Ivy Mead

21 April 2011

Bioinformatics Critique 2

I was particularly interested in the how the review explored the different genome-wide tests for finding selection in humans. This strikes me not only as relevant to better understand human evolution, but also as a particularly difficult question to tackle. The human genome is very large compared to the bacterial and viral genomes studied in similar papers, thus it is vastly more complicated and more difficult to sequence at a whole genome scale. As the review notes, because of the complexity, tests relying on snapshots of the genome (such as SNPs) are prone to particular sources of error. So, devising accurate models for the entire genome is necessary.

Because the human population as a whole has spread so far around the world, some tests would have to be calibrated to take into account recent genetic differences between well-established populations. This would not be as large of an issue in many of the other species being evaluated in genome-wide studies, because the species’ often do not have as large of a genome and the same history. Another complicating issue is that recently, with the event of global travel, human evolution and history is about to get much more complicated due to the amount of movement of alleles. It is too soon for that to have a huge effect, but in the future it could change the way in which analysis of human evolution needs to be done, as well as how selective sweeps are resolved.

Another interesting question is how to calibrate the tests to differentiate between recent selection on certain genes from selection in the far past in our ancestors’ genes. It is important that many of the selective sweeps be caught early enough or the signal can fade to the point where the frequency is not high enough and selection cannot be verified as acting on a gene that it had shaped in the past. This will likely be a problem in all studies, regardless of the species being studied. It is also not likely a problem that can be fixed because once enough evidence fades there will be no way to bring it back without introducing more error into the study. One study mentioned in the review found that 1.6% genes in the human genome are presently going through selective sweeps, and that more than one population shares those particular genes. It is possible that that percentage is slightly higher than they found because the genes not showing up in the study were not as strongly selected for or are not showing up in the analyses for other reasons that are not yet known.

The tests based on site frequency spectrum (SFS) could be another way of better understanding the role of selective sweep in humans because it can detect a bias formed in the distribution of alleles so that there are more alleles at lower frequencies. Thus it may be a way to compensate for the downsides in afore mentioned haplotype scans. As a result of the SFS analysis it was found that there may be an increased number of selective sweeps in the genes surrounding the centromeres, which is an exciting result that is not easily explained.

Another way in which selective sweep in humans can be studied is to look at differences in allele frequencies concerning different populations of humans. The review mentions that selection present in a certain population may be due to environmental pressures or due to splitting of a population into separate groups. This makes sense when migration patterns of humans are taken into account. The frequency of the allele in question would be higher in a population that did not move because they stayed in a more steady state while those that moved may have lost that allele through adoption of new ones in populations in the new area moved to. Thus, human history could be very important when looking at evolutionary questions in the human race.