

Seminar Response

This is the most substantial critique I've read in this class—ever! Yours is a serious evaluation of the issues and written very constructively. You have a real future as a reviewer!

On November 15, 2010 Liz Hadly gave two seminar presentations on the program SerialSimCoal (SSC) and the research she did using the program. While she didn't directly cover population genetic concepts in her afternoon seminar for the department she did cover these in the earlier seminar; which I will focus on for this response. Hadly explained that there are generally two categorical methods to reconstruct a population history, frequentist and Bayesian; SSC incorporates the Bayesian method to make reconstructions of molecular histories, this Bayesian method views all the data as real and the mean as abstract, and seeks to utilize all other known facts surrounding the data for analysis. In SSC, the user inputs their data collection parameters (not their data) and the program creates simulated data sets from those parameters. Unknown parameters in the history are then manipulated until the simulated data matches the true data. The main idea is that when the simulated data matches the true data then the parameters needed to create that in the model (bottleneck, drift, population size shifts, etc) are what happened in reality to create the changes in genetic structure that we see in the data. Using these models SSC attempts to present the genetic history that is no longer present when we look at a contemporary population. *Nice explanation!*

I am personally impressed with the fundamental concepts behind this program; finding ways to make (educated) hypothesis when addressing a question with missing data like ancient DNA studies. Hadly gave an example of her study where they matched their data by entering a huge bottleneck into the program during a certain year; when the data matched they investigated further and found that a volcano erupted that year in another country and decided that the population had been affected by that eruption. While initially this conclusion seems logical, this specific example also gives me pause. It seems that SSC might force some heavily subjective evaluations of significance in regards to events or forces acting on the data. While a volcano did erupt somewhere in the same year that the population may have had a bottleneck that factor could be coincidental, and a much smaller change might have forced the bottleneck – or a large but undocumented change – or perhaps SSC is incorrect and the population bottlenecked the year before the volcano erupted. Also, we must remember that the idea that the population history was created by a bottleneck is a theory in itself (of simulated data in SSC) so now we have our analysis based on 2 theories that are interrelated. Hadly did address this a little when she discussed the incorporation of isotope data into the SSC analysis. The group has been using spatial isotope analysis to determine population ranges and other data to mix with the data that SSC produces. However, this argument creates a few problems for me. First, isotope analysis is based on heterozygosity in the data. Analysis of isotope data from homogenous landscapes or migratory species could create inaccurate pictures. Additionally, the reconstruction of an ecosystem based on isotope data also seems subjective and based heavily on theoretical and model approaches. Using one relatively new and subjective science to provide rigor for another relatively new and subjective science does not seem that plausible to me. I did ask Hadly about this briefly and she acknowledged some of the issues with the infancy of this program. She indicated that currently they working through this by employing many scientists in the independent evaluation of data, and using the peer review process to check assumptions. However, science is not a democracy; it is, ideally, based upon facts not consensus. Not too long ago one could have gotten most scientists to agree that the world is flat, but it is not.

While there are still some issues with SSC it is not my purpose to argue it is completely useless. I think the program has great potential and is actually one of the only tools we have at this time to look at the molecular histories of populations. The current disadvantages and weaknesses of SSC can be overcome in several ways. On a programming level, as more fossil and molecular data is acquired this can be incorporated into the logic of SSC which will improve the rigor of the program in itself. On an analysis level, a cautious and diligent approach to analysis can help prevent overreaching conclusions. When using SSC, the question should be meticulously defined as should the definition of "population" for the data parameters. Working to carefully define these bounds will help create clear assumptions that can be validated for analysis. Also, I think SSC should only be used in conjunction with other analysis to create more checks on the data. At this point, SSC alone does not seem to have enough power to verify it is not violating the basic assumptions needed to analyze population structures.

*This is a big drawback of Bayesian analyses—your conclusions depend on your "priors"!*