

# Measuring the genetic structure of the pollen pool as the probability of paternal identity

PE Smouse<sup>1,2</sup> and JJ Robledo-Arnuncio<sup>2,3</sup>

<sup>1</sup>Department of Ecology, Evolution & Natural Resources, Cook College, Rutgers University, New Brunswick, NJ 08901, USA;

<sup>2</sup>CIFOR-INIA, Ctra. de la Coruña km 7,5, 28040 Madrid, Spain; <sup>3</sup>Unidad de Anatomía, Fisiología y Genética Forestal, ETSI de Montes, Ciudad Universitaria s/n, 28040 Madrid, Spain

Contemporary pollen flow in forest plant species is measured by the probability of paternal identity (PPI) for two randomly sampled offspring, drawn from a single female, and contrasting that with PPI for two random offspring, drawn from different females. Two different estimation strategies have emerged: (a) an indirect approach, using the 'genetic structure' of the pollen received by different mothers and (b) a direct approach, based on parentage analysis. The indirect strategy is somewhat limited by the assumptions, but is widely useful. The direct approach is most appropriate where a large majority of the true fathers can be identified exactly, which is sometimes possible with high-resolution SSR markers. Using the parentage approach, we develop estimates of PPI, showing that the obvious estimates are

severely biased, and providing an unbiased alternative. We then illustrate the methods with SSR data from a 36-tree isolated population of *Pinus sylvestris* from the Meseta region of Spain, for which categorical paternity assignment was available for over 95% of offspring. For all the females combined, we estimate that  $PPI = 0.0425$ , indicating uneven male reproductive contributions. Different (but overlapping) arrays of males pollinate different females, and for the average female,  $PPI = 0.317$ , indicating substantial 'pollen structure' for the population. We also relate the direct measures of PPI to those available from indirect approaches, and show that they are generally comparable.

*Heredity* (2005) **94**, 640–649. doi:10.1038/sj.hdy.6800674  
Published online 11 May 2005

**Keywords:** paternal analysis; paternal identity; pollen structure; pollen flow; genetic diversity; correlated paternity

## Introduction

The movement of genes across the landscape is of critical interest to evolutionary and conservation biologists. The long-term evolutionary theory governing this situation dates to Wright (1943, 1946), and there are numerous indirect estimation methods for effective gene flow rates (eg, Malécot, 1969; Morton, 1977; Cockerham and Weir, 1993; Slatkin, 1993; Rousset, 1997). The centerpiece of this theory is the probability of identity by descent (IBD) for two uniting alleles, gauging the level of 'population structure' resulting from restricted dispersal and limited reproductive density. For continuous populations, the theory is couched in terms of IBD at any particular point location on that landscape. In the case of subdivided populations, the theory is couched in terms of IBD within and among discrete subpopulations. It is also customary to invoke the *inbreeding effective population size*,  $N_e = 1/IBD$ , defined as the number of idealized reproductive adults (contributing equally and randomly) that would yield the estimated level of IBD (Wright, 1943; Crawford, 1984), a familiar population size representation of population structure, but the parameter of measurable interest is IBD itself.

By contrast, our ability to measure contemporary gene flow is rudimentary, particularly where recent anthro-

pogenic disturbance has altered the pattern of gene flow across the landscape, thus altering the existing pattern of genetic affinity. Moreover, contemporary gene flow in forest plant species is thought to be controlled more by pollen than by seed dispersal, and recent attention has focused on pollen flow (see review in Smouse and Sork, 2004). One way to characterize the level of pollen structure is to estimate the probability of paternal identity (PPI) of two offspring, drawn at random from a single female, and to contrast that with PPI for two offspring, drawn at random from two different females (Austerlitz and Smouse, 2001a). We use such measures as a gauge of restricted pollen flow across the landscape (the failure of panmictic pollination). As with its IBD analogue, we can describe pollen structure as the effective number of pollen donors,  $N_{ep} = (1/PPI)$ , but PPI itself is the real target of inquiry.

For estimating contemporary pollen structure, two different strategies have emerged: (a) an indirect approach, based on the 'genetic structure' of the pollen clouds received by different mothers and (b) a more direct approach, based on formal parentage analysis. The indirect strategy compares genetic similarity of offspring within and among maternal sibships, without relying on exact paternal designations. Genetic inference in such cases is necessarily limited and heavily assumption-dependent, but these methods are very general and have proven to be widely useful (see Ritland, 1989, 2002; Hardy *et al*, 2004; Smouse and Sork, 2004). The direct (paternity-based) approach is better for the case where a large majority of the true fathers of the offspring from a

set of maternal sibships can be identified exactly. With traditional allozyme markers, that level of resolution was almost never achievable (Chakraborty *et al*, 1988), but multiallelic microsatellite markers have increased the parental resolution to a level that would support the enterprise (eg, Streiff *et al*, 1999). There has been some effort devoted to translating paternity data into an estimate of PPI (see Pamilo, 1993; Burczyk *et al*, 2002; Nielsen *et al*, 2003), but the matter needs some further attention.

For illustration, it would also be good to have some real data sets, for which assignment of paternity could be done categorically for the vast majority of offspring, but there are still very few data sets that qualify. We are fortunately able to access an extremely unusual data set from Scots pine (*Pinus sylvestris* L) that meets the need. In the Meseta region of Spain, there is a relict population at Coca, consisting of exactly 36 mature individuals and no younger recruits. The nearest conspecific populations are 30–60 km distant, and the population is reproductively quite isolated. The paternity of over 95% of the seeds can be categorically assigned to exactly one of the 36 resident males (Robledo-Arnuncio and Gil, 2005); this population will support the direct estimation of PPI, using a paternity-derived approach.

Our objectives are: (1) to develop appropriate estimates of the PPI for individual mothers, for pairs of mothers, and for the entire population, using direct paternal designations from the offspring of maternal sibships, and (2) to illustrate the usage of these methods with microsatellite data from a highly isolated relict population of *P. sylvestris* from the Meseta region of Spain.

## Theoretical development

### PPI

It will be convenient to use the shorthand term ‘mother’ for the seed-parent and ‘father’ for the pollen donor, with the understanding that the organism in question might be monoecious and that selfing is sometimes a possibility. The essence of our problem is the need to determine the probability that two seedlings, drawn at random from the same mother, have the same father, which we will label as the *probability of paternal identity*, PPI. Assume that there is a specific array of males ( $F_k$ :  $k=1, \dots, K$ ) that could father offspring for the collection of mothers under study ( $M_g$ :  $g=1, \dots, G$ ). In a small, isolated population,  $K$  might be small ( $K \leq N$ , where  $N$  is the number of reproductive adults in the population), but in a large, extensive population,  $K$  is larger than the modest sample of offspring drawn from any single mother, and it may be larger than the total number of offspring sampled.

Concentrating here on the PPI values for single mothers, we define the probability of the  $k$ th male contributing paternity to the  $g$ th mother as  $\lambda_{gk}$ , where

$$\sum_{k=1}^K \lambda_{gk} = 1 \quad (1)$$

The  $\lambda$ 's will not generally be the same for the different fathers, nor will they be the same for the same father but with different mothers. Identifying the sources of the inequalities is sometimes the motivation for the

study, but our concern here is solely with the set of  $\lambda$ 's themselves, from which we can obtain an estimate of PPI.

Assuming that we have an assay battery that provides sufficient genetic resolution to designate the male parents of all  $n_g$  offspring from the  $g$ th mother, an obvious estimate of the probability of a random pair of these progeny having the same father, a measure of PPI, is

$$q_{gg} = \sum_{k=1}^K p_{gk}^2 = \sum_{k=1}^K \lambda_{gk}^2 \quad (2a)$$

where  $p_{gk} = (x_{gk}/n_g)$  and  $x_{gk}$  is the number of offspring from the  $g$ th mother for whom the  $k$ th male is the designated father. Now,  $p_{gk}$  is an unbiased estimate of the parametric probability of the  $k$ th father contributing to the  $g$ th mother ( $\lambda_{gk}$ ), but unfortunately,  $q_{gg}$  is upwardly biased to a degree that depends on the sample size ( $n_g$ ) and on the unknown parameter (PPI) itself. For very large sample sizes, the bias is small, but we usually sample modest numbers of offspring for single mothers, and the bias can be large (Nei and Roychoudhury, 1974; Pamilo, 1993).

It is possible to obtain an unbiased estimate of PPI from the array of paternal contributions, but we need to extract it in slightly different fashion. Begin with the array of  $n_g$  offspring from the  $g$ th mother, exhibiting an array of fathers, in numbers  $x_{g1}, x_{g2}, \dots, x_{gK}$ , where  $K$  is the total number of realized fathers for the study. For any single sibship, many of the  $x$ 's will be observably '0', but that will not affect what follows. The parametric criterion PPI is a measure of ‘paternal matching’, and we can estimate it by sampling offspring without replacement. We estimate PPI as the rate of paternal matching within the  $g$ th sibship. Formally, we need

$$\begin{aligned} r_{gg} &= \frac{\binom{x_{g1}}{2} + \binom{x_{g2}}{2} + \dots + \binom{x_{gK}}{2}}{\binom{n_g}{2}} \\ &= \frac{(n_g q_{gg} - 1)}{(n_g - 1)} \end{aligned} \quad (2b)$$

using combinatorial enumeration of pairs sampled without replacement, for example,

$$\binom{x_{g1}}{2} = \frac{x_{g1}!}{2!(x_{g1} - 2)!} = \frac{x_{g1}(x_{g1} - 1)}{2} \quad (3)$$

and so on. Nei and Roychoudhury (1974) have shown, in the context of a genetic diversity measure for a single population, that the far right-hand side of Equation (2b) provides an unbiased estimate of the probability of allelic identity (PAI), the population analog of PPI, as defined here. Similar derivation shows that  $r_{gg}$  is an unbiased estimator of PPI. In other words, to estimate the *probability of paternal identity*, we simply tally the paternal matches directly, rather than estimate them indirectly from the array of estimated paternal contributions.

To compare the statistical properties of  $q_{gg}$  and  $r_{gg}$  as estimators of PPI, it will be useful to determine their respective mean-squared errors (MSE), while separately considering two components, variance (Var) and (Bias)<sup>2</sup>. Using Equations (1) and (12) in Nei and Roychoudhury

(1974), the expected MSE of  $q_{gg}$  is

$$\begin{aligned} \text{MSE}(q_{gg}) &= \frac{2(n_g - 1)}{n_g^3} \\ &\times \left[ (3 - 2n_g) \text{PPI}^2 + 2(n_g - 2) \sum_k \lambda_{gk}^3 + \text{PPI} \right] \\ &+ \left( \frac{1 - \text{PPI}}{n_g} \right)^2 \end{aligned} \tag{4a}$$

where  $\lambda_{gk}$  is the relative paternal contribution of the  $k$ th male to the  $g$ th female. The first component of the right-hand side is  $\text{Var}(q_{gg})$ , and the second is  $\text{Bias}^2(q_{gg})$ , both defined in terms of the unknown parametric values of the  $\lambda_{gk}$  and PPI. For the case of  $r_{gg}$ , the bias is (fortunately) zero, and  $\text{MSE} = \text{Var}(r_{gg})$  is given by (Eq. 8.12 in Nei, 1987)

$$\begin{aligned} \text{MSE}(r_{gg}) &= \frac{2}{n_g(n_g - 1)} \\ &\times \left[ 2(n_g - 2) \left( \sum_k \lambda_{gk}^3 - \text{PPI}^2 \right) + \text{PPI} - \text{PPI}^2 \right] \end{aligned} \tag{4b}$$

In order to investigate the MSE of the two estimators for different parametric values of PPI and sample size ( $n_g$ ), it is necessary to assume a distribution for the relative paternal contributions ( $\lambda_{gk}$ ). For simplicity, we use a geometric distribution, with dominance parameter  $\alpha$ , under which

$$\lambda_{gk} = \alpha(\alpha - 1)^{k-1} \text{ and } \sum_k \lambda_{gk}^\mu = \alpha^\mu / [1 - (1 - \alpha)^\mu] \tag{5}$$

where the  $\lambda_{gk}$ -values are again the proportional contributions of the various pollen donors to the  $g$ th female, and where  $\mu$  is any chosen exponential integer. Using these relationships, together with Equations (2a) and (2b), we generated  $\lambda_{gk}$  distributions, with corresponding PPI values ranging from 0 to 1, and then calculated the variance, bias and MSE of  $q_{gg}$ , as well as the  $\text{MSE} = \text{Var}$  of  $r_{gg}$ , for three different numbers of sampled offspring,  $n = 10, 20$ , and  $80$  (Figures 1 and 2).

The bias of  $q_{gg}$  increases as PPI and the number of offspring sampled per mother tree ( $n$ ) decrease, becoming large when PPI is small (many, relatively equal paternal contributors) for  $n$  values commonly used in pollen flow studies ( $n = 10$  or  $20$ ; Figure 1). As an example, for PPI values below 0.2, the relative bias of  $q_{gg}$  would be larger than 20% when  $n$  is below 20, and would become as large as 200% for  $\text{PPI} = 0.05$  and  $n = 10$ . By contrast, the variances of both  $q_{gg}$  and  $r_{gg}$  reach maxima for PPI values of about 0.7, always being larger for  $r_{gg}$  than for  $q_{gg}$ , although the difference becomes small for large sample sizes (Figure 1). Thus,  $r_{gg}$  is a more accurate but less precise estimator than  $q_{gg}$ . The relative performance of the two estimators is determined by the parametric value of PPI itself,  $r_{gg}$  being better (smaller MSE) for PPI values below 0.4 (Figure 2). This threshold value ( $\text{PPI} = 0.4$ ) is virtually constant for any sample size above five, under the hypothesized geometric distribution of paternal contributions (data not shown), and may be used to choose between the two estimators. In many studies dealing with plant species, especially wind-pollinated trees, PPI has been found to fall below 0.4 (see review in Smouse and Sork, 2004), which suggests that  $r_{gg}$  will be a better estimator in most situations.

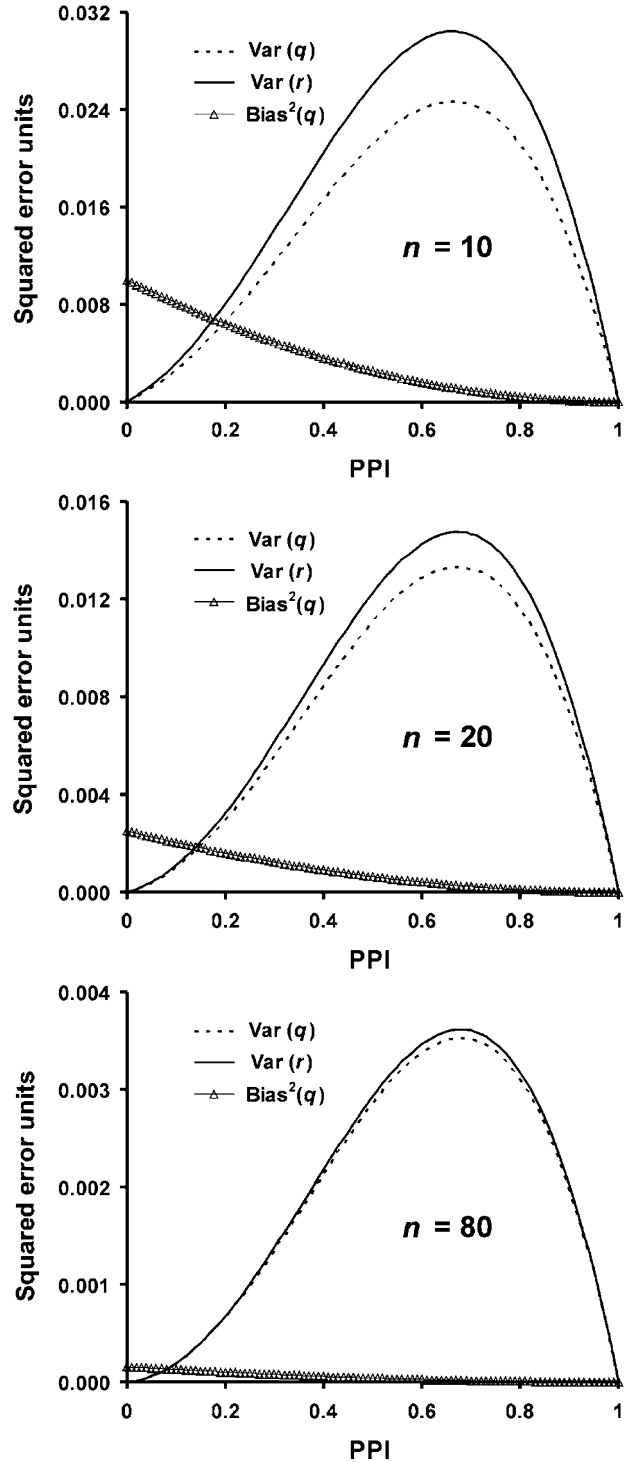
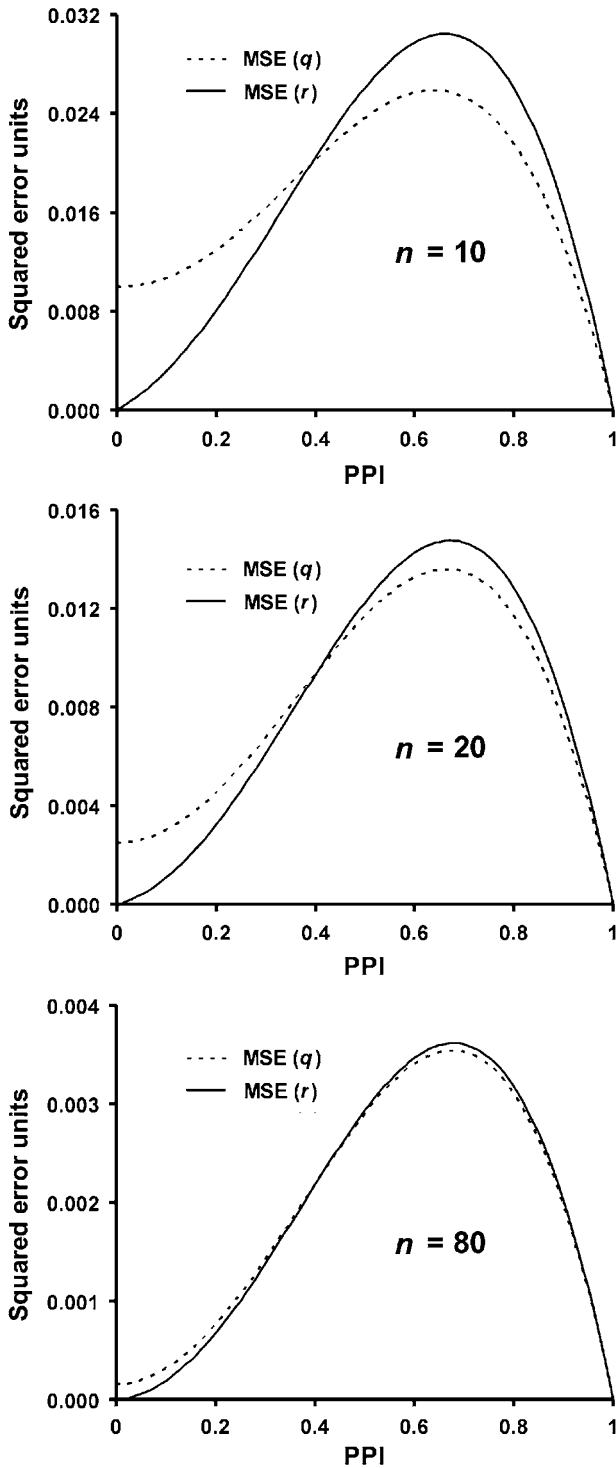


Figure 1 Expected variance (Var) and squared bias (Bias<sup>2</sup>) of  $q$  estimator and variance of  $r$  estimator (without bias) for the probability of paternal identity (PPI), as a function of the parametric value of PPI and for three different values ( $n = 10, 20, 80$ ) of the number of offspring per maternal sibship. Note that the vertical axis scale is different for each  $n$ .

In general, the  $g$ th and  $h$ th mothers will have different probabilities of being pollinated by any particular male,  $\lambda_{gk} \neq \lambda_{hk}$ , for any of a large number of interesting reasons. We can expect to encounter some variation in single-mother PPI values. It would be useful to obtain an



**Figure 2** The expected mean-squared error (MSE) of  $q$  and  $r$  estimators of the probability of paternal identity (PPI), as a function of the parametric value of PPI and for three different values ( $n = 10, 20, 80$ ) of the number of offspring per maternal sibship. Note that the vertical axis scale is different for each  $n$ .

estimate for the ‘average female’. If all the  $n_g$  are equal, then clearly we can use

$$\bar{r}_0 = \frac{r_{11} + r_{22} + \dots + r_{GG}}{G} \quad (6a)$$

but in the more usual case where the sample sizes are unequal ( $n_g \neq n_h$ ), we require a weighted average of the components of Equation (2b), specifically

$$\bar{r}_0 = \frac{\sum_{g=1}^G \sum_{k=1}^K \binom{x_{gk}}{2}}{\sum_{g=1}^G \binom{n_g}{2}} \quad (6b)$$

where the summations are over  $G$  mothers and  $K$  potential fathers.

Note that the expected variance of  $\bar{r}_0$  decreases with an increasing number of maternal sibships and has no bias. While the variance of  $q_{gg}$  is a bit smaller, it is biased, even when averaging across maternal sibships. As a practical matter, averaging  $r_{gg}$  (rather than  $q_{gg}$ ) across maternal sibships is likely to yield estimates of PPI with a smaller mean-squared error for the average sibship. For what follows, we will thus use Equation (2b), rather than (2a), and will extend that preference to Equation (6a) or (6b).

#### Global PPI

If the different mothers are drawing differentially from different fathers, then the global array of fathers will be larger for the collection of sampled mothers than for any one mother. If we ignore the question of which mothers are sampling which fathers for the moment, then we can label the total number of offspring fathered by the  $k$ th male as  $X_k$ , which is simply the sum of the  $x_{gk}$  over all sampled mothers, and can compute an unbiased estimate of global PPI for the total population as

$$R_0 = \frac{\binom{X_1}{2} + \binom{X_2}{2} + \dots + \binom{X_K}{2}}{\binom{N}{2}} \quad (7)$$

using combinatorial enumeration of matching paternal pairs, and with  $N$  as the total number of offspring examined, the sum of the  $n_g$ -values for the collection of  $G$  sampled mothers.  $R_0$  is a measure of the nonequal contributions of the  $G$  candidate males to the entire population, irrespective of the female-specific pattern of that pollination. There is also a relationship between the PPI value for a single mother and that for the collection of sampled mothers, but to elucidate it, we must first characterize the cross-maternal rate of paternal sharing.

#### Paternal overlap

To characterize the degree of paternal overlap among mothers, we need to consider the probability that two random offspring, one from the  $g$ th mother and one from the  $h$ th mother, have the same father, for which we compute the unbiased estimator,

$$r_{gh} = \frac{\binom{x_{g1}}{1} \binom{x_{h1}}{1} + \dots + \binom{x_{gK}}{1} \binom{x_{hK}}{1}}{\binom{n_g}{1} \binom{n_h}{1}}$$

$$= \sum_{k=1}^K p_{gk} p_{hk} \quad (8)$$

a tally of the proportion of pairwise paternal matches for the  $g$ th and  $h$ th mothers. With our complete battery of PPI estimates, we build a  $G \times G$  square symmetric matrix  $\mathbf{R}$ , having the within-mother ( $r_{gg}$ ) values down the diagonal and the among-mother ( $r_{gh}$ ) pairwise values off-diagonal

$$\mathbf{R} = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1G} \\ r_{21} & r_{22} & \cdots & r_{2G} \\ \vdots & \vdots & \ddots & \vdots \\ r_{G1} & r_{G2} & \cdots & r_{GG} \end{bmatrix} \quad (9)$$

To gauge the impact of paternal sharing among maternal sibships on the relationship between  $\bar{r}_0$  and  $R_0$ , it is useful to simplify the notation. Equation (2b) can also be written as

$$r_{gg} = \sum_{k=1}^K \frac{x_{gk}(x_{gk} - 1)}{n_g(n_g - 1)} \equiv \frac{Y_{gg}}{n_g(n_g - 1)} \quad (10a)$$

where  $Y_{gg}$  is the tally of paternal matches within the  $g$ th sibship. We can also write Equation (8) as

$$r_{gh} = \sum_{k=1}^K \frac{x_{gk}x_{hk}}{n_g n_h} \equiv \frac{Y_{gh}}{n_g n_h} \quad (10b)$$

where  $Y_{gh}$  represents the tally of paternal matches between the  $g$ th and  $h$ th sibships. To simplify further, assume that  $n_g = n$  for all  $g = 1, \dots, G$  maternal sibships. We can then write

$$\bar{r}_0 = \sum_{g=1}^G \frac{Y_{gg}}{N(n-1)} \quad (11a)$$

mathematically identical to Equation (4a), and the average rate of paternal sharing across mothers as

$$\bar{r}_{gh} = \sum_{g \neq h}^G \sum_{h=1}^G \frac{Y_{gh}}{N(N-n)}. \quad (11b)$$

Using the same notational simplifications, we can also rewrite  $R_0$  as

$$R_0 = \frac{\left[ \sum_{g=1}^G Y_{gg} + \sum_{g=1}^G \sum_{h=1, h \neq g}^G Y_{gh} \right]}{N(N-1)} \quad (12a)$$

which simplifies to

$$R_0 = \frac{(n-1)\bar{r}_0 + (N-n)\bar{r}_{gh}}{(N-1)} \quad (12b)$$

or in linear algebra terms, to

$$R_0 = \frac{n \cdot \text{Sum}(\mathbf{R}) - \text{Tr}(\mathbf{R})}{G(N-1)} \quad (12c)$$

where  $\text{Sum}(\mathbf{R})$  is the sum of all  $G^2$  elements of Equation (9), and  $\text{Tr}(\mathbf{R})$  is the sum of the  $G$  diagonal elements for that same matrix.

In any event, with no paternal overlap among mothers, the second term in the numerator of Equation (12b) becomes '0', and the equation degenerates to  $(N-1)R_0 = (n-1)\bar{r}_0$ , so that  $\bar{r}_0 \sim G \cdot R_0$ . On the other hand, if each mother receives the same array of paternal contributions (the same fathers and in the same proportions), then we should expect the global PPI value to be virtually the same as that for a single mother. Formally, if we set  $Y_{gh} = nY_{gg}/(n-1)$  in Equation (11b) to allow for the

fact that we are sampling without replacement within any single sibship, manipulation of Equation (12a) and (12b) yields  $R_0 = \bar{r}_0$ . In the nonoverlap case, we would need to sample all the mothers to characterize the global array of fathers, but in the case of total overlap, sampling one mother would be sufficient to characterize the entire population of fathers. In general, we observe a bounded ratio of within- to global-PPI estimates,  $1 < (\bar{r}_0/R_0) < (N-1)/(n-1)$ , tending to '1' as the degree of paternal overlap increases and to the number of different mothers ( $G$ ) as the overlap tends to '0'.

### Effective number of pollen parents, $N_{ep}$

It has become customary to extract an estimate of  $N_{ep}$ , the number of 'idealized pollen donors', each contributing equally (all  $\lambda_{gk} = \lambda$ ) and randomly to a single mother (or collection of mothers), which would yield the estimated value of PPI (see review in Smouse and Sork, 2004). The usual strategy is to use  $N_{ep} = (1/q_{gg})$ , but since  $q_{gg}$  is biased, Pamilo (1993) suggests using  $(1/r_{gg})$  instead. For the average mother, we use  $N_{ep} = (1/\bar{r}_0)$ , and for the total population of mothers, we use  $N_{ep} = (1/R_0)$ . Nielsen *et al* (2003) have pointed out that even the estimate  $N_{ep} = (1/\bar{r}_0)$  is biased. [The inverse of an unbiased estimate of a parameter is not, in general, an unbiased estimate of the inverse of that parameter.] For small sample sizes (say  $n_g < 20$ ), that bias can be substantial. For the total population, the sample size for  $R_0$  is much larger,  $N \sim G \cdot n$ , and the bias is small. Nielsen *et al* (2003) offer a more elaborate estimator with slightly smaller bias and mean-squared error, and we refer the readers primarily interested in an estimate of  $N_{ep}$  to that paper. As always with an effective number,  $N_{ep}$  is strictly a derivative construct, chosen to reflect some measurable 'population structure' criterion (here PPI) and to provide a familiar reference frame for thinking about the degree of subdivision (Austerlitz and Smouse, 2001a). The actual analytical target is PPI;  $N_{ep}$  is simply a useful derivative of the  $r$ - and  $R$ -estimates, which latter are unbiased by construction.

## An empiric example – pollen structure in Coca Scots Pine

### The coca remnant

Scots Pine (*P. sylvestris* L.) is a monoecious, wind-pollinated, predominantly outcrossing conifer. The species reaches the southern limit of its distribution in Spain. The singular Scots Pine population at Coca represents a small and isolated remnant of the formerly widespread populations of the species that grew in the Northern Meseta plains of the Iberian Peninsula (Franco-Múgica *et al*, 2001). Holocene climatic warming, along with a recent human-mediated lowering of the water table in the area, have reduced the Scots Pine population to its present size of 36 remnant adult trees. The Scots Pine adults at Coca are scattered through a larger and continuous Maritime Pine (*P. pinaster* Aiton) woodland, over approximately 15 ha (about 2.4 trees/ha; mean pairwise distance = 180 m). The closest conspecific populations are 30 and 60 km away, respectively.

**Paternal assay**

The population is ideal for an illustration of the analysis described here. It is quite isolated from conspecific pollen sources, and the vast majority of pollination should be very local. With only 36 candidate fathers to evaluate, exhaustive enumeration is possible, and categorical delineation of paternity is feasible. Robledo-Arnuncio and Gil (2005) have investigated mating patterns at Coca precisely. They used four chloroplast and two nuclear microsatellite loci to determine the paternal genotype of 813 open-pollinated seeds collected from 34 out of 36 trees in the stand (22–24 seeds per tree), with two trees contributing no seed. For details of laboratory techniques to identify the molecular markers, see Robledo-Arnuncio and Gil (2005). The expected paternal exclusion probability for the six-marker set was 0.996, but each of the potential fathers was uniquely identifiable, and it was possible to assign paternity categorically to exactly one of the 36 resident fathers for 95.7% of the offspring (778 of the 813). Those same 36 resident males were categorically excluded as the fathers of the remaining 35 offspring (4.3%), differing by one or two SSR markers each. These outlier offspring were ascribable to at least 30 different paternal donors, none present within the Coca isolate. Careful re-genotyping of these outliers showed consistent results; they were not a reflection of lab error.

This unusually powerful genetic assay provides a direct count of the pollinating males for each of 34 seed-trees, allowing for a direct estimation of PPI. Excluding the 35 outliers for the moment, we illustrate the methods introduced here with a balanced subset of 20 offspring from each of the 34 females, all from within the Coca population. We will return to the question of how to handle the outliers in Discussion. For the 680 offspring considered here, there is substantial variation in the paternal array for different mothers, and within any one mother, the different fathers contribute differentially. We present a partial tally of the results in Robledo-Arnuncio and Gil (2005) in Table 1, illustrating the type of raw paternity data we can expect from this sort of study. The full data set is available from JRA upon request. Using these tallies, we next compute the **R**-matrix (34 × 34) that contains the observed within-mother ( $r_{gg}$  – diagonal) and among-mother ( $r_{gh}$  – off-diagonal) paternal match fractions, the upper left corner of which is shown in Table 2.

The paternal matching rates within individual mothers range from  $0.047 < r_{gg} < 0.900$  (average  $\bar{r}_0 = 0.317 \pm 0.015$ ). That standard deviation is computed as

$$SD(\bar{r}_0) = \sqrt{Var(\bar{r}_0)} = \sqrt{\left(\frac{1}{G}\right)^2 \cdot \sum_{g=1}^G Var(r_{gg})} \quad (13)$$

**Table 1** Upper-left, 10 × 10 corner of the paternal tally matrix for Coca Scots pine; total for each mother is set at 20 offspring; total paternal tally, over 34 mothers (680 offspring) in the bottom row

Mothers	Contributing fathers											Mat. total
	1	2	3	4	5	6	7	8	9	10	...	
1	0	10	0	0	0	0	0	3	0	0	...	20
2	1	4	0	1	1	0	0	2	0	1	...	20
3	0	0	0	18	0	0	0	0	1	0	...	20
4	4	5	5	0	0	2	0	1	0	1	...	20
5	0	0	1	0	13	2	0	0	0	1	...	20
6	0	0	0	1	0	16	1	0	1	0	...	20
7	1	0	0	0	0	3	7	3	1	1	...	20
8	1	0	0	0	0	1	3	3	3	4	...	20
9	2	0	0	0	0	0	9	1	0	7	...	20
10	0	0	0	0	0	0	3	1	4	0	...	20
...	...	...	...	...	...	...	...	...	...	...	...	...
Pat.total	14	26	6	21	16	30	28	36	23	46	...	680

**Table 2** Upper-left, 10 × 10 corner of the **R**-matrix for Coca Scots pine, with diagonal terms calculated from  $20(19/2) = 190$  pairs of offspring and off-diagonal terms representing  $20(20) = 400$  pairs; each element is a measured rate of paternal sharing, an unbiased estimate of PPI

Mothers	Mothers										
	1	2	3	4	5	6	7	8	9	10	...
1	0.274	0.140	0	0.135	0	0.008	0.052	0.052	0.015	0.070	...
2	0.140	0.058	0.045	0.070	0.038	0.010	0.050	0.063	0.035	0.065	...
3	0	0.045	0.805	0	0	0.047	0.003	0.008	0	0.010	...
4	0.135	0.070	0	0.142	0.025	0.080	0.035	0.032	0.040	0.003	...
5	0	0.038	0	0.025	0.416	0.080	0.018	0.015	0.018	0.003	...
6	0.008	0.010	0.047	0.080	0.080	0.632	0.150	0.065	0.025	0.035	...
7	0.052	0.050	0.003	0.035	0.018	0.150	0.174	0.142	0.198	0.140	...
8	0.052	0.063	0.008	0.032	0.015	0.065	0.142	0.111	0.160	0.130	...
9	0.015	0.035	0	0.040	0.018	0.025	0.198	0.160	0.305	0.087	...
10	0.070	0.065	0.010	0.003	0.003	0.035	0.140	0.130	0.087	0.168	...
...	...	...	...	...	...	...	...	...	...	...	...

with each of the  $\text{Var}(r_{gg})$  estimated from Equation (4b), with estimated paternal contributions substituted for the parametric values.

This average value of PPI is larger than that (0.196) reported in Robledo-Arnuncio *et al* (2004), because in that study, the correlation of paternity was calculated without the selfs. To separate the impact of the selfs, one would remove them from the sample prior to formal analysis. Our objective here is to illustrate the translation of paternity information into inference on PPI, and for comparative purposes, we have chosen not to separate the selfs. The derivative estimates of  $N_{ep}$  for individual mothers range from  $1.11 < N_{ep} < 21.11$ , and the average (per mother) is  $N_{ep} = (1/\bar{r}_0) = 3.15 \pm 0.29$ , where the value of the root mean-squared error of ( $N_{ep}$ ) is provided by combining Equation (15) and the square of Equation (14) in Nielsen *et al*, 2003). We remind the reader that the observed PPI estimates are unbiased, but the derivative  $N_{ep}$ -estimates are not, particularly the larger estimates, because the array of male parentage is relatively even (minimal PPI) and the sample sizes ( $n_g = 20$ ) are the same order of magnitude as the estimated  $N_{ep}$ -values (Nielsen *et al*, 2003).

Two of the 36 males contribute no progeny at all to the 34 sampled mothers, and the total numbers of offspring from the others (summed over mothers) ranges from 1 to 74 (Robledo-Arnuncio and Gil, 2005); the observed  $\lambda$ 's are far from equal. The PPI for the entire sample of  $N = 680$  offspring, drawn from the 34 mothers, is  $R_0 = 0.0425 \pm 0.0021$  (computed as  $\sqrt{\text{Eq}[4b]}$ , with  $N = 680$  and observed global  $\lambda$ 's), larger than the value expected under the null hypothesis of homogeneous male contributions,  $R_{pan} = 1/36 = 0.0278$  ( $\pm 0.0004$ ). Clearly, males are contributing unequally, overall. On the other hand,  $R_0$  is only about 13% of the average for a single mother,  $\bar{r}_0 = 0.317$ , indicating that the total collection of mothers is drawing from a wider array of fathers than is any one female and that the pollen pool structure among those females is substantial. It is obvious from Table 2 that there is at least some overlap ( $\bar{r}_{gh} = 0.035$ ), which is why  $\bar{r}_0 < G \cdot R_0$ . We estimate the global effective number of pollen donors to be  $N_{ep} = [1/R_0] = 23.5 \pm 0.42$ , or about (2/3) of the adult stem count.

There is strong genetic isolation of the Coca population from other populations of Scots pine, but even local pollination departs substantially from panmixia. A simulation analysis has shown that a leptokurtic pattern of pollen dispersal, along with unequal pollen contributions of different males (quite noticeable here), and (possibly) phenological nonoverlap among certain pairs, might account for the degree of local pollen structure observed (Robledo-Arnuncio *et al*, 2004). The essential point is that the PPI for pairs of offspring from the average female is  $\bar{r}_0 = 0.317$ , an order of magnitude increase over panmictic expectation (1/36). The entire collection of offspring from the 34 mothers does yield a paternally more diverse sample ( $R_0 = 0.0425$ ) than any one female, but still indicates less than panmictic mating. Robledo-Arnuncio *et al* (2004) and Robledo-Arnuncio and Gil (2005) investigate the possible sources of the diversity in  $r_{gg}$ -values among mothers. Readers interested in additional discussion of the conservation and evolutionary issues of this particular example are referred to those papers. Our purpose here is primarily illustration of the methods themselves.

## Discussion

### Outlier pollen

For convenience of the Coca illustration, we have chosen to suppress the 35 outlier offspring that are not compatible with any of the 36 resident males. We note that SSR markers are subject to mutation, and the outliers could (in principle) include a few germline mutations in the pollen pool from the 36 resident males. Provan *et al* (1999) have reported an upper bound of  $5.5 \times 10^{-5}$  for the mutation rate of chloroplast microsatellites in *Pinus torreyana*, however, and in the absence of other information, the fraction of male gametes containing a mutation in the Coca sample would be about two orders of magnitude lower than our actual outlier rate (data not shown). The most parsimonious explanation for the outliers is that they represent pollen immigration from outside Coca.

Treating those outliers as immigrants, the 35 offspring were assignable to at least 30 separate male genotypes, none represented within the Coca population. To include them, we simply increase the number of potential fathers from  $K = 36$  to  $\bar{K} = 66$  and recompute the PPI-values for the total sample (including the incoming pollen) and repeat all the analyses described above. If  $m$  is the fraction of incoming pollen, that is tantamount to setting  $\tilde{\lambda}_{gk} = (1 - m)\lambda_{gk}$  for the  $k$ th local pollen donor and adding the incoming male contributors. Here, our best estimate of  $m = 35/813 = 0.043$ , and we obtain  $\bar{r}_0 = 0.317$  and  $R_0 = 0.0425$ , without allowing for outlier pollen and  $\bar{r}_0 = 0.294$  and  $R_0 = 0.0393$ , if we do allow for that outlier pollen. By including the rare outside contributors, almost doubling the total number of contributing males, we have increased the effective number of pollen donors by only 8%, either for a single female or for the population as a whole. If pollen can arrive from considerable distances, then there are clearly more than 30 potential outside donors, but their collective contribution to PPI is negligible (and not detectable) with 813 progeny. The genetic pollen structure for Coca is dominated by the internal pollen donors, rather than by the rare outside donors. For the present study, inference is affected little by the difference, since the outlier rate (treated as immigration) is minimal, but this is an atypically low rate of pollen immigration for a forest tree pollination study, and the usual situation involves a substantial fraction of offspring that cannot be assigned to one of the identifiable males within the stand (Smouse and Sork, 2004).

### Ambiguous paternity analysis

A great deal of work has been invested in improving the parental resolution of the genetic assay battery, on the premise that by reducing the paternal exclusion probability sufficiently, we ensure that any given offspring could have only a single father. For the SSR markers used in our Scots pine example, the father's genotype is exactly identifiable, although there may (in principle) be more than one male that could have contributed any particular haplotype. In many studies, especially when using exclusively Mendelian nuclear markers, 'paternal designation' is inherently noncategorical. An example should suffice. Imagine four nuclear microsatellite loci, each with nine alleles ( $A_1, A_2, \dots, A_9; \dots; D_1, D_2, \dots, D_9$ ).

The paternal contribution to a particular offspring, after subtraction of the maternal genotype, is judged to be (A<sub>1</sub>-B<sub>5</sub>-C<sub>2</sub>-D<sub>8</sub>). Any male genotype that could generate such a gamete is possible, 9<sup>4</sup> = 6561 candidate genotypes, although the posterior likelihoods will vary substantially among them. That is the typical result, and there is no escape.

Under such circumstances, there are two options: (a) assign the offspring to that male genotype with the greatest likelihood of having produced the offspring, provided that the likelihood is sufficiently higher than those of all other genotypes, and refrain otherwise (encapsulated in CERVUS 2.0 software, Marshall *et al*, 1998), or (b) assign paternity proportionally to all of the credible candidates, in proportion to their relative likelihood values (encapsulated in PATQUEST software, Smouse *et al*, 1999). Strategy (a) has the advantage that while we have to compute the likelihoods for the full array, we can choose the most likely candidate and assign it to that father. That leads to biased estimates of the  $x_{gk}$ 's (the number of offspring of the  $g$ th mother sired by the  $k$ th father), which will affect the PPI estimates (see Devlin and Ellstrand, 1990). If the threshold value is sufficiently stringent, the bias is tolerable for any given offspring, but the problem is that virtually every one of the assayed offspring yields such an outcome, and the cumulative impact can be nontrivial. Strategy (b) is less biased, but requires substantially larger amounts of book-keeping, since all 6561 male genotypes acquire noninteger  $x_{gk}$ -values (Smouse and Meagher, 1994). It is possible to convert these noninteger  $x_{gk}$ -values into estimates of PPI (cf, Burczyk *et al*, 2002), but available algorithms will probably need some further refinement.

With small numbers of credible fathers, the problem is manageable (at Coca, the number is never more than 36), but if immigration is considerable, then limiting the list of candidates to local males is problematic, and choosing between an observable local male and a set of cryptic outsiders (unsampled, but producing compatible immigrant gametotypes) becomes problematic. A high degree of genetic resolution is valuable, but it is no substitute for being able to locate, sample, and genotype the relevant males. We can adjust for cryptic immigration in a general way (Harju and Nikkanen, 1996), and for cases where immigration ( $m$ ) is substantial and where determination of individual donors for the incoming pollen is beyond reach, we can follow Burczyk *et al* (1996, but in our own notation) in providing a lower bound estimate of PPI, namely

$$P\hat{P}I \approx (1 - \hat{m})^2 PPI \quad (14)$$

on the assumption that each of the incoming male gametes will be from a different male, and that there will be virtually no duplication of male parentage in the immigrant subset of the pollen. That assumption is less attractive as  $\hat{m}$  increases, because high-likelihood external donors have an elevated probability of contributing multiple paternal gametes.

#### Indirect estimation

In those situations where accurate paternal inference is not feasible, either because the genetic battery does not provide sufficient resolution or because the spatial scale of pollination is simply too large to allow genetic assay of

the majority of fathers, discarding the portion of offspring for which we cannot assign the father, will yield biased estimates of PPI. Most of the unassigned offspring have probably been sired by males external to the study area, while the assigned fathers will always lie within the boundaries of the sampling plot. If the probability of an (unknown) distant father siring more than one offspring (ie full-sibs) in a focal female is smaller than that of an identified father from within the site, as we reasonably expect, discarding unassigned offspring would likely yield an upwardly biased estimate of PPI.

In these situations, we can still choose between three indirect estimation methods, any one of which will provide at least a rough estimate of PPI. The first of these is TWOGENER analysis (Smouse *et al*, 2001; Austerlitz and Smouse, 2001a), the basic idea of which is that different mothers, scattered across the landscape, sample from different arrays of fathers. By comparing the diploid nuclear genotypes of offspring with those of their mothers, we can extract the haploid paternal genotypes, and can partition the male gametic variation into its within-mother and among-mother components, using AMOVA (Excoffier *et al*, 1992). We then estimate the intraclass correlation of male gametes,  $\hat{\Phi}_{ft}$ , the proportion of the total paternal variation accounted for by the divergence of pollen pools of maternal strata. We translate our estimate of  $\hat{\Phi}_{ft}$  into a statement about  $\bar{r}_0$  and  $\bar{r}_{gh}$  (Austerlitz and Smouse, 2001a), and noting that the observable here is a haploid cpDNA genome of the male parent, we can write

$$E(\hat{\Phi}_{ft}) = \frac{\bar{r}_0 - \bar{r}_{gh}}{1 - \bar{r}_{gh}} \quad (15)$$

Lacking paternal designation, we have no directly observable measures of  $\bar{r}_0$  and  $\bar{r}_{gh}$ , but we can extract indirect estimates from the  $G \times G$  matrix of pairwise  $\phi_{gh}$  estimates produced by AMOVA (Austerlitz and Smouse, 2002; Austerlitz *et al*, 2004). Using the TWOGENER approach for Coca, we obtain an intraclass correlation of male cpSSR haplotypes within maternal sibships of  $\hat{\Phi}_{ft} = 0.280$ . Using the iterative estimation techniques in Austerlitz and Smouse (2002), we obtain an adjusted value of  $\bar{r}_0 = 0.304 \pm 0.060$ , within estimation error of the estimate of 0.317 obtained from the categorical assignment of paternity.

The second method, denoted MLTR, is provided by Ritland (1989, 2002), who derives maximum likelihood estimates of the probability of 'correlated paternity' ( $r_p$ ), the probability that the paternal allele of one offspring is identical to either of the two alleles contributed by the father of a maternal half-sib. MLTR software (ver 2.4) yields maximum-likelihood estimates of individual (per mother) estimates for  $r_p$ , with  $r_p \sim 2\Phi_{ft}$  for the diploid Mendelian case and  $r_p \sim \Phi_{ft}$  for the haploid (cpDNA) case. This method uses the assayed genetic profiles of the offspring, and assumes that noninbreeding pollination is a panmictic draw from the population of available males. The estimate for Coca is  $r_p = 0.310 \pm 0.046$ , once we set MLTR software to include selfs in the calculation of  $r_p$  by setting the selfing rate to zero, a result that also lies very close to the PPI value of 0.317 obtained from the categorical paternity assay. The major difference between MLTR and TWOGENER is that MLTR also estimates the selfing rate and adult inbreeding, whereas TWOGENER



can use an externally provided estimate of adult inbreeding (Austerlitz and Smouse, 2001b), but (so far) has no provision for selfing.

The third of the indirect methods is based on the analysis of 'relationship coefficients' between the male gametes inferred for pairs of offspring, say  $F_{ij}$  for the  $i$ th and  $j$ th offspring. We can define  $F_{ij}$  for offspring pairs, both of which are within the same maternal sibship, and for those from different maternal sibships (see Loiselle *et al*, 1995). Hardy *et al* (2004) show that the average within-sibship relationship,  $F_S \sim \Phi_{fr}$ , although with a mean-squared error that is slightly smaller. For the Coca illustration, the estimated average  $F_S$  for a single female is  $0.275 \pm 0.016$ . Numerical values for the three indirect methods are usually within measurement error of each other (Hardy *et al*, 2004; Smouse and Sork, 2004). These methods extend our analytical capability to less information-rich real world situations, providing rough estimates of PPI, and are most useful where paternal designation for most offspring is beyond genetic or sampling reach.

#### Pollen flow

There has also been considerable attention devoted to indirect estimation of pollen flow, the shape of the pollen dispersal curve, and the average distance of pollination (review in Smouse and Sork, 2004). It is clear that the same paternal identification procedures used for an estimation of PPI can also be deployed for the estimation of pollen dispersal parameters, given appropriate measures of interparental distance. Whether the resolution is better than that already offered by paternity analysis, encapsulated in Neighborhood models (Burczyk *et al*, 2002), or by TWOGENER analysis (Austerlitz *et al*, 2004), remains to be seen. Such questions represent still unmet challenges for the future.

#### Conclusion

In summary, we have developed estimates of the PPI for the case where the vast majority of male parentage can be specified unambiguously. We show that the obvious estimate can be severely biased under typical sampling conditions, but develop an alternative that is unbiased and that has smaller MSE for the frequent cases in which paternal contributions are not severely skewed (PPI smaller than ca. 0.4). The pattern of paternal identity can be captured in a  $G \times G$  matrix of paternal matching fractions,  $\mathbf{R} = \{r_{gh}\}$ , which can be translated into statements about the average within-mother value ( $\bar{r}_0$ ) and a total collection value ( $R_0$ ), as well as a statement of the rate of intermother paternal matching ( $\bar{r}_{gh}$ ), providing a detailed picture of the 'pollen pool' of the population.

We then illustrate the methods with SSR data from a 36-tree isolated population of *P. sylvestris* from the Meseta region of Spain, where categorical paternity assignment was available. We find that  $R_0 = 0.0425$ , indicating uneven male reproductive contributions (ranging from 0 to 74 offspring each). Different but overlapping ( $\bar{r}_{gh} = 0.035$ ) arrays of males pollinate different females, and for the average female,  $\bar{r}_0 = 0.317$ , indicating substantial 'pollen structure' for the population. We also relate the direct measures of PPI to those available from indirect methods that are theoretically compatible.

Where we can mount both sets of methods, we discover that the results are comparable.

#### Acknowledgements

We thank F Austerlitz, SC González-Martínez, VL Sork, AJ Irwin, and a pair of anonymous reviewers for voluminous helpful critique on the manuscript. The work was primarily accomplished while the senior author was a sabbatical visitor at the Centro de Investigación Forestal (CIFOR) – Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA) in Madrid. The work reported here was funded by the Secretaría de Estado de Educación y Universidades (España), and PES thanks both the Secretaría and INIA for their courtesy and cooperation. PES was also supported by USDA and the NJ Agricultural Experiment Station (USDA-NJAES-17111) and by the National Science Foundation (NSF-BSR-0211430 and NSF-BSR-0089238). JJR-A was supported by DGCN-ETSIM project *Conservación y Mejora de Recursos Genéticos de Coníferas* (2000–2003), NSF-BSR-0211430, and an MEC/Fulbright postdoctoral fellowship.

#### References

- Austerlitz F, Dick CW, Dutech C, Klein E, Oddou-Muratorio S, Smouse PE *et al* (2004). Using genetic markers for the estimation of the pollen dispersal curve. *Mol Ecol* **13**: 937–954.
- Austerlitz F, Smouse PE (2001a). Two-generation analysis of pollen flow across a landscape. II. Relation between  $\Phi_{fr}$ , pollen dispersal, and inter-female distance. *Genetics* **157**: 851–857.
- Austerlitz F, Smouse PE (2001b). Two-generation analysis of pollen flow across a landscape. III. Impact of within-population structure. *Genet Res* **78**: 271–280.
- Austerlitz F, Smouse PE (2002). Two-generation analysis of pollen flow across a landscape. IV. Estimating the dispersal parameter. *Genetics* **161**: 355–363.
- Burczyk J, Adams WT, Moran GF, Griffin AR (2002). Complex patterns of mating revealed in a *Eucalyptus regnans* seed orchard using allozyme markers and the neighborhood model. *Mol Ecol* **11**: 2379–2391.
- Burczyk J, Adams WT, Shimizu J (1996). Mating patterns and pollen dispersal in a natural knobcone pine (*Pinus attenuata* Lemmon.) stand. *Heredity* **77**: 251–260.
- Chakraborty R, Meagher T, Smouse PE (1988). Parentage analysis with genetic markers in natural populations. I. The expected proportion of offspring with unambiguous paternity. *Genetics* **118**: 527–536.
- Cockerham CC, Weir BS (1993). Estimation of gene flow from F-statistics. *Evolution* **47**: 855–863.
- Crawford TJ (1984). The estimation of neighborhood parameters for plant populations. *Heredity* **52**: 273–283.
- Devlin B, Ellstrand NC (1990). The development and application of a refined method for estimating gene flow from angiosperm paternity analysis. *Evolution* **44**: 248–259.
- Excoffier L, Smouse PE, Quattro JM (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**: 479–491.
- Franco-Múgica F, García-Antón M, Maldonado-Ruiz J, Morla-Juriasti C, Sainz-Ollero H (2001). The Holocene history of *Pinus* forests in the Spanish Northern Meseta. *Holocene* **11**: 343–358.
- Hardy OJ, González-Martínez SC, Colas B, Fréville H, Mignot A, Olivieri I (2004). Fine-scale genetic structure and gene dispersal in *Centaurea corymbosa* (Asteraceae) II. Correlated paternity within and among sibships. *Genetics* **168**: 1601–1614.

- Harju AM, Nikkanen T (1996). Reproductive success of orchard and nonorchard pollen during different stages of pollen shedding in a Scots pine seed orchard. *Can J For Res* **26**: 1096–1102.
- Loiselle BA, Sork VL, Nason J, Graham C (1995). Spatial genetic structure of a tropical understory shrub, *Psychotria officinalis* (Rubiaceae). *Am J Bot* **82**: 1420–1425.
- Malécot G (1969). *The Mathematics of Heredity*. Freeman: San Francisco.
- Marshall TC, Slate J, Kruuk LEB, Pemberton JM (1998). Statistical confidence for likelihood-based paternity inference in natural populations. *Mol Ecol* **7**: 639–655.
- Morton NE (1977). Isolation by distance in human populations. *Ann Hum Genet* **40**: 361–365.
- Nei M (1987). *Molecular Evolutionary Genetics*. Columbia University Press: New York.
- Nei M, Roychoudhury AK (1974). Sampling variances of heterozygosity and genetic distance. *Genetics* **76**: 379–390.
- Nielsen R, Tarpay DR, Reeve K (2003). Estimating effective paternity number in social insects and the effective number of alleles in a population. *Mol Ecol* **12**: 3157–3164.
- Pamilo P (1993). Polyandry and allele frequency differences between the sexes in the ant *Formica aquilonia*. *Heredity* **70**: 472–480.
- Provan J, Soranzo N, Wilson NJ, Goldstein DB, Powell W (1999). A low mutation rate for chloroplast microsatellites. *Genetics* **153**: 943–947.
- Ritland K (1989). Correlated-matings in the partial selfer *Mimulus guttatus*. *Evolution* **43**: 848–859.
- Ritland K (2002). Extensions of models for the estimation of mating systems using  $n$  independent loci. *Heredity* **88**: 221–228.
- Robledo-Arnuncio JJ, Alía R, Gil L (2004). Increased selfing and correlated paternity in a small population of a predominantly outcrossing conifer, *Pinus sylvestris*. *Mol Ecol* **13**: 2567–2577.
- Robledo-Arnuncio JJ, Gil L (2005). Patterns of pollen dispersal in a small population of *Pinus sylvestris* L. revealed by total-exclusion paternity analysis. *Heredity* **94**: 13–22.
- Rousset F (1997). Genetic differentiation and estimation of gene flow from  $F$ -statistics under isolation by distance. *Genetics* **145**: 1219–1228.
- Slatkin M (1993). Isolation by distance in equilibrium and nonequilibrium populations. *Evolution* **47**: 264–279.
- Smouse PE, Dyer RJ, Westfall RD, Sork VL (2001). Two-generation analysis of pollen flow across a landscape. I. Male gamete heterogeneity among females. *Evolution* **55**: 260–271.
- Smouse PE, Meagher TR (1994). Genetic analysis of male reproductive contributions in *Chamaelirium luteum* (L.) Gray (Liliaceae). *Genetics* **136**: 313–322.
- Smouse PE, Meagher TR, Kobak CJ (1999). Parentage analysis in *Chamaelirium luteum* (L.): why do some males have disproportionate reproductive contributions? *J Evolut Biol* **12**: 1056–1068.
- Smouse PE, Sork VL (2004). Measuring pollen flow in forest trees: a comparison of alternative approaches. *Forest Ecol Manag* **197**: 21–38.
- Streiff R, Ducousso A, Lexer C, Steinkellner H, Gloessl J, Kremer A (1999). Pollen dispersal inferred from paternity analysis in a mixed oak stand of *Quercus robur* L. and *Q. petraea* (Matt.) Liebl. *Mol Ecol* **8**: 831–841.
- Wright S (1943). Isolation by distance. *Genetics* **28**: 114–138.
- Wright S (1946). Isolation by distance under diverse systems of mating. *Genetics* **31**: 39–59.