PROGRAM NOTE

POLDISP: a software package for indirect estimation of contemporary pollen dispersal

JUAN J. ROBLEDO-ARNUNCIO,* FRÉDÉRIC AUSTERLITZ† and PETER E. SMOUSE‡

*Laboratoire Génétique et Environnement, Université de Montpellier II, Institut des Sciences de l'Evolution, 34095 Montpellier cedex 05, France, †Laboratoire de Ecologie, Systématique et Evolution, UMR CNRS 8079, Université Paris-Sud, F-91405 Orsay cedex, France, ‡Department of Ecology, Evolution and Natural Resources, School of Environmental and Biological Sciences, Rutgers University, New Brunswick, NJ 08901-8551, USA

Abstract

POLDISP 1.0 is a free software package to estimate the distribution of pollen dispersal distances from mother–offspring diploid genotypic data. It requires the spatial coordinates and genotypes of a sample of seed plants and their respective maternal progenies, providing estimates of the average, variance and kurtosis of the pollen dispersal curve. POLDISP also estimates the effective reproductive density of pollen donors and the correlation of paternity within and among maternal sibships. POLDISP is useful for characterizing the spatial scale of pollen dispersal, for assessing the variation in male fertility and for investigating biological factors affecting correlated paternity in plants.

Keywords: correlated paternity, dispersal kernel, effective density, genetic markers, KINDIST, TWOGENER

Unlike direct methods based on genetic paternity assignment, indirect methods of contemporary pollen dispersal estimation draw inference on male gamete movement solely from pollen-pool genetic structure measures, obtained from mother–offspring genotypic data. They do not require the costly mapping and genotyping of all potential pollen donors (Austerlitz & Smouse 2001, 2002; Smouse *et al.* 2001). By allowing an extension of the spatial scale of sampling at low cost, relative to paternity-based approaches, indirect methods represent a valuable tool for investigating pollen flow patterns within natural plant populations. Given sufficient genetic resolution and enough replication of offspring and mothers, indirect methods yield accurate estimates of the average distance of dispersal and other moments of the pollen dispersal distribution under a range of flowering conditions, as well as approximate estimates of the effective reproductive density of the pollen donors (Austerlitz *et al.* 2004; Robledo-Arnuncio *et al.* 2006). Here we present POLDISP, a software package that implements the two available indirect methods for contemporary pollen flow estimation, KINDIST (Robledo-Arnuncio *et al.*

Correspondence: Dr. Juan J. Robledo-Arnuncio,
Fax: +33 467143622; E-mail: jjrobledo@gmail.com

2006) and TWOGENER (Austerlitz & Smouse 2001, 2002; Smouse *et al.* 2001). The POLDISP package permits complementary use of the two methods to obtain estimates of both the pollen dispersal distribution and effective pollen donor density from mother–offspring diploid codominant genotypic data.

POLDISP requires the following input information: (i) two spatial coordinates for each mother plant; (ii) genotypes for each of n_L codominant diploid marker loci for each mother plant; (iii) genotypes for the same n_L loci for each seed collected from mother plants before dispersal (seeds must be of known maternal origin); and (iv) an among-mother threshold distance (d_u), used to define 'unrelated' pollen pools in the sample (see below). The sampled mothers should cover as many pairwise-distance classes as possible, from neighbouring to long-distance pairs. Ideally, a total of at least $N = 800$ seeds should be analysed in order to obtain accurate estimates of the dispersal parameters. Accuracy will be improved more by increasing the number of mothers than by sampling more offspring per mother, but a minimum of 10–20 offspring/mother should be observed.

KINDIST, the first procedure implemented in POLDISP package, estimates the distribution of pollen dispersal distances. The approach is based on the expected decay

with spatial distance of a normalize measure of correlated paternity between female pairs, $\Psi(z)$, defined as the ratio of the correlation of paternity between maternal sibship pairs a distance z apart to the average correlation of paternity within single maternal sibships (Robledo-Arnuncio *et al.* 2006). Assuming a given dispersal function (e.g. an exponential power) with parameter set θ, KINDIST yields θ estimates by means of nonlinear least-square regression of observed $\Psi(z)$ on expected $\Psi(z,\theta)$ for pairs of female parents. The expected $\Psi(z,\theta)$ values are formulated as a function of the dispersal distribution, using the same isolation-by-distance theoretical framework developed originally for the TWOGENER model (Austerlitz & Smouse 2001, 2002). In contrast to TWOGENER's Φ_{FT} statistic, however, $\Psi(z,\theta)$ is independent of the effective density of the pollen donors, which is generally unknown, so KINDIST does not require a joint estimation of this quantity, yielding more precise and accurate estimates of the dispersal parameters (θ), under a wide range of sampling and flowering scenarios (see Robledo-Arnuncio *et al.* 2006 for details on the model and its statistical properties).

KINDIST proceeds in three steps: first, it infers the male gametic contribution to each seed and estimates the global allele frequencies of the pollen pool for each locus. In case of ambiguity (i.e. both mother and offspring share the same heterozygous state), paternity is assigned fractionally to each of the two possible alleles according to their posterior likelihood value, given the pollen pool frequencies estimated from the unambiguous cases (Smouse *et al.* 2001; Irwin *et al.* 2003). Second, it computes the correlation of paternity (probability of paternal identity) within and among maternal sibships, using an estimation procedure that involves the calculation of pairwise kinship coefficients between the paternal gametic genotypes of offspring pairs, using the global pollen pool allele frequencies as a reference (for details see Hardy *et al.* 2004). Finally, the program uses the estimated correlated paternity rates to derive the observed pairwise $\Psi(z)$ values, which, along with the set of expected $\Psi(z,\theta)$ values, allows a least-square regression estimation of the dispersal parameters (θ).

The user should follow two stages through the analysis while using the KINDIST program: initially, after loading the input data file with the spatial coordinates and genotypes of the samples, the program generates an initial text output file containing the estimated pollen pool allele frequencies at each locus, the estimated correlated paternity rate within each maternal sibship, and a table with the among-sibship correlated paternity estimates and the intermaternal separation distance for all pairs of maternal sibships. This file is in tab-delimited format and can be edited on any standard worksheet program such as OpenOffice.org Calc or Microsoft Excel, allowing an inspection of the relationship between among-sibship correlated paternity and distance.

If no decrease in among-sibship correlated paternity is detected with distance, this pattern indicates an absence of significant pollen pool spatial genetic structure (Fig. 1a), and subsequent estimations using the POLDISP package become pointless, as they may suffer inflated biases. By contrast, if the correlation of paternity among sibships decreases significantly with separation distance (our empirical results suggest that a negative Spearman's rank correlation coefficient $r_s < -0.1$ will yield acceptable dispersal estimates; Fig. 1b), the user may continue with the second step of the analysis, setting the threshold distance D_u at the approximate observed value beyond which correlated paternity rates among mothers stabilize (they typically stabilize at a slightly negative value), which allows KINDIST to estimate the dispersal function parameters. Although the method is reasonably insensitive to the precise value of D_u (Robledo-Arnuncio *et al.* 2006), the user may set different values of this parameter (in consecutive runs of the program) to test the stability of the results.

The second and final output file of KINDIST contains the parameter estimates of the pollen dispersal curve. KINDIST currently fits two (Gaussian and exponential) one-parameter dispersal distributions and three (exponential-power, geometric and bivariate Student's t –2Dt-) two-parameter dispersal distributions. These distributions accommodate a wide range of kurtosis values and satisfactorily fit pollen dispersal data (Clark 1998; Austerlitz *et al.* 2004). The program provides the least-square estimates of the dispersal curve parameters, as well as estimates of its average, variance and kurtosis. The average and the variance provide an intuitive characterization of the spatial range of dispersal, while the kurtosis is a measure of the peakedness of the curve, indicating the contribution of extreme deviations to the total variance of the distribution. Higher estimates of kurtosis will correspond to increasing frequency of rare long-distance dispersal events, which represent a fundamental component of dispersal, regulating the spatial scale of ecological and evolutionary processes (Nathan 2005).

While the KINDIST procedure provides the most accurate indirect method available for estimating the pollen dispersal curve (Robledo-Arnuncio *et al.* 2006), it does not provide an estimation of the effective reproductive density of pollen donors (d_e), which may have biological interest in its own right. This effective density is the equivalent of the effective male population size for a spatially explicit continuous population (Austerlitz *et al.* 2004). It is usually lower than the census density (d) of adult trees because of the reproductive heterogeneity of pollen donors (Oddou-Muratorio *et al.* 2005). The ratio d_e/d provides a useful measure of this heterogeneity. TWOGENER provides joint estimates of d_e and the parameters of the dispersal curve, but empirical tests show that substantial data are required to yield accurate results, limiting practical application (Austerlitz *et al.* 2004).
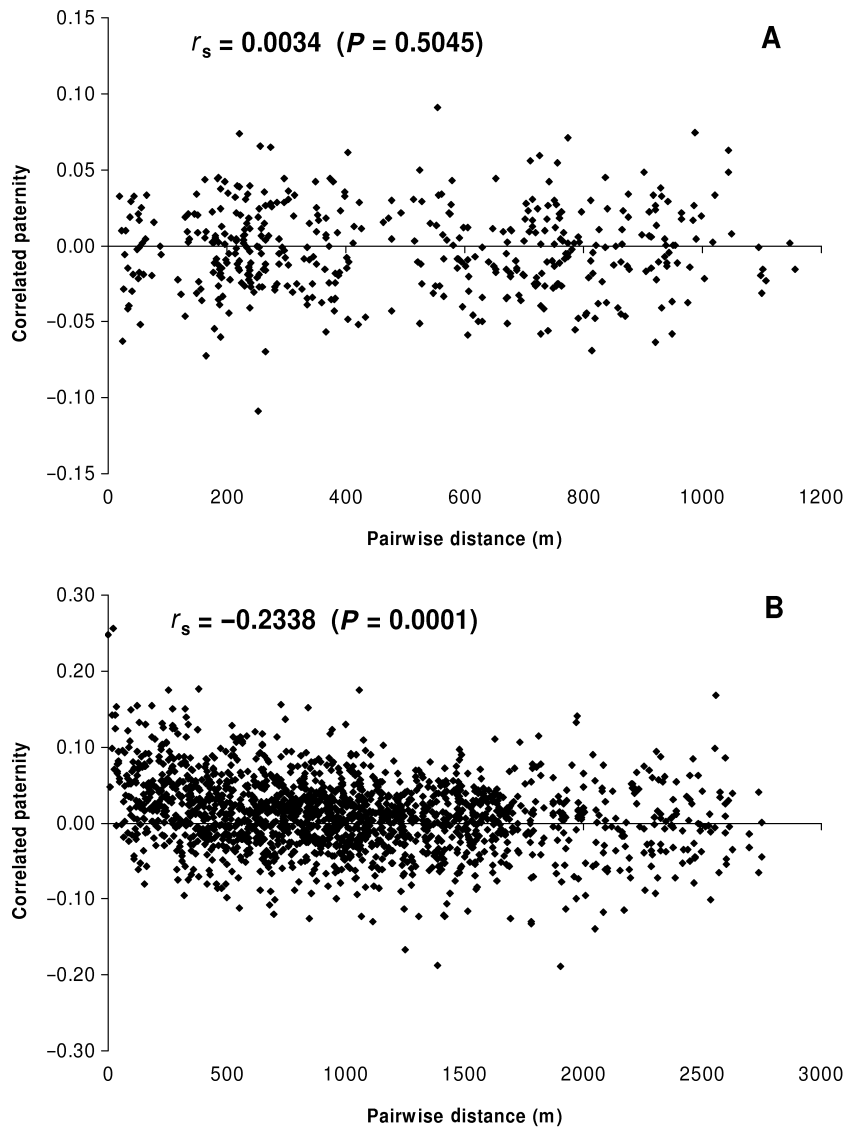
Fig. 1 Contrasting examples of the estimated relationship between correlated paternity among sibships and pairwise distance. Panel A exemplifies a lack of spatial genetic structure of the pollen pool (no decay of correlated paternity with distance), corresponding to 30 *Pinus sylvestris* L. seed-trees in a dense monospecific stand in Central Spain (re-analysed from Robledo-Arnuncio *et al.* 2004). By contrast, panel B displays a significant negative correlation between pairwise distance and correlated paternity estimates among 60 *Sorbus torminalis* L. maternal sibships, sampled from a low-density population in a mixed forest in Northern France (re-analysed from Oddou-Muratorio *et al.* 2005). Note that correlated paternity estimates stabilize at 1500 m in panel B, defining the threshold distance beyond with pollen pools are unrelated. r_s is Spearman's rank correlation coefficient and p is the significance level obtained with a Mantel test performing 1000 permutations.

The POLDISP package allows more reliable estimation of d_e, combining KINDIST and TWOGENER procedures. Using the same mother–offspring genotypic inputs, the user can employ the dispersal curve estimators from KINDIST as input for TWOGENER, and then solving for d_e. By fixing the dispersal parameters in TWOGENER, the estimation of d_e improves substantially. For completeness and comparative purposes with previous studies, POLDISP also includes the necessary programs to allow the user perform the classical TWOGENER estimation procedure (joint estimation of effective density and the dispersal function, as in Austerlitz *et al.* 2004), also available in the FaMoz package (Gerber *et al.* 2003). Unless there is very substantial replication, however, we recommend the described sequential procedure in POLDISP.

Finally, the estimates of correlated paternity (within and among sibships) generated by the KINDIST program can be used to investigate biological factors determining effective pollination (Hardy *et al.* 2004; Oddou-Muratorio *et al.* 2006). Correlated paternity, a key statistic of plant mating system analysis, is a dynamic quantity subject to environmental effects and interindividual variation (Ritland 2002). In order to characterize this variation, correlated paternity among sibships can be regressed on explanatory variables such as overlap of flowering phenology, while the correlation of paternity within individual sibships might be associated with local (conspecific or total) density, plant height or length of the female receptivity period.

The POLDISP package, documentation and test files can be freely downloaded from http://poldisp.googlepages.com. It runs on a PC under Microsoft Windows. The source codes, available from the Correspondent upon request, are written in C/C++ language and can also be compiled on Linux/Unix platforms with a GCC compiler.

References

Austerlitz F, Smouse PE (2001) Two-generation analysis of pollen flow across a landscape. II. Relation between Φ_{ft}, pollen dispersal and interfemale distance. *Genetics*, **157**, 851–857.

Austerlitz F, Smouse PE (2002) Two-generation Analysis of pollen flow across a landscape. IV. Estimating the dispersal parameter. *Genetics*, **161**, 355–363.

Austerlitz F, Dick CW, Dutech C *et al.* (2004) Using genetic markers to estimate the pollen dispersal curve. *Molecular Ecology*, **13**, 937–954.

Clark JS (1998) Why trees migrate so fast: confronting theory with dispersal biology and the paleorecord. *American Naturalist*, **152**, 204–224.

Gerber S, Chabrier P, Kremer A (2003) famoz: a software for parentage analysis using dominant, codominant and uniparentally inherited markers. *Molecular Ecology Notes*, **3**, 479–481.

Hardy OJ, González-Martínez SC, Colas B *et al.* (2004) Fine-scale genetic structure and gene dispersal in *Centaurea corymbosa* (Asteraceae). II. Correlated paternity within and among sibships. *Genetics*, **168**, 1601–1614.

Irwin A, Hamrick J, Godt M, Smouse PE (2003) A multiyear estimate of the effective pollen donor pool for *Albizia julibrissin*. *Heredity*, **90**, 187–194.

Nathan R (2005) Long-distance dispersal research: building a network of yellow brick roads. *Diversity and Distributions*, **11**, 125–130.

Oddou-Muratorio S, Klein EK, Austerlitz F (2005) Pollen flow in the wildservice tree, *Sorbus torminalis* (L.) Crantz. II. Pollen dispersal and heterogeneity in mating success inferred from parent-offspring analysis. *Molecular Ecology*, **14**, 4441–4452.

Oddou-Muratorio S, Klein EK, Demesure-Musch B, Austerlitz F (2006) Real-time patterns of pollen flow in the wild-service tree, *Sorbus torminalis* (Rosaceae). III. Mating patterns and the ecological maternal neighborhood. *American Journal of Botany*, **93**, 1650–1659.

Ritland K (2002) Extensions of models for the estimation of mating systems using *n* independent loci. *Heredity*, **88**, 221–228.

Robledo-Arnuncio JJ, Smouse PE, Gil L, Alía R (2004) Pollen movement under alternative silvicultural practices in native populations of Scots pine (*Pinus sylvestris* L.) in central Spain. *Forest Ecology and Management*, **197**, 245–255.

Robledo-Arnuncio JJ, Austerlitz F, Smouse PE (2006) A new method of estimating the pollen dispersal curve independently of effective density. *Genetics*, **173**, 1033–1045.

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.