Statistical Methods for the Detection of Positive Selection using Genome Maps

David M. Ramsey[1]

[1] Department of Mathematics and Statistics, University of Limerick, Limerick, IRELAND

The genetic makeup of a population reflects its demographic history and the effects of selection. An expanding population will show less genetic variation than a population of the same, but constant, size. This effect is visible across the genome as a whole. A selective sweep occurs when a positive mutation spreads through a population. Neutral alleles which are strongly linked to this mutation also spread through the population. This process is known as hitchhiking and leads to low genetic variation around the site of the mutation. Classical tests, such as Tajima's [4], are based on genetic information from a small section of the genome and thus cannot differentiate between the effects of demography and positive selection.

Advances in the mapping of single nucleotide polymorphisms (SNPs) enable us to differentiate between the local action of positive selection and the genome wide effect of demographic changes. Nielsen et al. [2] propose a test for detecting selective sweeps by looking for appropriate changes in genetic variation from the background level. Such a test has been shown to be powerful. However, the level of mutation varies across the genome and without comparing with related populations or subpopulations it may be unclear whether a very low level of variation in a section of the genome indicates the positive selection has occurred or that there is a low mutation rate (possibly mutations do not appear due to strong purifying selection).

It has been observed that the human subpopulations, which migrated relatively recently in evolutionary time from Africa, will have been subject to different selective pressures in their new environments and so we should observe the effects of selective waves in the European and Asian subpopulations. Schlötterer [3] presents a test to detect such selective waves based on a comparison of repeat numbers in the test population and an ancestral population. Such a test is robust to variation in the mutation rate over the genome.

Kimura et al. [1] propose a simple method based on SNP maps, which detect areas of low variance in a test population and compare these regions with an ancestral population. We present an approach which combines these approaches of genome scanning and comparison, in order to give a firmer mathematical foundation for comparing the SNP maps of two populations.

References

[1] Kimura R., Fujimoto A., Tokunaga K., Ohashi J. (2007) A practical genome scan for population-specific strong selective sweeps that have reached fixation. PLos ONE 2(3); e286. doi:10.1371/journal.pone.0000286.

[2] Nielsen R., Williamson S., Kim Y., Hubisz M. J., Clark A. G. and Bustamente C. (2005) Genomic scans for recently sweeps using SNP data. Genome Research, 15:1566-1575.

[3] Schlötterer C. (2002) A microsatellite-based multilocus screen for the identification of local selective sweeps. Genetics, 160:753-763.

[4] Tajima F. (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics, 123:585-595