**Computer Lab 12 – Population Structure**

**Conservation Genetics (BIOL 4174 / 5174)**

Due date: 4/15/2014 (Graduates) or 4/17/2014 (Undergraduates) at Midnight.

Submit this file via Blackboard.

Part I – Structure

1. Which model (Admixture or No Admixture) seems appropriate given your background knowledge of the populations being studied?
2. Your lab\_12 folder contains the “msat-data.xlsx” file from the previous two labs. A sheet containing graphs of allele frequencies has already been generated for you within this file. Study these frequencies. Based on the similarities/differences among populations, which allele frequency model would you select? (correlated or independent).
3. Provide your bar plots for *K*=2, *K*=3, *K*=4, and *K*=5 here. For each plot, group samples by POP Id. You can save images of each plot by clicking the “Save” button in the bar plot window, and then insert the saved files into your homework document.
4. Briefly describe each plot. Discuss if there are any individuals with ancestry derived from multiple populations. Are there any sampling locations that have been split among multiple populations?
5. Overall, do you see any sampling locations that have consistently clustered together across all four *K* values? Do any sampling locations cluster together in most of the plots? How does this compare to your PCA results from Lab 11?
6. Insert the time series plot of alpha for one of your runs. State whether or not convergence occurred before the end of the burnin period.
7. Insert the triangle plot for *K*=3. Pick two of the original six sampling locations. Provide the color representing each of these two sampling locations in the plot. Based on the triangle plot, describe in your own words how the samples collected at these localities have been assigned (i.e., unambiguous assignment to a cluster, partial assignment to clusters 1 and 2, etc.).
8. A problem that is commonly encountered in Structure is the possibility of converging on different answers for the same *K* value. Run the “exploratory” parameter set at *K*=3 for 3 iterations. When you get a bar plot that differs from your previous run at *K*=3, insert it into your homework document and stop the analysis (in other words, you don’t need to run all three iterations if on your first iteration you receive a different answer than you had previously). Make sure your plot is “grouped by POP Id.”
9. Describe how this bar plot differs from the previous run at *K*=3.
10. Why is it possible for Structure to find different clusters for the same *K* value?

Part II – Structure Harvester

1. Explore the results in Structure. Do independent runs at *K*=2 still converge on multiple different answers? What about *K*=3?
2. Control+click this image to bring up the shortcut menu. Copy this image and paste it into your homework document. Based on this graph, how many populations are represented by the data?
3. Copy and paste the Delta*K* plot into your homework document. Based on this graph, how many populations are represented by the data? Does the Delta*K* plot have a secondary peak at any other *K* value? If so, provide that value as well.
4. Go back to Structure. Copy and paste the bar plot that corresponds to the *K* value selected by the L(*K*) plot and the Delta*K* plot. It is only necessary to copy one of each, since independent runs converged on the same results for these *K* values.
5. Compare the two plots to each other, and to the background information provided on the first page of this lab. Which *K* value seems to make more sense biologically given this information? Explain why.
6. Look at the bar plot for *K*=5, specifically the DES population. How would you interpret what is happening here?