

Detecting selection using Extended Haplotype Homozygosity-based statistics on unphased or unpolarized data

Alexander Klassmann¹ and mathieu gautier²

¹University of Cologne

²INRAE

October 30, 2020

Abstract

Analysis of population genetic data often includes the search for genomic regions with signs of recent positive selection. One of the approaches involves the concept of Extended Haplotype Homozygosity (EHH) and its associated statistics. These statistics typically need phased haplotypes and, some of them, polarized variants. Here, we unify and extend previously proposed modifications to loosen these requirements. We compare the modified versions with the original ones by measuring the False Discovery Rate on simulated whole-genome scans and quantifying the overlap of inferred candidate regions on empirical data. We find that phasing information is indispensable for the accurate estimation of within-population statistics for all but very large samples and of cross-population statistics for small samples. Ancestry information, in contrast, is of lesser importance for both. Our publicly available R package `rehh` incorporates the modified statistics presented here.

Hosted file

`rehh3_v14.pdf` available at <https://authorea.com/users/308075/articles/489834-detecting-selection-using-extended-haplotype-homozygosity-based-statistics-on-unphased-or-unpolarized-data>