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

Alexander Klassmann[1], Renaud Vitalis[2], and mathieu gautier[3]

[1]University of Cologne
[2]INRA Centre de Montpellier
[3]INRA

April 28, 2020

Abstract

Analysis of population genetic data often includes the search for genomic re- gions with signs of recent positive selection. One of the approaches involves the concept of Extended Haplotype Homozygosity and its associated statistics. These statistics presume that haplotypes are phased and some of them that variants are polarized. Here we assess the consequences if one of the two con- ditions is not fulfilled. We find that phasing information is indispensable for the accurate estimation of within-population statistics and, if sample sizes are small, for cross-population statistics, too. Ancestry information, in contrast, is of lesser importance for both. We make use of a publicly available update of our R package rehh which, among other features, incorporates the adapted statistics presented here.

Hosted file

`rehh3_v8.pdf` available at https://authorea.com/users/308075/articles/439071-detecting-selection-using-extended-haplotype-homozygosity-ehh-based-statistics-on-unphased-or-unpolarized-data