

## Program "AutocorQ" ver. 2.00

Olivier Hardy  
Laboratoire de Génétique et d'Ecologie Végétales  
Université Libre de Bruxelles  
1850 Chaussée de Wavre  
B-1160 Bruxelles Belgium  
e-mail: ohardy@ulb.ac.be

### GENERAL INFORMATION

This program estimates autocorrelation coefficients (Moran's I statistic) between pairs of individuals according to geographical distance for quantitative traits (or individual allele frequencies). It is designed to characterize the spatial structure in a continuous space. It also compute correlation coefficients among characters. It performs several statistical tests using numerical resampling methods.

To launch the program, you can call the input file "qin.txt" (default name) and double click on the software icon. "qin.txt" and "autocorQ.exe" must then reside on the same directory. The default output file is then called "qout.txt".

An alternative is to create a shortcut to the program file "autocorQ.exe" (by dragging the program file after right-clicking and choosing CREATE SHORTCUT) and then type the instruction line above on the shortcut sheet of the PROPERTIES associated to the shortcut file/object (reached by right-clicking on the shortcut file/object). The shortcut file/object does not need to reside on the same directory as the program file. The names of inputfile and outputfile on the instruction line are necessary.

### INPUT FORMAT

Inputfile must be in the following format, with each piece of information within a line being separated by a tab :

- first line: number of individuals, number of variables, value given for a missing data, name of each variable (up to 10 characters)
- second line: number of classes of distance intervals, the maximum distances for each interval
- starting third line up to the last individual, each line = 1 individual:
  - name of the individual (up to 15 characters),
  - position of the individual along x axis (integer or floating point),
  - position of the individual along y axis (integer or floating point),
  - value of the quantitative trait for each variable (for missing data, use the value predefined on the first line)
- line after the last individual of the last population: the word END

Here is an example of inputfile:

```
5      3      -99    var1   var2   var3
4      10     20.5   50     100
ind1   7.3     21      0      1.3   -99
ind2   8.4     52      1      -99    2
3      5.11   103     2      -1.2   2.3
5      1.0    13.2    0      2      -1.5
lastind 1.94   129     2      0      3
END
```

If the individuals are not independent from each other but come from different families/populations (e.g. progenies from different mothers, individuals from different populations), they will be treated as non independent by giving them the same x,y coordinates when they belong to the same family/population (moreover the members of a same family/population must follow each other in the input file, otherwise they would be treated as independent individuals that share the same position). The program will display the number of different families/populations and will add a class of distance "0" for the autocorrelation coefficients among members of the same family/population.

The program also check that the maximal distance between two individuals is included into the last distance class, if not it creates an additional distance class.

It is important that between each value there is a tab!! and we suggest that you use a worksheet program in order to generate and edit these inputfiles (in Microsoft EXCEL save the file as "text file with tab delimited"). Extra blanks are allowed anywhere in the inputfile.

Once the program works, it first displays the basic information on the screen, including those relative to the distance classes, and wait for a RETURN. If you want to modify the input file at this stage, stop the program with "Ctrl" + "c". When the program has computed the statistics it asks for the numerical resampling sample size in order to perform the statistical tests.

If a mistake occurs the program gives an error message in the file "ERROR.TXT".

#### PRESENT LIMITATIONS OF THE PROGRAM:

max. 1000 individuals

max. 100 variables

max. 10 characters for the locus names

max. 15 characters for the population names

max. 10000 resamplings

Please contact us if these limitations are a problem for you, we may be able to send you a recompiled version with other specifications

#### RESULTS WRITTEN IN FILE OUTPUTFILE:

The output file can be read as a text file or as an EXCEL worksheet; in the latter case you can change the extension ".txt" into ".xls" and open the file by double-clicking on its icon.

First are written the basic information: names of input and output files, numbers of individuals, variables, families/populations (if existing) and classes of distance intervals, names of variables. The number of missing data, the average value, and the variance are given for each variable.

Next for each distance class are figured: the maximum distance, the average distance between all the pairs of individuals belonging to that class, the number of pairs of individuals belonging to that class, the percentage of individuals that are represented at least once in that class, the coefficient of variation of the number of times each individual is represented into that class. This information is useful to choose the distance classes.

The following results are then figured for each variable and for the average estimates:

- for each distance class, the average autocorrelation coefficients between all the pairs of individuals belonging to the distance class (=Moran's I),
- the correlation coefficients ("r-") and regression coefficients ("b-") between the matrix of pairwise autocorrelation coefficients and the matrix of pairwise geographical distance computed on a linear ("-lidist") or a logarithmic (base e) ("-logdist") scale. Those estimates do not depend on the classes of distance interval.

Two average estimates are given (they are only meaningful if the different variables are of the same type, as in the case of individual allele frequencies): one is an unweighted average, the other is a weighted one where the variance of the variable is used as weight.

Correlation coefficients among variables are figured in a matrix. If individuals are grouped into families/populations, an analysis of variance and covariance provide estimators for the intraclass correlation coefficients of each variable as well as the intraclass correlation coefficients among variables. No test is yet available for those estimates.

Random permutations provide probabilities (bilateral tests) that the observed values of the different statistics belong to the corresponding distribution when random permutations of the data are performed.

These tests are performed for the autocorrelation, correlation and regression coefficients, and for a chi-square statistic that measures the global deviation of Moran's I per distance class (out of intraclass) from the mean values obtain by resampling. For the intraclass autocorrelation coefficients, the individuals are permuted among the families/populations. For the autocorrelation coefficients at distance larger than zero, as well as the

correlation and regression coefficients and the chi-square statistic, the families/populations (or simply the individuals if they are not grouped) are permuted among the existing positions.

The program is regularly modified for further improvements. Any suggestion for improvement is welcome. Also if you have trouble with some data sets you can send us the data set by e\_mail and we could try to fix the problem.

Good luck!