**Prep-work:**

***Install software***:

We will need up-to-date versions of **R, Rstudio**, **STRUCTURE, GENECLASS, CLUMPP**,and **Distruct** installed for the workshop. You should install all of these on your machine, in roughly that order.

**R:** The actual programming language! R can be installed via the links (or instructions in the case of a Linux machine) on [this page](https://cran.rstudio.com/).

**Rstudio:** An Integrated Development Environment (IDE) designed specifically for R, but that also works for python, bash, and a handful of other languages. IDEs make coding *much* more comfortable. They are incredibly useful--I would never try to do any serious coding without one. You can download Rstudio [here](http://www.rstudio.com/products/rstudio/download/). The free version is totally fine.

**STRUCTURE:** STRUCTURE is a tool for finding and visualizing population structure that works with both microsatellite, RAD or whole-genome SNP data, and quite a few other types of data as well! **We need to install both the command line and graphical user interface (GUI) version of the program.** Both are available [here](https://web.stanford.edu/group/pritchardlab/structure_software/release_versions/v2.3.4/html/structure.html). Download the versions both with and without the graphical front-end for your operating system.

* Run the installer for the graphical version and follow the prompts to install STRUCTURE.
* Unzip the zipped file for the version without the graphical front end—no need to install!

**Geneclass**: Geneclass is a program for assigning individuals to their population-of-origin. The program is available in both Windows and Unix software, and is available [here](https://gaow.github.io/genetic-analysis-software/g/geneclass/).

**CLUMPP**: CLUMPP is a command-line program that combines information from replicate STRUCTURE runs (e.g., with the same parameters, but different starting seeds) that can then be used to make publication-quality barplots. This program creates output files that are then used to make the actual plot in the program Distruct (mentioned below). CLUMPP has great documentation and example files, and can be downloaded [here](https://rosenberglab.stanford.edu/clumpp.html).

**Distruct**: Distruct is a command-line program used to produce publication-ready barplot figures that display results from the program STRUCTURE. The program, example files, and documentation can be downloaded [here](https://rosenberglab.stanford.edu/distruct.html).

***Linux extra steps***:

If you have a linux machine, you'll probably also need to install a handful of other tools: r-base-dev, libxml2-dev, and libcurl4-openssl-dev. The install method will depend on your distro, so you'll need to google it. I used

sudo apt-get install r-base-dev

sudo apt-get install libxml2-dev

sudo apt-get install libcurl4-openssl-dev

on my up-to-date Ubuntu distro without issue, but your mileage may vary. These are needed to install some of the packages we'll be using in R, and there may be a few others you'll end up needing as well.

**Schedule:**

**Saturday, July 22nd (08:30 – 12:00): Introduction to R and overview of genetic data types**

* Overview of genetic data types
* Introduction to R and Rstudio
* File structures and reading in data
* Objects, classes, object access, and vectorized operations
* Logical operators and subsetting
* Functions and getting help
* Overview of genetics tools in R (as time permits)

**Saturday, July 22nd (13:00 – 17:00): Inferring population structure**

* PCA, DAPC, and STRUCTURE – wolf microsatellite data
  + PCA in Adegenet
  + DAPC in Adegenet
  + STRUCTURE via graphical user interface (GUI)
* PCA, DAPC, and STRUCTURE – monarch butterfly RAD data
  + PCA in snpR
  + DAPC in Adegenet
  + STRUCTURE via snpR
  + Phylogeny via snpR and ape

**Sunday, July 23nd (08:30 – 12:00): Genetic diversity estimation**

* Concepts and background – wolves and bighorn microsatellite data
  + Genetic diversity estimation in GENEPOP online
    - Heterozygosity, allelic diversity, linkage disequilibrium
* Concepts and background – monarch butterfly RAD data
  + Genetic diversity estimation in snpR
    - *F*IS, *F*ST, site frequency spectra, Tajima’s *D*

**Sunday, July 23nd (13:00 – 16:00): Inferring population connectivity and migration**

* Population division – wolves and microsatellite data
  + *F*ST in GENEPOP
* Population division – monarch butterfly RAD data
  + *F*ST in snpR
* Assignment tests – wolves and microsatellite data
  + Geneclass
* Assignment tests – monarch butterfly RADdata
  + Ranger (R package)

Looking forward to seeing all of you on Monday!

Best,

Will Hemstrom (on behalf of Jared Grummer and Gordon Luikart)