

Spatial assignment test.

Patrick G. Meirmans

patrick.meirmans@unil.ch

The package

This package contains an R-script for performing a spatial assignment test that uses Kriging to interpolate the allele frequencies. The code consists of two different files: “spatialAssignment.R” which contains all the code to perform the test, and “run.R” which contains an example of how to use it with a file that contains genetic data. The zip-file also contains an example data-file, a file with co-ordinates, and this short manual.

Requirements

- Before running this script, it is necessary to install the “R” statistical framework, which can be downloaded for free from <http://www.r-project.org/>. The test also needs the "hierfstat" and "RandomFields" packages and their dependencies. Some basic knowledge of “R” is required.
- A genetic data file in Fstat format. At the moment, the script only works with diploid diallelic data. Haploid data, or data from dominant markers can also be used by coding all alleles or bands as diploid homozygotes. The alleles should be coded as 1 and 2, so not the 0/1 which is normally used for dominant data. A suitable large number of individuals (hundreds) are necessary to perform the test.
- The co-ordinates of all sampling locations, in the same order as the populations in the genetic data file. The co-ordinates should be in a normal XY co-ordinate system, so not in longitude-latitude.
- The co-ordinates of all possible source locations of migrants. Usually, these locations form a continuous area (e.g. the distribution area of a species) over which the sampling locations have been suitable distributed. This can overlap with the sampling locations, but this is not really necessary. For example, it is often handy to use possible source locations that form a evenly spaced grid, though the sampling locations may not exactly correspond to the exact grid points.

The method

- Assignment tests (Paetkau et al. 1995) are often used for detecting dispersers as they can infer the source population of individuals based on genetic markers. However, this class of tests suffers from the problem of “ghost populations” since populations that are not included in the sample may be the source population of sampled migrants. This means that extensive population sampling is required for most applications.
- We therefore developed a spatial extension of the classical assignment test, where we use kriging to interpolate allele frequencies on which the assignment is performed.
- Unfortunately, the performance of the assignment test is in general very poor, and many individuals are falsely inferred to be long-distance dispersers, unless there is a very strong spatial autocorrelation in the distribution of the genetic variation.
- However, many of those false positives can be filtered out by calculating a likelihood ratio with the likelihood that the individual belongs to the location where it was sampled as the numerator, and the maximum likelihood found as the denominator. The great majority of the false positives have very low likelihood ratios, so only keeping the dispersers with very high ratios greatly improves the accuracy of the test, though some false positives may still remain and some actual dispersers may be lost.
- Fortunately, it is possible to get a good idea of the total rate of long-distance dispersal by applying a threshold in the likelihood ratio that is a fraction of the maximum likelihood ratio found. The size of the fraction depends on the sample size: with 500 individuals a fraction of 0.75 works best, with 1000 individuals a fraction of 0.65, and with 2000 individuals a fraction of 0.6.
- The method is described in detail in our manuscript: “Using assignment tests to infer long-distance dispersal in the presence of ghost populations”. P. Meirmans, O. Gaggiotti, & J. Goudet (*in prep*).

More info

- The code is quite heavily annotated, so that should help in understanding how the method works. The best locations to start are the code in the “run.R” file, and the “source.pops” function in the “spatialAssignment.R” file, which is the function that actually performs the interpolation of allele frequencies and the assignment test.
- If you have any question regarding the use of this script or about the method in general you can of course also always email me at patrick.meirmans@unil.ch