PROGRAM NOTE

SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels

OLIVIER J. HARDY and XAVIER VEKEMANS

Laboratoire de Génétique et Ecologie végétales, Université Libre de Bruxelles, 1850 chaussée de Wavre, B-1160 Bruxelles, Belgium

Abstract

SPAGeDi version 1.0 is a software primarily designed to characterize the spatial genetic structure of mapped individuals or populations using genotype data of codominant markers. It computes various statistics describing genetic relatedness or differentiation between individuals or populations by pairwise comparisons and tests their significance by appropriate numerical resampling. SPAGeDi is useful for: (i) detecting isolation by distance within or among populations and estimating gene dispersal parameters; (ii) assessing genetic relatedness between individuals and its actual variance, a parameter of interest for marker based inferences of quantitative inheritance; (iii) assessing genetic differentiation among populations, including the case of haploids or autopolyploids.

Keywords: software, isolation by distance, relatedness coefficients, genetic differentiation, polyploidy, genetic structure

Received 11 June 2002; revision received 26 July 2002; accepted 26 July 2002

SPAGeDi (Spatial Pattern Analysis of Genetic Diversity) is a software which estimates genetic distances between populations or relatedness coefficients between individuals using data from codominant genetic markers. It is designed specifically to characterize association between genetic and spatial distances, permitting investigation of isolation by distance processes, but should also find major applications in studies focusing on sibship genetic structure, genetic differentiation in the case of autopolyploids, marker based inference of quantitative inheritance, or phylogeny reconstruction of populations. SPAGeDi is the successor of a previous program called AUTOCORG version 3.0 that was distributed to a limited extent but used in several published studies (e.g. Bonnin *et al.* 2001; Hardy & Vekemans 2001; Dutech *et al.* 2002). SPAGeDi does all what AUTOCORG 3.0 was able to do, but many additional types of analyses have been implemented, in particular analyses at the population level.

The program requires the following information for each individual: (i) one to three spatial coordinates (facultative); (ii) value of a categorical variable (facultative); and (iii) genotype at each locus of a codominant marker, where missing data are allowed. The categorical variable can be

Correspondence: Olivier Hardy. Fax: + 32 2650 91 70; E-mail: ohardy@ulb.ac.be

used to define populations or to restrict analyses within or among categories. The spatial coordinates permit to compute pairwise distances between individuals or populations (Euclidian distances), but can also be used to define populations as sets of individuals with identical coordinates. Alternatively, pairwise distances between individuals or populations can be defined in a separate matrix, so that one can also consider: (i) nonEuclidian spatial distances (e.g. distances taking into account the earth curvature or distances more related to the putative paths for gene flow); (ii) nonspatial distances (e.g. a morphological distance between individuals or populations that one wishes to correlate with genetic differentiation); (iii) categorical pairwise comparisons (e.g. to get average values of the statistics for predefined types of pairwise comparisons between individuals or populations).

The program first computes pairwise statistics describing genetic relatedness or differentiation for all pairs of individuals or populations (available statistics described below). The association between these values and pairwise spatial distances (or other types of distances) is then characterized by: (i) averaging the pairwise statistics for a set of predefined distance intervals, in a way similar to a spatial autocorrelation analysis; and (ii) regressing them on spatial (or other) distances, and on their logarithm. When spatial information is provided, the regression slopes can

potentially be used to obtain indirect estimates of gene dispersal distances parameters (e.g. neighbourhood size) in the context of an isolation by distance process (Rousset 1997, 2000; Hardy & Vekemans 1999). Upon request, the program provides the values of the statistics for all pairwise comparisons between individuals or populations in various formats, including PHYLIP format (Felsenstein 1993).

At the individual level, the following statistics can be computed for each pair of individuals: (i) two estimators of the kinship coefficient described in Loiselle *et al.* (1995) and Ritland (1996); (ii) four estimators of the relationship coefficient (a 'two-genes' relatedness coefficient closely related to the kinship coefficient) described in Queller & Goodnight (1989), Hardy & Vekemans (1999), Lynch & Ritland (1999) and Wang (2002); (iii) two estimators of the fraternity coefficient (a 'four-genes' relatedness coefficient defined in Lynch & Walsh 1998) described in Lynch & Ritland (1999) and Wang (2002); (iv) a genetic distance between individuals described in Rousset (2000); and (v) an analogue of the kinship coefficient based on allele size (that can be used for microsatellites) and described in Streiff *et al.* (1998). Inbreeding coefficients are computed as kinship coefficients between genes within individuals.

At the population level, statistics designed for pairwise comparisons include F_{ST}, *Rho* (an intraclass relationship coefficient which permits comparison among different ploidy levels), R_{ST} (an F_{ST} analogue but based on microsatellite allele sizes, Slatkin 1995), *Ds* (Nei's 1978 standard genetic distance), and $(\delta\mu)^2$ (a *Ds* analogue based on microsatellite allele sizes, Goldstein & Pollock 1997). F_{ST}, *Rho*, and R_{ST} are computed using an ANOVA approach following, respectively, Weir & Cockerham (1984), Ronfort *et al.* (1998), and Rousset (1996). Global *F*-statistics (F_{IT}, F_{IS}, F_{ST}) or *R*-statistics (R_{IT}, R_{IS}, R_{ST}) are also provided.

All individual or population based statistics are computed for each locus and a multilocus weighted average. A jackknife procedure over loci (i.e. deleting information from one locus at a time, Sokal & Rohlf 1995) provides approximate standard errors for the multilocus estimates. In the case of pairwise statistics, mean values per distance intervals and regression slopes on spatial (or other) distances are given. Spatial genetic structure, population differentiation, and inbreeding coefficients are tested by numerical resampling procedures whereby spatial locations, individuals, or genes are permuted. Permuting locations is equivalent to carrying out a Mantel test between the matrices of pairwise genetic statistics and pairwise spatial (or other) distances. In addition, the program permits to estimate the actual variance (i.e. the remaining variance when the sampling variance has been removed) of these coefficients following Ritland (2000). The actual variance of kinship (or relationship) coefficients and of pairwise F_{ST} is useful for marker based inference of the heritability and the Q_{ST} of quantitative traits (Ritland 2000) respectively.

SPAGEDI is available free from *http://www.ulb.ac.be/sciences/lagev/spagedi.html* where a user manual and examples of data files are also available. The program runs on PC under MICROSOFT WINDOWS 95 or later versions. It also runs on a Macintosh under VIRTUAL PC. The code, written in C language, can be obtained from O. Hardy by writing to ohardy@ulb.ac.be. The data file and the single output file containing the results are text files with tab delimited pieces of information, so that they are best edited on a worksheet program such as MICROSOFT EXCEL. Data files formatted for the software GENEPOP (Raymond & Rousset 1995) or FSTAT (Goudet 1995) can be imported directly.

References

Bonnin I, Ronfort J, Wozniak F, Olivieri I (2001) Spatial effects and rare outcrossing events in *Medicago truncatula* (Fabaceae). *Molecular Ecology*, **10**, 1371–1384.

Dutech C, Seiter J, Petronelli P, Joly HI, Jarne P (2002) Evidence of low gene flow in a neotropical clustered tree species in two rainforest stands of French Guiana. *Molecular Ecology*, **11**, in press.

Felsenstein J (1993) PHYLIP (*Phylogeny Inference Package*), *Version 3.5c. Distributed by the Author*. Department of Genetics. University of Washington, Seattle.

Goldstein DB, Pollock DD (1997) Launching microsatellites: a review of mutation processes and method for phylogenetic inference. *Journal of Heredity*, **88**, 335–342.

Goudet J (1995) FSTAT Version 1.2.: a computer program to calculate F-statistics. *Journal of Heredity*, **86**, 485–486.

Hardy OJ, Vekemans X (1999) Isolation by distance in a continuous population: reconciliation between spatial autocorrelation analysis and population genetics models. *Heredity*, **83**, 145–154.

Hardy OJ, Vekemans X (2001) Patterns of allozymic variation in diploid and tetraploid *Centaurea jacea* at different spatial scales. *Evolution*, **55**, 943–954.

Loiselle BA, Sork VL, Nason J, Graham C (1995) Spatial genetic structure of a tropical understory shrub, *Psychotria officinalis* (Rubiaceae). *American Journal of Botany*, **82**, 1420–1425.

Lynch M, Ritland K (1999) Estimation of pairwise relatedness with molecular markers. *Genetics*, **152**, 1753–1766.

Lynch M, Walsh B (1998) *Genetics and Analysis of Quantitative Traits*. Sinauer Associates, Inc, Sunderland.

Nei M (1978) Estimation of average heterozygosity and genetic distance for small number of individuals. *Genetics*, **89**, 583–590.

Queller DC, Goodnight KF (1989) Estimating relatedness using genetic markers. *Evolution*, **43**, 258–275.

Raymond M, Rousset F (1995) GENEPOP, Version 1.2. A population genetic software for exact tests and eucumenism. *Journal of Heredity*, **86**, 248–249.

Ritland K (1996) Estimators for pairwise relatedness and individual inbreeding coefficients. *Genetical Research, Camb.*, **67**, 175–185.

Ritland K (2000) Marker-inferred relatedness as a tool for detecting heritability in nature. *Molecular Ecology*, **9**, 1195–1204.

Ronfort J, Jenczewski E, Bataillon T, Rousset F (1998) Analysis of population structure in autotetraploid species. *Genetics*, **150**, 921–930.

Rousset F (1996) Equilibrium values of measures of population subdivision for stepwise mutation processes. *Genetics*, **142**, 1357–1362.

Rousset F (1997) Genetic differentiation and estimation of gene flow from *F*-statistics under isolation by distance. *Genetics*, **145**, 1219–1228.

Rousset F (2000) Genetic differentiation between individuals. *Journal of Evolutionary Biology*, **13**, 58–62.

Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, **139**, 1463–1463.

Sokal RR, Rohlf FJ (1995) *Biometry*. W.H. Freeman, New York.

Streiff R, Labbe T, Bacilieri R, Steinkellner H, Glössl J *et al.* (1998) Within-population genetic structure in *Quercus robur* L. & *Quercus petraea* (Matt.) Liebl. assessed with isozymes and microsatellites. *Molecular Ecology*, **7**, 317–328.

Wang J (2002) An estimator for pairwise relatedness using molecular markers. *Genetics*, **160**, 1203–1215.

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.