

# A method for estimating outcrossing rates in natural populations of plants

D. Charlesworth

Department of Ecology and Evolution,  
University of Chicago, 915 E. 57th St., Chicago,  
Illinois 60637, U.S.A.

**A method is described for estimating outcrossing rates in natural populations of self-compatible plants, by comparing the viability of zygotes produced by natural open pollination with that of the products of artificial self-fertilisation and outcrossing.**

Outcrossing rates are of interest to population biologists for several reasons, and there have been attempts to estimate these rates by using genetic markers to score the percentage of progeny produced under natural pollination conditions (*e.g.*, Brown and Allard, 1970; Ritland, 1986). These methods have generated a great deal of useful data, but they suffer from two difficulties. The first is that one cannot use these methods when no genetic marker is available in the population. This is quite unlikely to be the case for electrophoretic markers in a partially or highly outbreeding population, but it sometimes happens that the species is difficult to electrophorese, so that no good systems can be obtained. Even when there is no problem of this kind, the method of scoring genetic markers in progeny uses viable seeds or progeny, and so gives an estimate of the fraction of plants that are products of self-fertilisation at the stage when the progeny are scored. In some situations, this is not what is really required. For example, one sometimes wants to know the selfing rate at the time of fertilisation. If there is any difference between the progeny of selfing and of outcrossing, before the time when progeny can be scored, for example if there is differential viability of seeds of these two origins, then the outcrossing rate at fertilisation will differ from that estimated using viable seeds. For these reasons, it would be desirable to have another method.

A method that suggests itself is to use measurements of some character that can be scored on the seeds very soon after fertilisation, such as the viability of the fertilisation products. If the value of the measured character differs between selfed and

outcrossed seeds, the frequency of self pollination could be estimated by comparison of the values of seeds from naturally pollinated flowers, with those of hand self-fertilisation and hand outcrossing. Writing  $p_x$  for the value from crossing,  $p_s$  for the value from selfing, and  $p_o$  for the open-pollinated value, the selfing rate,  $S$ , can be estimated by

$$S = \frac{p_x - p_o}{p_x - p_s} \quad (1)$$

(Charlesworth and Charlesworth, 1987).

The variance of the estimate of  $S$  can be obtained by using the "delta method" described by Bulmer (1980, pp. 82-83). The general formula for a function  $S$  of three variables ( $p_x$ ,  $p_s$  and  $p_o$ ) is as follows

$$\begin{aligned} V(S) = & V(p_x) \left( \frac{\partial S}{\partial p_x} \right)^2 + V(p_o) \left( \frac{\partial S}{\partial p_o} \right)^2 \\ & + V(p_s) \left( \frac{\partial S}{\partial p_s} \right)^2 \\ & + 2 \text{Cov}(p_x p_o) \frac{\partial S}{\partial p_x} \cdot \frac{\partial S}{\partial p_o} \\ & + 2 \text{Cov}(p_x p_s) \frac{\partial S}{\partial p_x} \cdot \frac{\partial S}{\partial p_s} \\ & + 2 \text{Cov}(p_o p_s) \frac{\partial S}{\partial p_o} \cdot \frac{\partial S}{\partial p_s} \end{aligned} \quad (2)$$

where  $V$  stands for variance and  $\text{Cov}$  for covariance. Given the formula above for  $S$ , one gets the following expressions for the derivatives needed in equation (1).

$$\frac{\partial S}{\partial p_x} = \frac{p_o - p_s}{(p_x - p_s)^2} \quad (3a)$$

$$\frac{\partial S}{\partial p_o} = -\frac{1}{p_x - p_s} \quad (3b)$$

and

$$\frac{\partial S}{\partial p_s} = \frac{S}{p_x - p_s} \quad (3c)$$

which finally yields the variance formula

$$V(S) = \frac{1}{(p_x - p_s)^2} \left\{ V(p_x) \left( \frac{p_o - p_s}{p_x - p_s} \right)^2 + V(p_o) + V(p_s) \cdot S^2 - 2 \text{Cov}(p_x p_o) \cdot \frac{p_o - p_s}{p_x - p_s} - 2 \text{Cov}(p_x p_s) \cdot S \cdot \frac{p_o - p_s}{p_x - p_s} - 2 \text{Cov}(p_o p_s) \cdot S \right\}$$

Naturally, it is important that the different  $p$  values should be measured as similarly as possible, so that there are as few differences as possible between them, apart from those due to the different frequencies of self and other pollen. For example, each of the three treatments should be done on each plant in the study, using the same number of flowers on each plant, if possible. All the flowers should be treated as far as possible in the same way, with any bagging that is needed to ensure that there is no contamination in the hand pollination treatments being done for the same periods, except for the time when pollination is happening. Clearly, the estimates made in this way assume that seeds are produced in only two ways, by selfing or outcrossing, but in reality it is likely that crosses between plants may sometimes involve some degree of inbreeding and these inbred seeds will generally have character values between those of selfed seeds and those of seeds from crosses between non-relatives (Layton and Goddard, 1983; Griffin and Lindgren, 1985). There are certainly cases in which crosses between nearby plants yield progeny with lower values of some fitness components than crosses between plants that are growing further apart (Levin, 1984; but see Fenster, 1988). If this occurs, and if the crosses used to get the value of  $p_x$  are exclusively done between non-related individuals (rather than by using a sample of the plants in the natural neighbourhoods of the recipients),  $S$  would be overestimated. The degree of error is unlikely to be very large. With the

maximum likelihood methods based on electrophoretic markers, there is a similar problem that the selfing rate is overestimated, due to the tendency for relatives to have similar genotypes. However with that method one might have enough information to estimate the extent of matings between relatives as a third parameter in the estimation (see Ritland, 1986).

The method can be applied to gymnosperm species, using the fraction of filled seeds as a measure of the viability of seeds that were fertilised. It could also be used in angiosperm species, using the fraction of non-aborted seeds out of all fertilised seeds, in species where this differs between selfing and crossing. It might also be possible to use other kinds of information. Number of viable seeds per flower is one possibility, but if the fraction of ovules fertilised is lower for selfing than crossing one would get an estimate that confounds the fraction of self pollen in the pollen load and the fraction of zygotes produced by selfing.

Another possibility is to use seed size, which often differs between selfing and crossing. However, there may be difficulties with this, because the numbers of seeds in fruits may often be negatively correlated with their size (reviewed in Mazer *et al.*, 1986). If selfing tends to give few seeds per fruit, this could lead to the size of seeds from selfing being overestimated compared with their value when matched for seed number with crossed fruits. One could correct for this, using the regression of seed size on seed number per fruit. If one had this type of seed size data, it could be used for estimating this regression, and the correction could be made provided that the regression was similar for selfed and outcrossed seeds. Without correction, however, there would be a bias in the estimate. This would most likely result in too high an estimate of the selfing rate, because the denominator in the estimation equation would be too small due to the overestimation of the size of selfed seeds. It would be possible for the opposite effect to occur, leading to an underestimate of  $S$ , if the main effect of the size/number relationship were that the size of outcrossed seeds in the hand outcross treatment were much smaller than in the open pollinated treatment, but this seems less likely.

Clearly, this method is restricted to populations in which inbreeding depression is detectable in the early stages of zygote development. It is not at present clear how often this would be the case, but it seems likely that early acting inbreeding depression may not be uncommon, and may sometimes be hard to distinguish from self-incompatibility (see Seavey and Bawa, 1986).

## REFERENCES

- BROWN, A. H. D. AND ALLARD, R. W. 1970. Estimation of the mating system in open-pollinated maize populations using isozyme polymorphisms. *Genetics*, 66, 133-145.
- BULMER, M. G. 1985. *The Mathematical Theory of Quantitative Genetics*. Clarendon Press, Oxford.
- CHARLESWORTH, D. AND CHARLESWORTH, B., 1987. Inbreeding depression and its evolutionary consequences. *Ann. Rev. Ecol. Syst.*, 18, 237-268.
- FENSTER, C. J. 1988. Gene Flow and Population Differentiation in *Chamaecrista fasciculata*. Ph.D. Dissertation, University of Chicago, Chicago, IL., U.S.A.
- GRIFFIN, A. R. AND LINDGREN, D. 1985. Effect of inbreeding on production of filled seed in *Pinus radiata*—experimental results and a model of gene action. *Theoret. Appl. Genet.*, 71, 334-343.
- LAYTON, P. A. AND GODDARD, R. E. 1983. Low level inbreeding effects on germination, survival and early height growth of slash pine. *Proc. Southern For. Tree Improv. Conf.*, 17, 106-115.
- LEVIN, D. A. 1984. Inbreeding depression and proximity-dependent crossing success in *Phlox drummondii*. *Evolution*, 39, 116-127.
- MAZER, S. J., SNOW, A. A. AND STANTON, M. L. 1986. Fertilization dynamics and parental effects upon fruit development in *Raphanus raphanistrum*: consequences for seed size variation. *Amer. J. Bot.*, 73, 500-501.
- RITLAND, K. 1986. Joint maximum likelihood estimation of genetic and mating structure using open-pollinated progenies. *Biometrics*, 43, 25-43.
- SEAVEY, S. R. AND BAWA, K. S. 1986. Delayed self-incompatibility in angiosperms. *Bot. Rev.*, 52, 195-219.