

**\*Note: Since you will be working in several different programs, it helps to create individual subfolders for each program to store imported and exported data files, software manuals, and resources.**

### **Overall Pipeline**

1. Format data from GeneMapper for import into statistical analyses software using GenAlEx.
2. Identification of clones and subsequent removal from dataset of unique MLGs.
3. Run frequency-based statistics on unmanipulated data in GenAlEx.
  - a. Assumptions testing with HWE
  - b. Genetic diversity with allele frequencies
  - c. Population differentiation with Fst
  - d. Nei genetic distance for PCoA and Mantel test
  - e. Analysis of within vs among population variation with AMOVA
  - f. Isolation by distance with Mantel test
4. Identify if clones were the result of asexual reproduction with MLGsim.
5. Generate Fst values corrected for null alleles using ENA correction in FreeNA.
  - a. Determine if null alleles are skewing Fst calculations.
6. Run tests to identify and correct null alleles with MicroChecker if needed.
7. Create a new dataset with null-corrected genotypes if needed.
  - a. Reimport into GenAlEx and rerun frequency-based statistics.
8. Run population differentiation analyses in Arlequin.
  - a. Population differentiation with Fst
  - b. Assumptions testing with HWE
  - c. Assumptions testing with linkage disequilibrium
  - d. Analysis of within vs among population variation with AMOVA
9. Run population structure analyses in Structure and Structure Harvester.
  - a. Identification of genetic clusters
  - b. K determination
  - c. Combine simulation plots with CLUMPP and Distruct
10. Assess isolation by distance with allele permutation test in Spagedi.
11. Assess genetic bottlenecks with Bottleneck.
12. Determine population relatedness using PopTree.
13. Determine subset migration models based on research questions and Structure results.
14. Determine gene flow and migration rates using GeneClass, BayesAss, and Migrate.

### **GenAlEx 6.503 (Windows 7)**

1. Once you have a .csv exported from GeneMapper, open the file and resave as an .xlsx. Make sure all cells are coded as Numeric in the format options.
2. Remove unnecessary columns:
  - a. This leaves columns Sample Name, Marker, Allele 1, and Allele 2.
3. Use Pivot tables to sort all markers by sample and double check repeated runs for consistency.
  - a. Eventually, you will have one table including all the allelic data for each sample as an individual row.
  - b. Once all screening is completed, delete any samples with more than 3 missing loci. It's a good idea to keep these data somewhere on another sheet.

4. Insert a population column, using familiar site codes (CBC, FGB-EB, FGB-WB, PR, TER). Sort by Pop according to geographic location, then Sample.
5. Create a Pivot Table (Data, Pivot Table) that sums the counts of each pop to get the pop size.
  - a. Next, go to GenAlEx, Create, Codominant Template, and fill out the required information:
    - i. # Loci: 9
    - ii. # Samples: total sample rows
    - iii. # Pops
    - iv. Pop Size: enter each pop size from the Pivot Table in geographical order, then select Add Pop
  - b. This will create a preset template based on your sample data parameters.
  - c. Copy your data into the template and change Pop and Locus names.
6. Any blank cells due to lack of allelic data for a particular locus must be coded as 0 for codominant data (microsatellites are codominant).
  - a. Go to GenAlEx, Manage Data, Edit Raw Data, Empty>0, or use the Replace function in Excel.
7. Check the data format using GenAlEx, Data, Check Raw Data. Look at the two output sheets for any warnings or errors.
8. To identify any clones in your data, use Frequency-Based, Multilocus, Matches...
  - a. Summary of Matches
  - b. Data Subset without Matches
  - c. # Loci to Evaluate for Near Matches: 9
  - d. Ignore missing data when finding matches
    - i. Repeat the same test, but select Consider missing data when finding matches. Double check number of clones is the same.
  - e. Clone samples are listed at the top of the first sheet.
  - f. Use a Pivot Table on the subset dataset to determine new population sizes.
  - g. Create a new Data Template using the pop size information and copy over the subset allele data as the new GenAlEx template and use it going forward.
9. To test for Hardy-Weinburg Equilibrium, use GenAlEx, Frequency-Based, Disequal, HWE. Check that population parameters are set correctly, then click OK.
  - a. Check the options Obs. V Exp. Values, and Summary.
  - b. Rename sheets.
  - c. Export the HWE table (including p values) as a .csv file and run the associated portion of the FDRtool.R script. Copy the FDR-corrected q values back into your GenAlEx spreadsheet.
    - i. Double check that the q values make sense! Small datasets reduce the reliability of FDR correction!
  - d. Look for overall patterns of HWE violations – do the same loci violate HWE across populations? If not, you can proceed with analyses.
10. To calculate allele frequency statistics, use GenAlEx, Frequency-Based, Frequency. Make sure the population parameters are correct and press OK.
  - a. The next window will give options for which tests are to be run:
    - i. Frequency by Pop
    - ii. Het, Fstat & Poly by Pop
    - iii. Nei Distance
    - iv. Pairwise Fst

- b. Each analysis will show up as a separate sheet. Rename the sheets so that they are familiar to you.
- 11. To generate a Principal Coordinates Analysis (PCoA) figure showing the genetic distance between populations, keep the Nei output sheet active and use GenAlEx, Distance-Based, PCoA, Analysis.
  - a. Tri Distance Matrix
  - b. Number of Samples: pop size
  - c. Covariance-Standardized
  - d. Data Labels
  - e. Color Code Pops
  - f. Rename the axes of the resulting figure to Coordinate 1 and 2, and add the % variation from the table above to the axes labels.
- 12. Create an XY template sheet using GenAlEx, Create, XY.
  - a. Select the number of coordinates to be the number of sites you have.
  - b. Replace the data in the created template with your site names and geographical locations (Lat, Long as X, Y) in decimal degrees.
    - i. Copy the XY data from this template over to two columns after your allelic data. Leave a one-column gap between the genotype and XY data.
  - c. Next, create a geographical distance matrix from the XY sheet using GenAlEx, Distance-Based, Distance, Geographic Distance.
    - i. Select X,Y coordinates in Cols 3 & 4, Decimal Lat/Long, As Tri Matrix, Label Matrix, and None for transformation.
- 13. With the Geographic Distance matrix sheet active, select GenAlEx, Distance-Based, Mantel, Paired.
  - a. Select Tri Distance Matrix, 9999 permutations, set the Y Matrix to be your Nei genetic distance sheet.
  - b. Once finished, it will output a new sheet with the results of the permutation test (p value) and a scatterplot comparing geographic and genetic distance with a trendline. Geographic scale is in kilometers.
  - c. To set custom data labels for the plot, insert a new column before the X and Y columns. Create a pairwise comparison label from the labels on the geographic matrix.
  - d. Format the current data labels by right clicking any data point, Format Data Labels, deselect X and Y values, select Value From Cells, and highlight your custom data labels.
  - e. You can repeat this analysis with the D<sub>A</sub> matrix generated by Poptree.
    - i. Copy over the matrix, reformat to match the GenAlEx Nei template, and rerun the same Mantel test parameters.
- 14. To run an analysis of molecular variance (AMOVA), select GenAlEx, Distance-Based, AMOVA.
  - a. Select Raw Data, OK.
  - b. Select Codom-Allelic, Interpolate Missing, Analysis for Total Only.
  - c. Select 9999 permutations, Pie graph, Standard Permute, Output for Total Only, 9999 pairwise permutations, Output Labeled Pairwise Fst Matrix.
  - d. Once the AMOVA is finished, export the pairwise Fst matrix. The bottom values are Fst, the top values are p values.
  - e. Create separate matrices for both Fst and associated p values, making sure the correct order is retained when you transpose the p value matrix.

- i. Save each as .csv files, p values for FDR correction (FDRtool.R), and both Fst and p values for Fst\_heatmap.R
  - ii. After running the FDR correction, copy the FDR-corrected q values back into your GenAlEx spreadsheet and reconstruct the matrix. Re-export p.csv for Fst\_heatmap.R
    1. Double check that the q values make sense! Small datasets reduce the reliability of FDR correction!
15. GenAlEx can export your data into formats used by other analysis programs.
  - a. For MicroChecker outputs, use GenAlEx, Import-Export, MicroChecker.
    - i. This will generate a new Excel workbook. Save as an Excel 2003 (.xls) file.
  - b. For GenePop outputs (FreeNA, Bottleneck, Poptree2, GeneClass2, Formatomatic), use GenAlEx, Import-Export, Export Data, GenePop.
    - i. The filename should not include spaces.
  - c. For Arlequin outputs, use GenAlEx, Import-Export, Export Data, Arlequin.
    - i. Change Options to MICROSAT.
  - d. For Structure outputs, use GenAlEx, Import-Export, Export Data, Structure.
    - i. Check the Use Pop option.
    - ii. If the .txt extension was not added, add it manually.
  - e. For Spagedi outputs, use GenAlEx, Import-Export, Export Data, Spagedi.
    - i. Save the resulting Excel sheet as a tab-delimited .txt file
  - f. For MSA-Analyzer outputs, use GenAlEx, Import-Export, Export Data, MSA.
    - i. Change the top row to: 2, 4, 4, 3, 6, 6, 4, 3, 3, 3 (two alleles, motif for each marker)
    - ii. Save as a tab-delimited .txt file.

### **FDRtools.R (R)**

1. Copy the .csv files needed for FDR correction of HWE, LD, and Fst p values into the FDRtool working directory with the FDRtool.R script
2. Run each block of code, exporting the FDR-corrected p values (q values) as a new column in the dataframe.
3. Check over the calculated q values to ensure they make sense. FDR correction becomes less reliable with fewer p values.
4. Re-export the .csv files and copy the tables back into your GenAlEx spreadsheet.

### **Fst\_heatmap.R (R)**

1. Copy the fst.csv and p.csv files into the heatmap directory with the R script.
2. Run all the lines in the script and save the output figure.
  - a. You will likely have to specify an upper threshold for the Fst color scale
  - b. Change the figure dimensions to fit your particular matrix.

### **MLGsim 2.0 (Windows 7)**

1. The data file (.txt) and program executable file (.exe) must be in a local directory (on the C: drive, not on a virtual machine drive).
2. Use the example data text file to create a project file with your allelic data. Keep the default settings, except for FREQUENCY = MLG.
  - a. Make sure you make the data file in Windows. Mac is adding a hidden line that does not all MLGsim to run.

- b. Follow the MLGsim manual for complete data layout.
    - c. Copy allelic data from the GenAlEx template sheet.
  3. Open the Command Prompt by searching for 'cmd' in the Start menu.
  4. Change directory to the Desktop (or whatever local directory your data file and .exe are located) by using cd Desktop.
  5. Run MLGsim with the following command:
    - a. MLGsim.exe your\_project\_file.text
    - b. The simulations will run, and will output two .csv files to the same directory.
      - i. simresults\_your\_project\_file.csv
      - ii. fulltable\_your\_project\_file.csv
      - iii. The simresults table includes the statistical tests of clones being the result of two sets of parents producing the exact same genotype through random chromosome assortment (statistically rare).

### **FreeNA (Windows 7)**

1. Put the FreeNA.exe file in your working directory with the GenePop file.
2. Open the program.
  - a. Specify the name of the input file (Genepop) including the extension.
  - b. Specify the name of the output file without the extension.
  - c. Specify the number of replicates: 25000.
3. The number of samples, populations, and loci will appear. Double check and hit enter to start the bootstrapping.
4. The program will output 5 files and the window will remain open and will indicate the bootstrap progress. Be patient, this also takes a long time.
  - a. your\_output\_file\_name.r
    - i. null allele frequency estimation
  - b. your\_output\_file\_name.fr
    - i. allele frequencies, genotype numbers
  - c. your\_output\_file\_name.gFst
    - i. global F<sub>ST</sub> values
  - d. your\_output\_file\_name.pFst
    - i. F<sub>ST</sub> values for each population pair
  - e. your\_output\_file\_name.dc
    - i. genetic distance for each population pair
5. Open the .gFst file with a text editor and copy the By Locus table comparing Fst values with and without ENA (null allele) correction.
  - a. Paste into the GenAlEx Excel workbook.
  - b. Insert an XY scatter, add a trendline and display the equation and R squared value.
  - c. If the Fst values are tightly correlated, null alleles are not impacting your Fst calculations and you can use the raw dataset for additional analyses.
6. Repeat the same correlation analysis with the .pFst file. Since these values are arranged in a matrix, arrange them in a single column for without and with ENA correction.

**Note:** If the FreeNA correlation analysis revealed your Fst calculations are not being affected by null alleles, skip the following MicroChecker analyses and proceed to Arlequin with the raw dataset.

### MicroChecker 2.2.3 (Windows XP)

1. Select File, Open, and select the .xls file generated by GenAlEx.
  - a. This will load the first population in the dataset. Select Next Population in the bottom of the window to load each population until they are all loaded.
2. Set the Repeat Motif for each of the microsatellite loci.
  - a. MC29: Tetra
  - b. MC41: Tetra
  - c. MC49: Tri
  - d. MC46: Hexa
  - e. MC65: Hexa
  - f. MC97: Di-tetra
  - g. MC4: Tri
  - h. MC18: Tri
  - i. MC114: Tri
3. Select Analyze for each population, and Proceed with Analysis with the Suspect Data.
  - a. Once the analysis has been completed, select Compare null alleles across loci.
  - b. If null alleles have been identified in some of the markers, select Display estimates of allele frequencies.
  - c. For each marker that had null alleles, select Adjusted Genotypes.
    - i. Create a duplicate dataset from the MicroChecker import file.
    - ii. Sort the population by allele size from smallest to largest (this sorts missing data, or zeroes, to the top).
    - iii. Copy the adjusted genotypes from MicroChecker into the corrected Excel sheet.
  - d. Repeat this process with all the populations, so that all null alleles are corrected.
  - e. This will, of course, prevent the identification of multi-locus genotypes since the alleles are ordered by size in MicroChecker and therefore not associated with a particular sample anymore. However, this corrected dataset should be used going forward for all other population analyses besides the identification of clones and MLGs, and analyses in MLGsim.
    - i. Repeat GenAlEx steps 5-11 and 13 with the corrected data.
    - ii. Continue with the analyses below.

**Note:** If the FreeNA correlation analysis revealed your  $F_{st}$  calculations are not being affected by null alleles, run the following analyses with the raw dataset.

### Arlequin 3.5.2.2 (Windows 7)

1. The first time opening Arlequin, you have to specify the location of your text editor program and Rgui.exe under Arlequin Configuration.
2. GenAlEx will export the Arlequin data as a unix file. On a Mac, changing it to an Arlequin project file is as simple as adding the file extension .arp to the filename.
  - a. Then use Open Project to open the project.
3. Select Structure Editor and change the Group ID by geographical region (for example, NW Caribbean, NW Gulf of Mexico, NE Gulf of Mexico, SE. Gulf of Mexico).
4. Analyses to be tested are found under Settings.
  - a. For simple analyses, use the following:

- i. AMOVA
    1. Locus by locus AMOVA
    2. Include individual level
    3. Compute population specific FIS's
    4. Number of permutations: 16000
  - ii. Population Comparisons
    1. Compute Pairwise FST
  - iii. Hardy-Weinburg
    1. Perform exact test
  - iv. Pairwise Linkage
    1. Linkage disequilibrium test
    2. No. permutations: 16000
    3. No. initial conditions for EM: 5
  - v. Molecular Diversity Indices
    1. Standard diversity indices
      - a. Output sample allele freqs
      - b. Molecular diversity indices
5. Press Start and be patient as it can take a while.
6. The outputs will come as two or more files:
  - a. AllAlleleFreqs.txt found in the working directory.
  - b. The remaining analysis output is the .xml file in the .res folder of the working directory.
7. Copy the linkage disequilibrium values into a new table and save as a .csv for FDR correction.
  - a. Run the associated portion of the FDRtool.R script. Copy the FDR-corrected q values back into your GenAEx spreadsheet.
    - i. Double check that the q values make sense! Small datasets reduce the reliability of FDR correction!
8. Create a linkage disequilibrium matrix in the GenAEx Excel document with all the pairwise loci comparisons and populations.
  - a. Mark the all the paired loci per population (p value <0.05, violating assumptions of linkage disequilibrium).
  - b. Make sure no pairs of loci are linked across all the populations.
  - c. If you are concerned about any trends, rerun the linkage disequilibrium analysis with a dataset corrected for null alleles and check if the trend disappears.
  - d. Keep in mind that highly inbred populations (I'm talking about you, Pulley Ridge) will violate linkage disequilibrium across most if not all pairs of loci.

### **Structure 2.3.4 (Mac)**

1. Open Structure and Select New Project.
  - a. Give a name and set the directory.
  - b. Browse for data file; choose the .txt file generated from GenAEx.
  - c. Step 2, fill in:
    - i. Number of individuals
    - ii. Ploidy: 2
    - iii. Number of loci: 9
    - iv. Missing data value: -9

- d. Step 3, check:
      - i. Row of marker names
      - ii. Data file stores data for individuals in a single line
    - e. Step 4, check:
      - i. Individual ID for each individual
      - ii. Putative population origin for each individual
      - iii. USEPOPINFO selection flag
  2. Go to Parameter Set, New.
    - a. You will be running 10 replicate simulations for hypothesized numbers of populations (K) with Markov Chain-Monte Carlo burnin replications.
    - b. Run Length
      - i. Length of Burnin Period: 1000
      - ii. Number of MCMC Reps after Burnin: 1000000
    - c. Ancestry Model
      - i. Use Admixture Model
      - ii. Use sampling locations as prior (LOCPRIOR)
    - d. Allele Frequency Model
      - i. Allele Frequencies Correlated
    - e. Advanced
      - i. Compute probability of the data (for estimating K)
    - f. Name the Parameter Set; it should appear on the left panel.
  3. Select Project, Start a Job
    - a. Select the parameter set.
    - b. To set the range of K, start at 1, and determine the upper value by adding 3 to the number of sampling sites (10 sampling sites = maximum K of 13).
    - c. Set 10 iterations tests per value of K (200 tests total).
  4. Settle in; the analysis will run for a longgg time.
    - a. The results will display at the bottom of the screen, as well as automatically outputted to your working directory.
    - b. The value to look for is labeled “Estimated Ln Prob of Data,” describing  $\ln \Pr(X|K)$ , or the likelihood that the number of populations (K) is correct.
      - i. The value should become higher as you get closer to the actual K value.
  5. Navigate to the Results folder in your working directory and compress it to a .zip file. The folder should include the output files from all the simulations.

**Structure Harvester (<http://taylor0.biology.ucla.edu/structureHarvester/>)**

1. Upload the Results.zip file created from Step 6 of the Structure protocol, and select Harvest!
2. Once the results page loads, use the download link at the top to download the results page.
3. This package analyzes the variance among simulations for different levels of K, and changes in  $\ln P(D)$  values as you increase K. As a result, it gives a better predictor of the true number of source populations K, versus just looking at the highest value of  $\ln P(D)$  for the simulations in Structure.
  - a. The shift in model probabilities can be seen in the Delta K figure, with the corresponding likely K highlighted in the Evanno output table.
  - b. You can also save the Evanno output table as a tab-delimited table using the link above the table.



### **CLUMPP (Mac)**

1. Navigate to the Results folder you downloaded to the Structure Harvester directory.
2. Copy the .indfile and .popfile associated with the correct number of K identified by Structure Harvester (if K=2, copy K2.indfile and K2.popfile) to your CLUMPP directory.
3. Edit the paramfile with a text editor.
  - a. Datatype 0
  - b. Indfile K2.indfile
  - c. Popfile K2.popfile
  - d. Outfile K2.indivq
  - e. Miscfile K2.miscfile
    - i. K=2
    - ii. C=number of samples
    - iii. R=10
    - iv. M=1
    - v. W=1
    - vi. S=1
4. Open Terminal and change directory to your working directory.
5. Type ./CLUMPP paramfile to run the program.
  - a. If your particular population structure requires that a lot of combinations need to be tested, change the Method (M) from 1 to 2, set the Greedy\_Option to 2, and set the Repeats to 100,000.
6. Repeat step 3 for the .popfile, with the following changes:
  - a. Datatype 1
  - b. Outfile K2.popq
  - c. Miscfile K2.miscfile
    - i. C=number of populations
7. Rerun ./CLUMPP paramfile

### **Distruct (Mac)**

1. Copy the K2.indivq and .popq files to the Distruct working directory.
  - a. There are so many things to change for the drawparams, .names, .perm, and .languages files, you'll just have to look at the ones given with this protocol.
2. Open terminal, navigate to the Distruct working directory, and type './distructMac2013'
3. The .ps file created is your new combined structure plot. Preview will convert it to a .pdf

### **Spagedi (Windows 7)**

1. Open the Spagedi .exe file in your working directory.
2. Drag your input file into the executable window when it asks for the input filename.
3. Set the output filename including the extension.
4. Double check the input file format, then process with Enter.
5. For Level of Analyses, select 2 – Populations as categorical group.
6. Under Statistics, select 6 – global R-statistics and pairwise Rst (ANOVA approach).
7. Under Computational Options, select 3 – Make permutation tests.
8. Under Permutation Options, select 2 – Tests of mutation effect on genetic structure (permuting alleles).
9. Under additional options, select 1 – Report only P-values.
10. Enter 20000 permutations.
11. Under Output Options, select Enter.

12. Once the permutation tests complete, open the result file with Excel.
  - a. The genetic distance (Rst and pairwise Rst) are found in the Global R-statistics table under All Loci.
  - b. The corresponding p values are found in the Alleles Permutation Tests table under All Loci.

### **Bottleneck (Windows XP)**

1. Import the Genepop-format datafile using Add data file.
2. Select the options for the test.
  - a. TPM, SMM, Variance for TPM: 0.36, Proportion of SMM in TPM 0, 1000 iterations, Wilcoxon sign rank test, Mode-shift.
3. Select GO! Analysis will take some time.
4. Once the analysis has finished, save the results as a .txt file.
  - a. Also, select Summary, then save summary as a .txt file.

### **Poptree2 (Windows 7)**

1. Open the Genepop-formatted datafile using a text editor.
2. On the "Pop" lines, add a population name after a space. Don't use spaces in the pop name.
3. Import the datafile using Data Input.
4. Select Distance/Phylogeny with the following settings:
  - a. D<sub>A</sub>
  - b. NJ
  - c. Bootstrap 2000 replicates
5. Select Run Poptree
6. Once the tree is finished, you can switch the order of the branches by first selecting the root, and hitting the switch branch icon.
  - a. The length of the horizontal branches shows the number of base substitutions leading to differences among populations (use the given scale to determine differences).
  - b. The number at each node is the percentage of the bootstrap support – the closer it is to 100, the higher the support for that relationship.
7. To save the tree, use the copy to clipboard icon and save in Paint.
8. You can also save the D<sub>A</sub> genetic distance matrix using Output, Save. This may be good to compare to the Mantel test in GenAlEx.

**Note:** My experience is that this did not represent my genotypic patterns well.

### **GeneClass2 (Windows XP)**

1. Open the Genepop-formatted datafile using Open.
2. Choose Detection of first generation migrants, and  $L=L_{\text{home}} / L_{\text{max}}$ .
3. On the 2) Criteria for Computation tab, select Paetkau et al. from the Frequencies-based method, and set the Default frequency for missing allele to 0.01.
4. On the 3) Probability computation tab, select Enable probability computation (Monte Carlo resampling), select Paetkau et al.
  - a. Set the number of simulated individuals to 10000 (1000 is the default, 10000 is more precise but will take 10x longer to analyze).
  - b. Set the Type I alpha to 0.01.

5. Select Start.
6. Once completed, potential new migrants are labeled in red. The most likely population is indicated in green.
  - a. Save the results using Export.
7. Rerun the analysis with all settings the same, except for L=L\_home under Detection of first generation migrants.
8. Open the results .csv files in Excel, use Text to Columns to separate the data, then sort by probability.
  - a. Color the sample ID red for any probabilities under 0.01.
  - b. Compare all the source populations  $-\log(L)$  values for the F0 migrants. The lowest value is the correct source population and should be colored green.

### **Formatomatic (Mac)**

1. This is only used to convert a Genepop-formatted datafile into a Immanc/BayesAss datafile.
2. Copy the resulting file into the BayesAss working directory and use a text editor to add population names. Remove the first line starting with “%Immanc.”
  - a. Also add a .txt file extension.
  - b. Create subset datafiles containing the simple migration models to be tested. For example, to assess migration across depth ranges, create subset datasets within regions.

### **BayesAss3 (Mac)**

1. Open Terminal and change directory to the BayesAss working directory for each region.
2. Enter the following text (this is an example):
  - a. `./BA3 -v regional_datafile.txt`
    - i. Look at the first, third, and fourth values given for % accepted. They should range between 0.20-0.60
    - ii. If they are higher, add the following syntax to change the acceptance rate until they fall in the desired range at the end of the run:
    - iii. `-a0.50 -m0.25 -f0.75` (These are a good start)
      1. Notice the acceptance rates may increase over time.
  - b. Use something like following for a full run:
    - i. CBC: `./BA3 -v -m0.25 -a0.40 -f0.35 -t -s100 -i10000000 -b1000000 -n100 -o CBC_MLG1out.txt`  
`MS_Mcav_msats_MLG_CBC_BayesAss.txt`
    - ii. EFGB: `./BA3 -v -m0.25 -a0.35 -f0.45 -t -s100 -i10000000 -b1000000 -n100 -o EFGB_MLG1out.txt`  
`MS_Mcav_msats_MLG_EFGB_BayesAss.txt`
    - iii. PRTER: `./BA3 -v -m0.25 -a0.30 -f0.35 -t -s100 -i10000000 -b1000000 -n100 -o PRTER_MLG1out.txt`  
`MS_Mcav_msats_MLG_PRTER_BayesAss.txt`
    - iv. WFGB: `./BA3 -v -m0.20 -a0.45 -f0.55 -t -s100 -i10000000 -b1000000 -n100 -o WFGB_MLG1out.txt`  
`MS_Mcav_msats_MLG_WFGB_BayesAss.txt`
    - v. NWGOM: `./BA3 -v -m0.25 -a0.40 -f0.50 -t -s100 -i10000000 -b1000000 -n100 -o NWGOM_MLG1out.txt`  
`MS_Mcav_msats_MLG_NWGOM_BayesAss.txt`

- vi. GOM: ./BA3 -v -m0.10 -a0.25 -f0.25 -t -s100 -i100000000 -b10000000 -n100 -o GOM\_MLG1out.txt  
MS\_Mcav\_msats\_MLG\_GOM\_BayesAss.txt
  - c. Remember to rename the trace files after each run so they don't get overwritten.
  - d. Change the seed number (-s100) and the output file each time for several runs.
    - i. 951
    - ii. 190
    - iii. 563
    - iv. 235
3. Analyze the trace files in Tracer (below) to assess model convergence across 5 runs.

### **Tracer (Mac)**

1. Copy all the tracer files to the Tracer working directory.
2. Open all replicate files from each model by selecting the + button under Trace Files:
3. Look at the Trace plot for convergence.
  - a. Burn-in period is in gray, remaining iterations in black.
  - b. There will be an initial spike during burn-in, but the remaining plot should be fairly flat with oscillations. Large peaks or valleys indicate a failure to converge.
4. Once you have confirmed convergence across all the models, select the Estimates tab, highlight the Combined trace model, then highlight each of the migration rates (m[]) in the Traces box.
  - a. Export using File, Export Data and add a .txt extension.

### **MSA-Analyzer (Windows 7)**

1. This software is only used to convert the datafile to one that can be imported into Migrate.
2. Copy the program executable and your datafile to a local directory, like the desktop.
3. Run the executable and select I for InputFile.
  - a. Drag the file from the desktop to the Please enter the filename: line. This will also add the path.
4. Select c for data conversion settings.
  - a. Deselect all of the active settings by selecting the corresponding number option, except for Make Migrate/LAMARC, and reduce to stepsize.
  - b. Select ! to run.
  - c. A new folder will appear on the local directory. Copy it back to your working directory.
  - d. Under the Formats&Data directory, copy the .txt.migrate file to your Migrate working directory.
    - i. Remove the .migrate extension.

**Note:** You can run the Migrate-n analysis on your own computer, or on the FAU HPC. The HPC is much faster and you can run as many models as you want at a time. Both sets of instructions are given below.

### Migrate-n 3.6.11 (Mac)

1. Drag the migrateshell application from the migrate-3.6.11 folder to your Applications directory.
2. Likely you will be testing four models for each region if you are modeling vertical connectivity across two depths, or horizontal connectivity across 2-3 sites.
  - a. A – symmetric migration across all sites
  - b. B – unidirectional migration from first (deep or upstream) to last (shallow or downstream) sites
  - c. C – unidirectional migration from last (shallow or downstream) to first (deep or upstream) sites
  - d. D – panmixia
3. Create subdirectories within your Migrate working directory if you have multiple regions you are modeling.
  - a. Copy over the template infile and parmfileA-D from the GitHub repository.
4. Open the infile with a text editor and edit it with the pertinent data from the .txt file:
  - a. Depending on the models you want to test, you may need to remove populations. For example, if testing for migration across depth ranges, remove populations that do not have both depths.
    - i. Remember to change the corresponding population sizes and pop codes (0000\_) for each model set.
5. Change the name of the file to “infilePOPNAME” with no extension.
6. Open the parmfile with a text editor. There’s a lot of stuff you can mess up here, so pay attention. Change only the following:
  - a. Under General options
    - i. Menu=YES
  - b. Under Input options
    - i. Infile=infileSITENAME
  - c. Under Output options
    - i. Outfile=outfileSITENAME-MODELLETTER
    - ii. Pdf-outfile=outfileSITENAME-MODELLETTER
7. Hoo-boy. Figuring out the codes for the custom migration models can be challenging. For simple vertical connectivity models, it is relatively straightforward. You create a model matrix using Excel, with the To population as the vertical axis, and the From population as the horizontal axis. Asterisks are used to denote migration and zeroes are used to denote a lack of migration. Look at the template model matrix spreadsheet associated with this tutorial.
  - a. To enter these data into the parmfile, read the combination of \* and 0 from left to right and combine to a single line. For example, the matrix below would be represented in the parmfile as \*000\*\*00\*\*\*0\*\*\*\*\*

B				
To/From	WFGB	EFGB	BRT	MCG
WFGB	0	EFGB-WFGB	BRT-WFGB	MCG-WFGB
EFGB	WFGB-EFGB	0	BRT-EFGB	MCG-EFGB
BRT	WFGB-BRT	EFGB-BRT	0	MCG-BRT
MCG	WFGB-MCG	EFGB-MCG	BRT-MCG	0
To/From	WFGB	EFGB	BRT	MCG
WFGB	*	0	0	0
EFGB	*	*	0	0
BRT	*	*	*	0
MCG	*	*	*	*

- b. Enter the corresponding migration model code under custom-migration={ }
  - c. For model D, the custom migration model is a single asterisk \*.
8. Next, navigate to your working directory, then open the migrateshell app from the Applications directory. This starts running migrate in the correct working directory that includes your infile and parmfile.
9. To speed up processing time, we want to use fastmigrate-n that utilizes multiple cores in parallel. Type it in and select Enter to bring up the main menu.
10. For models A-C, you can start running the program by typing Yes or Y. Just remember to rewrite the custom migration model in the parmfile using a text editor.
11. For model D, you first need to change the population code mapping so that there is only a single population.
  - a. Select P to enter the parameters menu.
  - b. Select 7 to enter the Sampling localities menu
  - c. Enter the number of sample sites.
  - d. Enter the remapping as 1 1
  - e. Select Y to go back to the main menu, then select W to rewrite the parmfile.
12. Run all four models for each region. Model A will possibly take several days on 8 cores running in parallel, so be patient. The remaining models get progressively faster as they have fewer migration routes to estimate.
13. The output PDFs will appear in the working directory. Open each one, and find the model log likelihood under the “Log-Probability of the data given the model (marginal likelihood)” page. The value you need is the Bezier approximation score under All.
14. Copy each of model log likelihoods to the template migration model ranking spreadsheet.
15. Determine which model has the highest log likelihood, and subtract it from all the other model log likelihoods.
16. Find the model likelihood using the =EXP() function. The sum of the model likelihood should equal 1.
17. Find the model probability by dividing each model’s likelihood value by the sum.
18. Rank the models based on probability from highest to lowest log likelihoods.

### **Migrate-n 3.6.11 (KoKo HPC)**

1. Likely you will be testing four models for each region if you are modeling vertical connectivity across two depths, or horizontal connectivity across 2-3 sites.
  - a. A – symmetric migration across all sites
  - b. B – unidirectional migration from first (deep or upstream) to last (shallow or downstream) sites
  - c. C – unidirectional migration from last (shallow or downstream) to first (deep or upstream) sites
  - d. D – panmixia
2. Create subdirectories within your Migrate working directory if you have multiple regions you are modeling.
3. Copy over the template infile and parmfileA-D from the GitHub repository.
4. Open the infile with a text editor and edit it with the pertinent data from the .txt file:
  - a. Depending on the models you want to test, you may need to remove populations. For example, if testing for migration across depth ranges, remove populations that do not have both depths.
    - i. Remember to change the corresponding population sizes and pop codes (0000\_) for each model set.

5. Change the name of the file to “infilePOPNAME” with no extension.
6. Open the parmfile with a text editor. There’s a lot of stuff you can mess up here, so pay attention. Change only the following:
  - a. Under General options
    - i. Menu=NO
  - b. Under Input options
    - i. Infile=infileSITENAME
  - c. Under Output options
    - i. Outfile=outfileSITENAME-MODELLETTER
    - ii. Pdf-outfile=outfileSITENAME-MODELLETTER
7. Hoo-boy. Figuring out the codes for the custom migration models can be challenging. For simple vertical connectivity models, it is relatively straightforward. You create a model matrix using Excel, with the To population as the vertical axis, and the From population as the horizontal axis. Asterisks are used to denote migration and zeroes are used to denote a lack of migration. Look at the template model matrix spreadsheet associated with this tutorial.
  - a. To enter these data into the parmfile, read the combination of \* and 0 from left to right and combine to a single line. For example, the matrix below would be represented in the parmfile as \*000\*\*00\*\*\*0\*\*\*\*\*

<b>B</b>				
<b>To/From</b>	<b>WFGB</b>	<b>EFGB</b>	<b>BRT</b>	<b>MCG</b>
<b>WFGB</b>	0	EFGB-WFGB	BRT-WFGB	MCG-WFGB
<b>EFGB</b>	WFGB-EFGB	0	BRT-EFGB	MCG-EFGB
<b>BRT</b>	WFGB-BRT	EFGB-BRT	0	MCG-BRT
<b>MCG</b>	WFGB-MCG	EFGB-MCG	BRT-MCG	0
<b>To/From</b>	<b>WFGB</b>	<b>EFGB</b>	<b>BRT</b>	<b>MCG</b>
<b>WFGB</b>	*	0	0	0
<b>EFGB</b>	*	*	0	0
<b>BRT</b>	*	*	*	0
<b>MCG</b>	*	*	*	*

  - b. Enter the corresponding migration model code under custom-migration={ }
  - c. For model D, the custom migration model is a single asterisk \*.
8. Duplicate the parmfile so there are four versions and rename each parmfileA-D.
  - a. Change the outfile file names in each parmfile corresponding to the correct model.
  - b. Change the migration model matrix for each parmfile.
  - c. For model D, change the population-relabel={ 1, 1, 1, 1 } (or however many populations you have)
9. Follow the instructions in microsat\_migrate\_README.txt for the HPC steps.
10. The output PDFs will appear in the working directory. Open each one, and find the model log likelihood under the “Log-Probability of the data given the model (marginal likelihood)” page. The value you need is the Bezier approximation score under All.
11. Copy each of model log likelihoods to the template migration model ranking spreadsheet.
12. Determine which model has the highest log likelihood, and subtract it from all the other model log likelihoods.
13. Find the model likelihood using the =EXP() function. The sum of the model likelihood should equal 1.
14. Find the model probability by dividing each model’s likelihood value by the sum.
15. Rank the models based on probability from highest to lowest log likelihoods.