

Lab 3: Structure

Nora Mitchell

February 2015

Contents

1	Getting Started	1
1.1	Installation	1
1.2	Mac Attack	1
2	Running an Analysis	2
2.1	Creating a Project	2
2.2	Parameter Set	3
2.3	Running the analysis	3
3	Looking at Results	3
3.1	Structure Results	3
3.2	Structure Harvester	4
4	References	4

1 Getting Started

Structure is a software package from Pritchard et al. (2000) that uses multi-locus genotype data and MCMC to perform individual assignment for population genetics analysis.

1.1 Installation

Please install Structure v2.3.4 from here http://pritchardlab.stanford.edu/structure_software/release_versions/v2.3.4/html/structure.html

1.2 Mac Attack

If you are installing Structure on a new Mac OS X, Kent has this advice: Running Structure on recent versions of Mac OS X: When you download Structure and try to run it on a

recent version of Mac OS X, you may encounter an error message saying that the file is corrupted and that the disk image it's on should be ejected. Don't worry. That message is misleading. What it really means is that you've downloaded an executable from a developer who hasn't registered with Apple, and you've run afoul of the enhanced security associated with Gatekeeper.

Here's how you get around it.

- Bring up "System preferences."
- Click on "Security and privacy" and make sure that "General" is selected.
- At the bottom of the panel you'll see three check boxes under "Allow apps downloaded from": (1) Mac App Store, (2) Mac App Store and identified developers, and (3) Anywhere. If you've run into the error, you almost certainly have either the first or the second button selected. Select "Anywhere" instead, copy Structure to your Applications folder (or someplace else that's convenient), and run it. Once you've run it once, you should be able to return your security settings to the way you had them before. If not, then just remember to change them before you try running Structure and change them back when you're finished.

2 Running an Analysis

2.1 Creating a Project

Open up Structure and go to **File >New Project**, which will open up a new window.

- Step 1: Structure will ask you to name the project, select a directory to navigate to (find the folder where you've stored your data file), and then to choose the data file. When you've done this, click "Next".
- Step 2: Fill in the number of individuals, ploidy, number of loci, and missing data values (typically "-9", for whatever reason). If you can't remember what the format looks like, click "Show data file format" which will show you the number of lines and columns. Click "Next".
- Step 3: Now pick the format of the data set. Check any that apply. For Project 2, Kent has given you a hint for this section. Click "Next".
- Step 4: More format input! Click those that apply. It should be evident from the format of your data. Click "Finish".

You will reach a confirmation window with everything you entered. If it checks out, hit "Proceed."

Now in the left-hand portion, you'll see a folder with your project, and the main window will have your Project Data. Make sure it looks okay!

2.2 Parameter Set

Now you need to create a parameter set for your MCMC settings.

Go to **Parameter Set >New...**

~

Enter your desired reps for the burnin and post-burnin MCMC reps. These numbers will depend on how complicated your data is. You can also navigate the tabs to adjust other settings. Click "OK" when done. It will ask you to name the parameter set. For this project, please name it "LastName" so I can easily compile the results and don't get duplicate names.

You will now have a Parameter Sets folder in the left hand portion, and the main window will have "Simulation Configuration-Last Name" and will list your settings.

2.3 Running the analysis

Go to **Project >Start a Job**. A window called "Structure Scheduler" will now open. Make sure to highlight your parameter set name (click on it), then adjust your K settings from 1 to the desired high number. Click "Start."

Now a Structure Job Log window will open, and the bottom portion of the screen will show you the status/reps that Structure is going through. It will scroll very quickly through the burnin and MCMC reps for each K-value.

You'll notice a "Results" subfolder in your "Parameter Sets" folder on the lefthand side, which will have new results for each K-value it runs through. It will be naming these "LastName_run_1 (K=1)" etc.

Depending on the size of the dataset and your iterations, the analysis could take just a few minutes up to several hours. You may have to let it run overnight—just make sure the computer does not turn off or go to sleep. When it's done, you'll get a pop-up window that says "Job is Completed!"

3 Looking at Results

3.1 Structure Results

You can look at some valuable plots within Structure. In the lefthand portion, you can click on any of the runs and access it's simulation results in the main window. You can then explore by looking at things like **Bar plot >Show** and play with the settings to see the

individual assignment results for that specific K-value. I like to “Group by POP Id” which delineates the original populations that you specified from the dataset.

You can also explore data plots, histograms, triangle plots, and tree plots in a similar way.

3.2 Structure Harvester

In terms of choosing K-values, the direct Structure output is not enough, since it is the result of a single run for each K-value. You should run this analysis many times (10+) and then upload the results as a .zip file to Structure Harvester to use Evanno et al.’s (2005) method for choosing K. Your results files will be stored in your original directory in a subfolder named “LastName”, then in a subfolder called “Results”. You can easily zip these files for use in Structure Harvester.

<http://taylor0.biology.ucla.edu/structureHarvester/>

Structure Harvester is very easy to use, and is all web-based! You simply upload your zip file and then click “Harvest!” It may take a few minutes to run.

The program will then give you several plots, including one for L(K), DeltaK, and .csv files with actual values for each of these. It’s up to you to decide what output you want and how to interpret it!

4 References

- Dent, A., and vonHoldt, B.M. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* 4(2):359-361.
- Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* 14:2611-2620.
- Matesanz, S., K. E. Theiss, K. E. Holsinger, and S. E. Sultan. 2014. Genetic Diversity and Population Structure in *Polygonum cespitosum*: Insights to an Ongoing Plant Invasion. *PLoS One* 9:e93217.
- Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155:945-959.