R-lequin - Description of the graphics

Table of contents:

Mismatch distribution

The mismatch distribution is the distribution of the observed number of differences between pairs of haplotypes. Populations at demographic equilibrium show a multimodal distribution. Unimodal mismatch distributions have been interpreted as being due to past demographic expansions (Slatkin and Hudson 1991, Rogers and Harpending 1992). But also spatial expansion can lead to the same unimodal mismatch distribution if neighboring sub-populations exchange many migrants (Ray et al. 2003, Excoffier 2004).
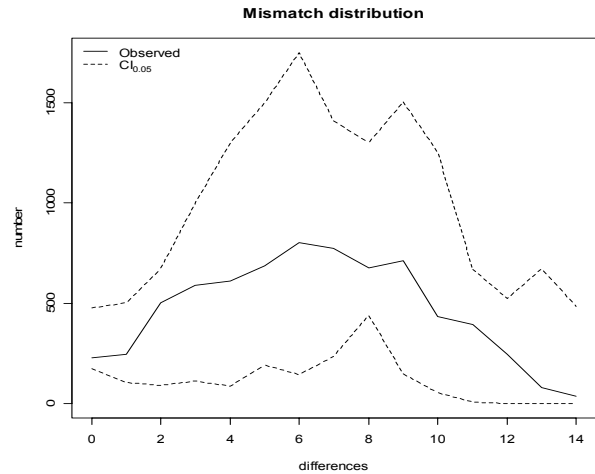
In this graphic you can see on the x axis the number of differences between pairs of haplotypes and on the y axis their frequencies. The solid line indicates the observed frequency and the dashed lines the 95% confidential intervals (α=0.05).

1. F_{ST} Matrix

The fixation index F_{ST} is a measure of population differentiation based on genetic polymorphism data. It compares the genetic variability within and between populations. A common definition given by Hudson et al. (1992) is:

F_{ST}= 1- H_w/H_b

Where H_w is the mean number of differences between different sequences sampled from the same subpopulation and H_b is the mean number of differences between sequences sampled from two different subpopulations sampled. The average pairwise difference

within a population can be calculated as the sum of the pairwise differences divided by the number of pairs. The pairwise F_{ST} can be used as short-term genetic distance between populations with a slight transformation to linearize the distance with population divergence time (Excoffier et al. 2005).

In this graphic you can see the pairwise F_{ST} values between each population. The F_{ST} values are coded with a color code with the legend on the right side.
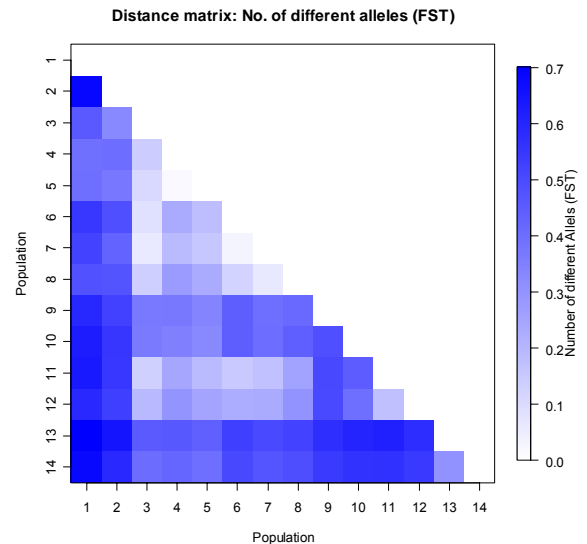
2. Population average pairwise difference

π is the mean number of pairwise differences. In this graphic you can see the average number of pairwise differences between each population in the upper half of the matrix (green). The average number of pairwise differences within each population is shown in the diagonal (orange). And the lower half of the matrix (blue) shows the corrected average pairwise difference between the populations ($\pi_{\text{between xy}} - (\pi_{\text{within x}} + \pi_{\text{within y}}) / 2$). The corresponding legends to the colour codes are on the right side.

Population average pairwise difference
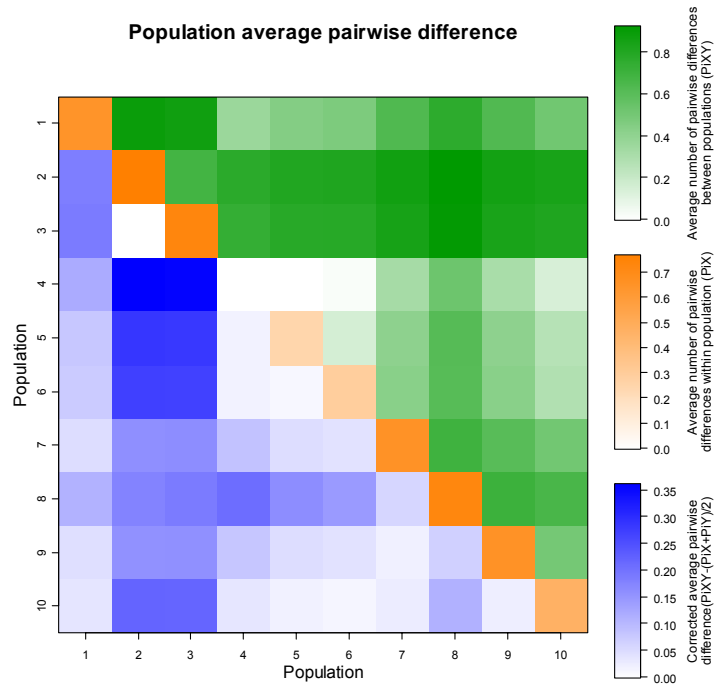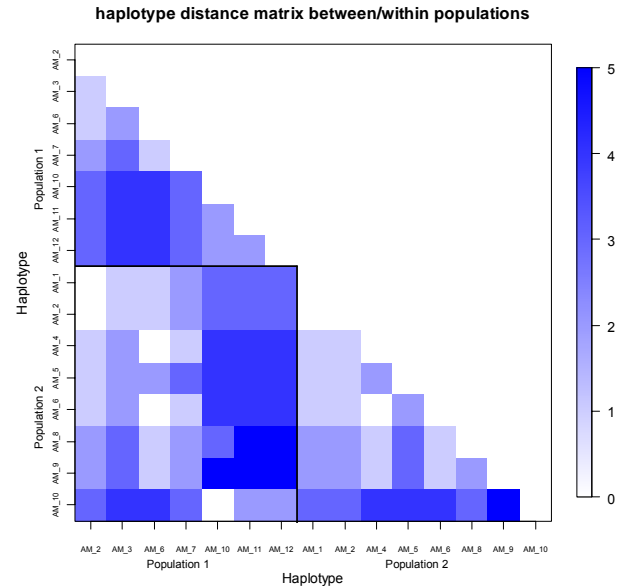
3. Haplotype distance matrix

The inter-haplotypic distance matrix is simply the number of different alleles between each haplotype. The legend of the colour code is on the right side.

Inter-haplotypic distance matrix
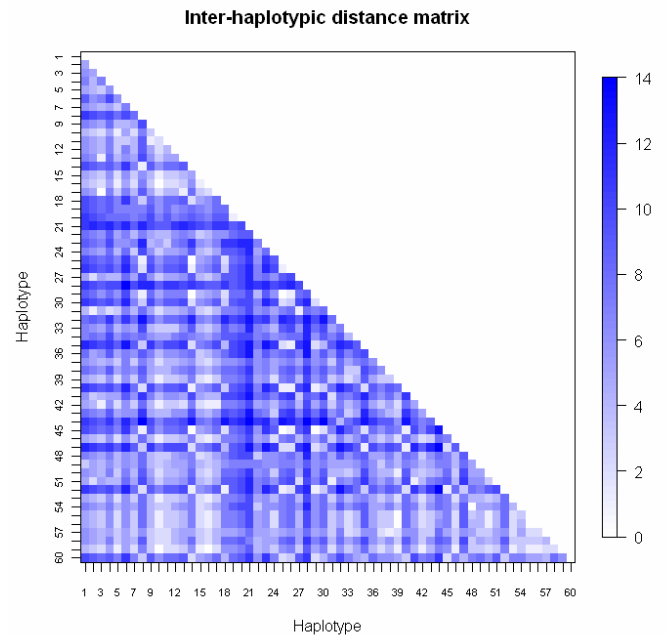
4. Haplotype distance matrix between/within two populations

The haplotype distance matrix is simply the number of different alleles between each haplotype within or between two populations.
Black solid lines separate the haplotypes of the two populations. In the lower left edge of the graphic you can see the haplotype distance matrix between the populations. In the upper left and lower right edge the haplotype distance within the populations is shown. The legend of the colour code is on the right side.

**haplotype distance matrix between/within populations**

5. Haplotype distance matrix between/within populations and groups

The haplotype distance matrix is simply the number of different alleles between each haplotype within or between two populations or groups.

This graphic consist of 3 different graphics. The graphic in the upper left edge shows the haplotype distance matrix within group 1. Black solid lines separate the haplotypes of the two populations. In the lower left edge of this graphic you can see the haplotype distance matrix between the populations and in the upper left and lower right edge the haplotype distance within each population is shown.
In the graphic in the lower left edge you can see the haplotype distance matrix between group 1 and group 2. The populations are also separated by black lines. The graphic in the lower right edge shows the haplotype distance matrix within group 1 and like before the populations are separated by solid black lines.
The legend of the colour code is in the upper right edge.

haplotype distance matrix between/within populations and groups
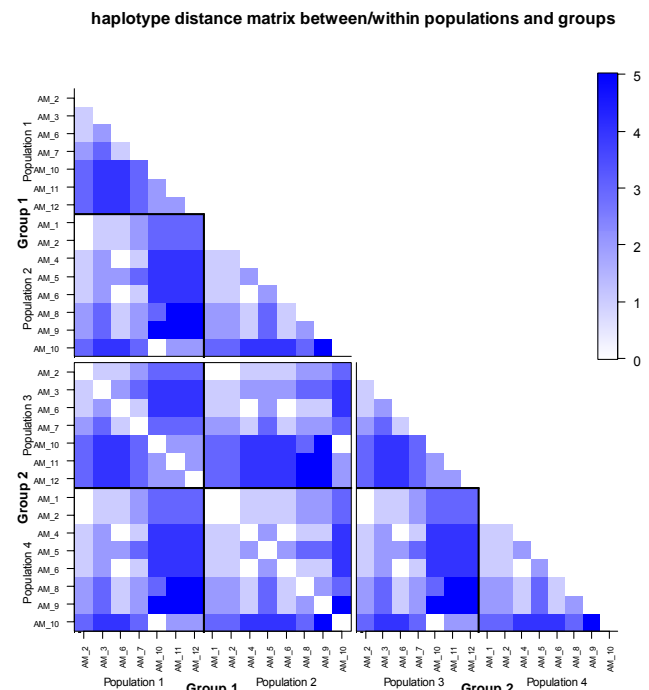
6. Expected/observed haplotype

This graphic shows the observed haplotype frequencies and the expected haplotype frequencies with their standard deviation at different alleles. The expected haplotype frequencies are calculated under the infinite-allele model as predicted by Ewens' (1972) sampling distribution. The null Distribution of the haplotype frequency is generated by simulating random neutral samples having the same number of genes and the same number of haplotypes using the algorithm of Stewart (1977). It can be used to test the hypothesis of selective neutrality and population equilibrium against either balancing selection or the presence of advantageous alleles (Excoffier et al. 2005). Watterson (1978) has shown that the homozygosity is a good statistic for testing departures from selective neutrality in the direction of heterozygote advantage or disadvantage.
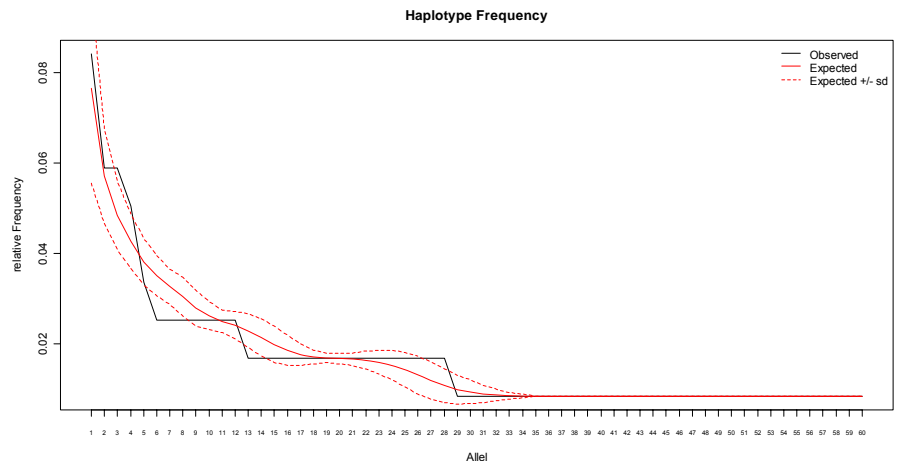
7. Haplotype frequencies in populations

This graphic shows the frequency of each haplotype in different populations. Each population has a line of different colour or style which you can see in the legend in the upper right edge.
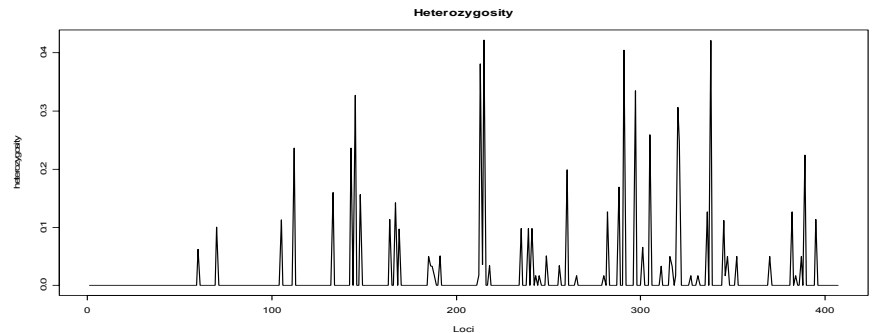With this graphic you can see the dominate haplotypes in different populations an you can therefore compare the different populations with each other.
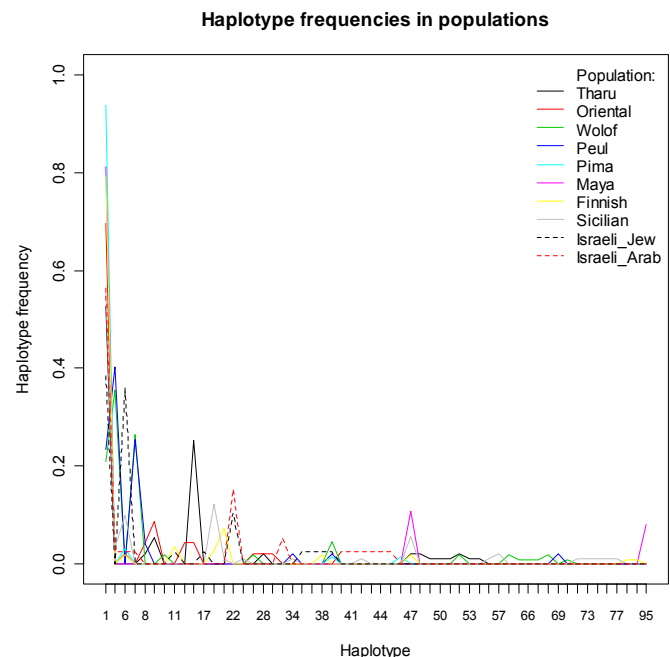
8. Heterozygosity

Heterozygosity is the fraction of individuals in a population that are heterozygous for a particular locus.

In this graphic you can see the heterozygosity of each observed locus.

9. Allelic size range

The allelic size range is the difference between the maximum and the minimum number of repeats for microsatellite data (Excoffier et al. 2005).

This graphic shows the allelic size for each population. The different colours indicate the allelic size for each population at different loci.
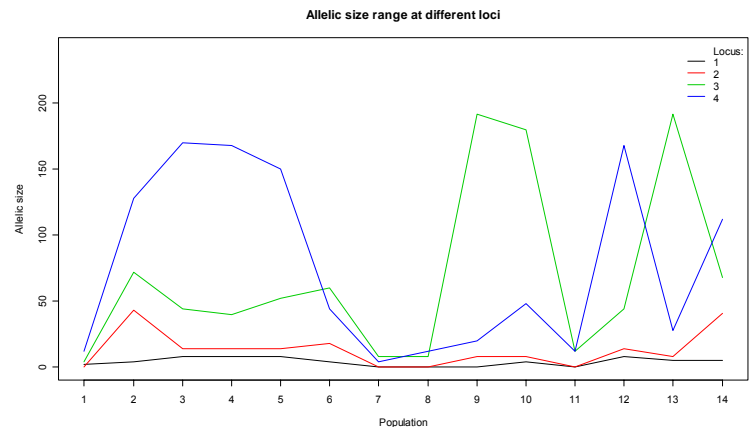
10. Molecular diversity indexes

This graphic shows you the values and the standard deviation of four different diversity indexes in different populations. The solid lines of different colours shows the different values of the diversity indexes and the dashed lines in the same colour the corresponding standard deviations.

Theta is population parameter of genetic differentiation.
$\theta = 2Mu$, where M is equal to 2N for diploid populations of size N or equal to N for haploid populations and u is the overall mutation rate at the haplotype level (Excoffier et al. 2005).
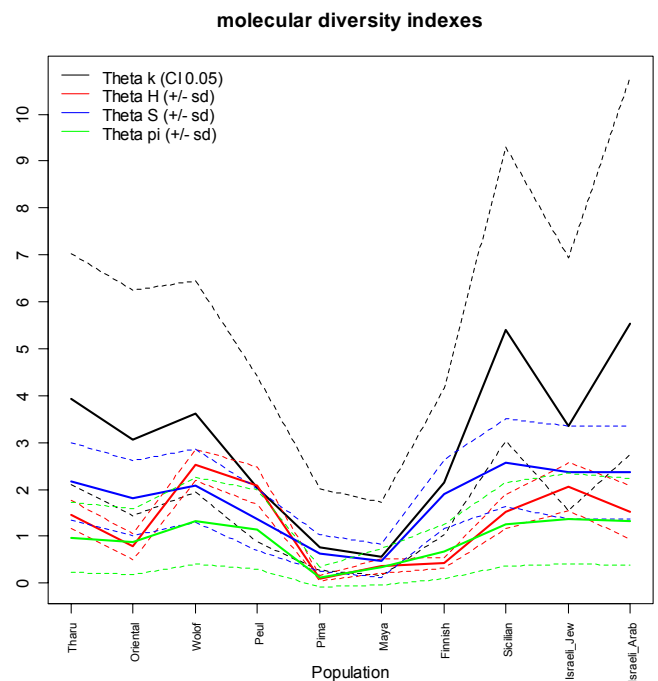
Theta H: is calculated from the expected homozygosity (H) in a population at equilibrium between drift and mutation (see Arlequin 3.1 manual p.92).

Theta S: is estimated from the infinite-site equilibrium relationship (Watterson, 1975) between the number of segregating sites (S), the sample size (n) and theta (θ) for a sample of non-recombining DNA (see Arlequin 3.1 manual p.93).

Theta k: is estimated from the infinite-allele equilibrium relationship (Ewens, 1972) between the expected number of alleles (k), the sample size (n) and theta (θ) (see Arlequin 3.1 manual p.94).

Theta π: is estimated from the infinite-site equilibrium relationship between the mean number of pairwise differences (π) and theta (θ) (Tajima, 1983) (see Arlequin 3.1 manual p. 94)

11. Divergence times allowing for unequal population sizes (tau)

The divergence time (tau) between populations of unequal size is estimated (Gaggiotti and Excoffier 2000). The model assumes that two populations have diverged from an ancestral population of size N_0 some T generations in the past and have remained isolated from each other ever since. The size of the daughter populations can be different but their sum adds up to the size of the ancestral population. From the average number of pairwise difference between and within populations the divergence time scaled by mutation rate is estimated. The estimated values should be interpreted with caution. The procedure implemented is based on the comparison of intra and inter-population diversities (π's) which have a large variance variance, which means that for short divergence times, the average diversity found within population could be larger than that observed between populations. This could lead to negative divergence times (Excoffier et al., 2005).


Divergence times allowing for unequal population sizes (tau)
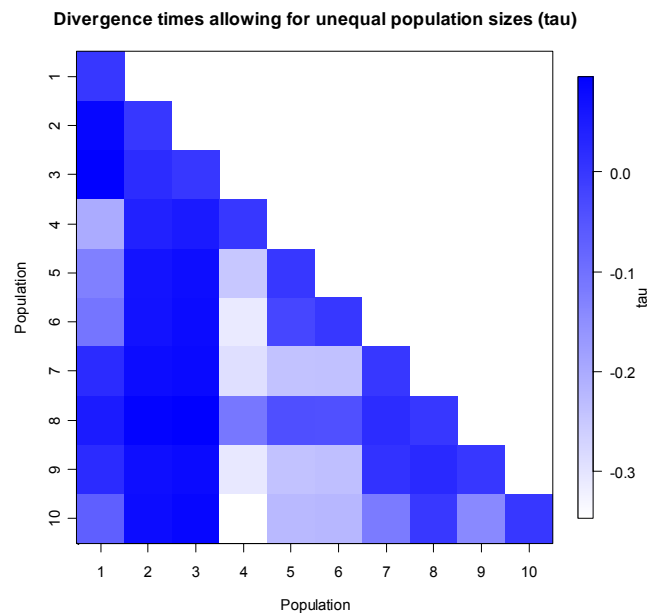
In this graphic you can see the divergence time (tau) between each population. The legend of the colour code is on the right side.

12.   Population assignment test

We may be interested to detect the number of migrants that are presently exchanged between populations, to detect recent immigrants in a given population or to determine the origin of particular individuals.

In the population assignment test determine the log-likelihood of each individual multi-locus genotype in each population's sample, assuming that the individual comes from that population. The allele frequencies estimated in each sample from the original constitution of the sample is used to calculate the likelihood. It is assumed that all loci are independent, such that the global individual likelihood is obtained as the product of the likelihood at each locus (Excoffier et al. 2005).

Population assignment test
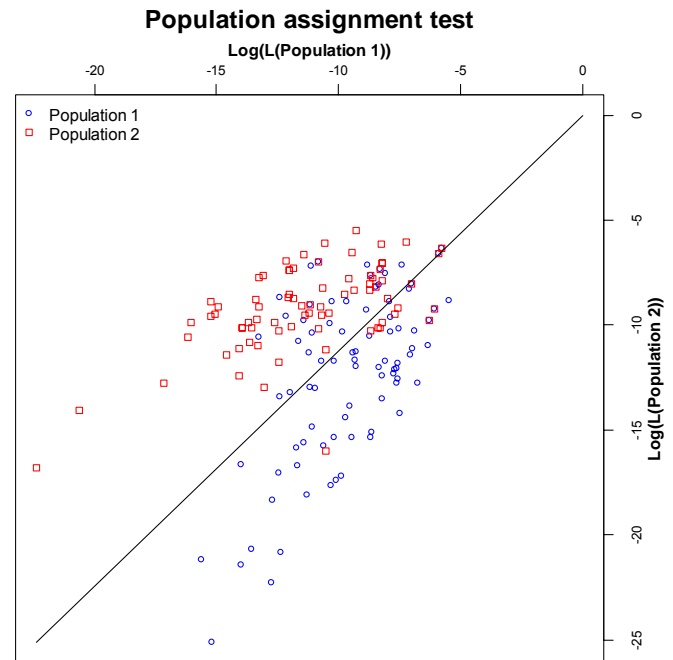
In the graphic the log-likelihood of two populations is shown on the two different axes. The line in the graphic indicates the equal probability to belong to population 1 or population 2. The more the genotypes of the individuals are above the line, the better they are explained by belonging to the population 2 than to population 1 and the more the genotypes of the individuals are below the line, the better they are explained by belonging to population 1.

With this graphic we try to detect outsider individuals from a given population, allocating migrant individuals to different source populations, inferring movements of animals over different years, detecting admixed populations, detecting hybridisation events and so on. But interpreting these results in term of gene flow is difficult. Additional there exist some problems for example for rare and private alleles, linkage disequilibrium, individuals may be chimeric for the source of their genes, etc.