**Alignment of sequence data and the conversion of alignments into different file formats**

Sequence alignment is an important first step in any downstream analysis of DNA sequence data. There are a number of different methods but all of them seek to maximize the “match” between two or more sequences whilst allowing for gaps (insertion or deletions) in the sequence alignment. In this exercise we will use the program MEGA to align mitochondrial sequence data from gorillas and also format the dataset for a hierarchical Analysis of Molecular Variance (AMOVA) to test what proportion of the total molecular variance can be attributed to differences within populations, among populations, and among groups of populations.

**Exercise 1. Carrying out a sequence alignment in MEGA using the MUSCLE algorithm**

The program MUSCLE uses a phylogenetic tree to guide the progressive alignment between multiple sequences in a dataset. Today we are going to align a dataset of 166 sequences from Mountain, Eastern Lowland and Western Lowland gorillas.

Note the following key to the different sites for the three subspecies of gorilla:

1. Mountain gorillas

Uganda (BW or BWD = Bwindi)

Rwanda (RW = Rwanda)

1. Eastern lowland gorillas

Democratic Republic of the Congo (KBG = Kahuzi-Biega, ITB = Itibero, TSH = Tshiambero, ITW = Itombwe Massif)

1. Western lowland gorillas

Central African Republic (CAR )

Equatorial Guinea (EQG )

Cameroon (LBK = Lobeke, DJA = Dja Faunal Reserve, YB = Ebo Forest)

Republic of the Congo (LOS = Lossi, NDK = Ndoki, CQY = Counquati, UEL = Gorilla gorilla uellensis, CAM = Cameroon museum specimen)

Nigeria (NIG or CRS = Cross River, Nigeria)

Northern Gabon (IPS = Ipassa, MAK = Makokou, MtCristal = Monts de Cristal, BEL or BIL = Belinga, MKB = Minkebe)

Southern Gabon (MfCh = Massif du Chaillu, MIK = Mikongo, LAS = Lastourville, PLO = Petit Loango, LOP = Lope reserve, MIK = Mikongo, MtDoudou = Mont Doudou, RAB = Rabi, Langoue, Mayumba, Lekedi, BKB = Bakoumba, COLA, Porthos, Mikongo, ITO = Itombe)

Download:[**http://www.megasoftware.net/**](http://www.megasoftware.net/)

Click: Download v5.1

1. Go to File>Open a File/Session
2. Select Gorilla\_Superfile\_2005\_166input.fas
3. Select the arm icon and select “DNA alignment”
4. Select “Align DNA” and “Select All”
5. Then select “Compute”
6. You can trim the excess sequences from the alignment by highlighting the section for deletion and hitting the “X” button
7. Export the data by using the command Data > Export Alignment > MEGA format or PAUP format
8. Examine the two output formats. They both convey the same information but use very different formats.

**Exercise 3. Formatting the data for an AMOVA analysis**

**To do this we will use the program DnaSP.** The programDnaSP (DNA Sequence Polymorphism) is a software package for the analysis of nucleotide polymorphism from aligned DNA sequence data. DnaSP can estimate several measures of DNA sequence variation within and between populations. It is also very helpful program for converting aligned sequence data files in MEGA format into a format for conducting AMOVA analyses in ARLEQUIN.

1. Click on DNAsp.exe
2. File > Open DataFile > Gorilla\_Superfile\_2005\_166input.meg
3. File > Save/Export Data > Arlequin File Format
4. Click “OK” to save the haplotype file

You can also explore the program – it is very flexible and can be used to compute a variety of molecular diversity indices