**Population genetic analyses using ARLEQUIN v3.5**

ARLEQUIN is a general multiple purpose software package that can be used to carry out a range of population genetic analyses of mitochondrial (haplotype) and microsatellite (genotype) data. In the first part of this exercise we will test for deviations from Hardy-Weinberg equilibrium and linkage equilibrium in a set of diploid nuclear microsatellite loci. This is a necessary first step in any population genetic analysis. In the second part of the exercise, we will use an Analysis of Molecular Variance (AMOVA) framework to assess what percentage of the total molecular variance can be attributed to (i) among groups of populations (ii) among populations and (iii) within populations. The program then assesses the significance of the variance components associated with different levels of genetic structure using non-parametric permutation procedures.

For the first exercise we will use the dataset “Gorilla\_nomatches\_data\_2013.txt” which is made up of diploid genotypes from 7 microsatellite loci typed in the following six gorilla populations:

1. Western gorillas:

LOP (Lope, Gabon)

LOS (Lossi, Republic of the Congo)

CAR (Central African Republic)

NIG (Cross River area, Nigeria)

1. Eastern gorillas

BW (Bwindi, Uganda)

KBG (Kahuzi-Biega, Democratic Republic of the Congo)

Before we begin any ARLEQUIN analyses (or in fact any kind of population genetic analyses) we need to convert the raw microsatellite data into a format that can be read by the corresponding program. For this exercise we will use the program CONVERT.

To download: <http://www.agriculture.purdue.edu/fnr/html/faculty/rhodes/students%20and%20staff/glaubitz/software.htm>

A. Open Convert.exe

B. File > Load Data File > Gorilla\_nomatches\_data\_2013.txt

C. Select “Convert input data file format”

D. Selection ARLEQUIN .arp

E. Select “standard” data file format

F. Repeat for STRUCTURE .str

To download ARLEQUIN: <http://cmpg.unibe.ch/software/arlequin35/>

Installation:

1. Download Arlequin35.zip to any temporary directory.
2. Extract all files contained in Arlequin35.zip in the directory of your choice.
3. Start Arlequin by double-clicking on the file WinArl35.exe, which is the main executable file.

Data types DNA sequences

RFLP data

Microsatellite data

Allele frequency data

General format for sequence data

Gorilla1 ACGCGGGAATTAGCGCTA

Gorilla2 ACGTTTTAGTCAGTCTTTA

Microsatellite data

Gorilla1 63 24 32

62 24 30

Gorilla2 63 25 32

63 24 32

What can Arlequin do?

Standard indices Polymorphic sites, number of haplotypes, nucleotide diversity, genetic differentiation

Linkage disequilibrium Tests of non-random association of alleles at different loci

Hardy-Weinberg equilibrium Tests whether genotypes conform to Hardy-Weinberg equilibrium.

Neutrality tests Test of selective neutrality and sample heterogeneity

Mismatch distribution Distribution of population pair-wise sequence differences (unimodal distribution indicates recent population expansion)

Network Computes a minimum spanning network of haplotypes

Assignment Assignment of individual genotypes to specific populations

AMOVA Analysis of molecular variance: hierarchical analysis of molecular variance to quantify the amount of differentiation within/among populations and among regions or taxonomic groups

Population comparisons Pair-wise measures of genetic distance including Fst values between populations

Mantel Tests for a significant association between geographic and genetic distance

1. **Testing for deviations from Hardy-Weinberg equilibrium:**
2. Open Project > Gorilla\_nomatches\_data\_2013.arp

Note that in this dataset there are six populations:

1. Settings > Hardy Weinberg > Click “Perform Exact Test of Hardy-Weinberg Equilibrium”
2. Settings > Molecular Diversity Indices > Click “Standard Diversity Indices” and “Molecular Diversity Indices”
3. Click “Start Computations”
4. To find the results go to a new folder created in the same directory as your input file called “Gorilla\_nomatches\_data\_2013.res”
5. View the results in “Gorilla\_nomatches\_data\_2013.htm”
6. Review the results in the html file

As we are testing for Hardy-Weinberg equilibrium in 7 loci, we will use a conservative correction for testing multiple hypotheses (where p = 0.05/7 = 0.007) to reject the Null hypothesis that loci are in Hardy-Weinberg equilibrium.

1. For each population, which loci are not in Hardy-Weinberg equilibrium?
2. What is the average expected heterozygozity for each population?
3. Which locus has the highest number of alleles?
4. **Testing for deviations from linkage equilibrium**

To test for linkage disequilibrium we will use a likelihood ratio test where the likelihood of the sample evaluated under the hypothesis of no association between loci (linkage equilibrium) is compared to the likelihood of the sample when association is allowed. The significance of the observed likelihood ratio is found by computing the null distribution of this ratio under the hypothesis of linkage equilibrium using a permutation procedure.

1. Settings > Pairwise Linkage > Click “Linkage equilibrium between all pairs of loci”

Review the analysis and record the pairs of loci that appear to be in linkage disequilibrium, again using a conservative correction for multiple hypothesis testing where p = 0.05/number of comparisons within each population. Note that this test is also very sensitive to deviations from Hardy-Weinberg equilibrium

1. How many significant associations can you detect within each population?
2. Are there any systematic associations?
3. **Conducting an AMOVA**

We will now carry out an AMOVA to determine what proportion of the total molecular variance in the nuclear microsatellite dataset is due to differences between Western (Lope, Lossi, Nigeria, CAR) and Eastern (Bwindi, Kahuzi-Biega) gorilla populations. To do this, you need to specify the hypothesized genetic structure in the .arp file as follows:

[[Structure]]

StructureName = "Two groups"

NbGroups = 2

#West

Group = {

"Lope"

"Lossi"

"Republic\_Afrique\_Centrale"

"Nigeria"

}

#East

Group = {

"Kahuzi-Biega"

"Bwindi"

}

1. Unselect all other options except for Genetic structure > AMOVA > Standard AMOVA computations
2. At what level is the largest percentage of the total genetic variance?
3. Is the group component significant (p < 0.05)?

Let us contrast this result with a similar analysis using the mitochondrial dataset that we aligned earlier. In this example, I have provided the file already formatted for you. This file is called: “Gorilla166\_2pop.arp”. Note that there is also an associated datafile defining the different sequence types (haplotypes) labeled as “Gorilla166\_allpops.hap”. This file contains 166 mitochondrial DNA sequences and 83 unique haplotypes. The regional genetic structure has been organized into two groups made up of Eastern and Western gorilla populations.

1. At what level is the largest percentage of the total genetic variance?
2. Is the group component significant (p < 0.05)?
3. Why do you think that these results contrast with the microsatellite results?
4. Can you think of any alternative regional genetic structures to test?