Spatial Genetic Software (SGS)

Version 2

2021

User's manual

From:

Bernd DEGEN

Thuenen-Institute of Forest Genetics
Sieker Landstrasse 2
22927 Grosshansdorf
Germany
e-mail: bernd.degen@thuenen.de

Last updated: 28 May 2021

How to cite the program?

Degen, B., Petit, R. and Kremer, A. (2001) SGS - Spatial Genetic Software: A computer program for analysis of spatial genetic and phenotypic structures of individuals and populations. Journal of Heredity 92: 447-448.

How to cite the manual?

Degen, B. (2021) SGS: Spatial Genetic Software 2021. Computer program and user's manual. https://www.thuenen.de/en/fg/software/

_____

Contents

1   Introduction

During the last decades the analysis of small and large scale genetic structures of herbs and trees have become an important issue (Epperson, 1992). An increasing number of experimental studies on spatial genetic structure of plant populations has been published. The first ones were made on the small scale using allozymes (Geburek, 1993; Merzeau *et al.*, 1994; Leonardi *et al.*, 1996; Chung *et al.*, 1998). Then lots of studies using microsatellites, RAPDs, AFLPs and recently SNPs followed (Streiff *et al.*, 1998; Degen *et al.*, 2001; Hardy *et al.*, 2006). Studies using uniparentally inherited molecular markers were manly concentrated on large scale genetic pattern. For example, maternally inherited chloroplast polymorphism has been utilised to reconstruct the postglacial remigration pathways of European oak species (Petit *et al.*, 1997; Ferris *et al.*, 1998). As an example for an investigation on animals, mitochondrial DNA variation has been used to study the spatial genetic autocorrelation between European otter (*Lutra lutra*) populations (Cassens *et al.*, 2000).

Different statistics are used to describe and quantify the spatial genetic structure. Moran's Index and Geary's index are among the most frequent used measures (Cliff and Ord, 1973; Sokal and Oden, 1978). Other approaches used the "number of alleles in common" or kinship-coefficients as a measure of spatial genetic structure (Boshier *et al.*, 1995; Hardy and Vekemans, 2002). Also F_{ST} and G_{ST} statistics have been used to quantify small and large scale genetic structures (Streiff *et al.*, 1998). Further multilocus measures of spatial autocorrelation based on genetic distances were introduced (Degen and Scholz, 1998; Smouse and Peakall, 1999; Vendramin *et al.*, 1999; Cassens *et al.*, 2000).

The program Spatial Genetic Software (SGS) provides a user-friendly WINDOWS tool to analyse small and large scale genetic and phenotypic structures. The program can deal with nearly all types of genetic data such as co-dominant marker (nuclear SNPs, nuclear microsatellites, allozymes), dominant marker (ISSRs, RAPDs, AFLPs) and uniparentally inherited markers (cp- and mt-haplotypes). Most of those data can be analysed on both scales, as individual genotypes of one population (local scale) and as frequencies of alleles or haplotypes for different populations (broad scale). Furthermore, a simple approach to analyse the spatial structure of continuous quantitative traits is integrated. The program offers various statistics to analyse the spatial genetic and phenotypic structure: Morans's index, Geary's index, number of alleles in common and approaches using genetic distances and F_{ST}-values. The statistical significance of all measures is verified by use of permutation test. The results can be stored as csv- or text-files.

2 Obtaining the program and manual

SGS is written in Visual Basic V. 6.0 (Professional Edition) and has been compiled as 32 bit versions for the operating system Microsoft Windows (Windows 10 and earlier versions). The program and the User's Manual are available on our internet homepage:

https://www.thuenen.de/en/fg/software.html

3 Installation

Download the file "SGS_2021_setup.exe". You start the setup procedure with double mouse click on the file " SGS_2021_setup.exe ". After successful setup the program "Spatial Genetic Software" will be added to your program list.

Data sets

After installation you will find in the program directory files with examples for each category of analysable data:

- allele frequencies of different populations: "Test_Allele_Frequencies.txt"
- multilocus-genotypes of different individuals: "Test_Multilocus_Genotypes.txt"
- quantitative traits of different individuals: "Test_Quantitative_Trait.txt"
- RAPDs of different individuals: "Test_RAPDs.txt"

4 Structure of the program

The program has four different menus (see figure 1):

- "File" => to open the input files and to exit the program
- "Analysis" => to specify the general settings and to select the calculated measure
- "Results" => to view the number of pairs in the distance classes and to create output files of the results
- "Help" => to be informed about the program and to get examples for the format of the input files

The structure represents the order of user's activities. First, he has to load an input file ("File"). Then he defines the general settings and selects the calculated measures ("Analysis"). Thereafter, he views the results ("Results"). The program follows these steps by activating (black colour of letters) or deactivating its functions (grey colour of letters).
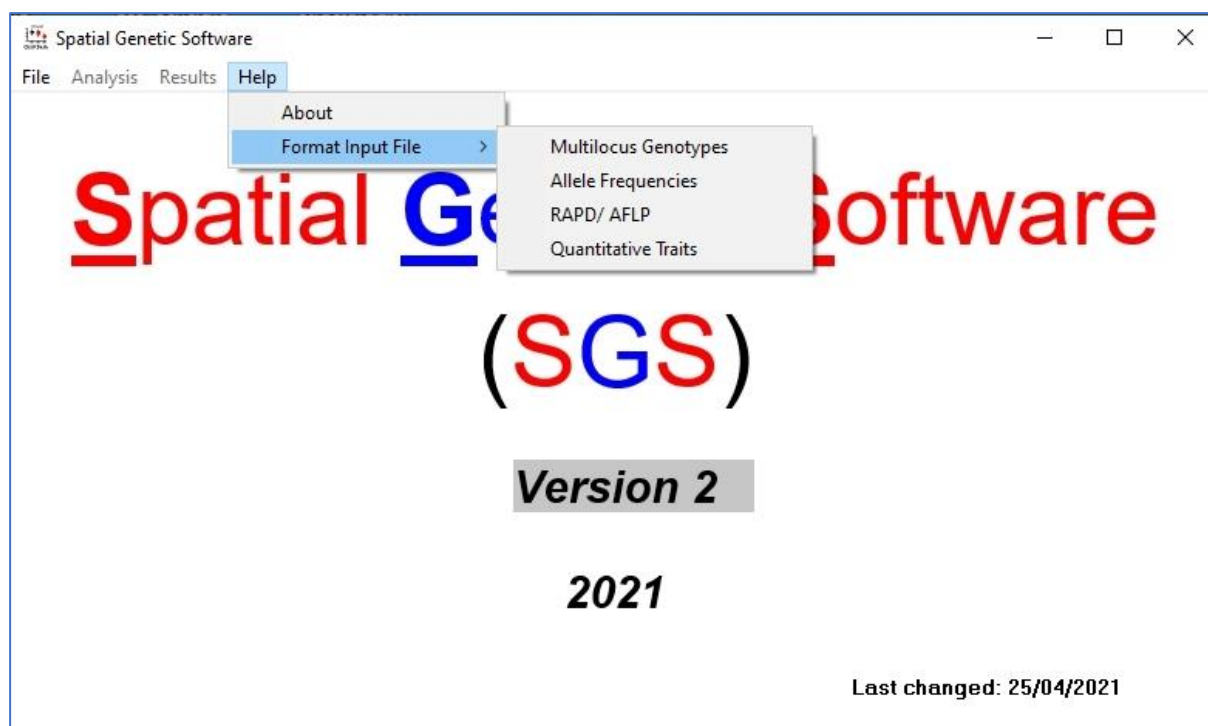
Figure 1: Structure of the program SGS

5 Input data files

The path "File/Load/" leads to the selection of four different types of input data (figure 2). The data can be multilocus genotypes or haplotypes of different individuals from one or several populations ("Multilocus Genotypes/ Haplotypes of Individuals"), relative frequencies of alleles or haplotypes of different populations ("Allele/ Haplotype Frequencies of Populations"), one ore more continuous quantitative traits of individuals or populations ("Quantitative Traits of Individuals or Populations") and RAPD or AFLP data of different individuals of one population ("RAPD/AFLP Data of Individuals"). All input files are simple ASCII files. They can be created with any editor e.g. WinWord, WRITE. The format of each type of input files is explained in figure 3 to figure 6.

Figure 2: Path to open the input files

```
Oaks Steinhorst    <= name of the population
8                  <= number of individuals
2                  <= number of loci
AAP-B              <= name of the first locus
PGI-B
         3     154.69    148.55    2 4   6 6   <= number of the individual
         7     161.04    134.95    3 4   6 6   <= x-position (or longitude)
        14     164.34    106.15    2 3   2 6   <= y-position (or latitude)
        20     155.84    117.76    4 4   6 6   <= 1. allele of the 1. locus
        25     144.86    146.15    2 4   6 6   <= 2. allele of the 2. locus
        28     141.35    138.13    2 3   6 6
        45      113.3    156.73    4 4   2 6
        47     116.31    148.77    2 4   6 6
```

Figure 3: Format "Multilocus Genotypes/ Haplotypes of Individuals"

Note: The different alleles must be coded by numerical values. There is no restriction in terms of number of alleles. Haplotypes must be coded as homozygotes. Don't mix biparentally and uniparentally inherited marker in one file.

```
Beech            <= title
2                <= number of populations
3                <= number of loci
PGDH_A, 3        <= name of the first locus
GOT_B, 3         <= number of alleles at the first locus
LAP_A, 3
6    6.310   51.685    0.965    0.035    0.000 <= number of the population
6    6.310   51.685    0.450    0.550    0.000 <= x-position (or longitude)
6    6.310   51.685    0.435    0.350    0.215 <= y-position (or latitude)
7    6.655   51.541    0.900    0.100    0.000 <= frequency of the 1. allele
7    6.655   51.541    0.315    0.685    0.000  at the 1. locus of the 2 population
7    6.655   51.541    0.440    0.265    0.295 <= frequency of the 2. allele
                                                  at the 3. locus
```

Figure 4: Format "Allele/ Haplotype Frequencies of Populations"

Note: The allele frequencies of each locus are stored in one line. Hence, the data of a population are stored in three lines if you have frequencies for three loci. The number of the population and the x- y-position is repeated in each of those lines.

```
Diversity Beech                      <= title
10                                   <= number of data sets
2                                    <= number of traits
Diversity                            <= name of the 1. trait
Eveness
 6   6.310  51.685  1.760  0.52  <= number of 1. data set
 7   6.655  51.541  1.800  0.68  <= x-position (or longitude)
 8   7.170  51.626  1.674  0.74  <= y-position (or latitude)
 9   7.452  52.006  1.700  0.45  <= value of the trait
10   7.494  52.006  1.776  0.88
11   8.716  51.433  1.913  0.45
12   7.532  50.661  1.398  0.36
13   7.628  50.971  1.719  0.89
14   8.926  51.825  1.858  0.58
15   8.860  51.556  1.761  0.91
```

Figure 5: Format "Quantitative Traits of Individuals or Populations"

Note: The data of quantitative traits can be given for individuals or for populations. The values represent means if they are given for populations. Any data on a continuous trait can be used.

```
RAPD Data Set                                        <= title
10                                                   <= number of individuals
5                                                    <= number of bands
1      12.2  5.02  1     0     1     0     1         <= No. of the 1st individual
2      19.2  7.6   0     0     1     0     1         <= x-position (or longitude)
3      44.1  22.1  1     1     1     0     1         <= y-position (or latitude)
4      35    10    1     0     1     0     1         <= presence of first band
5      41    99    1     1     1     0     1         <= absence of forth band
6      33    22    1     0     1     1     1
7      10    54    1     0     1     0     1
8      40    40    1     1     1     1     1
9      10    70    1     0     1     0     1
10     24    22    1     1     1     1     1
```

Figure 6: Format "RAPD/AFLP Data of Individuals"

Note: The data on RAPDs and AFLPs are given as a matrix of "0" and "1". The value "1" indicates presence of a given fragment and "0" indicates absence of the fragment.

6   Menu "Analysis"

6.1   Window "General Settings"

After reading the input data you select the path "Analysis/General/" (figure 7) and reach the window "General Settings" (figure 8). The window "General Settings" contains four different frames.

Spatial distances between samples

The frame "Spatial distance between samples" shows results of an analysis of the distances between data points of your input file. The information "Minimum of distance" and "Maximum of distance" are thought as a help for the definition of the parameters in the next frame "Spatial distance classes". Please note that the values will be re-calculated if you switch among the options "Type of co-ordinates" (see below).

Spatial distance classes

All calculated statistics are computed for pairs of data points belonging to a series of spatial distance classes. The Euclidean distance (x- and y-co-ordinates) and the distance at the sphere surface (longitude, latitude) are used as a measure of spatial distance between two data points. The user defines the number of spatial distance classes ("No. of distance classes") and the size = dimension of each spatial distance class ("Size of distance classes"). The

dimension of all distance classes is equal. To check the influence of definition of distance classes on the results, it is wise to repeat the calculations with different scales of distance classes.

Note: For statistical reasons you should choose the number and dimension of spatial distance classes so that each class includes at least 30 pairs of data points (Doligez and Joly, 1997). You can verify this number by viewing the window "Pairs of data per distance class" (Path: "Results/Pairs of data per distance class/").
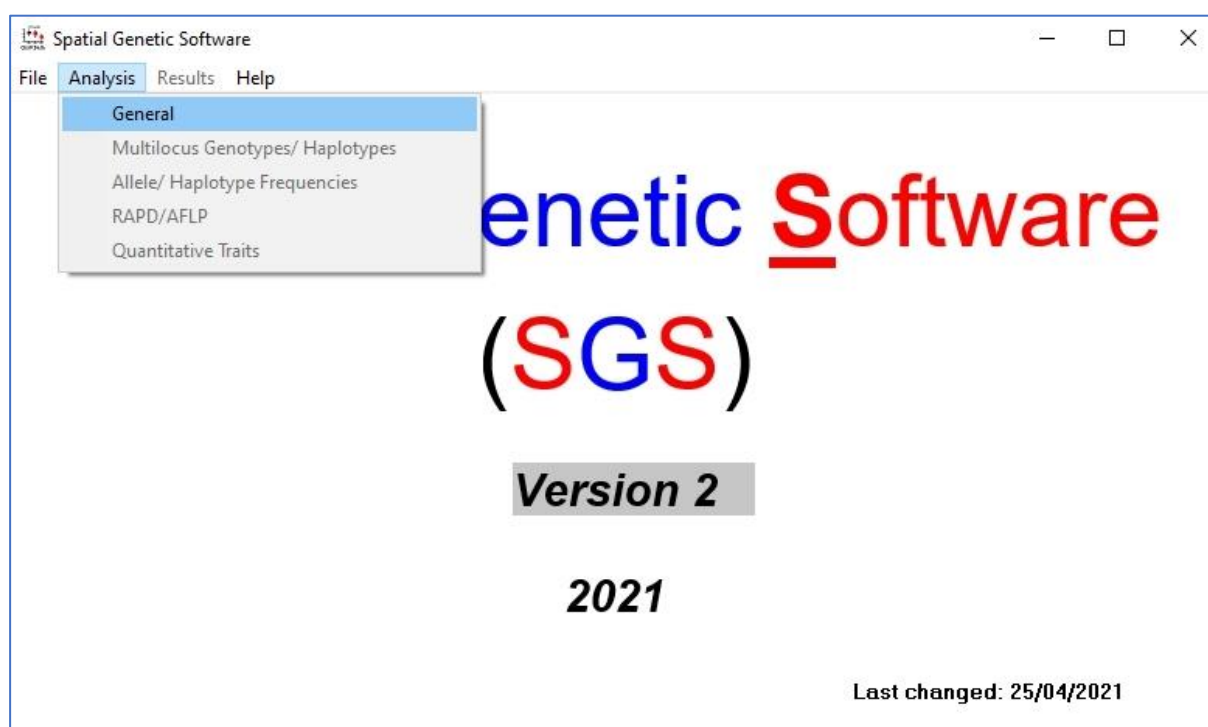

Figure 7: Path to select the analysis options

Permutation test

A permutation procedure using Monte-Carlo simulations is applied to test significant deviation from spatial random distribution of each calculated measure (Manly, 1997). Each permutation consists of a random redistribution of genetic or phenotypic data over the spatial co-ordinates of the sampled points. For each of the spatial distance classes, observed values were compared with the distribution obtained after N permutations. Then a X % confidence interval for the parameters is constructed as the interval from the CI 1. to the CI 2. ordered permutation estimates (Streiff *et al.*, 1998).

N: number of permutations => you should make at least 500 permutations, better 1000

X%: definition of the confidence interval => the 95% confidence interval is often used

CI 1: lower limit of the confidence interval

CI 2: upper limit of the confidence interval

For example: for N = 1000 permutations and X% = 95: CI 1 and CI 2 = 27th and 975th ordered permutation estimates.

Moreover, the user can select the option "No permutation test" in order to compute just the measures without test.
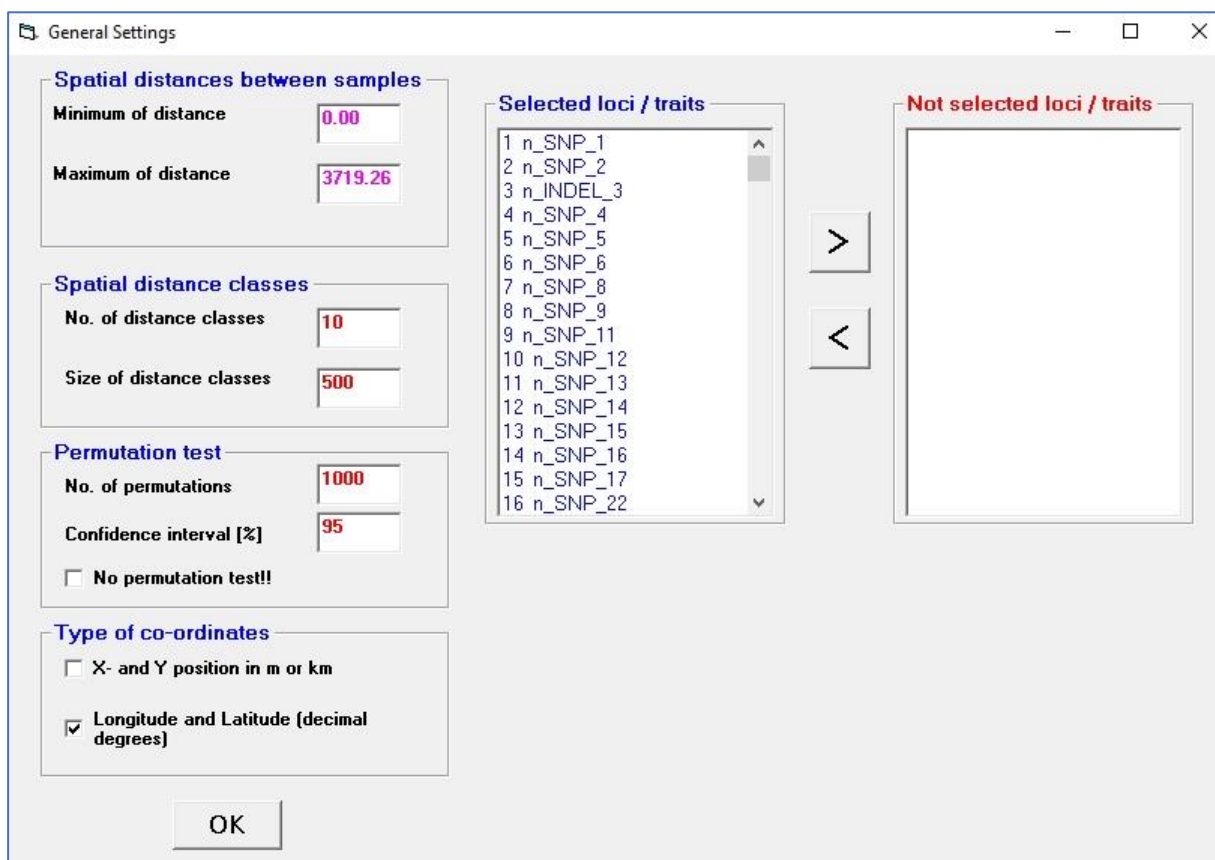

Figure 8: Window "General settings"

Selected loci / traits – Not selected loci /traits

The user can run the calculation for all loci or traits or he can restrict it for a selected part of them. Use the two arrows to select or unselect loci /traits.

Type of co-ordinates

The standard setting is that co-ordinates represent x- and y-values (m, km, miles). The spatial distance among points is then calculated as Euclidian distance. The use can change this and select "Longitude and Latitude (decimal degrees)". The spatial distance (dist) between two geographic co-ordinates [(lon1, lat1), (lon2, lat2)] is then computed in km using the formula for the distance calculation on the sphere surface:

dist = 6378.388 * acos(sin(lat1) * sin(lat2) + cos(lat1) * cos(lat2) * cos(lon2 - lon1))

6.2   Window "Analysis Spatial Structure of Allele/Haplotype Frequencies"

For data sets with allele or haplotype frequencies, the user can choose among three different options (figure 9), selecting a genetic distance measure, either D_N (Nei, 1972) or D_G (Gregorius, 1984), for calculation of genetic distograms (Degen and Scholz, 1998; Vendramin *et al.*, 1999); or selecting a measure based on pair-wise F_{ST} measures (Wright, 1984). Genetic distograms represent graphs where mean genetic distances between all pairs of individuals belonging to a spatial distance class (s_q) are plotted against the spatial distance classes.

The user can calculate genetic distograms, where the dissimilarity of each pair of populations is calculated as (Degen and Scholz 1998):

$$d(s_q) = \frac{1}{N_q} \sum_{i=1}^{N_q-1} \sum_{j=i+1}^{N} d_{ij} \,,$$

where d_{ij} represents the genetic distance between the allele or haplotype frequencies of the *i*-th and *j*-th populations. The estimate $d(s_q)$ represents the average genetic distance of all pairs of populations belonging to the spatial distance interval s_q. The construction of genetic distograms has two advantages. First, they describe spatial patterns for multiple variables simultaneously (not only the frequency of one allele). Second, they apply established concepts of genetic distance to measure dissimilarities. The user can compute the genetic distance d_{ij} using either D_N and D_G:

$$D_N(i, j) = -\ln\left(\frac{\sum_{k=1}^{n} (p_{ik} \cdot p_{jk})}{\sqrt{\sum_{k=1}^{n} p_{ik}^2 \cdot \sum_{k=1}^{n} p_{jk}^2}} \right) \qquad D_G(i, j) = \frac{1}{2} \cdot \sum_{k=1}^{n} \left| p_{ik} - p_{jk} \right|,$$

where *i* and *j* represent two populations, *n* is the number of alleles or haplotypes, p_{ik} is the relative frequency of the *k*-th allele or haplotype.
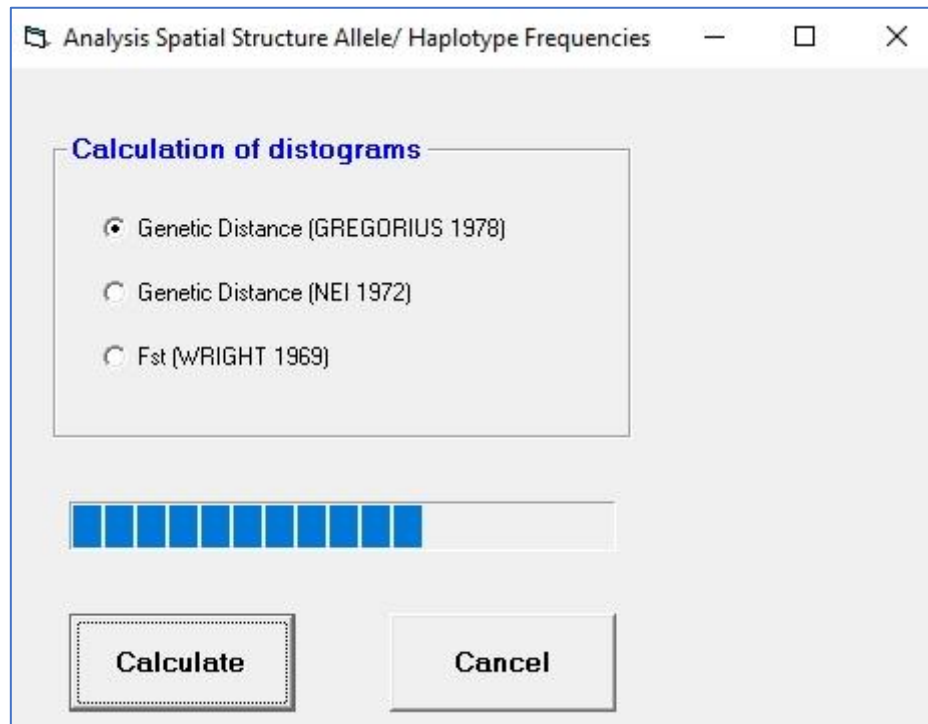
Figure 9: Window "Analysis Spatial Structure Allele/ Haplotyp Frequencies"

The distograms can also be computed from F_{ST} measures. For simplicity, the program is based on allele frequencies, rather than on direct counts of alleles, so that no correction for low sample sizes within populations can be made. Bias should be minimal, however, except when samples sizes per population are very low (<10 individuals). An F_{ST} estimate is obtained for each distance class, as described below (Pons and Petit, 1995). A positive spatial structure is indicated by F_{ST} values that are significantly smaller than the overall F_{ST} value (Dumolin-Lapegue *et al.*, 1997). The calculation consists of the following steps:

For all pairs i, j of populations: The first step is to compute the diversity (expected heterozygosity) H, based on allele frequencies for each population, separately (H_i and H_j), followed by

- Calculation of the average heterozygosity within a population, *i* and *j*, H_{Sij}
- Computation of the total divergence of the two populations, $H_{Tij} = 1 - \sum_k (a_{ki} \bullet a_{kj})$, where *k* indexes the alleles.

For each distance class s_q: The second step is to compute H_{Sq} and H_{Tq}, the means, over all population pairs (*i, j*), of H_{Sij} and H_{Tij}, respectfully. In addition, compute

$$F_{STq} = 1 - \frac{H_{Sq}}{H_{Tq}}$$

The next step is to get the mean F_{ST} over different loci (indexed by m), for each distance class s_q:

$$F_{STq} = 1 - \frac{\sum_m H_{Sm}}{\sum_m H_{Tm}}$$

Reference value: Third, we compute the overall F_{ST}–value, as follows:

$$H_S = \frac{1}{n} \sum_{i=1}^{n} H_i \qquad \text{and} \qquad H_T = 1 - \frac{1}{n(n-1)} \sum_{i \neq j} \sum_k a_{ik} a_{jk},$$

where *n* is the number of populations studied. Finally,

$$F_{ST} = 1 - \frac{H_S}{H_T}$$

6.3 Window "Analysis Spatial Structure of Multilocus Genotypes and Haplotypes"

The user has the choice of either calculating distograms or correlograms (see figure 10). The genetic distance D_G and the number of alleles/haplotypes in common (Hamrick *et al.*, 1993) are available as measures for the distograms. The correlograms are computed using the MORAN's index or GEARY's index (Sokal and Wartenberg, 1983). Please note that for large data sets with many individuals and many gene markers (e.g. SNPs) the calculation of distograms is much faster than the calculation of the correlograms!

Calculation of distograms using genetic distance D_G

Mean genetic distance D_G between all pairs *k, l* of multilocus genotypes (haplotypes) belonging to a given spatial distance class s_q serves as measure for spatial genetic structure. The mean value, over all pairs, constitutes a reference value, against which spatial genetic structure is measured. Distances below the reference value indicate positive spatial structure, where proximal individuals are genetically more similar than expected for a spatially random distribution of genotypes (positive structure), and distances above the reference indicate negative structure, where proximal individuals show greater genetic divergence than expected for a spatially random distribution of genotypes. The multilocus distance measure D_{Gm} between the genotypes *i, j* is defined as:

$$0 \le D_{Gm} = \frac{1}{M} \sum_{i=1}^{M} D_G(i,j) \le 1,$$

where M represents the number of loci. It is the arithmetic average of the single locus values of D_G. For a the calculation of D_G between two individuals, the frequencies of alleles at a certain locus are either 1 (homozygote for that allele), 0.5 (heterozygote with one copy of the allele) or 0 if the allele it is absent.
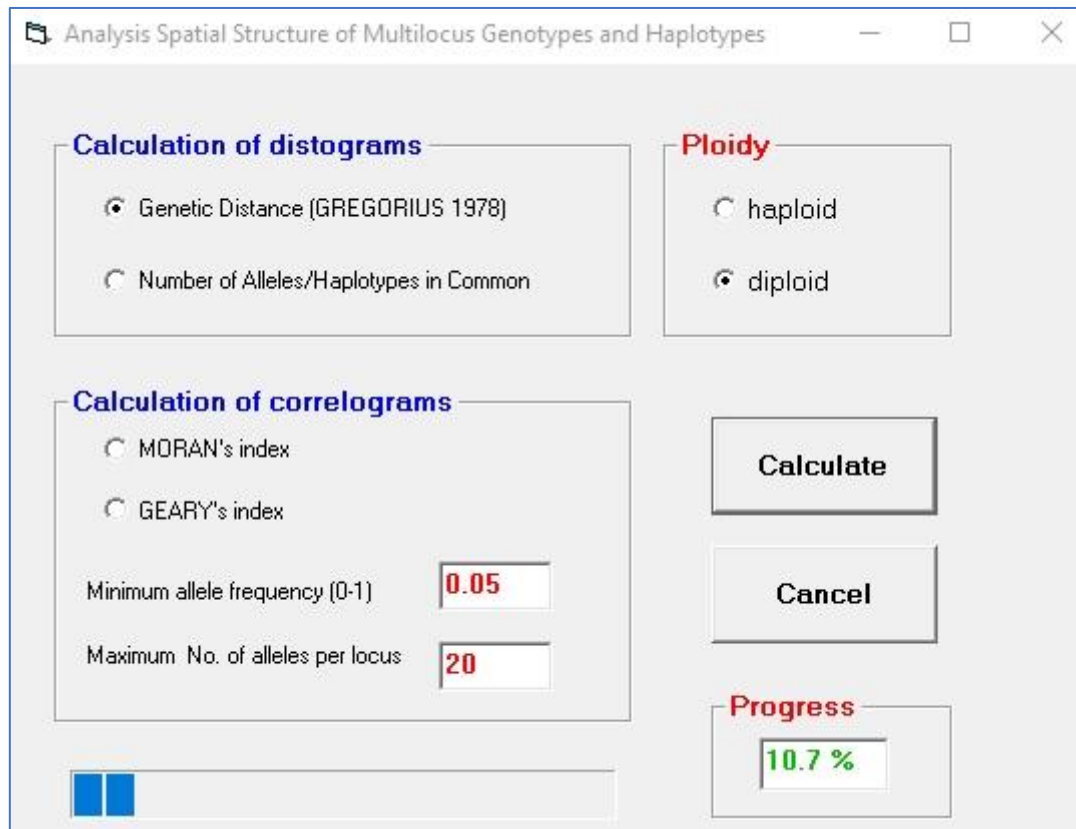

Figure 10: Window "Analysis Spatial Structure of Multilocus Genotype

Calculation of distograms using the number of alleles/haplotypes in common
This measure is adopted from an approach of Surles *et al.* (1990) and Hamrick *et al.* (1993). It is simply the mean number of alleles (haplotypes) in common, over all loci, between pairs of individuals belonging to a given spatial distance class. It's value ranges between 0 and 2 for the diploid case. The mean over all pairs of individuals serves as reference value. Values higher than the reference indicate positive spatial genetic structure, and values below the reference are obtained in case of negative spatial structure.

Calculation of correlograms using MORAN's index

Moran's Index I_q can be computed for both diploid multilocus genotype data and haplotypes (Sokal and Oden, 1978; Sokal and Wartenberg, 1983). The index is calculated for a given distance class s_q. For each allele or haplotype having higher frequency than some arbitrary threshold (X%) in all samples, I_q is calculated as follows:

$$I_q = \frac{n \sum_{i=1}^{n} \sum_{j \neq i}^{n} w_{ij}(a_i - \bar{a})(a_j - \bar{a})}{W \sum_{i=1}^{n} (a_i - \bar{a})^2} \qquad W = \sum_{i=1}^{n} \sum_{j \neq i}^{n} w_{ij}$$

n is the total number of samples and $w_{ij} = 1$ if the individuals i and j belong both to the spatial interval s_q, otherwise $w_{ij} = 0$. For diploid data, a_i is 1 if the i-th individual is homozygous for that allele, 0.5 if heterozygous, and 0 if the individual has no copy of the allele. For haplotypes, a_i is 1 if the i-th individual presents the haplotype and 0 otherwise. The value $\bar{a}$ corresponds to the mean value of a_i over all n individuals. Following Streiff *et al.* (1998), autocorrelation is calculated over all selected loci summing the numerator and denominator of the first equation over the total number of alleles. The expected values for the case of no autocorrelation are – $1/(n-1)$ (Sokal and Wartenberg, 1983). Higher values indicate positive spatial autocorrelation and smaller values indicate negative spatial autocorrelation.

The user can define the "Minimum allele frequency (0-1)" (see figure 10). This value determines the alleles which are included in the calculations. Usually, values not lower than 0.05 are recommended. The value "Maximum No. of alleles per locus" helps to reduce the computing time.

Calculation of correlograms using GEARY's index

The formula for the Geary's index [c] is:

$$c = (n-1) \frac{\sum_{ij} w_{ij}(z_i - z_j)^2}{2W \sum_{i=1}^{n} z_i^{\,2}} \quad \text{with } z_i = a_i - \bar{a}$$

Geary's index is computed over all loci, summing the numerator and denominator of the first equation over the total number of alleles. The index ranges from 0, for perfect positive autocorrelation, to an unbounded positive value for negative autocorrelation. The expected value, in the absence of autocorrelation, is 1.

6.4 Window "Analysis Spatial Structure of Quantitative Traits"

The spatial structure of continuous quantitative traits can be analysed in analogous fashion, by using a distance measure (figure 11). Again, the mean distance between all pairs of individuals belonging to a given distance class serves as a measure of the spatial structure, with the mean over all pairs providing the reference value. A simple city-block distance measure [D] is applied (Deichsel and Trampisch, 1985):

$$t_{im} = \frac{a_{im} - \overline{a}_m}{st_m} \qquad\qquad D_{ij} = \frac{\sum\limits_{m=1}^{n} \left| t_{im} - t_{jm} \right|}{n}$$

Here a_{im} is the value of individual i for trait m, $\overline{a}_m$ represents the mean value for trait m, st_m is the standard deviation of trait m, t_{im} gives the value of individual i for trait m after z-transformation and n is the number of traits. The program computes transformed values of each trait, using the z-transformation, necessary to avoid problems with variable scales among traits (Deichsel and Trampisch 1985, page 22).
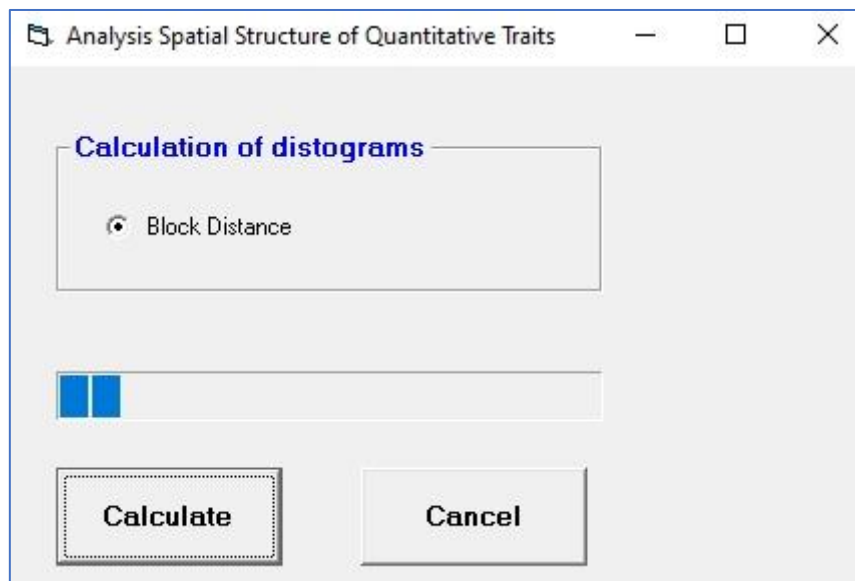

Figure 11: Window "Analysis Spatial Structure of Quantitative Traits"

6.5 Window "Analysis Spatial Structure of RAPD/AFLP data"

RAPD or AFLP data for each individual are transformed into a matrix of 1's and 0's, with 1 indicating presence of a certain fragment and 0 indicating its absence. The genetic dissimilarity [D_{ij}] between the binary vectors of two individuals, i and j, is computed (figure 12), using the Tanimoto distance (Deichsel and Trampisch 1985, page 24):

$$0 \le D_{ij} = 1 - \frac{v_{ij}}{v_{ij} + y_i + y_j} \le 1$$

where v_{ij} represents the number of common fragments of individual *i* and *j*, the numbers of fragments which exist only in individuals *i* or *j*, respectively, are given by y_i and y_j. The overall mean of genetic distances between all individuals serves as a reference for random spatial structure.
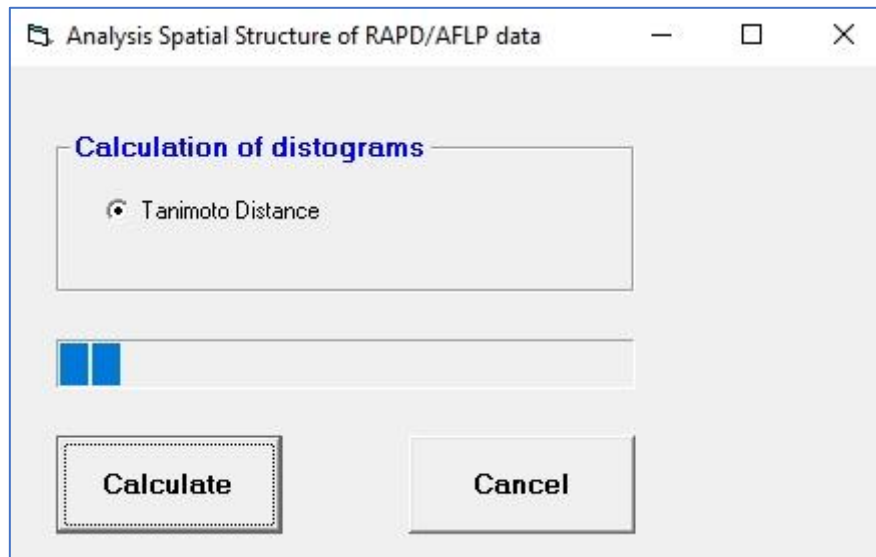

Figure 12: Window "Analysis Spatial Structure of RAPD/ AFLP data"

7   Menu "Results"

The menu "results" offers possibilities to view the number of pairs in the different distance classes and to save the results of the calculations (figure 13). The selection "Pairs of data per distance class" calls a table showing the frequencies of pairs in all spatial distance classes (figure 14). As a critical limit the size and number of spatial distance classes should be adjusted guaranteeing a minimum of 30 pairs for each distance class. Lower values may case statistical problems with the permutation test.

The user can save the results as comma separated (csv) and as and text file by selecting "Results/ Save as CSV-File" (figure 15) and "Results/Save as Text-File". The CSV-file can be used to create the distograms or correlograms (figure 16). Additional information is given in the legends of the figures.

Figure 13: Menu "Results"

Figure 14: Window "Pairs of data per distance class"

	A	B	C	D	E	F	G	H
1	Distance	N Pairs	CI_1	D_ob	D_ex	CI_2	P(D)<-CI	P(D)>+CI
2	15	1100	0.261091	0.247636	0.2695	0.277682	1	0
3	30	2447	0.263649	0.256784	0.2695	0.27546	1	0
4	45	3066	0.264416	0.263405	0.2695	0.274967	0.991	0.009
5	60	3259	0.264974	0.269024	0.2695	0.274609	0.572	0.428
6	75	3354	0.264758	0.266339	0.2695	0.27424	0.912	0.088
7	90	2964	0.265047	0.271407	0.2695	0.274106	0.21	0.79
8	105	2455	0.26442	0.276171	0.2695	0.275173	0.011	0.989
9	120	1867	0.262908	0.27812	0.2695	0.275683	0.002	0.998
10	135	1399	0.259471	0.2802	0.2695	0.278592	0.015	0.985
11	150	1132	0.257951	0.277429	0.2695	0.281979	0.107	0.893

Figure 15: Results stored as CSV-file after using the path "Results/Save as CSV-File/"; "Distance" = upper limit of each spatial distance classes"; "N Pairs" = number of data pairs in each distance class; "CI_1" = below limit of the confidence interval; "D_ob" = observed values for the used measure for spatial structure; "D_ex" expected value of the used measure of spatial autocorrelation for random structures; "CI_2" = above limit of the confidence interval;; "P(D)<-CI" = probability that the observed value is lower than the below limit of the confidence interval; "P(D)>+CI)" = probability that the observed value is greater than the above limit of the confidence interval
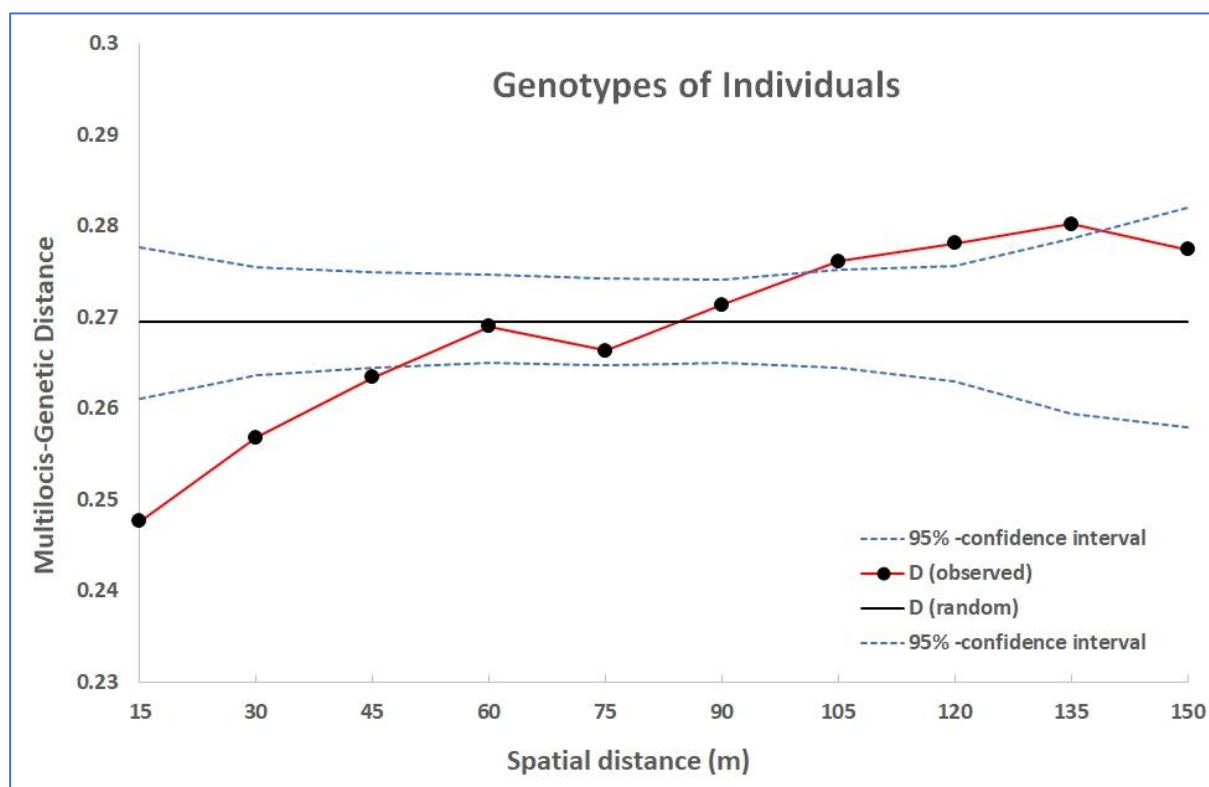

Figure 16: Distogram showing the data of figure 15

8 Acknowledgement

I am grateful to Remy Petit, Antoine Kremer, Hans-Rolf Gregorius, Peter Smouse, Giovanni Vendramin and Birgit Ziegenhagen for valuable discussions and suggestions about the program. Furthermore, I would like to thank Henri Caron, Cyril Dutech, Brigitte Demesure and Sylvie Oddou for critical testing of the program and for helpful suggestions of its improvement.

9 Literature

Boshier, D.H., Chase, M.R., Bawa, K.S., 1995. Population genetics of Cordia alliodora (Boraginaceae), a neotropical tree. 3. Gene flow, neighborhood and population substructure. Am. J. Bot. 82, 484-490.

Cassens, I., Tiedemann, R., Suchentrunk, F., Hartl, G.B., 2000. Mitochondrial DNA variation in the European otter (Lutra lutra) and the use of spatial autocorrelation analysis in conservation. J. Hered. 91, 31-35.

Chung, M.Y., Chung, G.M., Chung, M.G., Epperson, B., 1998. Spatial genetic structure in populations of Cymbidium goeringii (Orchidaceae). Genes Genet. Syst. 73, 281-285.

Cliff, A., Ord, J., 1973. Spatial autocorrelation. Pion Limited, London.

Degen, B., Caron, H., Bandou, E., Maggia, L., Chevallier, M.H., Leveau, A., Kremer, A., 2001. Fine-scale spatial genetic structure of eight tropical tree species as analysed by RAPDs. Heredity 87, 497-507.

Degen, B., Scholz, F., 1998. Spatial genetic differentiation among populations of European beech (Fagus sylvatica L.) in western Germany as identified by geostatistical analysis. Forest Genetics 5, 191-199.

Deichsel, G., Trampisch, H., 1985. Clusteranalyse und Diskriminanzanalyse. Gustav Fisher Verlag, Stuttgart.

Doligez, A., Joly, H.I., 1997. Genetic diversity and spatial structure within a natural stand of a tropical forest tree species, Carapa procera (Meliaceae), in French Guiana. Heredity 79, 72-82.

Dumolin-Lapegue, S., Demesure, B., Fineschi, S., LeCorre, V., Petit, R.J., 1997. Phylogeographic structure of white oaks throughout the European continent. Genetics 146, 1475-1487.

Epperson, B., 1992. Spatial structure of genetic variation within populations of forest trees. New Forests 6, 257-278.

Ferris, C., King, R.A., Vainola, R., Hewitt, G.M., 1998. Chloroplast DNA recognizes three refugial sources of European oaks and suggests independent eastern and western immigrations to Finland. Heredity 80, 584-593.

Geburek, T., 1993. Are Genes Randomly Distributed Over Space in Mature Populations of Sugar Maple (Acer saccharum Marsh.)? . Ann. Bot. 71, 217-222.

Gregorius, H.R., 1984. A unique genetic distance. Biom. J. 26, 13-18.

Hamrick, J.L., Murawski, D.A., Nason, J.D., 1993. The influence of seed dispersal mechanisms on the genetic structure of tropical tree populations. Vegetatio 107-108, 281-297.

Hardy, O.J., Maggia, L., Bandou, E., Breyne, P., Caron, H., Chevallier, M.H., Doligez, A., Dutech, C., Kremer, A., Latouche-Halle, C., Troispoux, V., Veron, V., Degen, B., 2006. Fine-scale genetic structure and gene dispersal inferences in 10 Neotropical tree species. Mol. Ecol. 15, 559-571.

Hardy, O.J., Vekemans, X., 2002. SPAGEDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. Mol. Ecol. Notes 2, 618-620.

Leonardi, S., Raddi, S., Borghetti, M., 1996. Spatial autocorrelation of allozyme traits in a Norway spruce (Picea abies) population. Can. J. For. Res.-Rev. Can. Rech. For. 26, 63-71.

Manly, B., 1997. Randomization, Bootstrap and Monte Carlo Methods in Biology. Chapmann & Hall, London.

Merzeau, D., Comps, B., Thiebaut, B., Cuguen, J., Letouzey, J., 1994. Genetic structure of natural stands of Fagus sylvatica L. (beech). Heredity 72, 269-277.

Nei, M., 1972. Genetic distance between populations. Am. Nat. 106, 283-+.

Petit, R.J., Pineau, E., Demesure, B., Bacilieri, R., Ducousso, A., Kremer, A., 1997. Chloroplast DNA footprints of postglacial recolonization by oaks. Proc. Natl. Acad. Sci. U. S. A. 94, 9996-10001.

Pons, O., Petit, R.J., 1995. Estimation, variance and optimal sampling of gene diversity. I. Haploid locus. Theor. Appl. Genet. 90, 462-470.

Smouse, P.E., Peakall, R., 1999. Spatial autocorrelation analysis of individual multiallele and multilocus genetic structure. Heredity 82, 561-573.

Sokal, R.R., Oden, N.L., 1978. Spatial autocorrelation in biology. 1. Methodology. Biol. J. Linnean Soc. 10, 199-228.

Sokal, R.R., Wartenberg, D.E., 1983. A test of spatial autocorrelation analysis using an isolation-by-distance model. Genetics 105, 219-237.

Streiff, R., Labbe, T., Bacilieri, R., Steinkellner, H., Glossl, J., Kremer, A., 1998. Within-population genetic structure in Quercus robur L. and Quercus petraea (Matt.) Liebl. assessed with isozymes and microsatellites. Mol. Ecol. 7, 317-328.

Surles, S.E., Arnold, J., Schnabel, A., Hamrick, J.L., Bongarten, B.C., 1990. Genetic relatedness in open-pollinated families of two leguminous tree species, Robinia pseudoacacia L. and Gleditsia triacanthos L. Theor. Appl. Genet. 80, 49-56.

Vendramin, G.G., Degen, B., Petit, R.J., Anzidei, M., Madaghiele, A., Ziegenhagen, B., 1999. High level of variation at Abies alba chloroplast microsatellite loci in Europe. Mol. Ecol. 8, 1117-1126.

Wright, S., 1984. Evolution and the Genetics of Populations, Volume 2: Theory of Gene Frequencies. University of Chicago Press.