

**Additional Paper for “Human Evolution”****Paper reference**

“How reliable are empirical genomic scans for selective sweeps?” Kosuke M. Teshima, Graham Coop and Molly Przeworski; *Genome Res.* 2006 16: 702-712;

**Abstract**

There is currently great interest in using scans of variability across the genome, to look for patterns that suggest a recent selective sweep. Studying such patterns could lead to the identification of mutations that have been favored by selection. Empirical approaches, where candidate loci that have been subjected to positive selection are identified to be in the extreme tail of empirical distributions, are increasingly being used for such studies. In this paper, the authors present an analysis of what empirical genomic scans for selective sweeps are and are not likely to find. They focus on applications to human data and to maize.

This paper suggests that the reliability and performance of empirical approaches depends on a number of factors such as population demographic history, whether selection acts upon a newly arisen advantageous allele or standing variation and the degree of dominance of the advantageous allele.

The authors simulated data sets for the evolution of modern humans and the domestication of maize. They made their data realistic by including bottlenecks and constant population sizes; strong selection on previously neutral alleles compared with weaker selection on new mutations. They found that identification of genuine selections worked best when the selection acted on new mutations in a population of constant size.

One subject of study was the non-sub-Saharan African human population, who are believed to have originated from ancestral African populations and only recently became cosmopolitan. They must have adapted to their new environments and it should have been possible to see which regions of the genome show evidence of selective sweeps. However, they have also gone through bottlenecks of small population size, which has similar effects to sweeps, but are distributed over the whole genome. As opposed to a model of constant population, the population bottleneck reduces polymorphism - confounding the observable effects of selection at surrounding sites.

Another subject of study was maize and the effects were better observed in this case. The results confirmed that if selection commonly acts on previously neutral mutations at appreciable frequency in the population, targets of selection are harder to distinguish from neutrally evolving loci.

**Discussion**

We studied in class that statistical tests can be applied to empirical genome-wide data sets to find selective sweeps. However, the results from this paper show that the reliability of empirical approaches to assessing genomic data sets depends on demography and the nature and history of the allele being selected. While empirical scans detect many good candidate sites for selection, they also miss many. They will miss recessive alleles and alleles that were selected after having previously been neutral. The false discovery rate is also higher when the population has experienced a bottleneck rather than maintained a constant size.

The findings of this paper would be useful to anyone who is using empirical approaches to come up with theories about the genetics of adaptation or about the types of mutations that selection acts on.