

Introduction to Arlequin 3.1 – Barber Lab (DD & EC)

- 1) Getting your Sequence Data Ready for Arlequin
 - 1) Getting your Sequence Data Ready for Arlequin
 - 2) Create a blank Arlequin project and import your data
 - 3) Setup and Run your Tests
 - 4) Analyze your results
- 1) Getting your Sequence Data Ready for Arlequin
 - a) Open your data set in MacClade. Make sure that the sequence data is fairly even on both ends. Save (Save as...) the file with a new name, so that you don't overwrite the dataset.
 - b) Remove the redundant taxa.
 - i) *Shift-click* the row numbers on the left side to highlight all of your taxa.
 - ii) *Taxa* → *Redundant Taxa...*
 - iii) Make sure that the “*Consider taxa redundant...*”, “*Select redundant taxa in list*”, and “*Merge Names*” are all selected. Then select “*Search and Merge*.”
 - iv) You should get a list of all the unique haplotypes. The list of taxa should also only have those sequences that are unique. All taxa that share a haplotype should be lumped into the same name. This is now your Haplotype List.
 - (1) Make a note of how many taxa share each haplotype.
 - v) You need to make sure that MacClade saves the file in a particular way.
 - (1) *File* → *Options for Saving* → *Nexus format*. Make sure that Interleave data is NOT selected.
 - c) Format missing data and gaps so that Arlequin can read them.
 - i) *Display* → *Missing, Gap, Multistate...* Make sure that the symbol for Gap is “-“ and the symbol for Missing Data is “?”
 - ii) *Utilities* → *Change all* → *Terminal gaps to missing*.
 - iii) *Save* your file and quit.
 - 2) Create a blank Arlequin project and import your data
 - a) Startup Arlequin
 - b) Select the *Project Wizard Tab*
 - c) Make sure the *DNA Type* is set to DNA
 - d) Under *Controls*, set the “Number of samples” to the total number of different localities or populations that you have. Set the “Missing Data” to be “?”
 - e) Under *Optional Section*, select “*Include haplotype list*” and “*Include genetic structure*.”
 - f) Select the *browse* button under *Data File*. Then select a name and location to start your blank Arlequin project. The filename must end in “.arp”
 - g) Finally select *Create Project*. Then quit Arlequin.
 - h) Open your Arlequin project file in the Windows text editor, or TextWrangler on the Mac side. In a separate window open up your nexus haplotype datafile that you created in MacClade.
 - i) Find the block of haplotypes in your datafile and select just the lines that contain your taxa names and the sequence data. Copy this to the clipboard and then paste this block below the haplotypes in the Arlequin project file where it lists some dummy haplotypes (h1 2 ATCT....). Delete the dummy haplotypes.
 - j) Remove any extra characters from your haplotype names as well as simplifying the names (changing to H1, H2... if you wish). Make sure your sequences stay in alignment as you do

this. Remove any extra characters at the end of your sequence data as well (“[452]”). Make sure that there are only “?” for missing data (on the ends) and “-“ for gaps (in the middle).

- k) Format each of your Sample sections for each locality. Change the *Samplename* to be appropriate for each. The *Samplesize* should be the total number of individuals that you have sequence data in that locality. This is NOT the number of different haplotypes (which will be equal to or less than your sample size).
- l) Under the *Sample Data* section you should have a lists of haplotypes and the number of individuals with that haplotype for that locality. Delete the example haplotypes. Only include haplotypes that have more than 0 representatives. For example:

| | |
|----|---|
| H1 | 2 |
| H2 | 1 |
| H8 | 3 |

- m) Complete this for all Samples that you have.
 - n) Format your *Structure* section. Choose a structure name. The Nb Groups should be 2 or more. In the *Group* section list the exact name of each of your samples, and divide the samples between the groups according to the hypothesis you wish to test.
 - o) Double check everything in this section. It is also crucial that extra characters are not floating around anywhere. Save your work and you are ready for the next section.
- 3) Setup and Run your Tests
- a) Startup Arlequin
 - b) Select *Open Project* and choose your project file.
 - c) Select OK and if your file is ok, it should load and you should see the different samples in the project window.
 - d) Select the structure editor and make sure that the structure is set up for the hypothesis you wish to test. You can change the group number for each sample to test different hypotheses. Make sure you select “update project” after making any changes.
 - e) Select the Settings Tab and then the following options:
 - f) In *AMOVA*:
 - i) Check *Standard AMOVA computations*
 - ii) Check *Compute Minimum spanning network (MSN)*
 - (1) Compute distance matrix
 - (2) Pairwise differences
 - (3) 10,000 permutations
 - g) In *Population Comparisons*:
 - i) Check *Compute pairwise Fst*
 - (1) Compute pairwise differences (pi)
 - (2) Compute distance matrix
 - (3) Pairwise differences
 - (4) 10,000 permutations
 - h) In Molecular diversity indices
 - i) Check *Standard diversity indices*
 - ii) Check *Molecular diversity indices*
 - iii) Check *Theta (S)* and *Theta (pi)*
 - i) In Neutrality tests
 - i) Check *Tajima’s D*
 - ii) Check *Fu’s Fs*
 - (1) 10,000 replicates

- j) Select Start and you should see Arlequin start to work. If it is taking too long, you can stop it and reduce the number of replicates for your tests. If everything went well you should see your results pop up in a browser window.

4) Analyze your results

- a. Pairwise Fst values, look for the following labels:

```
-----
Population pairwise FSTs
-----
-----
FST P values
-----
```

- b. Theta values, look for the following labels:

```
Theta(S) :
S.D. Theta(S) :
```

- c. Haplotype/gene diversity, look for the following labels:

```
=====
== Standard diversity indices :
=====
Gene diversity      :
```

- d. Nucleotide diversity, look for the following labels:

```
Mean number of pairwise differences      :
Nucleotide diversity                     :
```

- e. Hierarchical diversity measures (AMOVA)

```
-----
AMOVA design and results :
-----
```

- f. Building Minimum Spanning Trees

- i. Locate the MST block:

```
=====
MINIMUM SPANNING TREE between      # OTUs
=====
```

- ii. You can use these values to draw your minimum spanning tree by hand or you can find the nexus block and copy that into a new tree file (.tre) and open it in TreeView:

```
-----
NEXUS notation for MST
-----
#NEXUS
begin trees; [NEXUS Treedata section generated by Arlequin]
tree believe_MST = [&U] ((UXI1:0, (((UXI5:0, UXI2:7.00000):0,
UXI11:4.00000):0, UXI7:3.00000):0, UXI9:1.00000):5.00000):0, (UXI12:0,
UXI11:1.00000):2.00000);
end;
```