**P/BIO 381 Spring 2017**

**Assignment #3: Population genomic diversity and structure**

Your assignment is to assess the sensitivity of our inferences of population genomic structure in our SSW data for two different SNP filtering strategies (not including the one we’ve used in the tutorials). You may want to consider some, though not necessarily all, of the following variables for filtering the SNP data:

* presence of multiple alleles/locus (--min-alleles; --max-alleles)
* read depth (--minDP)
* site missingness (--max-missing)
* minor allele frequency (--maf)
* deviation from Hardy Weinberg equilibrium (--hwe)
* removing individuals with large amounts of missing data (--remove-indv)

You should then analyze the 2 resulting datasets with one of the 3 population genomic techniques we’ve covered: PCA, DAPC, or ADMIXTURE.

Please use 2 pages maximum to demonstrate your understanding of the conceptual background and technical details for using SNPs derived from RNA sequencing to analyze population diversity and structure. You should include relevant tables or figures (within the two-page limit) with legends.

* Clear statement of objective (1 sentence).
* Conceptual background on what the analysis does (2-3 sentences).
* Verbal description of the mechanics of the pipeline (3-4 sentences).
* Present results (3-5 sentences).
* Tables and figures with legends.
* Interpretation (3-5 sentences). ***Critique the filtering strategies and their effects on your inference;*** ***place your interpretation of the structure of the SSW data into context based on what you know of their natural history and the sampling design.***
* Critical thinking (2-3 sentences). What would you do differently? What would you do next?
* Include a link to your code on github!

You may discuss the assignment with classmates, but the assignment should be prepared individually. Due **Wednesday, April 5th**.