

## Data accompanying “hapbin: An efficient program for performing haplotype based scans for positive selection in large genomic datasets”

These files contain genome-wide integrated haplotype scores (iHS) for each of the 26 populations in the phase 3 release of the 1000 genomes project. iHS were calculated using the hapbin program that can be downloaded here <https://github.com/evotools/hapbin> (the version used in this analysis was commit eeda14bbb1, downloaded 8<sup>th</sup> February 2015). The 1000 genomes phased haplotypes were obtained from

<https://mathgen.stats.ox.ac.uk/impute/1000GP%20Phase%203%20haplotypes%206%20October%202014.html> and hapbin was run with default parameters. The iHS are provided in two formats; BED (<http://genome.ucsc.edu/FAQ/FAQformat.html#format1>) and bedGraph (<http://genome.ucsc.edu/FAQ/FAQformat.html#format1.8>). For each SNP the iHS for the allele with the positive iHS is shown. In the BED format file the allele in question (0 or 1 as annotated in the original IMPUTE data) is indicated in the fourth column following the “:”. The bedGraph formatted data can be easily viewed along the genome at the UCSC genome browser by specifying the URL to the corresponding file here (<http://genome-euro.ucsc.edu/cgi-bin/hgCustom?clade=mammal&org=Human&db=hg37>).

The population ids are:

CLM: Colombian in Medellin, Colombia  
MXL: Mexican Ancestry in Los Angeles, California  
PEL: Peruvian in Lima, Peru  
PUR: Puerto Rican in Puerto Rico  
BEB: Bengali in Bangladesh  
GIH: Gujarati Indian in Houston, TX  
ITU: Indian Telugu in the UK  
PJL: Punjabi in Lahore, Pakistan  
STU: Sri Lankan Tamil in the UK  
CDX: Chinese Dai in Xishuangbanna, China  
CHB: Han Chinese in Beijing, China  
CHD: Chinese in Denver, Colorado  
CHS: Southern Han Chinese, China  
JPT: Japanese in Tokyo, Japan  
KHV: Kinh in Ho Chi Minh City, Vietnam  
CEU: Utah residents with Northern and Western European ancestry  
IBS: Iberian populations in Spain  
FIN: Finnish in Finland  
GBR: British in England and Scotland  
TSI: Toscani in Italy  
ACB: African Caribbean in Barbados  
ASW: African Ancestry in Southwest US  
ESN: Esan in Nigeria  
GWD: Gambian in Western Division, The Gambia  
LWK: Luhya in Webuye, Kenya  
MSL: Mende in Sierra Leone  
YRI: Yoruba in Ibadan, Nigeria

(information taken from

<https://mathgen.stats.ox.ac.uk/impute/1000GP%20Phase%203%20haplotypes%206%20October%202014.html>)