

A Cross-Population Extended Haplotype-based Homozygosity Score Test to Detect Positive Selection in Genome-wide Scans

Ruzong Fan, Department of Statistics, Texas A&M University, July 2010

1 Overview

This document describes a C++ package to implement the models for “A Cross-Population Extended Haplotype-based Homozygosity Score Test (xp-EHHST) to Detect Positive Selection in Genome-wide Scans”. Section 2 briefly describes the installation and compile the program. Section 3 explains how to run the program for population data with one example of HapMap data.

The theoretical basis for this program is given in our research papers in **Reference**. Please refer to the reference if you use in any published work. In case of suggestions and questions and/or problems, you can contact us via e-mail (rfan@stat.tamu.edu).

2 Installation and Compile the Program

The package is good in Linux. First, download the package “xpEHHST.cc” and “xpEHHST.h” from xpEHHST.zip.

Use the following steps on Linux to install and compile the package:

- Put xpEHHST.cc and xpEHHST.h in a directory
- Compile the package by

```
g++ -O3 xpEHHST.cc -o xpEHHST
```

3 How to Run the Program for HapMap Data

This program requires two pairs of matched data files:

- “genotypes_chr2_CEU_r22_nr.b36_fwd.legend.txt” and “genotypes_chr2_CEU_r22_nr.b36_fwd.phase”;
- “genotypes_chr2_JPT+CHB_r22_nr.b36_fwd.legend.txt” and “genotypes_chr2_JPT+CHB_r22_nr.b36_fwd.phased”.

In each pair of matched files, the first one is SNP legend file, and the second is the haplotype phase file.

To simplify the notations, copy “genotypes_chr2_CEU_r22_nr.b36_fwd.legend.txt” to “legend1.dat”, “genotypes_chr2_CEU_r22_nr.b36_fwd.phase” to “phase1.data”, “genotypes_chr2_JPT+CHB_r22_nr.b36_fwd.legend.txt” to “legend2.dat”, and “genotypes_chr2_JPT+CHB_r22_nr.b36_fwd.phased” to “phase2.dat”.

To run analysis, simply type

```
xpEHHST legend1.dat phase1.dat legend2.dat phase2.dat xp-EHHST-CEU-JPT+CHB_chr2
```

Then, one output files are generated: “xp-EHHST-CEU-JPT+CHB_chr2.out”. Be careful, the computation can take a couple of days. Now it is time to go home and check if you have the results next day (even two days later depending one the speed of your computer)!

The file “xp-EHHST-CEU-JPT+CHB_chr2.out” contains the results of cross-population extended genotype-based homozygosity score test (xp-EHHST). For the theoretical description of the three score tests, refer to Zhong et al. (2010).

4 References

1. Zhong M, Zhang YW, Lange K, and Fan RZ (2010) A cross-population extended haplotype-based homozygosity score test to detect positive selection in genome-wide scans. *Statistics and Interface*, in revision.