

## Fastsimcoal Activity

Welcome to the Fastsimcoal Quarantine Thunderdome! Many teams will enter, but who will successfully infer the correct demographic model and parameters?? Bond with your teammates, choose your team name, fire up the command line, and get ready to rumble.

I have simulated a **folded site frequency spectrum** based on demographic parameters that only I know. 6Mb of neutral sequence was generated for 10 diploid individuals. You suspect this population may have had some changes in population size, but don't know for sure. Your task is to try to figure out:

- a) **Which demographic model (from 1-Epoch, 2-Epoch, and 3-Epoch) best fits the data**
- b) **Infer the maximum likelihood estimate (MLE) parameters for that model**

### *Instructions:*

0. Choose a worthy team name.
1. Download the activity folder from Github (or copy paste the "raw" scripts)  
<https://github.com/ab08028/DemographyTutorial/tree/master/fastsimcoalActivity>
2. Look at the SFS to make sure you understand the format  
[https://github.com/ab08028/DemographyTutorial/blob/master/fastsimcoalActivity/SFS/generic.SFSForDemographyWorkshop\\_MAFpop0.obs](https://github.com/ab08028/DemographyTutorial/blob/master/fastsimcoalActivity/SFS/generic.SFSForDemographyWorkshop_MAFpop0.obs)
3. Look through the three .tpl and .est files provided to make sure you understand each model  
<https://github.com/ab08028/DemographyTutorial/tree/master/fastsimcoalActivity/modelFiles>
4. Update the fastsimcoal wrapper script to use paths that are appropriate for your set up (give it the path to your copy of fastsimcoal and to wherever the fastsimcoalActivity directory is). Make sure you understand what the wrapper is going to do. For the sake of speed, I've lowered the number of ECM cycles and coalescent trees simulated. Note you'd want to increase those numbers if you were doing this for real.  
[https://github.com/ab08028/DemographyTutorial/blob/master/fastsimcoalActivity/step\\_1\\_fscWrapper.Demo.3Models.sh](https://github.com/ab08028/DemographyTutorial/blob/master/fastsimcoalActivity/step_1_fscWrapper.Demo.3Models.sh)

5. Use the fastsimcoal wrapper script to run 10 replicates of each model. You can run them as a group, or have each member run a different model, or each person run 10 replicates each model to have more replicates overall (more replicates will improve your chances of getting convergence). It should take ~20 mins to run all the replicates for all the models. Using the full set of cycles/trees it would take ~2 hours.
6. Once you've done all your runs, update the concatenation script to have the appropriate path to wherever the fastsimcoalActivity directory is on your computer/sever. Use the concatenation script to gather up all your replicates for each model.  
**[https://github.com/ab08028/DemographyTutorial/blob/master/fastsimcoalActivity/step\\_2\\_concatenateFSC2.Results.sh](https://github.com/ab08028/DemographyTutorial/blob/master/fastsimcoalActivity/step_2_concatenateFSC2.Results.sh)**
7. Copy paste the concatenated results into Excel or Google Sheets. I recommend using a separate tab for each model. If each person ran each model, you will need to combine your concatenated files from all the team members.
8. Sort your data by log-likelihood (LL) (be sure it's the Max Estimated Log-Likelihood) so that the largest LL's are at the top of your spreadsheet. (Use the automatic sort function)
9. Look for convergence. Does the log-likelihood change a lot between the top few runs? Do the parameters jump around a lot?
10. Choose the maximum likelihood estimate (MLE) for each of your models (should be the run with the highest LL), convert the population sizes from haploids → diploids (divide by 2; you do not need to make any conversion to the times). Enter each model's MLE in the activity summary sheet's tabs (choose the correct tab for each model).  
**[https://docs.google.com/spreadsheets/d/1vP4VvAaLm5GNq\\_3xJrMB0uiJ5oMzgaE8f9I34NLMr68/edit?usp=sharing](https://docs.google.com/spreadsheets/d/1vP4VvAaLm5GNq_3xJrMB0uiJ5oMzgaE8f9I34NLMr68/edit?usp=sharing)**
11. Calculate the Akaike Information Criterion for each model's MLE. ( $AIC = 2k - 2LL$ , where  $k$  is the number of estimated parameters.) Enter it in the activity sheet.
12. Decide which model you want to choose as the best-fit model. Enter it into the main summary page remember to convert haploids → diploids.

**[https://docs.google.com/spreadsheets/d/1vP4VvAaLm5GNq\\_3xJrMB0uiJ5oMzgaE8f9I34NLMr68/edit?usp=sharing](https://docs.google.com/spreadsheets/d/1vP4VvAaLm5GNq_3xJrMB0uiJ5oMzgaE8f9I34NLMr68/edit?usp=sharing)**

13. The group will reconvene and the simulation parameters will be revealed! Whichever group gets closest wins everlasting quarantine glory.

*Other notes:*

- **Remember to update your wrapper script and concatenation script with the appropriate paths for your computer/server. You need to provide the path to wherever fastsimcoal26 is installed on your computer, and a path to wherever you have put the fastsimcoalActivity directory. You shouldn't need to alter the model .tpl or .est files**
- **When entering your results, be careful not to overwrite anyone else's work!**
- **You can ask me for help in the Zoom chat or by requesting help in your break out group.**