PBL MODULE 2 - Population Structure, Genetic Variation, and Conservation

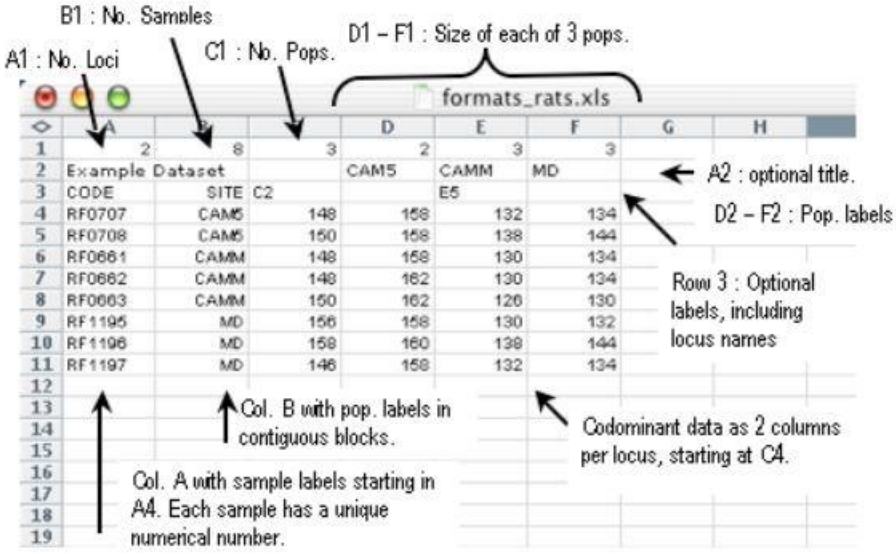
Sub-Module A: Population Size, Heterozygosity and F-Statistics

Data Files:

- File 1: Table 1: Sampling Data [Collection_Data_for_students.xlsx]
- File 2: Microsat_walrus_Genalex_by_location_students.xlsx
- File 3: Microsat_walrus_Genalex_by_year_students.xlsx
- File 4: mtdata_table_format_forstudents.xlsx

The microsatellite data is already formatted for Genalex (see uploaded files on webcourses.)

Example dataset



	A	B	C	D	E	F	G	H
1	2	8	3	2	3	3		
2	Example Dataset			CAM5	CAMM	MD		
3	CODE	SITE	C2		E5			
4	RF0707	CAM5	148	158	132	134		
5	RF0708	CAM5	150	158	138	144		
6	RF0661	CAMM	148	158	130	134		
7	RF0662	CAMM	148	162	130	134		
8	RF0663	CAMM	150	162	126	130		
9	RF1195	MD	156	158	130	132		
10	RF1196	MD	158	160	138	144		
11	RF1197	MD	146	158	132	134		
12								
13								
14								
15								
16								
17								
18								
19								

Codominant markers allow us to determine heterozygotes from homozygotes

Step 1: Examine the Data Files

1. In Collecting Data – note the information contained in the Genotype ID.
2. How many populations are in the Microsatellite by location Data? How many Loci? How many Samples?
3. How many "populations" are in the Microsatellite by year Data?
4. What is the F_{ST} between Hawani and Kinapak based on Mitochondrial CO1 data?

Step 2: Analysis: Running GenaLex (You will use files 2 and 3)

Open the excel file of interest. Then choose, Open file and choose the "GenAlEx 6.503.xlam" file from its location you downloaded it to (for version 6.503). Click "Enable Macros". This should generate a GenAlex ribbon at the top (for EXCEL2016).

Click "add-ins" tab

- From GenAlEx dropdown menu, select "*frequency*"
- Verify number of loci, samples, pops, and codominant data format.
- click OK
- Check only the following:
    - Het, Fstat & Poly by Pop
    - Step by step
    - Click 'ok'
- HFP tab contains estimates of *HO, HE,* mean *HO,* mean *HE, F, FIS, FIT, FST* and *Nm*

Note: You will repeat these steps for both the by location (spatial) and by year (temporal) data files (Files 2 and 3).

This process creates a tab called HFP. Examine the data carefully and check out the legend at the end.

You should make tables that summarize your results succinctly as you go.

1. Determine the observed (H_O) and expected (H_E) heterozygosities, and inbreeding coefficients (F) for each population (spatially) and temporal group by decade. Also determine the overall F_{IS} and F_{ST} values for the population as well as number of migrants (N_m). Note you can calculate the overall inbreeding coefficient (F_{IS})for each population using the mean observed and expected heterozygosities for that population.

Note: Because F_{ST} is influenced by the level of heterozygosity (H_S), another measure called D_{est} can be used, which corrects for genetic diversity and the number of subpopulations in the analysis (Jost, 2008). In Genalex you can assess significant departures from zero based on permuted datasets (e.g. 999 permutations is the default) using the G-statistics function. Here, we can just use Fst for the purposes of this class.

2. Determine the genetic distances, identity, and Fst values between the populations. Fill in the table provided (mtdata_table_format_forstudents.xlsx). You should copy the genetic distance and/or identity tables as well (they can be formatted in the same way as the provided table in the file). See the Note for interpreting genetic distance and identity.

Note: You will repeat these steps for both the by location (spatial) and by year (temporal) data files.

From GenAlEx dropdown menu, select "*frequency*"

- Verify number of loci, samples, pops, and codominant data format.
- click OK
- Click uncheck all
- Check only the following:
    - Nei Distance
    - Pairwise Fst
    - Step by step
    - Click 'ok'

This process will make the tabs NeiP, FstP, and SbySN. You need only data from the first two tabs.

Use the data you collected today to answer some of the questions in the main document.

Note*: F_{ST}* provides an overall estimate of the amount of differentiation among subpopulations (relative to the amount under complete fixation). However, when multiple populations are examined (>2), pairwise F_{ST} estimates are limited. This is because F_{ST} estimates the degree of genetic differentiation using only two populations at a time. It is best to examine genetic differentiation between multiple populations using all the data simultaneously. This can be accomplished using Nei's genetic distance and Nei's genetic identity. Genetic distance estimates range between zero (genetically identical populations) and 1 and can be interpreted as a percent. Nei's identity ranges between zero (genetically unrelated) and 1 (genetically identical). Nei's D can be useful when examining multiple, distantly related species.

References:

Jost, L. (2008). GST and its relatives do not measure differentiation. *Molecular Ecology, 17*(18), 4015-4026.