◇ indicate values obtained when this information is either ignored or unknown. See caption of Fig. 1 for the definition of MSE (GM). Each point represents a separate simulation run

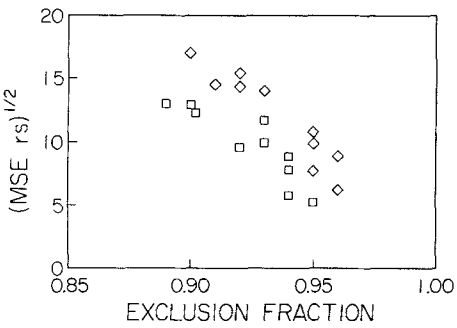

Fig. 8. The relative performance of the fractional method of paternity assignment in relation to the population's exclusion fraction – 4 of the 11 marker loci are linked and male parent fertilities vary as discussed in the caption of Fig. 3. □ indicate values obtained when the method is modified to take this linkage into account and ◇ indicate values obtained when this information is either ignored or unknown. See caption of Fig. 2 for the definition of MSE (RS). Each point represents a separate simulation run

paternity (theoretical and empirical results of the last section), we will not present the results of the most-likely technique, except to make two points: linkage has an effect on the accuracy of the technique and, in all replications of these simulations, the fractional paternity method proved to be more accurate than the most-likely method. With respect to the latter point, the magnitude of the differences between the two methods was similar to those of the previous section.

The effects of prior knowledge of linkage versus no knowledge of linkage when the loci are linked can be seen in Figs. 5 and 6. Prior knowledge of linkage and knowledge of the appropriate linkage phase of each individual in the population improve the accuracy of both estimators, as illustrated by the decrease in both the MSE_{GM} and the MSE_{AS}. Visually, it appears that the knowledge of linkage has its greatest effect on the accuracy of the estimators at higher exclusion fractions, although it has a strong effect throughout the range of exclusion fractions examined. The effect of the violation of the assumption of equal male fertilities, while still apparent in the increase in both MSE statistics (Figs. 7 and 8 versus Figs. 5 and 6), is somewhat ameliorated by the knowledge of the linkage phases of the individuals in the population. Finally, contrasting the case when 4 of the 11 marker loci are linked and the linkage phases are known to the case when all 11 marker loci are independent (Fig. 9), the former case leads to a significant improvement in the accuracy of the estimate of $\hat{F}_{ij}$.

While we have demonstrated that the knowledge of linkage and the linkage phases for members of a population can contribute considerable accuracy to the determination of populational patterns of paternity, it is not clear that this would necessarily be the case when the loci

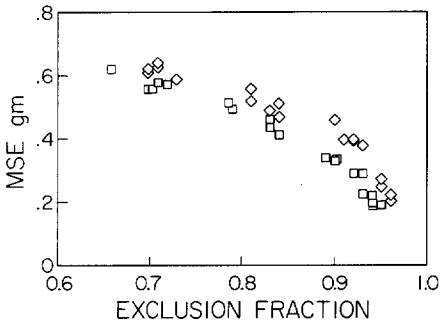

Fig. 9. The relative performance of the fractional method of paternity assignment in relation to the population's exclusion fraction. □ indicate values obtained when the method is modified to take into account that 4 of the 11 marker loci are linked and ◇ indicate values obtained when all loci segregate independently. See caption of Fig. 1 for the definition of MSE (GM). Each point represents a separate simulation run

are in linkage disequilibrium (Chakraborty and Hedrick 1983). Linkage disequilibrium is defined as a lack of independence of loci for a population, regardless of whether the loci occur on the same chromosome. Consider the case of 2 loci in linkage disequilibrium, with 2 alleles at each locus (say A and a, B and b). Further, assume for simplicity that A is always found in association with b and a with B. Then, in this case, rather than 2 independent loci, each with two alleles or even a pseudolocus with 4 possible gametes, the 2 loci represent a single pseudolocus with only two possible gametes, reducing the expected exclusion fraction relative to the case of linkage equilibrium. Therefore, increased linkage disequilibrium makes all methods of paternity analysis increasingly inaccurate.

Estimation of the linkage phase of hermaphrodites

Theory

We showed in the previous section that knowledge of the linkage of marker loci and knowledge of the individuals' linkage phases can significantly improve the measurement of populational patterns of paternity. It is reasonable to question, however, if one could ever obtain the necessary information on individual linkage phases. One could obtain the linkage phases of individuals in the usual way, by test crossing, although this may be more readily practicable for plant populations. Moreover, even in the case where the species is amenable to test-crossing, the effort involved would be formidable. In this section, we will demonstrate that this information is obtainable with little or no additional experimental effort if the species is hermaphroditic and each individual produces some progeny through the maternal function. Since approximately 80% of the seed plant species are hermaphroditic (Yampolsky and Yampolsky 1922), this technique should be particularly applicable to paternity studies of these species.

Suppose we have collected N offspring from a plant and we are interested in estimating the linkage phase of a pair of loci where each locus has 2 possible alleles (say A and a, B and b) and the recombination probability is r. Define n_{ab}, n_a, n_b and n as the number of offspring homozygous at both, only the first, only the second, and neither locus, respectively. We note that only for the latter class of progeny is the linkage phase of the progeny unknown. For completely homozygous progeny, we recognize only 4 linkage phases (AB/AB, Ab/Ab, aB/aB and ab/ab), but for the two single homozygote progeny genotypes, we recognize two linkage phases for each genotype (Ab/AB and aB/ab; AB/aB and Ab/ab). Finally, we can not meaningfully sub-divide the double heterozygote class. Let l denote the linkage phase of an offspring and X_l denote the number of offspring of each phase (e.g., $X_{AB/aB}$).

Let g denote the "gamete" (with respect to the linked loci) donated by the male parent. P(g) is the probability of this gamete from the population. Assuming gametic equilibrium, $P(AB)=p1 \cdot p2$, where p1 and p2 are the population gene frequencies of the A and B alleles at the respective loci. Further, let Y_g be the unknown number of gametes of type g contributed by the female parent to the progeny. Suppose that FP_i has a linkage phase Ab/aB, then FP_i will contribute gametes Ab and aB with probability $0.5(1-r)$ and gametes AB and ab with probability $0.5\,r$. Then the AB maternal gametes will appear as AB/AB, Ab/AB, aB/AB and ab/AB offspring with probabilities $0.5\,rP(AB)$, $0.5(1-r)P(Ab)$, $0.5(1-r)P(aB)$ and $0.5\,rP(ab)$, respectively. The data we have are the number of progeny in each distinct genotype class. Recall that the linkage phase of each genotype is known (except the n double heterozygotes). Let $\mathbf{X}=(X_{AB/AB}, \ldots, X_{Ab/ab}, n)$. Note that $\mathbf{X}$ has a multinomial distribution where the probability of each progeny genotype can be determined under the presumed linkage phase of FP (see Fig. 10). In Fig. 10 we see, for instance, that if the mother is in the repulsion phase the probability of an Ab/AB progeny is $0.5\,rP(Ab)+0.5(1-r)P(AB)$. Furthermore, the expected number of progeny for each genotype changes with the postulated linkage phase of the mother.

Using this model, we get the likelihood ratio of coupling versus repulsion by finding

$$\Lambda = \frac{L(\text{coupling} \mid \mathbf{X}=\chi)}{L(\text{Repulsion} \mid \mathbf{x}=\chi)} = \frac{P(\mathbf{x}=\chi \mid \text{coupling})}{P(\mathbf{X}=\chi \mid \text{Repulsion})}. \quad (17)$$

This expands to

$$\Lambda = \left[\frac{P^c(AB/ab)+P^c(Ab/aB)}{P^r(AB/ab)+P^r(Ab/aB)}\right]^n \prod_{l \in H}\left[\frac{P^c(l)}{P^r(l)}\right]^{\chi_l} \quad (18)$$

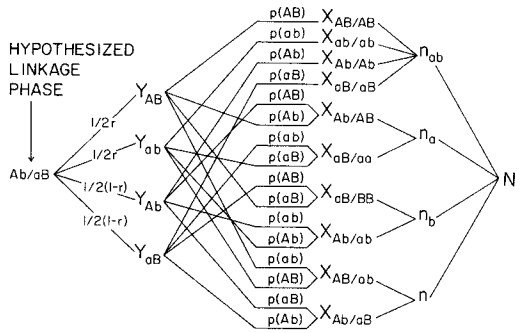

Fig. 10. Mapping of the expected number of progeny of each genotype under the presumed linkage phase of the mother (repulsion for this diagram). See text for definitions of the statistics

where $P^c(l)$ ($P^r(l)$) is the probability of the progeny's genotype being in linkage phase l ($l \in H$) when FP is in the coupling (repulsion) phase, and H is the set of all known linkage phases of the progeny (one/two homozygous loci). Decision:

$\Lambda \gg 1$ coupling phase

$\Lambda \ll 1$ repulsion phase

$\Lambda \doteq 1$ phase unclear.

The smaller r is, the more information we find in Λ and, obversely, as r approaches 0.5, the determination of the proper linkage phase becomes more difficult (at $r=0.5$, $\Lambda=1$, always). This fact, however, is not troublesome since the information contained in, and the effects of, linkage is greatest for small r and must go to 0 as r goes to 0.5. Also, when the alleles are all equally likely, Eq. (18) reduces to

$$\left(\frac{1-r}{r}\right)^{(X_{AB/AB}+X_{ab/ab})}\left(\frac{r}{1-r}\right)^{(X_{Ab/Ab}+X_{ab/ab})},$$

which behaves as we would expect. If we find many offspring in the coupling phase the first term is large and dominates, suggesting that FP is in the coupling phase. If a larger number of offspring are in the repulsion phase, the second term is extremely small and dominates, suggesting that FP is in the repulsion state. Consequently, in this special case of equal gene frequency, only the homozygous offspring offer any information.

If there are more than two loci in the linkage group, linkage can be determined in a pairwise manner. Since the order of the loci on the chromosome will generally be ascertained by some previous genetic analysis, this procedure is straightforward. Also, if more than two alleles are possible at the loci, the method can be extended directly.

Simulations

We examined the performance of this procedure for determining the linkage phase with the simulated data sets described in the previous section (expected exclusion fractions of 0.8, 0.9 and 0.96 only). After creating progeny arrays for each female parent with female and male parent's linkage phases taken into account, we then estimated each of the 48 hermaphrodite's linkage phase for the 4 linked loci, using the procedure outlined above. Recall that each hermaphrodite produced 60 progeny and that the recombination probabilities were 0.05 for each adjacent pair of loci. For these conditions, all linkage phases of all possible locus pairs were assigned correctly (4320/4320).

Discussion

In this paper, we develop a likelihood approach to ascertain populational patterns of paternity. In this method, some fraction between 0 and 1 of the progeny of a particular genotype from a female parent is assigned to each potential male parent, with the fraction calculated by dividing likelihood of paternity for each male parent by the sum of all male parents' likelihoods. We show that this method is superior to the most-likely and the simple exclusion methods.

While our conclusions concerning the relative performances of paternity methods apply to any analysis of populational patterns of paternity, we should point out that there are numerous situations where none of these methods would be appropriate. For instance, in our study, all simulation analyses were performed using a population of 48 hermaphroditic individuals functioning as both female and male parents. As the simulations indicate, the ability to accurately estimate paternal reproductive success is not very good at the lower exclusion fraction values (Figs. 2 and 4) and inaccuracy is exacerbated by violations of the model assumptions (e.g., Fig. 4). We note that while the fractional method minimizes the mean square error, it is always biased if individuals vary in fertility. If the purpose of the study is to determine if individuals in a population vary in fertility, then the fractional method is biased toward the null hypothesis, making it conservative. Moreover, the bias is predictable. On the other hand, under certain conditions, by considering the entire sample of progeny simultaneously, it is possible to obtain a Maximum Likelihood estimator of fertility (Roeder et al. 1988). Combining this approach with the fractional assignment of progeny (by using the fertility estimates as a prior) should lead to a better estimate of paternity.

A. H. D. Brown (personal communication) has pointed out that in those species where matings are known to be correlated, there is additional genetic information concerning paternal identity in the entire progeny array. Specifically, if we knew that all of the seed of a fruit were typically sired by one male parent, then the multilocus

genotypes of all progeny should be considered in the decision on paternal identity. This approach merits investigation; it should be particulary useful in situations where multiple paternity is infrequent (e.g., plants that produce pollinia and certain insects, such as dragonflies).

Our simulations also did not incorporate gene flow; recall that gene flow is defined as interpopulation gene movement. Gene flow is easily detectable when the successful paternal gamete could not be contributed by any member of the population (Ellstrand and Marshall 1985), but it is problematic when the gamete contributed by an individual from outside the population could also be contributed by an individual within the population. The problem of gene flow obscuring interpretations of the population patterns of paternity is made more severe by the fact that the boundary of a population is rarely circumscribed with certainty. For instance, Ellstrand and Marshall (1985) examined gene movement among locationally-circumscribed populations of wild radish, finding conservative values of "gene flow" of 8%–20% (conservative because they underestimated the actual value). While it could be argued that the values were too high to represent gene flow, they illustrated the importance of knowing the actual boundaries of the population when the question of interest is the populational pattern of paternity.

Experimental populations that are genetically structured so that paternity and gene flow are unambivalent can be important tools by which the genetic parameters, such as effective population size and genetic neighborhood area, are delimited (Levin and Kerster 1974; Schaal 1980; Levin 1981, 1983; Ennos and Clegg 1982; Handel 1983). Data on neighborhood size will be essential in order to study populations with large numbers of individuals. Moreover, structured populations should permit evaluation of the effects of plant spacing, male parent attractiveness, flower production, degree of relatedness and intrinsic features such as pollen production on paternal reproductive success and intra-versus interpopulation gene movement. Of course, populations need not be structured to obtain much of this information if the exclusion fraction is relatively high or the population size is small so that the genetic information is sufficient to track patterns of paternity in natural populations.

Appendix

Theorem. If the assumptions of random mating and equal male fertility are met then $\hat{F}_{ij}$ is an unbiased estimator of populational patterns of paternity, namely $E[\hat{F}_{ij}] = P(MP=j|FP=i) \times N$, where N is the number of offspring. Otherwise

$$E[\hat{F}_{ij}] = N \sum_k \{\sum_j P(0=k|MP=j, FP=i)\, P(MP=j|FP=i)\} \cdot \left\{\frac{P(0=k|MP=j^*, FP=i)}{\sum P(0=k|MP=j, FP=i)}\right\}. \quad (A1)$$

Proof. First note that clearly,

$$E[X_i(\gamma)] = N(P(0=k|FP=i) \quad (A2)$$

$$= N \sum_j P(0=k|MP=j, FP=i)\, P(MP=j|FP=i),$$

by the law of total probability and the multiplicative rule. By Bayes Theorem

$$P(MP=j^*|FP=i, 0=k) \quad (A3)$$

$$= \frac{P(0=k|MP=j^*, FP=i)\, P(MP=j^*|FP=i)}{\sum P(0=k|MP=j, FP=i)\, P(MP=j|FP=i)}.$$

From Eq. (7),

$$E[\hat{F}_{ij}] = \sum_k E[X_{ik}]\, P(MP=j|FP=i, 0=k). \quad (A4)$$

Substituting in equations A2 and A3 we get

$$E[\hat{F}_{ij}] \quad (A5)$$

$$= N \sum_k \{\sum_j P(0=k|MP=j, FP=i)\, P(MP=j|FP=i)\} \cdot \left\{\frac{P(0=k|MP=j^*, FP=i)\, P(MP=j^*|FP=i)}{\sum_j P(0=k|MP=j, FP=i)\, P(MP=j|FP=i)}\right\}.$$

By cancellation this reduces to

$$N \sum_k P(0=k|MP=j^*, FP=i)\, P(MP=j^*|FP=i)$$

$$= N \sum_k P(0=k, MP=j^*|FP=i) \quad (A6)$$

by the multiplicative rule. Then by the law of total probability this reduces to

$$N\, P(MP=j^*|FP=i), \quad (A7)$$

which was our claim.

Thus, we can conclude that $\hat{F}_{ij}$ is an unbiased estimator of F_{ij}, provided that we could estimate $P(MP=j|FP=i, 0=k)$ by equation 15. However, we determined that this was impossible except in the trivial case specified by our assumptions since it would require knowledge of the probability, $P(MP=j|FP=i)$, which is in fact unknown (and the parameter of interest). Given the trivial case, this probability is constant over j and i and cancels out of our calculations. However, if we estimate $P(MP=j|FP=i, 0=k)$ by equation 8 as recom-

mended by the fractional method, we introduce the following bias.

$$\mathbf{E}[\hat{F}_{ij}] = N \sum_k \{\sum_j P(0=k \mid MP=j, FP=i)\, P(MP=j \mid FP=i)\} \cdot \left\{ \frac{P(0=k \mid MP=j^*, FP=i)}{\sum P(0=k \mid MP=j, FP=i)} \right\}. \quad (A8)$$

Now, the sums over j no longer cancel unless $P(MP=j \mid FP=i)$ is a constant over j.

Acknowledgements. We gratefully acknowledge the support of the Academic Computing Center (UCR), who provided the facilities necessary for the simulations. This research was funded by National Science Foundation grant BSR-8505982 to N. C. Ellstrand.

References

Brown AHD, Barrett SCH, Morgan GF (1985) Mating system estimation in forest trees: models, methods and meanings. In: Gregorius HR (ed) Population Genetics in Forestry. Springer, Berlin Heidelberg New York pp 32–49

Brown AHD, Grant JE, Pullen R (1986) Outcrossing and paternity in *Glycine argyrea* by pair fruit analysis. Biol J Linn Soc 29:283–294

Chakraborty R, Hedrick PW (1983) Paternity exclusion and the paternity index for two linked loci. Hum Hered 33:12–23

Chakraborty R, Shaw M, Schull WJ (1974) Exclusion of paternity: the current state of the art. Am J Hum Genet 26:477–488

Cheliak WM, Skroppa T, Pitel JA (1987) Genetics of the polycross. I. Experimental results from Norway spruce. Theor Appl Genet 73:321–329

Devlin B, Stephenson AG (1987) Sexual variations among plants of a perfect-flowered species. Am Nat 130:199–218

Elandt-Johnson RC (1971) Probability models and statistical methods in genetics. Wiley, New York

Ellstrand NC (1984) Multiple paternity within the fruits of the wild radish, *Raphanus sativus*. Am Nat 123:819–828

Ellstrand NC, Marshall DL (1985) Interpopulation gene flow by pollen in wild radish, *Raphanus sativus*. Am Nat 126: 606–612

Ellstrand NC, Marshall DL (1986) Patterns of multiple paternity in populations of *Raphanus sativus*. Evolution 40:837–842

Ennos RA, Clegg MT (1982) Effect of population substructuring on the estimates of outcrossing rate in plant populations. Heredity 48:283–292

Ennos RA, Dodson RK (1987) Pollen success, functional gender and assortative mating in an experimental plant population. Heredity 58:119–126

Friedman ST, Adams WT (1985) Estimation of gene flow into two seed orchards of loblolly pines (*Pinus taeda* L.). Theor Appl Genet 69:609–615

Hamrick JL, Schnabel A (1985) Understanding the genetic structure of plant populations: some old problems and a new approach. In: Gregorius HR (ed) Population Genetics in Forestry. Springer, Berlin Heidelberg New York, pp 50–70

Handel SN (1983) Contrasting gene flow patterns and genetic subdivision in adjacent populations of *Cucumis sativus* (Cucurbitaceae). Evolution 37:760–771

Hanken J, Sherman PW (1981) Multiple paternity in Belding's ground squirrel litters. Science 212:351–353

Levin DA (1981) Dispersal versus gene flow in plants. Ann Mo Bot Gard 68:233–253

Levin DA (1983) An immigration-hybridization episode in *Phlox*. Evolution 37:575–582

Levin DA, Kerster HW (1974) Gene flow in seed plants. Evol Biol 7:139–220

Meagher TR (1986) Analysis of paternity within a natural population of *Chamaelirium luteum*. I. Identification of most-likely male parents. Am Nat 128:199–215

Meagher TR, Thompson EA (1986) The relationship between single parent and parent pair genetic likelihoods in genealogy reconstruction. Theor Popul Biol 29:87–106

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

Muller-Starck G, Ziehe M (1984) Reproductive systems in conifer seed orchards. 3. Female and male fitnesses of individual clones realized in seeds of *Pinus sylvestris* L. Theor Appl Genet 69:173–177

Neale DB (1983) Population genetic structure of the shelterwood regeneration system in southwest Oregon. PhD thesis, Oregon State University, Corvallis

Neel JV, Schull WJ (1954) Human Heredity. University of Chicago Press, Chicago

Roeder KM, Devlin B, Lindsay BG (1988) Application of maximum likelihood methods to population genetic data for the estimation of individual fertilities. Biometrics (in press)

Ryman N, Chakraborty R (1982) Evaluation of paternity-testing data from the joint distribution of paternity index and the rate of exclusion. Hereditas 96:49–54

Schaal B (1980) Measurement of gene flow in *Lupinus texensis*. Nature 284:450–451

Schoen DJ, Stewart SC (1986) Variation in male reproductive investment and male reproductive success in White Spruce. Evolution 40:1109–1120

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

Smith DB, Adams WT (1983) Measuring pollen contamination in clonal seed orchards with the aid of genetic markers. Proc 17th So Tree Improvement Conf, University of Georgia, Athens/GA, pp 69–77

Smouse PE, Chakraborty R (1986) The use of restriction fragment length polymorphism in paternity analysis. Am J Hum Genet 38:918–939

Thompson EA (1986) Likelihood inference of paternity. Am J Hum Genet 39:285–287

Trivers R (1985) Social Evolution. Benjamin/Cummings, Menlo Park

Valentin J (1980) Exclusion and attributions of paternity: practical experience of forensic genetics and statistics. Am J Hum Genet 32:420–431

Yampolsky E, Yampolsky H (1922) Distribution of sex forms in the phanerogamic flora. Bibliogr Genet 3:1–62