User's Manual

for

HAPLOTYPE ANALYSIS

Software for Analysis of Haplotype Data

Version 1.05

Prepared by N.-G. Eliades and D. G. Eliades

Georg-August University Goettingen, Germany
Section of Forest Genetics and Forest Tree Breeding

PHYSIS - Centre for Research and
Conservation of Forest Lands, Cyprus

June 2009

Table of Contents

1. Introduction

1.1 Overview

HAPLOTYPE ANALYSIS is a new software for analysis of data from organelle genomes (chloroplast or mitochondrial) which are observed from microsatellites (SSR) or PCR-RFLP markers. The main advantage of this software is that the analysis is performed based on the frequency of haplotypes, which was identified by combining the size variants at the investigated microsatellites (SSR) loci or by combining the restricted fragments of PCR-RFLP loci.

The analysis initiates by estimating haplotype frequencies in each population and, in continuation, it calculates genetic diversity within each population (intra-population analysis); in specific:

- the number of different haplotypes (A)
- the number of private haplotypes (P_h)
- the effective number of haplotypes (N_e)
- the haplotypic richness (H_R) was estimated using the rarefaction method (El Mousadik and Petit, 1996).
- the Nei's index of genetic diversity (H_E) estimated without bias (Nei, 1973)
- the mean genetic distance between individuals (D^2_{sh}) –only for *cpSSR (Goldstein et al., 1995)*

NOTE: For microsatellite markers (chloroplast, mitochondrial) the average genetic distances among the individuals within datasets were estimated using the value of D^2_{sh} (Goldstein et al., 1995), based on the average squared sum of all length differences at microsatellites and with the assumption of the stepwise mutation model (Morgante et al., 1998).

Population genetic structure from the population samples (inter-population analysis) is computed, utilizing:

- Nei's minimum genetic distance (Nei, 1987)

- Genetic differentiation among the populations and contribution of each of them to the total diversity (Finkeldey and Murillo, 1999)

1.2 Disclaimer

HAPLOTYPE ANALYSIS v1.0b is a first test version of software written in Visual Basic for Applications (VBA) within Excel. On the PC, it is compatible with Excel 97 upwards under Windows 95, 98 and 2000, while it has been extensively tested and run in Excel 2002 and Excel 2003 under Windows XP. This first version is possible to contain bugs, some of which are probably easy to detect. However, the authors would appreciate to be informed of any detected bug; in which case please send a sample data file and type of software used, while any suggestions for improvement of the scope and/or of the functionality of the program are welcome.

1.3 How to cite this work

How to cite HAPLOTYPE ANALYSIS:

- Eliades N-G., Eliades D. G. (2009). HAPLOTYPE ANALYSIS: software for analysis of haplotypes data. Distributed by the authors. Forest Genetics and Forest Tree Breeding, Georg-Augst University Goettingen, Germany

2. Working with HAPLOTYPE ANALYSIS

2.1 Installation

HAPLOTYPE ANALYSIS is provided as an Excel add-in, a compiled module. The associated HAPLOTYPE ANALYSIS menu that can be configured to load automatically when Excel is launched or simple loaded as needed.

2.1.1 Installation for automatic loading when Excel is launched

- Copy the HAPLOTYPE ANALYSIS add-in (called *HaplotypeAnalysis.xla*) to your choice of location on your computer. This should preferably be in a dedicated folder.
- Launch Excel. From the menu **Tools** choose **Add-ins** and then click the **Browse** button to locate the HAPLOTYPE ANALYSIS add-in. Once HAPLOTYPE ANALYSIS has been loaded into the dialog list, click the checkbox for HAPLOTYPE ANALYSIS, then click the **Ok** button. Hence, **HAPLOTYPE ANALYSIS** menu will appear in the Excel menu bar just before the **Help** menu.

2.1.2 Loading HAPLOTYPE ANALYSIS only as required

- Copy the HAPLOTYPE ANALYSIS add-in (called HaplotypeAnalysis.xla) to your choice of location on your computer.
- Launch Excel. From the menu **File** choose **Open**, locate the HAPLOTYPE ANALYSIS add-in then click the **Ok** button. Depending on the settings of your Excel program, Excel may warn you that HAPLOTYPE ANALYSIS contains macros. Click the **Enable** button to proceed. In a few seconds the HAPLOTYPE ANALYSIS menu will appear in the Excel menu bar, just before the **Help** menu.

*NOTE: A previous modification of the Security level of Macro extension of Excel may be needed, thus from menu **Tools** choose **Macro** and the click on the Security option. New windows open, where the increase of security level from High to Medium will be necessary.*

2.2 Input Data

2.2.1 Preparing Data

The input data-file is essentially a standard Excel worksheet, where each unit of information is inserted within a cell. In order for the software to recognize the input file properly, certain formatting rules are required. Please certify that all requirements are satisfied before running the software, otherwise error messages will instruct the user for instructions:

- The input data must be entered on the **first** worksheet (i.e. on the left).
- In cell **A1** the word **Cote** must be written, which corresponds to the number of each sample (individual ID); under this column (column A) the sample name must be written.
- In cell **B1** the word **Dataset** must be written, which corresponds to the name of population in which each sample belongs (population ID); under this column (column B) the population name must be written.
- The name of each locus investigated, is entered at the cells of the first row (e.g. **C1, D1, ...**).
- Under each locus name, the actual data are entered, in specific detected fragments size for microsatellites (i.e. as numeric data), or as binary haploid data observed for PCR-RFLP (i.e. **1** if present, **0** if absent).

Example of the input table for organelle genome microsatellite data, with loci scored as the fragment size:

	A	B	C	D	E
1	**Cote**	**Dataset**	**Primer1**	**Primer2**	**Primer3**
2	1	Pop1	123	97	144
3	2	Pop1	125	98	145
4	3	Pop1	125	97	144
5	4	Pop2	124	97	145
6	5	Pop2	124	97	144
7	6	Pop3	123	98	144

Example of the input table for organelle genome PCR-RFLP data, with loci scored as observed binary fragment size:

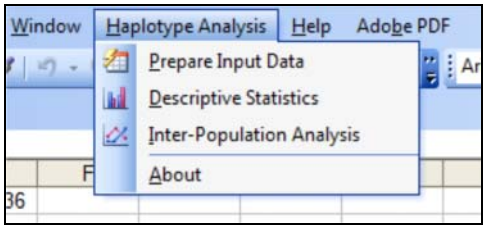
	A	B	C	D	E
1	**Cote**	**Dataset**	**Locus1**	**Locus2**	**Locus3**
2	1	Pop1	1	1	0
3	2	Pop1	1	0	0
4	3	Pop1	0	0	1
5	4	Pop2	1	1	1
6	5	Pop2	0	1	0
7	6	Pop3	1	0	1

2.2.2 Data limitations

Due to Excel's constraints (256 columns and 65,536 rows), the maximum number of haploid loci that can be considered is 254 for 65,535 samples.

2.3 Running HAPLOTYPE ANALYSIS

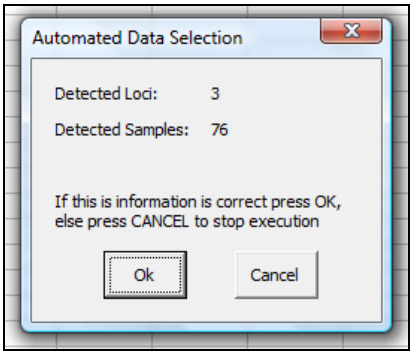
The software, when loaded, creates in the standard Excel menu a new choice, **Haplotype Analysis**, whose choices are depicted in the figure:



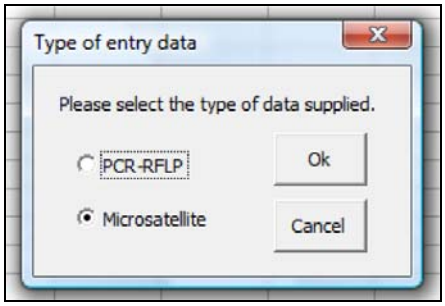
- **Prepare Input Data**, a pre-processing stage that must be executed prior to the other analyses. The user is asked to describe the nature of the data, and a hidden worksheet is created to be used in the following analysis methods.
- **Descriptive Statistics** computes the fragment and haplotype frequencies and performs Intra-population analysis.
- **Inter-population Statistics** are used to compute pairwise Nei's genetic distances and genetic differentiation (F_{ST}) and the total contribution of each population to the total diversity.

2.3.1 Prepare Input Data

The software automatically detects the number of loci and the number of samples, and asks the user for verification, as in the following figure. In case there are missing values, an error message will prompt the user to resolve the problem.



Afterwards, the user must select the type of data used. The software verifies the declared input type with the data imported.



A warning may appear that any previous worksheets will be deleted, which have been created in previous runs. The user can choose not to continue and save the previous worksheets.

In the end, a message notifies the user to run the **Descriptive Statistics**. There are no output worksheets from this action, other than a hidden worksheet.

NOTE: Though it is not necessary, to reveal the hidden worksheet, go to ***Format****,* ***Sheet****,* ***Unhide****, and select* ***"Input".***

2.3.2 Descriptive Statistics

Running this will not require any user interaction. The following worksheets are created:

- Fragments Frequencies
- Haplotypes
- Frequencies
- Private Haplotype
- Intra-Population Analysis (IPA)

"Worksheet **Fragments Frequencies**"

- This Worksheet presents the relative and absolute frequency of the detected fragments size in each population. The table results are also illustrated using Column-graph.

	A	B	C	D	E	F
1	**Fragments Frequencies and Sample Size by Population**					
2						
3			Pop_1	Pop_2	Pop_3	Pop_4
4	Locus1	N	18	20	20	18
5		114	0.111	0.100	0.300	0.000
6		115	0.889	0.900	0.700	1.000
7						
8	Locus2	N	18	20	20	18
9		146	0.167	0.000	0.000	0.000
10		147	0.222	0.750	0.850	0.111
11		148	0.611	0.200	0.100	0.000
12		149	0.000	0.050	0.050	0.000
13		150	0.000	0.000	0.000	0.778
14		151	0.000	0.000	0.000	0.111
15						
16	Locus3	N	18	20	20	18
17		110	0.889	0.800	0.700	0.778
18		111	0.111	0.200	0.300	0.000
19		112	0.000	0.000	0.000	0.222

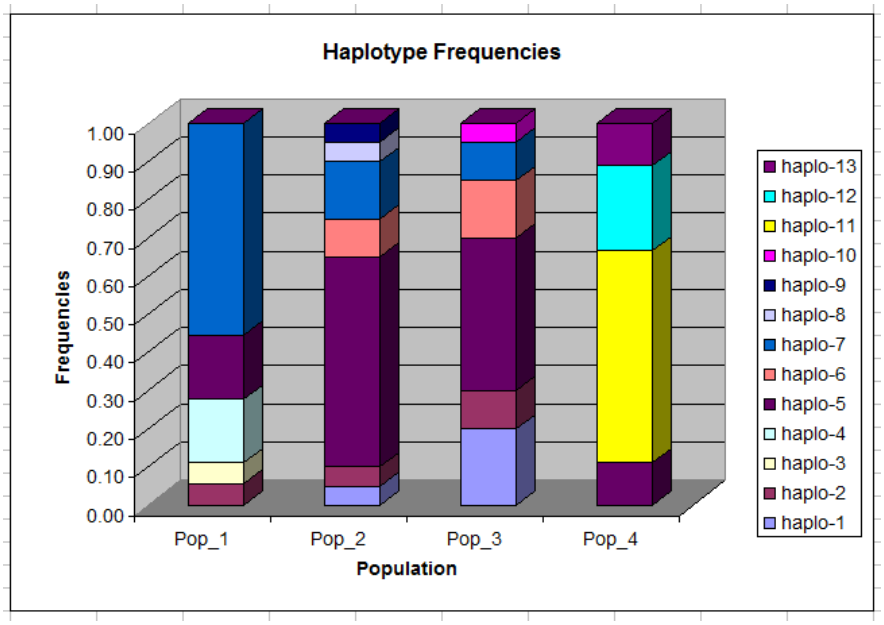
"Worksheet **Haplotypes**"

- This worksheet presents the *Haplotypes code* of each sample as this is identified by the combination of the detected fragments of each locus.

	A	B	C	D
1	Sample	Dataset	Haplotype Code	Haplotype
2	1	Pop_1	3	114 148 111
3	2	Pop_1	4	115 146 110
4	3	Pop_1	7	115 148 110
5	4	Pop_1	7	115 148 110
6	5	Pop_1	7	115 148 110
7	6	Pop_1	7	115 148 110
8	7	Pop_1	7	115 148 110
9	8	Pop_1	7	115 148 110
10	9	Pop_1	7	115 148 110
11	10	Pop_1	7	115 148 110
12	11	Pop_1	5	115 147 110
13	12	Pop_1	4	115 146 110
14	13	Pop_1	2	114 147 111
15	14	Pop_1	7	115 148 110
16	15	Pop_1	5	115 147 110
17	16	Pop_1	5	115 147 110
18	17	Pop_1	4	115 146 110
19	18	Pop_1	7	115 148 110
20	19	Pop_2	5	115 147 110

"Worksheet **Haplotypes Frequencies**"

- This worksheet shows the frequency of the identified Haplotypes within each population; these results are also illustrated in a Column-graph. In addition, the number of expected haplotypes and the number of observed haplotypes are presented at the top of this worksheet.



	A	B	C	D	E	F	G
1	Number of loci			3			
2	Number of expected haplotype			36			
3	Number of observed haplotypes			13			
4							
5							
6	Counts	Haplotype	Haplotype Code	Pop_1	Pop_2	Pop_3	Pop_4
7	5	114 147 110	haplo-1	0.00	0.05	0.20	0.00
8	4	114 147 111	haplo-2	0.06	0.05	0.10	0.00
9	1	114 148 111	haplo-3	0.06	0.00	0.00	0.00
10	3	115 146 110	haplo-4	0.17	0.00	0.00	0.00
11	24	115 147 110	haplo-5	0.17	0.55	0.40	0.11
12	5	115 147 111	haplo-6	0.00	0.10	0.15	0.00
13	15	115 148 110	haplo-7	0.56	0.15	0.10	0.00
14	1	115 148 111	haplo-8	0.00	0.05	0.00	0.00
15	1	115 149 110	haplo-9	0.00	0.05	0.00	0.00
16	1	115 149 111	haplo-10	0.00	0.00	0.05	0.00
17	10	115 150 110	haplo-11	0.00	0.00	0.00	0.56
18	4	115 150 112	haplo-12	0.00	0.00	0.00	0.22
19	2	115 151 110	haplo-13	0.00	0.00	0.00	0.11
20							
21							
22							
23							
24	Counts	Haplotype	Haplotype Code	Pop_1	Pop_2	Pop_3	Pop_4
25	5	114 147 110	haplo-1	0	1	4	0
26	4	114 147 111	haplo-2	1	1	2	0
27	1	114 148 111	haplo-3	1	0	0	0
28	3	115 146 110	haplo-4	3	0	0	0
29	24	115 147 110	haplo-5	3	11	8	2
30	5	115 147 111	haplo-6	0	2	3	0
31	15	115 148 110	haplo-7	10	3	2	0
32	1	115 148 111	haplo-8	0	1	0	0
33	1	115 149 110	haplo-9	0	1	0	0
34	1	115 149 111	haplo-10	0	0	1	0
35	10	115 150 110	haplo-11	0	0	0	10
36	4	115 150 112	haplo-12	0	0	0	4
37	2	115 151 110	haplo-13	0	0	0	2

"Worksheet **Private Haplotypes**"

- This worksheet presents the relative frequency of private haplotypes for each population.

	A	B	C	D
1	Private Haplotypes by Population			
2				
3	Haplotype Code	Haplotype	Population	Frequencies
4	haplo-3	114 148 111	Pop_1	0.056
5	haplo-4	115 146 110	Pop_1	0.167
6	haplo-8	115 148 111	Pop_2	0.050
7	haplo-9	115 149 110	Pop_2	0.050
8	haplo-10	115 149 111	Pop_3	0.050
9	haplo-11	115 150 110	Pop_4	0.556
10	haplo-12	115 150 112	Pop_4	0.222
11	haplo-13	115 151 110	Pop_4	0.111

"Worksheet **Intra-population analysis**"

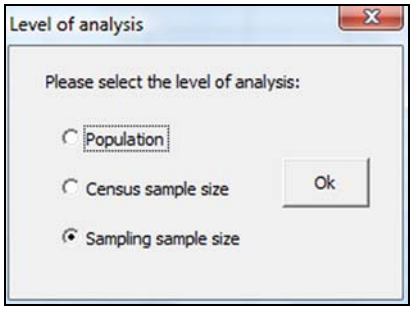
- The intra-population analysis comprises of:
  - **Column A:** The population name or ID
  - **Column B:** The sample size (N) in each population.
  - **Column C:** The number of haplotypes (A) detected in each population.
  - **Column D:** The number of private haplotypes (P) in each population.
  - **Column E:** The effective number of haplotypes (N_e) in each population; the range of values is from $[0, n]$, where n is the detected number of haplotypes within each population. This measure enables meaningful comparisons of haplotypic diversity to be made across populations with diverse distributions of haplotypes frequency.
  - **Column F:** The haplotypic richness (R_h) in each population; the range of values is from $[0, n]$, where n is the detected number of haplotypes within each population. This measure is similar to N_e but is calculated using the rarefaction analysis based on the lowest number of sample size (N) from **Column B.**
  - **Column G:** The genetic diversity (H_e) in each population, ranging from 0 to 1, presents the amount of the diversity within each population. Zero value indicates no diversity.

- **Column H:** The mean genetic distance between individuals (D^2_{sh}) within each population. This measurement takes values $D^2_{sh} \geq 0$, and is affected by the number of fragments detected in each locus (microsatellite), as well as the size of each fragment. The population with the same fragment size in each loci will have zero value, while if fragments size and number increase, D^2_{sh} also increases.

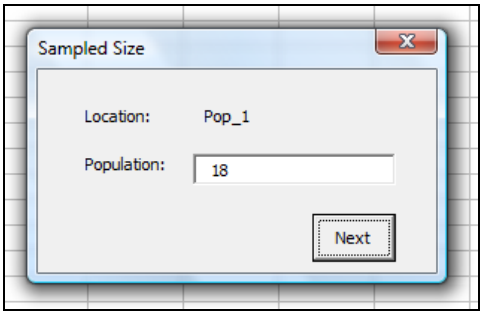
	A	B	C	D	E	F	G	H
1	Population	N	A	P	N_e	R_h	H_e	D^2_sh
2	Pop_1	18	5	2	2.700	4.000	0.667	0.880
3	Pop_2	20	7	2	2.899	5.595	0.689	0.596
4	Pop_3	20	6	1	4.082	4.889	0.795	0.814
5	Pop_4	18	4	3	2.613	3.000	0.654	1.516
6	**Mean**	**19.000**	**5.500**	**2.000**	**3.073**	**4.371**	**0.701**	**0.952**

2.3.3 Inter-population Statistics

Before the program starts the calculations for ***Inter-population statistics***, the program asks the user to specify the level of analysis (i.e. population-based, census or sample size), since that is needed for calculating the differentiation and contribution to the total differentiation.



When the analysis is chosen to be performed based on sample size, the following interface is prompted where the user enters the number of samples for each population (location).



During the analysis two worksheets are created:

"Worksheet **Nei GD" (Genetic Distances)**

- This worksheet presents the pairwise matrix table of Nei's minimum genetic distance, which ranges from 0 (for absolute genetic similarity between two populations) to 1 (for absolute dissimilarity between two populations). In addition, this worksheet includes the normalizations matrix pairwise table, based on the deviation of each value with the maximum value from the above table of **Nei GD**.

	A	B	C	D	E	F
1	**Pairwise Population Matrix of Nei Genetic Distance**					
2		Pop_1	Pop_2	Pop_3	Pop_4	
3	Pop_1	0.000				
4	Pop_2	0.180	0.000			
5	Pop_3	0.180	0.030	0.000		
6	Pop_4	0.358	0.303	0.269	0.000	
7						
8						
9	**Normalized Pairwise Population Matrix of Nei Genetic Distance**					
10		Pop_1	Pop_2	Pop_3	Pop_4	
11	Pop_1	0.000				
12	Pop_2	0.503	0.000			
13	Pop_3	0.503	0.084	0.000		
14	Pop_4	1.000	0.846	0.753	0.000	

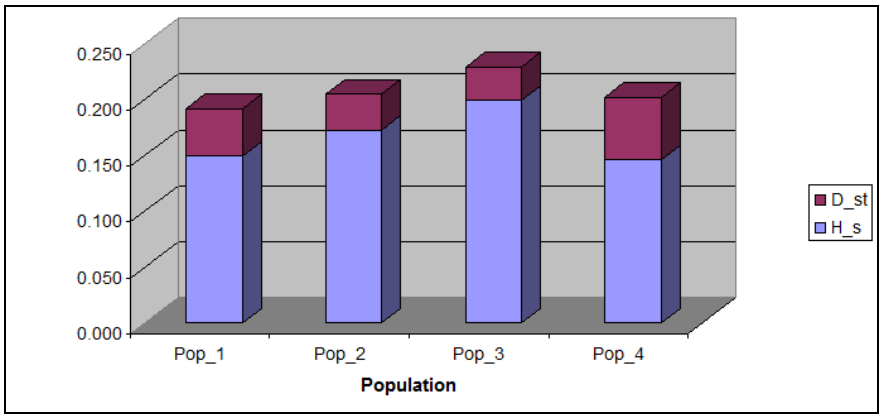
"Worksheet **Fst Differ" (F_{ST} Differentiation)**

- This worksheet shows a table with the estimations of divergence of each population, a graph illustrating the contribution of each population to the total diversity and a pairwise matrix table of genetic differentiation among populations.
    - **Column A:** The population name or ID
    - **Column B:** The sample size (N) in each population.
    - **Column C:** The *Relative contribution* (c_j) the ratio of each population sample size to the total sample size (i.e. number of populations, census or sample size).
    - **Column D:** The G*ene diversity* within each population (H_S), with the correction of the population weights (c_j); the range is from 0 to 1 (as the within population genetic diversity).

- **Column E:** The *Diversity due to differentiation* (D_{ST}) shows the relative differentiation of each population compared to the other (rest) populations; the range is from 0 to 1 (0: no differentiation between populations, 1: completely different populations).
- **Column F:** The *Total gene diversity* (H_T) is calculated as the global genetic diversity of sample gene pool; the range is from 0 to 1 and is also equal to the sum of **Column D** and **Column E**.
- **Column G:** The *Proportion of the total genetic differentiation* (F_{ST}) which in general terms provides a measure of the genetic differentiation among populations. This comprises of a component due to differentiation within populations and of one due to differentiation among populations; it is typically greater than or equal to zero.

	A	B	C	D	E	F	G
1		N	c(j)	H_s(j)	D_st(j)	H_t(j)	F_st(j)
2	Pop_1	18	0.237	0.149	0.043	0.192	0.222
3	Pop_2	20	0.263	0.172	0.032	0.205	0.157
4	Pop_3	20	0.263	0.199	0.030	0.229	0.131
5	Pop_4	18	0.237	0.146	0.056	0.202	0.276
6		*Census sample size					
7				**Hs**	**Dst**	**Ht**	**Fst**
8	**Total**	**76**	**1.000**	**0.666**	**0.160**	**0.827**	**0.194**
9							
10							
11	**Fst Table**						
12		Pop_1	Pop_2	Pop_3	Pop_4		
13	Pop_1	0.000					
14	Pop_2	0.028	0.000				
15	Pop_3	0.027	0.005	0.000			
16	Pop_4	0.051	0.046	0.039	0.000		
17							
18							
19	**Normalized Fst Table**						
20		Pop_1	Pop_2	Pop_3	Pop_4		
21	Pop_1	0.000					
22	Pop_2	0.146	0.000				
23	Pop_3	0.137	0.025	0.000			
24	Pop_4	0.263	0.239	0.201	0.000		
25							

The results from **Column D**, **Column E** and **Column F** are illustrated in a **Column-graphic**, where each column corresponds to a single population. Each column presents the within population genetic diversity (H_S: blue colour) and the amount of population genetic differentiation among the other populations (D_{ST}: mauve color), while the sum of these two components presents the total diversity of each population (H_S: blue color) and the amount of single population genetic differentiation among the other populations ($H_T = D_{ST} + H_S$).



This graph can be a powerful tool for conservation measurements, because it can be used for selecting the candidate population to be given priority for conservation, since conservation strategies must focus on the population with the lowest total diversity.

In addition, the pairwise matrix table of F_{ST} record in this Worksheet takes values from 0 to 1; the same is true for the pairwise matrix table of F_{ST} after the normalization of all values with the maximum value.

3. Implemented Algorithms

3.1 Intra-Population Methods

3.1.1 Effective number of haplotypes N_e

For the j-th dataset, the effective number of haplotypes $N_e(j)$ is given by

$$N_e(j) = \frac{1}{\sum_{i=1}^{N_h} \overline{P}_{(i,j)}^2},$$

where:

N_h is the number of observed haplotypes (or loci)

$\overline{P}_{(i,j)}$ is the relative frequency of the i-th haplotype for the j-th dataset

The effective number of haplotypes is the inverse probability that two randomly chosen haplotypes are identical.

3.1.2 Haplotypic richness R_h

For the j-th dataset, the haplotypic richness $R_h(j)$ is given by

$$R_h(j) = \sum_{i=1}^{N_h} \left[1 - \frac{\binom{N(j)-P_{(i,j)}}{g}}{\binom{N(j)}{g}} \right],$$

where:

$N(j)$ is the sample size of population j

$g = \min_j N(j)$ is the rarefied sample size, i.e. the size of smallest population of the dataset

$P_{(i,j)}$ is the absolute frequency of haplotype i in dataset j

N_h is the number of observed haplotypes (or loci)

NOTE: The implementation was based on the algorithm proposed by Remy Petit in the open source software RAREFAC (April 1995).

References: Hurlbert, 1971

3.1.3 Genetic diversity H_e

The Nei's (1973) unbiased genetic diversity index $H_e(j)$ for dataset j, is given by

$$H_e(j) = \frac{N(j)}{N(j)-1}\left(1 - \sum_{i=1}^{N_h} \overline{P}_{(i,j)}^2\right),$$

where:

$\overline{P}_{(i,j)}$ is the relative frequency of the i-th haplotype

N_h is the number of observed haplotypes (or loci)

$N(j)$ is the sample size within a population.

3.1.4 Average genetic distances among individuals D_{sh}^2

The average genetic distances among individuals metric (Goldstein et al. 1995), which was applied to ccSSRs by Mongante et al. (1998), is given by

$$\begin{aligned} D_{sh}^2(j) &= \frac{2}{N(j)(N(j)-1)}\frac{1}{L}\sum_{i=1}^{N(j)}\sum_{j=i+1}^{N(j)} d_{(i,j)}^2, \\ d_{(i,j)} &= \sum_{k=1}^{L} |a_{(i,k)} - a_{(j,k)}|, \end{aligned}$$

where:

L is the number of loci simulated

$d_{(i,j)}$ is the genetic distance calculated based on: $a_{(i,j)}$, which is the size (measured in repeat units) of the allele for the i-th individual and at the k-th locus, and $a_{(i,k)}$, which is the size of the allele for the j-th individual and at the k-th locus

3.2 Inter-Population Methods

3.2.1 Nei's minimum genetic distance

The genetic distance among populations $j = 1, ..., N_s$ of th i-th haplotype is given by

$$D_{(i,j)} = \frac{1}{2}\sum_{k=1}^{N_h}[\overline{P}_{(k,i)} - \overline{P}_{(k,j)}]^2$$

3.2.2 Differentiation among populations and contribution to the total diversity

3.2.2.1 Relative contribution c

The relative contribution $c(j)$, for each population j can be computed using three different methods:

- Number of populations, where $c(j) = \frac{1}{N_s}$; where N_s is the number of populations,
- Sampling sample size, where $c(j) = \frac{N(j)}{\sum_{i=1}^{N_s} N(i)}$;
- Census sample size, where $c(j)$ is defined by the user.

3.2.2.2 Gene diversity within population H_S

The total gene diversity due to variation within populations is given by

$$H_S(j) = c(j)\left(1 - \sum_{i=1}^{N_h} \overline{P}^2_{(i,j)}\right)$$

3.2.2.3 Diversity due to differentiation among the other populations D_{ST}

The diversity due to differentiation among population $j = 1, ..., N_s$ is given by

$$D_{ST}(j) \quad = \quad c(j) \sum_{i=1}^{N_s} D_{(i,j)} c(i)$$

3.2.2.4 Total gene diversity H_T

The gene diversity of the population is given by

$$H_T(j) = H_S(j) + D_{ST}(j),$$

and the total gene diversity is given by

$$H_T = \sum_{j=1}^{N_s} H_T(j).$$

3.2.2.5 Proportion of the total genetic variance F_{ST}

The proportion of the total genetic variance is given by

$$F_{ST}(j) = \frac{D_{ST}(j)}{H_T(j)}$$

and the total is given by

$$F_{ST} = \sum_{j=1}^{N_s} F_{ST}(j).$$

References

El Mousadik A, Petit RJ (1996). High level of genetic differentiation for allelic richness among populations of the argan tree [*Argania spinosa* (L.) Skeels] endemic to Morocco. Theoretical and Applied Genetics: 92, 832-839.

Finkeldey R, Murillo O (1999). Contributions of subpopulations to total gene diversity. Theoretical and Applied Genetics: 98, 664-668.

Goldstein DB, Linares AR, Cavallisforza LL, Feldman MW (1995). An evaluation of genetic distances for use with microsatellite loci. Genetics: 139, 463-471.

Hurlbert SH (1971). The nonconcept of species diversity: a critique and alternative and alternative parameters. Ecology: 52, 577-586.

Morgante M, Felice N, Vendramin GG (1998) Analysis of hypervariable chloroplast microsatellites in *Pinus halepensis* reveals a dramatic bottleneck. In: Karp A eds Molecular tools for screening biodiversity: plants and animals, Chapman and Hall, London, pp. 402-412.

Nei M (1973). Analysis of gene diversity in subdivided populations. Proc. Natl. Acad. Sci. USA, 70, 3321-3323.

Nei M (1987). Molecular Evolutionary Genetics. New York, Columbia University Press, pp. 219-220.

Petit RJ (2003). Manuscript of RAREFAC. http://www.pierroton.inra.fr/genetics-/labo/Software/Rarefac/