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

Received 18 September 2006; revision accepted 8 January 2007

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. 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 \( \theta \), \textsc{kindist} yields \( \theta \) 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 \textsc{twogener} model (Austerlitz & Smouse 2001, 2002). In contrast to \textsc{twogener}'s \( \Phi_{PT} \) statistic, however, \( \Psi(z, \theta) \) is independent of the effective density of the pollen donors, which is generally unknown, so \textsc{kindist} does not require a joint estimation of this quantity, yielding more precise and accurate estimates of the dispersal parameters \( \theta \), under a wide range of sampling and flowering scenarios (see Robledo-Arnuncio et al. 2006 for details on the model and its statistical properties).

\textsc{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 \( \theta \).

The user should follow two stages through the analysis while using the \textsc{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 \textsc{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 \textsc{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 \textsc{kindist} contains the parameter estimates of the pollen dispersal curve. \textsc{kindist} currently fits two (Gaussian and exponential) one-parameter dispersal distributions and three (exponential-power, geometric and bivariate Student’s \( t \)-2D-) 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 \textsc{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 \( d_f/d \) provides a useful measure of this heterogeneity. \textsc{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).
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.
Acknowledgements

J. J. Robledo-Arnuncio was supported by a postdoctoral fellowship from the Spanish Secretaria de Estado de Educacion y Universidades, financed in part by the European Social Fund. P. E. Smouse was supported by NJAES/USDA-17111, by NSF-DEB-0211430 and NSF-DEB-0514956. We thank Sylvie Oddou-Muratorio for kindly providing an example data set for Fig. 1b. This paper is ISEM 2007-008

References


