Supplemental Material

Limited Evidence for Classic Selective Sweeps in African Populations

Julie M. Granka[1,*], Brenna M. Henn[2], Christopher R. Gignoux[3], Jeffrey M. Kidd[4],

Carlos D. Bustamante[2], Marcus W. Feldman[1]

1. Department of Biology, Stanford University, Stanford, CA, USA

2. Department of Genetics, Stanford University, Stanford, CA, USA

3. University of California, San Francisco, CA, USA

4. Department of Human Genetics, University of Michigan, Ann Arbor, MI, USA

$*$ 385 Serra Mall, Stanford University, Stanford, CA, 94305; jgranka@stanford.edu

Supplemental Methods

Samples

Table S1 lists details of the samples used for each analysis. HapMap Maasai trio parents from Kinyawa, Kenya (MKK) (46 individuals) were included for population differentiation and haplotype analyses (CONSORTIUM, 2005; INTERNATIONAL HAPMAP CONSORTIUM *et al.*, 2007; INTERNATIONAL HAPMAP 3 CONSORTIUM *et al.*, 2010). As PEMBERTON *et al.* (2010) found evidence of undocumented relatedness between individuals in the Maasai, we removed a subset of the related individuals (NA21384, NA21475, NA21399, NA21365, NA21362, NA21382, NA21423, NA21453, NA21615, and NA21634), leaving a total of 46 Maasai individuals.

For haplotype statistic calculations, SNPs with a missing genotype rate $>$ 10%, SNPs with minor allele frequency $<$ 0.5% (i.e. singletons), and SNPs out of Hardy-Weinberg equilibrium (assessed in each population independently) with a $p >$ 0.001 were removed.

Population Differentiation

We first analyzed patterns of population differentiation of all autosomal SNPs for evidence of local adaptation driven by positive selection. Differences in allele frequencies between populations measured by statistics such as F_{ST} may have the power to detect selective events that have occurred less than 50,000 to 75,000 years in the past (AKEY *et al.*, 2002; SABETI *et al.*, 2006).

We calculated a per-SNP F_{ST} over all 11 populations (all populations except the HapMap YRI; see Methods,

main text), a per-SNP derived allele frequency difference (δ) between all pairs of populations, and mean F_{ST} averaged over all genotyped SNPs between all pairs of populations. For these calculations, several admixed individuals were removed based on their inferred proportions of ancestry in ancestral clusters (see Figure 1 of HENN *et al.* (2011)). Excluded individuals in the Hadza were BAR09, END13, END15, END21, END22, and END07; in the $\neq$Khomani Bushmen, excluded individuals were SA25, SA59, SA39, SA58, SA45, SA49, SA40, and SA03 (leaving 11 Hadza and 23 $\neq$Khomani Bushmen) (Table S1).

Enrichments of genic and non-genic SNPs (see main text) were calculated for the absolute value of derived allele frequency difference ($|\delta|$) between pairs of populations for each pair separately (55, or $\binom{11}{2}$, pairs), as in COOP *et al.* (2009) (using a bin size of 0.1). An enrichment in a given bin was calculated as the proportion of genic (or nongenic) SNPs in that bin divided by the proportion of genic (or nongenic) SNPs among all genotyped SNPs. For a clearer picture, we counted the number of population pairs for which each given $|\delta|$ value was defined (some closely-related populations, for instance, may not have SNPs with a $|\delta|$ as high as 0.9.) We then counted the number of pairs for which, at each $|\delta|$ value, there appeared to be an enrichment of genic SNPs (an enrichment value > 1). We also averaged the enrichment values of genic and non-genic SNPs in each $|\delta|$ bin over all 55 population pairs as another summary of enrichment. Here, we examined the absolute value of $|\delta|$, as when examining multiple population pairs, the polarization of a SNP is arbitrary.

We then examined more closely population differentiation between several individual population pairs (out of the 55 total pairs). For each pair, we calculated δ (since we examined each pair individually, we did not take the absolute value) and assessed the significance of genic and non-genic enrichments using bootstrapping. Briefly, bootstrapping involved resampling 200kb regions of the genome with replacement, and re-calculating enrichment values in each bootstrap sample to generate confidence intervals (see main text).

Over all 55 pairs of populations, we examined the relationships of the maximum (over all SNPs) per-SNP F_{ST} and $|\delta|$ values between population pairs (as well as the top 99.99% tail values of F_{ST} and $|\delta|$), versus mean pairwise F_{ST}. A best fit curve between the extreme F_{ST} (or $|\delta|$) values and mean F_{ST} was drawn using the lowess function in R (*www.r-project.org*, as in Coop et al. (2009)).

We simulated the expected relationship under neutrality between the top 99.99% tail of $|\delta|$ values and mean F_{ST} using the ``beta-binomial'' method of BALDING (2003) (also used in COOP *et al.* (2009)). The model assumes that a diverged pair of populations were once part of a common ancestral population, with the ancestral frequency of some allele equal to p. In the model, the allele frequency in each population follows a beta distribution with mean p and variance $p(1-p)F_{ST}$, where F_{ST} is that between the two populations. Within each population,

the sampled allele frequencies follow a binomial distribution, with the success probability parameter determined by the value drawn from the beta distribution. Since the ancestral allele frequency of each SNP is unknown, we first sampled an ancestral allele frequency p for each SNP from a uniform distribution on $[0, 1]$ (as in COOP *et al.* (2009)). We then simulated the sample allele frequencies in each population using the beta-binomial distribution and this value of p to calculate $|\delta|$. We repeated this simulation of $|\delta|$ for the number of SNPs in our dataset to obtain a simulated value of the 99.99% tail (see main text Methods).

Choice of Several Haplotype Statistic Parameters

We used several haplotype statistics to search for patterns of selective sweeps: the Integrated Haplotype Score (iHS) (VOIGHT *et al.*, 2006), and the Cross Population Extended Haplotype Homozygosity Test (XP-EHH) (SABETI *et al.*, 2007). For the XP-EHH, we used the HapMap CEU, HapMap YRI, HapMap MKK, and $\neq$Khomani Bushmen as reference populations.

For the results presented in the main text, we binned the genome into 100 kb non-overlapping windows, and assigned to each a statistic based on the per-SNP haplotype statistics within that window. For XP-EHH, the window statistic used was the maximum XP-EHH value within the window (as in PICKRELL *et al.* (2009)), shown to be a powerful summary statistic to identify selection. For iHS, we used the proportion of SNPs in that window with $|\text{iHS}| > 2$ (as in VOIGHT *et al.* (2006)).

We assigned empirical p-values to windows based on their numbers of SNPs by binning windows in increments of 10 SNPs per window. A p-value for a given window w was obtained as the proportion of windows within that bin with a statistic greater than or equal to the statistic of window w. We assessed the sensitivity of results to the size of genomic windows used, to the bin sizes (in SNPs per window) for assigning p-values to windows, and to the XP-EHH reference population. Our selection of these parameters is described in detail below.

100 kb vs. 200 kb windows size

We studied two sizes for genomic windows: 200 kb, following PICKRELL *et al.* (2009), and 100 kb, following VOIGHT *et al.* (2006) and due to the lower linkage disequilibrium (LD) in African populations.

To explore how strongly the choice of window size affected which windows appeared in the most extreme tails of the empirical distribution, we compared the top empirical windows for those obtained using each of the two window sizes. We examined iHS and XP-EHH using HGDP Europeans as a reference population (as in PICKRELL *et al.* (2009)). (Note that although this particular reference population was not used for our main results, these

analyses of window size should be insensitive to the reference population used.) We performed calculations only for the $\neq$Khomani Bushmen, Sandawe, and Hadza, since genome-wide scans for selection in these populations have not been conducted previously.

We obtained an empirical p-value for each window based on the statistic value of each window and accounting for the number of SNPs within each window. For both 100 kb and 200 kb windows, p-values were obtained by binning windows in increments of 20 SNPs (as in PICKRELL *et al.* (2009)). Windows with $\geq$ 100 SNPs were combined into 1 bin, as few genomic windows contained $\geq$ 100 SNPs.

The top 1% of genomic windows were compared for each population between the method using 200 kb windows and the method using 100 kb windows. As expected, the top 1% of 100 kb windows contained approximately twice as many windows as the top 1% of 200 kb genomic windows. The proportion of 200 kb windows that did not overlap with at least one 100 kb genomic window was near 0% for XP-EHH, and was $\approx$40% for the iHS. The proportion of 100 kb windows without an overlap in at least one 200 kb window (in other words, windows not identified by the 200 kb approach) was $\approx$30-50% for XP-EHH, and $\approx$40-70% for the iHS. Results did not change greatly for XP-EHH when we examined the top 0.1% or top 5% of genomic windows. When only the top 0.1% of windows were examined for iHS, the percentage of 100 kb windows that did not overlap a 200 kb window increased to $\approx$80%.

The results from XP-EHH indicate that selection signals can be localized to 100 kb within nearly all 200 kb windows (indicated by the $\approx$0% of 200 kb windows without an overlapping 100 kb window). The results from iHS suggest that this statistic is more influenced by the choice of window size; this is likely because the window statistic is the fraction of SNPs in a window with $|\text{iHS}| > 2$ (rather than the maximum value of the statistic, as it is for XP-EHH). Because of the lower levels of LD typically found in African populations (HENN *et al.*, 2011), we chose the smaller size of 100 kb.

Binning by 20 vs. 10 SNPs per window

Another parameter of interest is the most appropriate bin size (in numbers of SNPs) within which to calculate empirical p-values for genomic windows. As in the previous analysis, we first binned windows in increments of 20 SNPs per window, combining all windows with $\geq$ 100 SNPs into one window. As an alternative, we also binned windows in increments of 10 SNPs per window; for these bins, all windows with $\geq$ 50 SNPs were combined into one window, because few 100 kb genomic windows contain $\geq$ 50 SNPs. Again, the only populations for which we calculated these statistics were the $\neq$Khomani Bushmen, Sandawe, and Hadza, and we used HGDP Europeans as

a reference for XP-EHH (as previously).

For each population, the top 1% of genomic windows using increments of 20 SNPs per bin were compared with those using increments of 10 SNPs. The binning strategy does seem to have an effect on the empirical tail windows. For both XP-EHH and iHS calculations, ≈10% of the top 1% genomic windows from the 20 SNP binning method did not overlap with any of the top 1% of windows obtained using the 10 SNP method, and vice versa.

With increments of 20 SNPs per window, a majority of windows were in the lowest SNP frequency bin (0-20 SNPs), while with increments of 10 SNPs the number of 100 kb windows per bin was more consistent across bins. Thus, we decreased the SNP increment of bins to 10 SNPs per bin.

Selection of XP-EHH Reference Populations

Finally, we selected the most appropriate populations to use as reference populations for the XP-EHH statistic. Originally, we selected from the following reference populations: (i) all HGDP European populations, pooled together (152 individuals), as in PICKRELL *et al.* (2009); (ii) HapMap CEU Trio Parents (88 individuals); (iii) HGDP Yorubans (21 individuals); (iv) HapMap YRI Trio Parents (100 individuals). The motivation for including both a European population and one of Yoruban origin was to search for selective sweeps of different ages. With trio information, HapMap individuals have more reliable phase inferences than those of the HGDP.

As with our other analyses of haplotype statistic parameters (window size and bin size), we compared the top 1% of 100 kb genomic windows for the XP-EHH between those obtained using different reference populations. Again, we calculated the statistics only for the ≠Khomani Bushmen, Sandawe, and Hadza.

HGDP Europeans vs. HapMap CEU. We first compared the top 1% of windows for XP-EHH using HGDP Europeans as a reference population with XP-EHH using HapMap CEU as a reference population. For all populations, ≈20-30% of the top 1% of windows were unique when using HapMap CEU as a reference, while ≈25-35% of the top 1% were unique when using HGDP Europeans. Thus, while a majority of the windows did show consistent signals using either reference population, ≈25% of the top windows depended on the reference population used.

HGDP Yorubans vs. HapMap YRI. When we compared the calculations using either HGDP Yorubans or HapMap YRI as reference populations, between 40-50% of windows in the top 1% were unique to the calculations using each reference population. The difference in empirical tails between calculations using each of the two Yoruban reference populations was much greater than the difference between calculations using each of the two European reference populations (described above); this could be due to differences in the quality of phase information.

HGDP Europeans vs. Yorubans. We also assessed the overlap between the top 1% of windows obtained using

HGDP Europeans as a reference population versus those using HGDP Yorubans as a reference population. The overlap between the two methods was not high, ranging from 7-20% for all populations. As expected, using Europeans as a reference population resulted in different signals from those when using an African reference population.

Due to greater confidence in phasing, we selected the HapMap CEU and HapMap YRI as the European and Yoruban reference populations, respectively. We also used the HapMap Maasai (MKK) and ≠Khomani Bushmen (KHB) as reference populations. For these two reference populations, we had no alternative population against which to compare results (as we did for the HapMap CEU vs. HGDP Europeans and HapMap YRI vs. HGDP Yorubans).

Variation of Statistics within Windows

To ensure that breaking the genome into windows (as done by VOIGHT *et al.* (2006) and PICKRELL *et al.* (2009)) was a valid approach, we determined whether the variance of statistics within 100 kb windows was less than the variance among randomly selected SNPs. We examined the XP-EHH statistic in the HapMap YRI using HapMap CEU as a reference (as conducted in our main study); similar results should hold for all other sampled populations and statistics.

We used two methods to assess the validity of the windowing approach. First, we performed a Kruskal-Wallis test independently for each chromosome. We assigned to each SNP a categorical variable to indicate the identity of the 100 kb window in which it was located. We then determined whether there was a significant effect of window on the XP-EHH value of each SNP. We found the window effect to be significant, with a p-value of 0 for all chromosomes.

We also used a permutation approach to assess whether the variance of statistic values among SNPs within the same window was significantly less than the variance among groups of SNPs that were not within the same genomic window. To do so, we obtained the variance of the SNP XP-EHH values within each window, and then averaged the variance over all windows (average variance = 0.0554). For a null distribution, we sampled groups of non-adjacent SNPs on the same chromosome, with replacement, to mimic windows chromosome-wide. Since each genomic window contained a different number of SNPs, we used the same distribution of numbers of SNPs per window when drawing random groups of SNPs. We then computed a variance within each random group of SNPs, and averaged the variances across all random groups to obtain one draw of the average variance under the null hypothesis. We repeated this procedure 1,000 times. Our observed average variance within 100 kb windows

was significantly less than all of the 1,000 draws of average variances under the null hypothesis (p-value = 0).

Thus, from both analyses, it appears that breaking the genome into 100 kb windows effectively groups together SNPs with more similar statistic values, as desired.

Comparison of Empirical Tail Windows

As a summary, we obtained the five genomic windows with the lowest empirical p-values in each population, and their p-values in all other populations, for iHS, XP-EHH CEU, XP-EHH YRI, XP-EHH MKK, and XP-EHH KHB. To do so, we first merged adjacent windows in the top 10% of each population using the program *mergeBed* in the program suite BEDTools (QUINLAN and HALL, 2010). If a top window in one population overlapped with a window in at least the top 10% of another population, we noted its p-value.

For each of iHS, XP-EHH CEU, YRI, MKK, and KHB, we performed a Mantel correlation between a matrix of mean F_{ST} over all autosomal SNPs between all pairs of populations and matrices of (i) correlations of window statistics across all genomic windows between pairs of populations, (ii) overlap in the top 1% empirical tail of windows between pairs of populations, and (iii) overlap in the top 0.1%. We also performed Mantel correlations between a matrix of the correlations of window statistics across all genomic windows between pairs of populations ((i) above), and matrices of both (iv) the overlap in the top 1% between pairs of populations and (v) overlap in the top 0.1%. We made scatter-plots of these five comparisons for each statistic for a visual representation of the Mantel correlations (see Supplemental Results).

We also performed Mantel correlations between a matrix of the average Bantu ancestry of each population pair (see Methods, main text), with matrices of (vi) overlap in the top 1% tail between pairs of populations, and (vii) the overall correlation of window statistics between pairs of populations. Note that the Namibian San are not included in this analysis, as Bantu ancestral proportions for all individuals are not available in HENN *et al.* (2011).

Neutral Coalescent Simulations

We performed neutral coalescent simulations of a three-population model and compared empirical tail windows for iHS and XP-EHH between populations (see Figure S12). We assumed an N_e in each population of 12,000, and obtained divergence times under an equilibrium model given values of F_{ST} using the formula $(1 - F_{ST}) = (1 - \frac{1}{2N_e})^t$, where t is the divergence time in generations (HOLSINGER and WEIR, 2009). Divergence time between the reference population and populations 1 and 2 was 3900 generations (F_{ST} = 0.15); divergence times between

populations 1 and 2 were: 2529 (F_{ST} = 0.10), 1231 (F_{ST} = 0.05), and 241 (F_{ST} = 0.01) generations. We assumed a uniform mutation rate of 1.5 x 10^{-8} per bp per generation (see SCHAFFNER *et al.* (2005)), and a uniform recombination rate of 10^{-8} per bp per generation. We also assumed uniform recombination to translate the *msms* results to physical and genetic positions, required as input for XP-EHH and iHS calculations.

An example command line for *msms* corresponding to F_{ST} = 0.10 between the two daughter populations (Populations 1 and 2 in Figure S12) is below. The simulation is for one ten Mb segment, where population ``1'' is the reference population, and populations ``2'' and ``3'' are the two daughter populations. Scaling is in units of $4N_e$ generations.

```
msms -ms 60 1 -t 7200 -I 3 20 20 20 0 -ej 0.053 3 2 -ej 0.081 1 2 -r 4800 10000000
```

Selective Sweep Coalescent Simulations

For simulations under positive natural selection, we simulated a selective sweep in the first daughter population, and allowed the parent population and second daughter population to evolve neutrally (all parameters remained as in the neutral simulations, see above and main text). Segments of five Mb were simulated, since generally the signals of long haplotypes were contained within this region (the selected allele was in the center of each segment). In order to ensure that each simulated sweep was a hard sweep, we retained only the simulations for which the ``origin count'' of the selected allele was 1, and in which the selected allele was not lost.

We ran preliminary analyses and simulations in *msms* to determine which parameters of selection were the most appropriate. First, we chose a selection coefficient of s = 0.2. Smaller selection coefficients resulted in a large number of unusable simulations (with $<$ 5% of all simulations fitting the above criteria); $\approx$ 5-10% of all simulations fit the above criteria for s = 0.2.

We examined several different timings for the start of selection of the allele. The first was 144 generations ago. For XP-EHH, this timing resulted in a very high true positive rate (see description of calculation below) for detecting selective sweeps, but for iHS, the true positive rate was much lower (see Supplemental Results). This is because in nearly all simulations, the selected allele had reached fixation, resulting in lower power for iHS (see Results and VOIGHT *et al.* (2006); SABETI *et al.* (2007)). For these simulations, we simulated 80 five Mb segments under selection, and 260 ten Mb segments under neutrality (for a total genomic length of 3000 Mb).

Through additional simulations, we found that a more recent selection time of 62.4 generations ago led to a more intermediate final frequency (10-90%) of the selected allele, allowing iHS to have greater power (SABETI *et al.*, 2007). As a result, the true positive rate for both iHS, as well as XP-EHH, was high with this timing (see Sup-

plemental Results). For these simulations, we simulated 300 five Mb segments (50% of the simulated ``genome'') to ensure that we simulated enough genomic regions with selection; the remaining 1500 Mb were simulated under neutrality.

The selection options in *msms* that were used to simulate selection in the first daughter population (62.4 generations ago) are below (where timing is in units of $4N_e$ generations; arguments are used in conjunction with the demographic parameters, specified above, modified to simulate five Mb):

```
-SI 0.0013 3 0 0.00001 0 -Sp 0.5 -Sc 0 2 9600 4800 0 -N 12000 -oOC -oTrace -Smu 0
```

As for the neutral coalescent simulations, for each of the 100 genome-wide simulations we calculated iHS and XP-EHH, binned the genome into windows of 100 kb, and ranked windows to obtain empirical tails. To examine the performance, or true positive rate, of the iHS and XP-EHH statistics, for a given empirical tail we calculated the proportion of its 100 kb windows that were within a selective sweep (non-neutral) segment. This rough measure indicates the ability of the statistics to detect signals of selection in 100 kb windows that are within 2.5 Mb of a selected site (since the selected site was located in the center of each simulated five Mb region). Depending on the exact lengths of haplotype signals produced from the sweeps, this may be an overestimate of the true positive rate, and an underestimate of the false positive rate.

Gene Ontology Enrichment

Replication of Voight et al. 2006 Results

We followed the procedure of VOIGHT *et al.* (2006), who searched for Gene Ontology (GO) term enrichment in the HapMap YRI for the iHS statistic. First, we obtained the 50 closest SNPs centered around each gene (obtained from the UCSC refFlat mappings of RefSeq genes to hg18; downloaded in May 2010, *www.genome.ucsc.edu*). If there were more than 50 SNPs located within a gene, we kept all of the SNPs. Then, we assigned to each gene a summary statistic, calculated as the proportion of its 50 SNPs with normalized $|\text{iHS}| > 2$.

We ranked the genes in order of their summary statistics, and examined enrichments in the top 10% of genes. We assigned GO terms to genes using the *biomaRt* package in R (*www.r-project.org*; described in more detail below). We also pruned terms to include only those present in the PANTHER database (*www.pantherdb.org*), as VOIGHT *et al.* (2006) use this dataset. Note that in our analysis, we examined $\approx$ 500,000 SNPs; this is only a subset of the SNPs analyzed by VOIGHT *et al.* (2006).

Then, we assessed enrichments for each PANTHER biological process GO term using the hypergeometric dis-

tribution. The hypergeometric distribution for a random variable X can be written as:

$$P(X = k) = \frac{\binom{r}{k}\binom{n-r}{m-k}}{\binom{n}{m}}. \quad (1)$$

Here, for a particular GO term, k is the number of genes in the top 10% of genes with that GO term, r is the number of genes annotated with the GO term in the genome overall, n is the total number of genes in the genome, and m is the number of genes in the top empirical 10%. To calculate a p-value for a particular GO term, given its count of occurrences in the top empirical 10% of genes (k), we calculate $P(X \geq k)$ by summing $P(X = i)$ for $i \in [k, r]$ if $r \leq m$ and $i \in [k, m]$ if $m \leq r$.

Haplotype Statistics

We explain our novel permutation approach for assessing significance of GO terms within populations in more detail, as in Figure 1 of the main text. The method is particularly designed for analyses of genomic windows, within which several genes can exist. Significance is assessed for the top x% of 100 kb genomic windows for a particular haplotype statistic (iHS or XP-EHH); in our analyses, we examine the top 0.1%, 1%, and 5%. (We focus on the top 0.1% in the main text.)

1. **Observed empirical tails.** For the top x% of genomic windows, we determined the GO terms associated with the genes in each genomic window. (The *biomaRt* package in R was used to perform these lookups, using ``ensembl_mart_51'' to ensure that the genomic coordinates (Build 36) were consistent). Adjacent genomic windows were merged, so that a signal $>$ 100 kb would have each of its GO terms counted only once. (This helps to reduce possible dependence between 100 kb windows.) For each unique GO term, we counted its number of occurrences.

2. **Genome-wide windows.** We determined the associated GO terms for each genomic window containing at least one SNP with a statistic value (using *biomaRt*, as above). (Some SNPs were removed in the XP-EHH and iHS calculations due to certain criteria. For instance, for iHS, the minor allele frequency of each SNP must be $\geq$ 5%; VOIGHT *et al.* (2006) also stated, ``if the region spanned by EHH $>$ 0.05 reached a chromosome end or the start of of a gap $>$ 200 kb, then no iHS value was reported for the core SNP.'' Similarly for XP-EHH, SABETI *et al.* (2007) stated, ``if there is no SNP with such an EHH between 0.03 and 0.05, the XP-EHH test is skipped.'' Thus, not all genotyped SNPs resulted in a calculated XP-EHH or iHS value.)

We then binned the above windows by the number of SNPs per window, as done when calculating p-values for the observed data (described in section ``Choice of Several Haplotype Statistic Parameters'').

3. **Randomization step**. In each SNP count bin, we randomly selected x% of windows, and combined the top windows from all bins to create a ``random top $x$%'' set of windows. Then, for each unique GO term, we recorded its number of occurrences across all of the windows in the ``random top x%.''

   We repeated this randomization step 20,000 times, recording for each GO term its number of occurrences in each of the randomizations to form the null distribution.

4. **P value calculation.** For each GO term G, we calculated a p-value for over-representation of G in the observed x% tail as the proportion of the 20,000 permutations for which the number of occurrences of G is greater than or equal to the observed number of occurrences.

5. **Multiple hypothesis testing correction.** To account for multiple hypothesis testing, we controlled the false discovery rate (FDR) at a 5% level using the procedure of BENJAMINI and HOCHBERG (1995). This is less conservative than the Bonferroni correction, but more conservative than using no multiple hypothesis testing correction.

In some cases, we analyzed only a subset of all GO terms that are found in the PANTHER database, as in VOIGHT *et al.* (2006). This only affects our assignment of GO terms to windows in steps 1 and 2 above.

Top F_{ST} Values

To assess whether any GO terms were enriched among SNPs with the most extreme values of F_{ST} (calculated across all 11 populations), we first followed the approach described previously for genomic windows. For each SNP, we found the gene, if any, in which it was located, and obtained the GO terms associated with that gene. We then randomly selected x% of all SNPs, and counted the occurrence of each GO term; we repeated this 20,000 times to obtain a null distribution for the counts of each GO term. We obtained p-values for observed counts of GO terms from this null distribution with an FDR correction. To be more conservative when identifying over-represented terms among top genomic windows, we restricted our results to terms associated with extreme F_{ST} SNPs occurring in different, not the same, genes.

In addition, we searched for significant GO terms with the program GREAT (MCLEAN *et al.*, 2010), which uses a binomial test to assess significance among a foreground set of genomic regions (in our case, SNPs with large

F_{ST} values). We used the SNPs genotyped in our dataset as a background set of genomic regions to control for non-uniform SNP density and ascertainment. Using the parameters described below, GREAT first infers gene regulatory domains, and then the GO terms associated with a given genomic region based on the domains it overlaps. It calculates the percentage of the genome annotated with each GO term; then, it counts the SNPs in the tails that are associated. A binomial test over the genomic regions is performed to obtain a p-value for each GO term, using an FDR correction (see McLean *et al.* (2010) for more details). (This program was not used for 100 kb genomic windows, as the genomic regions of GREAT are required to be smaller).

GREAT allows for a number of ``association rule'' parameters, which set the boundaries of gene regulatory regions. We first used the default settings of the ``basal plus extension'' rule, allowing each gene to be associated with a basal regulatory domain a minimum distance of 5 kb upstream and 1 kb downstream of the transcription start site (TSS). This domain is extended both upstream and downstream until the nearest gene's basal domain, and no more than 1000 kb away from the gene; any SNPs within this domain are associated with the gene. We then allowed the domain to be extended no more than 50 kb or 0 kb away (instead of 1000 kb).

Supplemental Results

Population Differentiation

Pairwise F_{ST} estimates averaged over all autosomal SNPs agreed with those of Henn *et al.* (2011), who calculated F_{ST} between inferred ancestral populations (Table S2). On average, our estimates were slightly lower, as we included admixed individuals with ancestry in several ancestral populations.

We calculated for each SNP the absolute values of derived allele frequency difference ($|\delta|$) between all (55, or $\binom{11}{2}$) pairs of populations, following Coop *et al.* (2009). Not all population pairs exhibited high $|\delta|$ values (Figures S1A and C). For some pairs of populations, there was an enrichment of genic SNPs at high $|\delta|$ values, but for others, there was no enrichment (Figure S1A). Among the population pairs for which which genic SNPs were enriched among the highest $|\delta|$ values for the pair were the Kenya Bantu and Maasai; the South African Bantu and the $\neq$Khomani Bushmen and Sandawe; the Hadza and the Maasai; the Mbuti and the: Hadza, Maasai, Mandenka, Namibian San, $\neq$Khomani Bushmen, Sandawe, and Yoruba; and the $\neq$Khomani Bushmen and Yoruba. Interestingly, all population pairs had SNPs with allele frequency differences in the smallest bin ($0 < |\delta| < 0.1$), with an apparent raw genic enrichment of SNPs for all (Figure S1C). Though this enrichment of genic SNPs may not be significant in all populations, it may indicate the action of purifying selection. When enrichment values in each $|\delta|$

value bin (in Figure S1A) were averaged over all population pairs, there was only a slight increase in enrichment of genic SNPs at extreme values (Figure S1B).

We examined several of the population pairs individually for evidence of genic enrichment using a bootstrapping procedure to assign significance (mentioned in the Supplemental Methods and the Main Text) and present results in Figure S2. For individual population pairs, we examined the non-absolute value of δ. In most cases, the SNPs in the most extreme δ value bins were indeed genic -- but there were very few (< 5) SNPs. In the bootstrapping procedure, we found that this small number of SNPs with extreme δ values was problematic. In general, less than 95% of the 1,000 bootstrap simulations actually contained a SNP with a value of δ in the highest (or lowest) bin (Figure S2). Thus, the actual existence of SNPs in this bin was not significant, and so we could not assign significance to genic enrichment in these bins. (In Figure S2C and D, while $>$ 95% bootstrap samples were valid, enrichment was not significant). Interestingly, the next-most-extreme δ bin was never significant according to the 95% bootstrap confidence intervals. While it may be interesting that a majority of the most strongly differentiated SNPs were genic (and that the few bootstrap samples that actually contained these SNPs showed an enrichment), we cannot state with statistical certainty that there is evidence of genic enrichment of strongly differentiated SNPs among pairs of African populations.

We also examined the SNPs with $|\delta| > 0.8$ in at least one population pair. In most cases, we found that these SNPs were primarily due to one allele being completely fixed or absent from the Mbuti Pygmies or Hadza. However, these high-$|\delta|$ SNPs between the two populations did not show a genic enrichment (Figure S2D). The lack of genic enrichment among the many highly differentiated SNPs is likely due to the fact that they have both undergone severe bottlenecks in recent history (HENN *et al.*, 2011).

We also analyzed population differentiation between the HGDP Tuscans and Yorubans (Figure S3). Here, while we found that the bootstrap confidence intervals were valid, genic enrichment was not significant for either population. See main text for discussion of these results.

Although the analyses here (see also Figure S4) and in the main text (Figures 2 and 3) indicated that allele frequency differentiation is potentially strongly affected by population history, there were several interesting genes within which several SNPs lied in the empirical top 1% of all F_{ST} values. Of these, several SNPs were within the lactase gene (*LCT*), involved in lactase persistence, in several genes known to be involved in skin pigmentation (*HERC2, HPS2, KIT, MITF, NF1, OCA2, PTGER3*), and in other genes known to be involved in disease resistance (*HLA, LARGE, TLR5, TLR6*). However, as differentiation of these SNPs may not necessarily be due to positive selection (see main text and above), we did not explore these signals further.

Haplotype Statistics

Overlap of Windows Between Populations

There was considerable overlap between populations of top windows for the XP-EHH CEU statistic (Figure S6), but less overlap for iHS (Figure S5), and even less for XP-EHH MKK, YRI, and KHB (main text Figure 4 and Figures S7 and S8). Note that our results for the iHS and XP-EHH CEU were slightly different from those of PICKRELL *et al.* (2009) and HENN *et al.* (2011), as they examined genomic windows of 200 kb.

We examined the relationship between mean F_{ST} and extent of sharing of the top 0.1% of windows between pairs of populations (see Supplemental Methods). We saw results similar to the top 1% (see main text Results) for the XP-EHH CEU, XP-EHH MKK, and XP-EHH KHB, although the relationship between these two variables appeared weaker than for the top 1% of windows (Figure S9). Mantel correlations remained significant for these statistics, but not for the iHS or XP-EHH YRI (Table S3). (Since the maximum number of windows shared in the top 0.1% for any two populations was 1 for the iHS and XP-EHH YRI, there may be little power to detect any relationship.)

For each of iHS, XP-EHH CEU, YRI, MKK, and KHB, we also computed a correlation of the selection statistics per window across all genomic windows between all pairs of populations (denoted ``correlation of window statistics''). For all five statistics, including the XP-EHH YRI, the number of windows in the top 1% of both populations of a pair increased as the correlation of window statistics for that pair increased, with a significant Mantel correlation between these two variables (Table S3, Figure S10). For the top 0.1% of genomic windows, XP-EHH CEU, XP-EHH MKK, and XP-EHH KHB still exhibited significant relationships between the correlation of window statistics and the number of shared windows in the top 0.1%, though to a lesser degree than for the top 1% (Table S3). For the iHS and XP-EHH YRI, no more than 1 window was shared between any pair of populations, again resulting in non-significant Mantel correlations (Table S3).

We also found that more closely related population pairs (those with lower mean F_{ST}) had more highly correlated window statistic values for iHS, XP-EHH CEU, XP-EHH MKK, and XP-EHH KHB (all except for XP-EHH YRI). Mantel tests for these four statistics indicated a significant correlation between the two variables (Table S3). This negative trend between F_{ST} and correlations of window statistics suggests that values of these selection statistics in the genome as a whole correspond closely with population divergence.

However, the XP-EHH YRI statistic appeared to behave in a fundamentally different manner from the other statistics. We observed no relationship between F_{ST} and the correlations of window statistics between pairs of populations, nor between F_{ST} and the extent of overlap in empirical tails. Since the empirical tails did seem

to be related to overall correlations of window statistics, however, the different behavior of XP-EHH YRI may not necessarily indicate that between-population patterns are driven by shared selective pressures. Interestingly, we found that the correlations of window statistics across the whole genome appeared to be related to the average proportion of Bantu ancestry. The correlation of window statistics between pairs of populations for XP-EHH YRI decreased as the average proportion of inferred Bantu ancestry for each population in the pair increased (p-value = 0.009; see Figure S11 and Table S3). We found a similar result for sharing in the 1% empirical tail (Table S3). The greater amount of sharing of windows between populations with low amounts of Bantu ancestry appears to be an artifact of using a population with a large amount of Bantu ancestry as a reference.

Further Examination of Shared Windows

Population history appeared to be difficult to separate from the extent of sharing of candidate signals (see previous section). However, positive selection on the same genomic regions might be able to explain an increased number of shared windows beyond a ``baseline'' expected from shared ancestry. We examined the pairs of populations that appeared to have a greater than expected number of windows shared in their top 0.1% when compared to other population pairs with similar values of F_{ST} (Figure S9).

The Mandenka and Yoruba were an outlying population pair for the XP-EHH MKK, with five common windows in their top 0.1% (see Supplemental Results for more details). Of these, two adjacent regions on chromosome 11 (9,800,000-10,100,000) were outliers only in the Mandenka and Yoruba and the South African Bantu. This region contains the gene *SBF2*, mutations in which are known to be associated with Charcot-Marie-Tooth disease (characterized by abnormal myelin sheaths) (YAN *et al.*, 2011). For the XP-EHH YRI, the Sandawe, Hadza, and Maasai all shared a region on chromosome 2 (136,300,000-136,400,000) containing the genes *MCM6, LCT* and *DARS*. This region was also found in at least the top empirical 5% of the Mbuti and Biaka Pygmies, as well as the $\neq$Khomani Bushmen. Because this region appeared to be significant in hunter-gatherer, pastoralist, and agricultural populations, sharing may not be due to convergent adaptation of lactase persistence but by another unknown selective pressure or by a sweep at another gene, as suggested by TISHKOFF *et al.* (2007). A region shared in the top 0.1% of the Kenyan Bantu and Maasai for XP-EHH YRI was on chromosome 2 (38,800,000-38,900,000), and was otherwise found only in the top 5% tail of the Mandenka. Among other genes, this region contains *GALM*, involved in galactose metabolism. Thus, a potentially diet-related trait could be under selection in all three of these agricultural/pastoralist populations (Kenyan Bantu, Maasai, Mandenka).

There were several outlying population pairs for the XP-EHH MKK (Figure S9). For the Mandenka and Yoruba,

among the five shared windows in their top 0.1% was a window that was not in the top empirical 5% of any other population except the Kenyan Bantu; however, it contains no known genes (chromosome 13: 56,900,000-57,000,000). Two adjacent windows on chromosome 3 (56,500,000-56,700,000) were not in the top empirical 5% of any other tested population. This region contains the genes *CCDC66* and *C3orf13*, whose functions are unknown.

For iHS, one pair with a shared window was the Sandawe and Maasai; this window (chromosome 11: 84,600,000-84,700,000) was not an outlier in any other population and contains *DLG2*, potentially involved in synapse function. The ≠Khomani Bushmen and Maasai shared an empirical outlying region for XP-EHH YRI, found only also in the empirical tail windows of the Hadza. This region on chromosome 2 (196,900,000-197,000,000) contains the gene *HECW2*, possibly related to ubiquitination.

Coalescent Simulations

For data simulated under the neutral coalescent, in nearly all empirical tails for XP-EHH, there appeared to be a strong relationship between the F_{ST} of the two daughter populations and the extent of sharing of windows in their empirical tails (Figure S13). For iHS, this relationship was less strong, though still apparent (Figure S14).

For data simulated under selection for XP-EHH, there was generally no relationship between the number of overlapping windows and F_{ST}, regardless of timing of selection (Figures S15 and S17). In Figure S17, where selection is less recent, there was a more apparent relationship between the two quantities (especially for the 0.001 tail). While our calculated true positive rate was still 100%, this observation may be due to the fact that our estimate of the true positive rate is an overestimate (see Supplemental Methods). This observation may also be due to the inadequate number of regions simulated for this selection timing (80 five Mb regions, versus 300 five Mb regions for the more recent selective sweeps in Figure S15).

For data simulated under selection for iHS, there was a more apparent relationship of the number of overlapping tail windows with F_{ST} for older selective sweeps (Figure S18) than for recent sweeps (Figure S16). The true positive rate for iHS was nearly 100% for more recent sweeps (62.4 generations ago), where the selected alleles had not yet swept to fixation. However, for older sweeps (144 generations ago), the selected allele had generally risen to fixation, resulting in a drop of power (VOIGHT *et al.*, 2006; SABETI *et al.*, 2007), and decrease of the true positive for iHS to as low as 60%. This caused the empirical tails to generally contain more neutral false positive genomic windows. This existence of false positive signals in empirical tails likely explains the apparent relationship between extent of sharing of empirical tail windows with a population pairs' F_{ST} (Figure S18).

Gene Ontology Enrichments

Replication of Voight et al. 2006 Results

We replicated the procedure of VOIGHT *et al.* (2006) to assess enrichment of Gene Ontology (GO) terms in the top 10% of genes according to iHS calculations in the HapMap YRI. VOIGHT *et al.* (2006) reported enrichment for steroid metabolism (GO:0008202), MHC-I mediated immunity (GO:0032393 and GO:0032394), chemosensory perception (GO:0007606), olfaction (GO:0007608), and peroxisome transport (GO:0043574). Repeating their analysis with their methodology, we also found enrichments for MHC-I mediated immunity (p-value = 0.002) and olfaction (p-value = 0.02625). Although we did not find an enrichment for chemosensory perception, this is likely due to a change in the GO database (*www.amigo.geneontology.org*, accessed June 2011): surprisingly, we found very few genes in the GO database associated with chemosensory perception. We also found a higher p-value for steroid metabolism than do VOIGHT *et al.* (2006) (p-value = 0.6684). This is again likely due to the GO database: while we found many genes associated with the term when we look up the *term* in the database, the term was not associated with any of them when the *genes* were looked up in the database (as done for our analysis). Finally, the term peroxisome transport was not in the PANTHER database, though VOIGHT *et al.* (2006) reported it to be.

Thus, most differences in results when replicating the method of VOIGHT *et al.* (2006) were likely due to changes in the GO database since their study. In addition, we examined only a subset of the SNPs analyzed by VOIGHT *et al.* (2006), resulting in slightly different p-values. While we found several other significant terms when using the method of VOIGHT *et al.* (2006), none of them were significant when using our more rigorous permutation method (described in the Supplemental Methods and in the main text Methods and Results).

Haplotype Statistics

Using our permutation method, we found that all haplotype statistics (iHS, and XP-EHH CEU, YRI, MKK and KHB) exhibited at least some signatures of GO term enrichment. In addition, GO terms appeared to be enriched in all empirical tails for the statistics (the top 0.1%, 1%, and 5%) (Table S4 and main text Tables 1 and 2). However, terms that appeared enriched in a less extreme tail (i.e. top 1%) were not necessarily enriched in a more extreme tail (i.e. top 0.1%), potentially complicating the choice of the most appropriate cutoff. Our neutral simulations also highlight that if an improper empirical tail is analyzed, attributing sharing of signals between populations to positive selection could be largely inaccurate (see also Figures S13 and S14). In the main text (Tables 1 and 2), we present results from the top 0.1%, motivated by the results of Table S3.

F_{ST} Values

We performed a similar Gene Ontology (GO) term enrichment analysis for SNPs with the top empirical values of F_{ST} (calculated across all 11 populations). We first used a permutation method, analogous to the one performed for genomic windows and haplotype statistics, but found no significant results (see Supplemental Methods). We also used the program GREAT (McLean *et al.*, 2010) to search for enrichment of terms associated with SNPs with the largest values of F_{ST}. Our results varied depending on the ``extension'' rules determining the association of a SNP with a particular gene (see Supplemental Methods). When we allowed the domain associated with a gene to extend up to 1000 kb away beyond the basal domain, we found five significant GO terms with multiple gene associations: regulation of saliva secretion (GO:0046877), regulation of ATPase activity (GO:0043462), positive regulation of ATPase activity (GO:0032781), tricarboxylic acid cycle (GO:0006099), and coenzyme catabolic process (GO:0009109). However, when we allowed an extension of only 50 kb, none of these terms were significant, and instead we found associations of a different nature: spermatogenesis (GO:0007283), regulation of estrogen receptor signaling pathway (GO:0033146), chromatin modification (GO:0016568), regulation of cytokine biosynthetic process (GO:0042089), gamete generation (GO:0007276), muscle filament sliding (GO:0030049), atrial cardiac muscle tissue morphogenesis (GO:0055009), and male sex determination (GO:0030238). When we required that a SNP be located within a gene in order to be associated with it (0 kb extension) we found no enriched GO terms, consistent with our permutation analysis.

Since the GREAT analysis depended heavily on certain parameters, and our permutation procedure found no evidence of over-representation for any GO terms, there did not appear to be any particular functions associated with SNPs that were strongly differentiated between African populations (as measured by F_{ST}). This agrees with Figures 2 and 3 of the main text and Figures S1, S2, and S4, which suggest that selection is not a primary driver of extreme allele frequency differentiation between African populations for the SNPs in our dataset.

References

AKEY, J. M., G. ZHANG, K. ZHANG, L. JIN and M. D. SHRIVER, 2002 Interrogating a high-density SNP map for signatures of natural selection. Genome Res **12**: 1805--1814.

BALDING, D. J., 2003 Likelihood-based inference for genetic correlation coefficients. Theor Popul Biol **63**: 221--30.

BENJAMINI, Y. and Y. HOCHBERG, 1995 Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society, Series B **57**: 289--300.

CONSORTIUM, I. H., 2005 A haplotype map of the human genome. Nature **437**: 1299--1320.

COOP, G., J. K. PICKRELL, J. NOVEMBRE, S. KUDARAVALLI, J. LI *et al.*, 2009 The role of geography in human adaptation. PLoS Genet **5**: e1000500.

HENN, B. M., C. R. GIGNOUX, M. JOBIN, J. M. GRANKA, J. M. MACPHERSON *et al.*, 2011 Hunter-gatherer genomic diversity suggests a southern African origin for modern humans. Proc Natl Acad Sci USA **108**: 5154--62.

HOLSINGER, K. E. and B. S. WEIR, 2009 Genetics in geographically structured populations: defining, estimating and interpreting F(ST). Nat Rev Genet **10**: 639--50.

INTERNATIONAL HAPMAP 3 CONSORTIUM, D. M. ALTSHULER, R. A. GIBBS, L. PELTONEN, D. M. ALTSHULER *et al.*, 2010 Integrating common and rare genetic variation in diverse human populations. Nature **467**: 52--8.

INTERNATIONAL HAPMAP CONSORTIUM, K. A. FRAZER, D. G. BALLINGER, D. R. COX, D. A. HINDS *et al.*, 2007 A second generation human haplotype map of over 3.1 million SNPs. Nature **449**: 851--61.

MCLEAN, C. Y., D. BRISTOR, M. HILLER, S. L. CLARKE, B. T. SCHAAR *et al.*, 2010 GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol **28**: 495--501.

PEMBERTON, T. J., C. WANG, J. Z. LI and N. A. ROSENBERG, 2010 Inference of unexpected genetic relatedness among individuals in HapMap Phase III. Am J Hum Genet **87**: 457--464.

PICKRELL, J. K., G. COOP, J. NOVEMBRE, S. KUDARAVALLI, J. Z. LI *et al.*, 2009 Signals of recent positive selection in a worldwide sample of human populations. Genome Res **19**: 826--837.

QUINLAN, A. R. and I. M. HALL, 2010 BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics **26**: 841--842.

SABETI, P. C., S. F. SCHAFFNER, B. FRY, J. LOHMUELLER, P. VARILLY *et al.*, 2006 Positive natural selection in the human lineage. Science **312**: 1614--1620.

SABETI, P. C., P. VARILLY, B. FRY, J. LOHMUELLER, E. HOSTETTER *et al.*, 2007 Genome-wide detection and characterization of positive selection in human populations. Nature **449**: 913--8.

SCHAFFNER, S. F., C. FOO, S. GABRIEL, D. REICH, M. J. DALY *et al.*, 2005 Calibrating a coalescent simulation of human genome sequence variation. Genome Res **15**: 1576--83.

SCHUSTER, S. C., W. MILLER, A. RATAN, L. P. TOMSHO, B. GIARDINE *et al.*, 2010 Complete Khoisan and Bantu genomes from southern Africa. Nature **463**: 943--947.

TISHKOFF, S. A., F. A. REED, A. RANCIARO, B. F. VOIGHT, C. C. BABBITT *et al.*, 2007 Convergent adaptation of human lactase persistence in Africa and Europe. Nat Genet **39**: 31--40.

VOIGHT, B. F., S. KUDARAVALLI, X. WEN and J. K. PRITCHARD, 2006 A map of recent positive selection in the human genome. PLoS Biol **4**: e72.

YAN, G., G. ZHANG, X. FANG, Y. ZHANG, C. LI *et al.*, 2011 Genome sequencing and comparison of two nonhuman primate animal models, the cynomolgus and chinese rhesus macaques. Nat Biotechnol **29**: 1019--23.

Figures

List of Figures

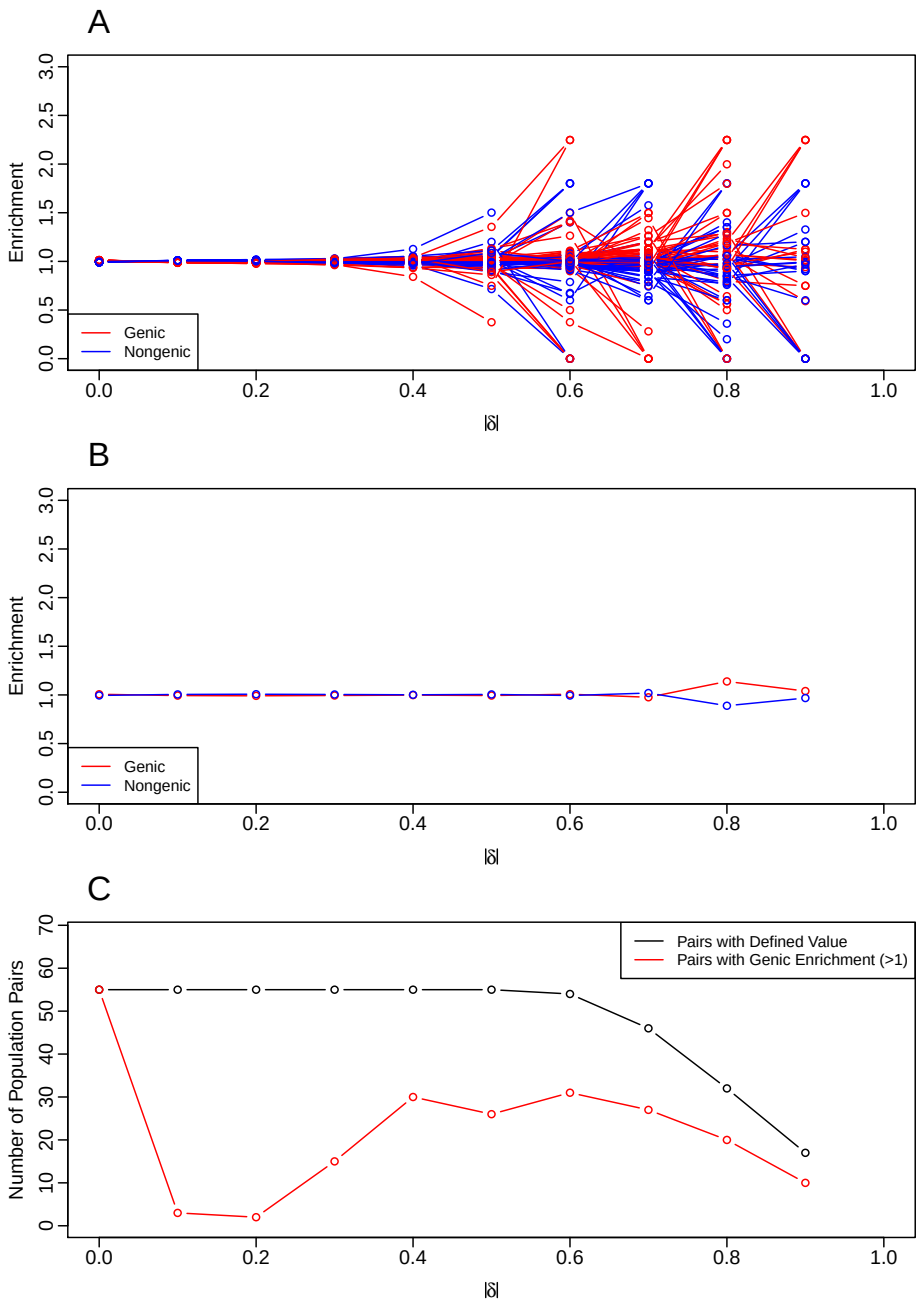

Figure S1. Enrichment of genic and non-genic SNPs for values of $|\delta|$ between pairs of populations. Enrichments of genic and nongenic SNPs are calculated within each $|\delta|$ value bin (width 0.1). (A) Enrichments calculated separately for each pair of populations among the 11 populations (55 pairs in total). Each line indicates a unique population pair. (B) Enrichments of genic and non-genic SNPs averaged over population pairs -- i.e, an average of the lines in panel (A), for genic and non-genic enrichments separately. (C) Summary of panel (A). Black line indicates the number of population pairs with a defined value of $|\delta|$ for each bin (i.e., the number of population pairs where there exists at least one SNP in the given $|\delta|$ bin). Red line indicates the number of population pairs where there appears to be an enrichment of genic SNPs (enrichment value > 1) in the given $|\delta|$ bin.

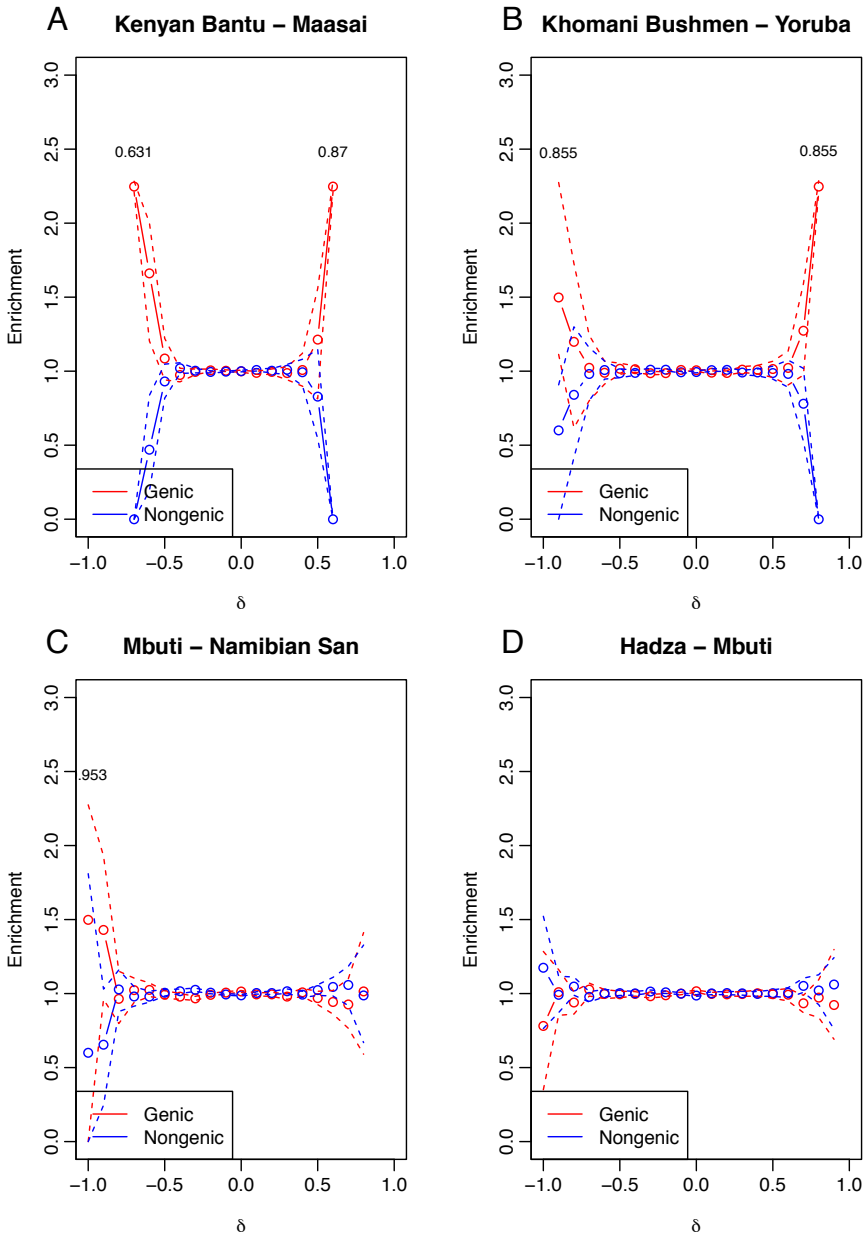

Figure S2. Enrichment of genic and non-genic SNPs for values of δ between selected pairs of populations. Enrichments of genic and nongenic SNPs are calculated within each δ value bin, as in Figure S1 and main text (width 0.1). Bootstrap confidence intervals are indicated by dotted lines (see main text). SNPs where the derived allele is fixed in the first population and absent in the second are depicted on the right side of the plot (near 1); SNPs where the derived allele is fixed in the second population and absent in the first are depicted on the left side of the plot (near -1). The number above the most extreme δ bins indicates the proportion of bootstrap simulations where an enrichment value was defined for the given bin (see Results); if no value is listed, the proportion is 1. Values less than 0.95 can be considered insignificant. (A) Kenyan Bantu and Maasai. (B) $\neq$Khomani Bushmen and Yoruba. (C) Mbuti and Namibian San. (D) Hadza and Mbuti Pygmies.

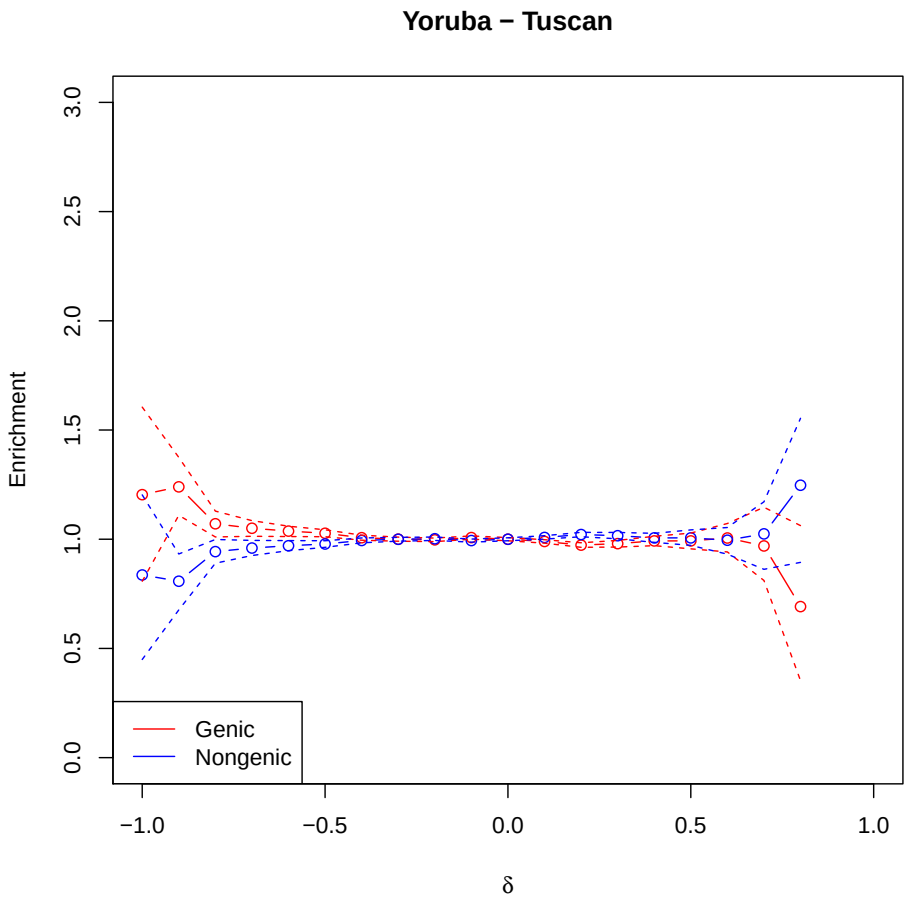

Figure S3. Enrichment of genic and non-genic SNPs for values of δ between the HGDP Yoruba and HapMap Tuscans. Enrichments of genic and nongenic SNPs are calculated within each δ value bin, with bootstrap confidence intervals, as in Figures S1 and S2 and main text. SNPs where the derived allele is fixed in the Yorubans and absent in the Tuscans are depicted on the right hand side of the plot (near 1); SNPs where the derived allele is fixed in the Tuscans and absent in the Yorubans are depicted on the left hand side of the plot (near -1).

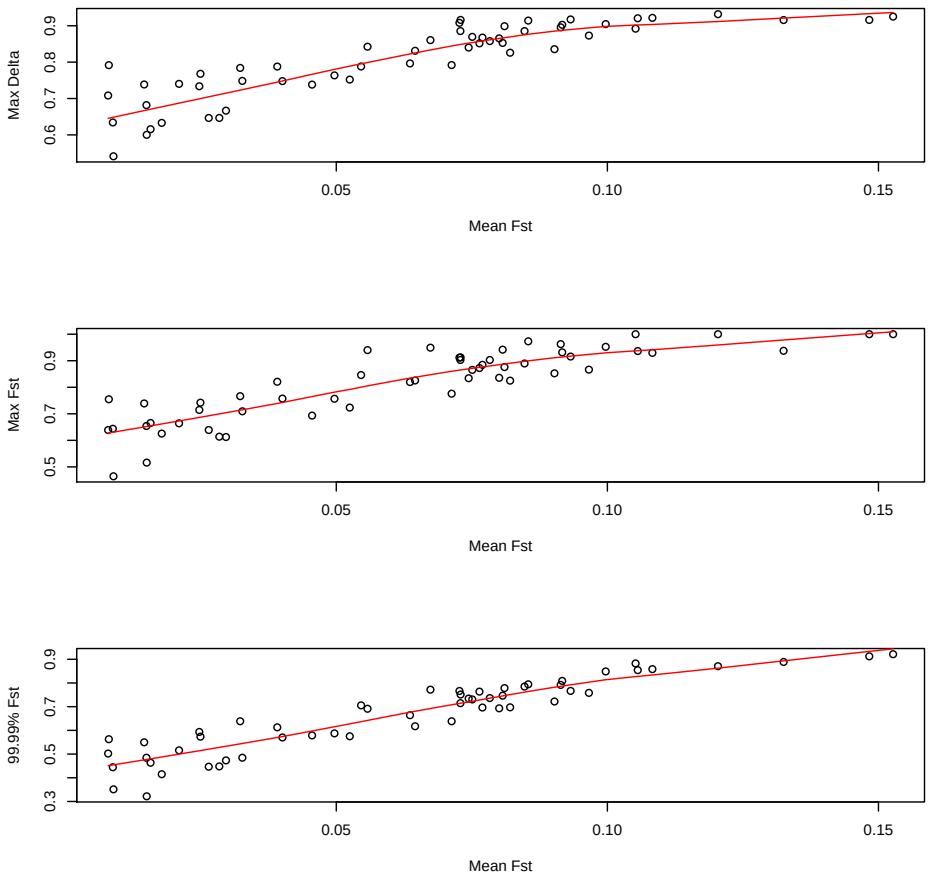

Figure S4. Extreme values of SNP differentiation over all SNPs vs. mean F_{ST} for all pairs of populations. Each point is a population pair; there are 55 pairs in total. Line is the best-fit *lowess* curve through the points. Top panel Y-axis: maximum value over all SNPs of derived allele frequency difference ($|\delta|$) between pairs of populations; Middle panel Y-axis: maximum value over all SNPs of F_{ST} between pairs of populations; Bottom panel Y-axis: upper 99.99% tail cutoff value over all SNPs of F_{ST} between pairs of populations.

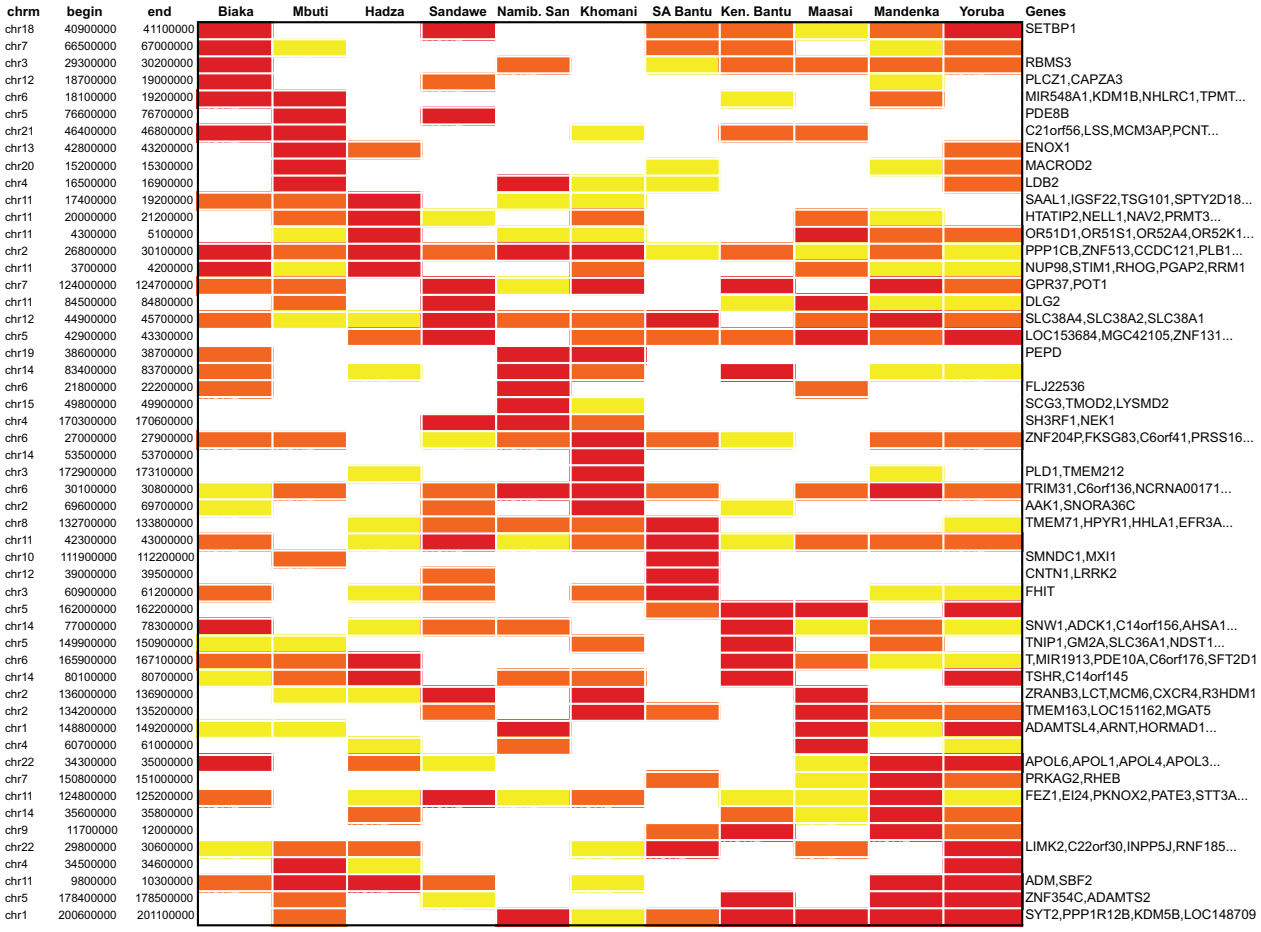

Figure S5. P-values for the 5 most extreme 100 kb genomic windows according to the iHS statistic in each population. Red indicates $p < 0.01$, orange indicates $p < 0.05$, yellow indicates $p < 0.10$, and white indicates $p > 0.10$. ``Genes'' column lists the genes located within the indicated windows.

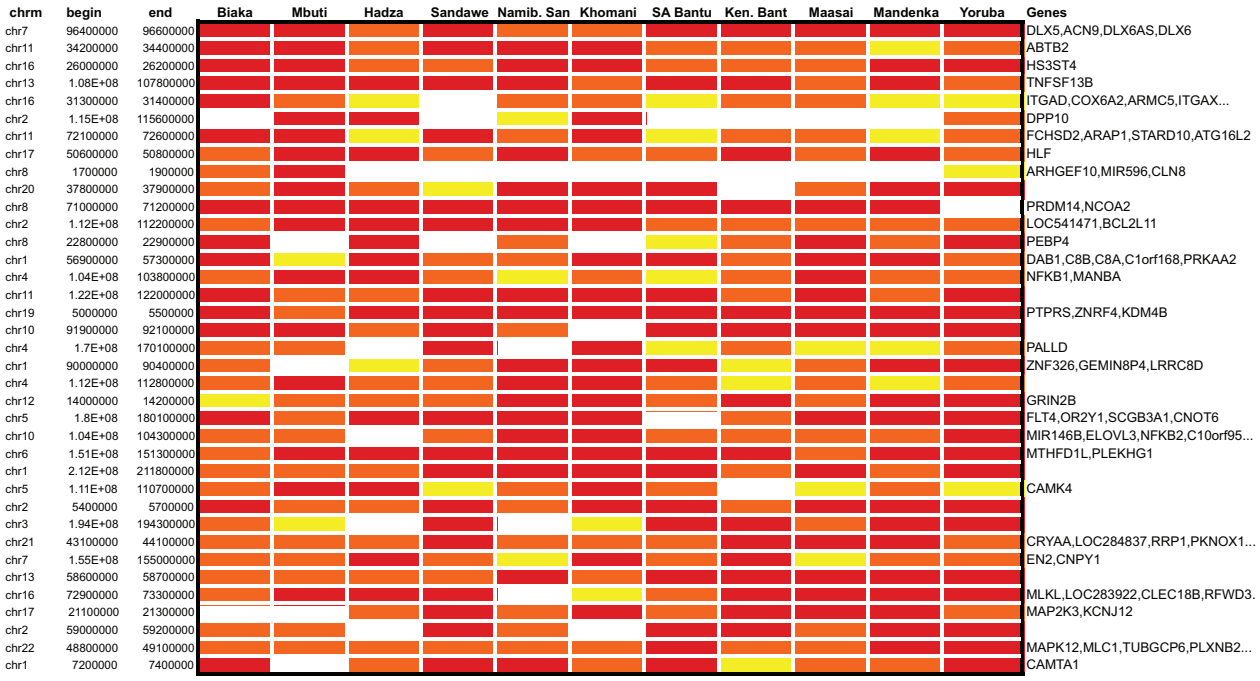

Figure S6. P-values for the 5 most extreme 100 kb genomic windows according to the XP-EHH CEU statistic in each population. Red indicates $p < 0.01$, orange indicates $p < 0.05$, yellow indicates $p < 0.10$, and white indicates $p > 0.10$. ``Genes'' column lists the genes located within the indicated windows.

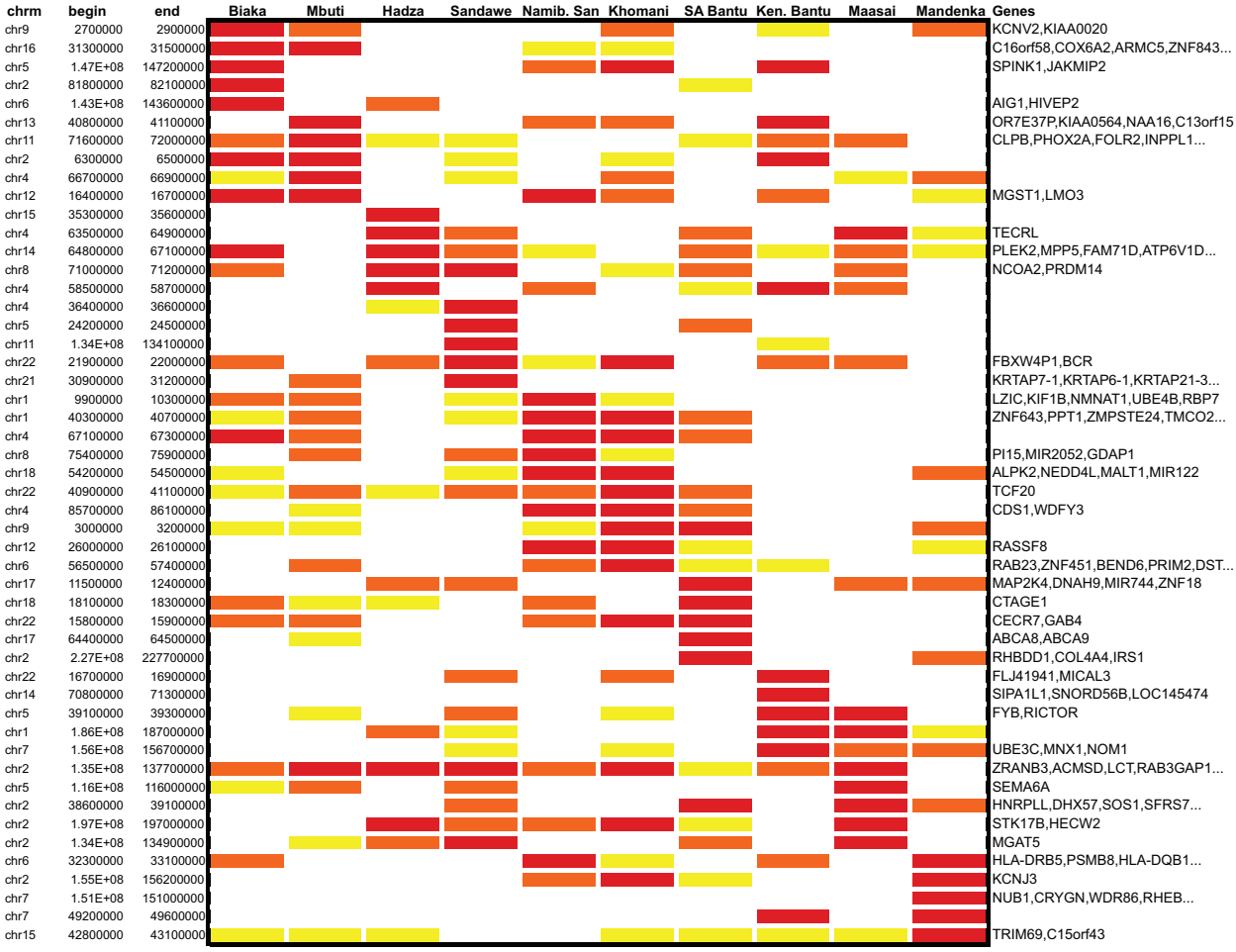

Figure S7. P-values for the 5 most extreme 100 kb genomic windows according to the XP-EHH YRI statistic in each population. Red indicates $p < 0.01$, orange indicates $p < 0.05$, yellow indicates $p < 0.10$, and white indicates $p > 0.10$. ``Genes'' column lists the genes located within the indicated windows.

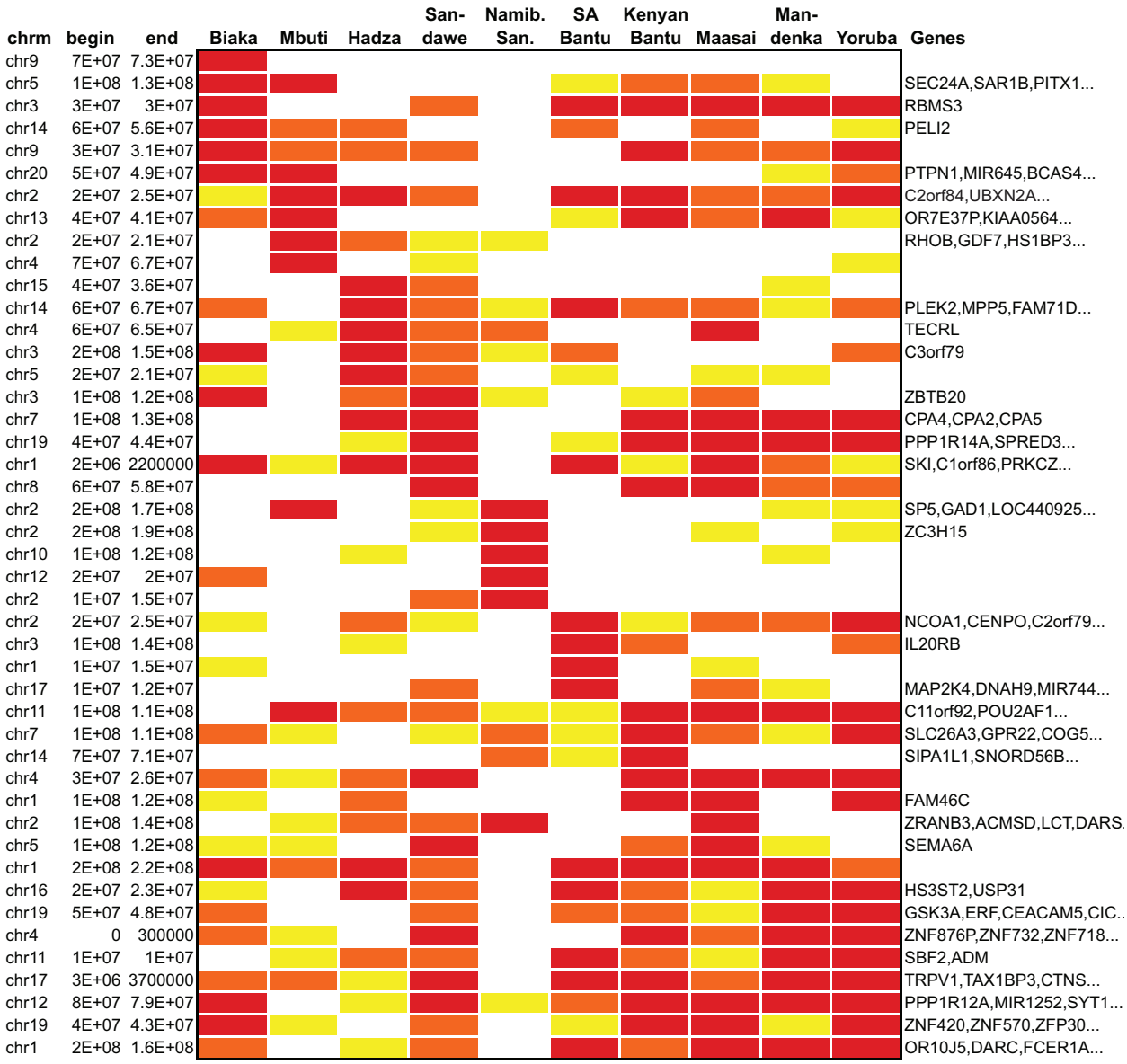

Figure S8. P-values for the 5 most extreme 100 kb genomic windows according to the XP-EHH KHB statistic in each population. Red indicates $p < 0.01$, orange indicates $p < 0.05$, yellow indicates $p < 0.10$, and white indicates $p > 0.10$. ``Genes'' column lists the genes located within the indicated windows.

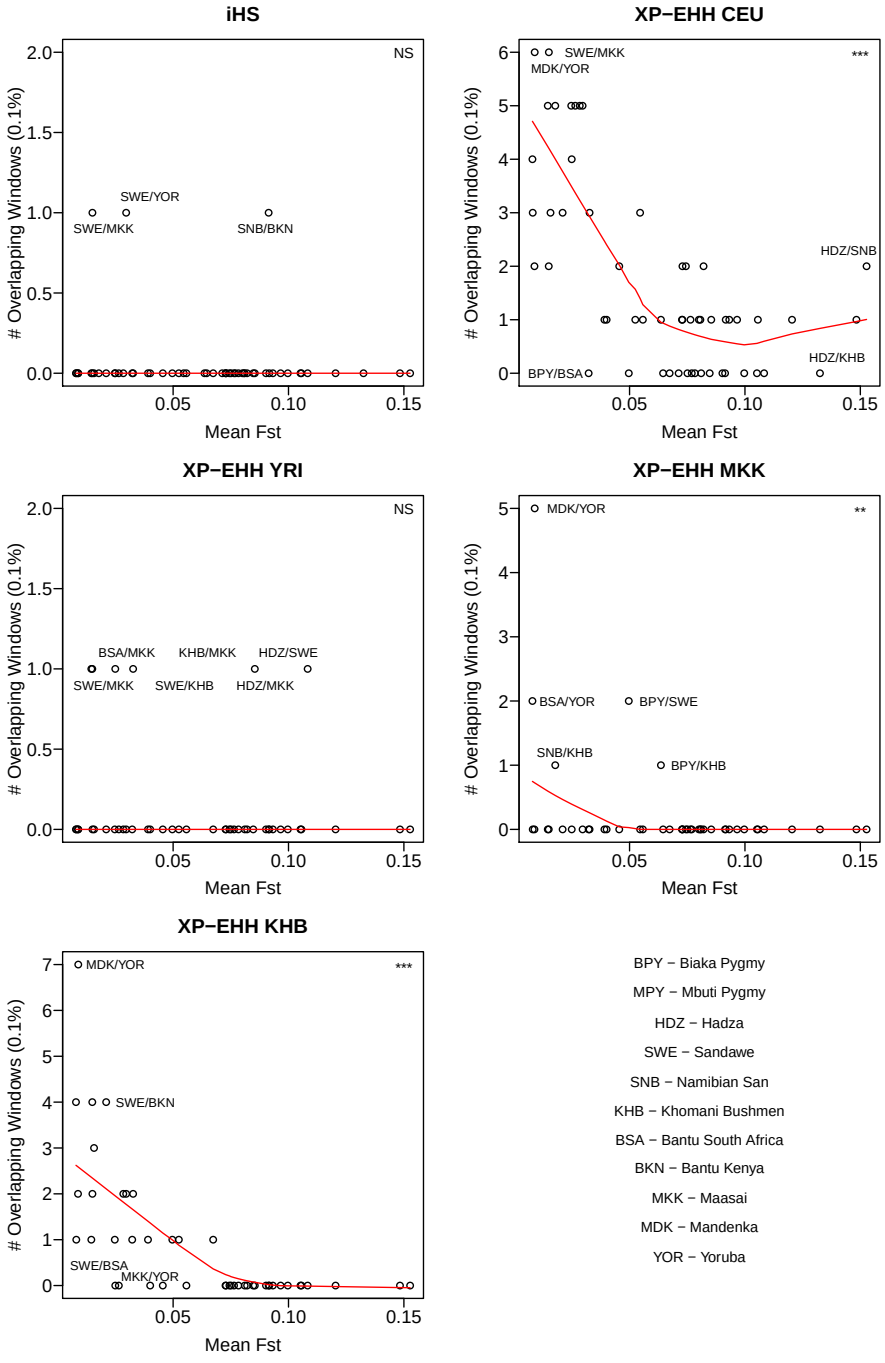

Figure S9. Number of shared 100 kb genomic windows in the top empirical 0.1% vs. mean F_{ST} for pairs of populations. Each point is a different population pair (some points are labeled, see key). Titles indicate the haplotype statistics for which the shared windows occur. Line is a best-fit *lowess* curve through the points. Significance of the p-value for the Mantel correlation between the x and y variables is indicated in the upper right corner: ``NS'': not significant, ``*'': < 0.05, ``**'': $< 0.01$, ``***'': < 0.001.

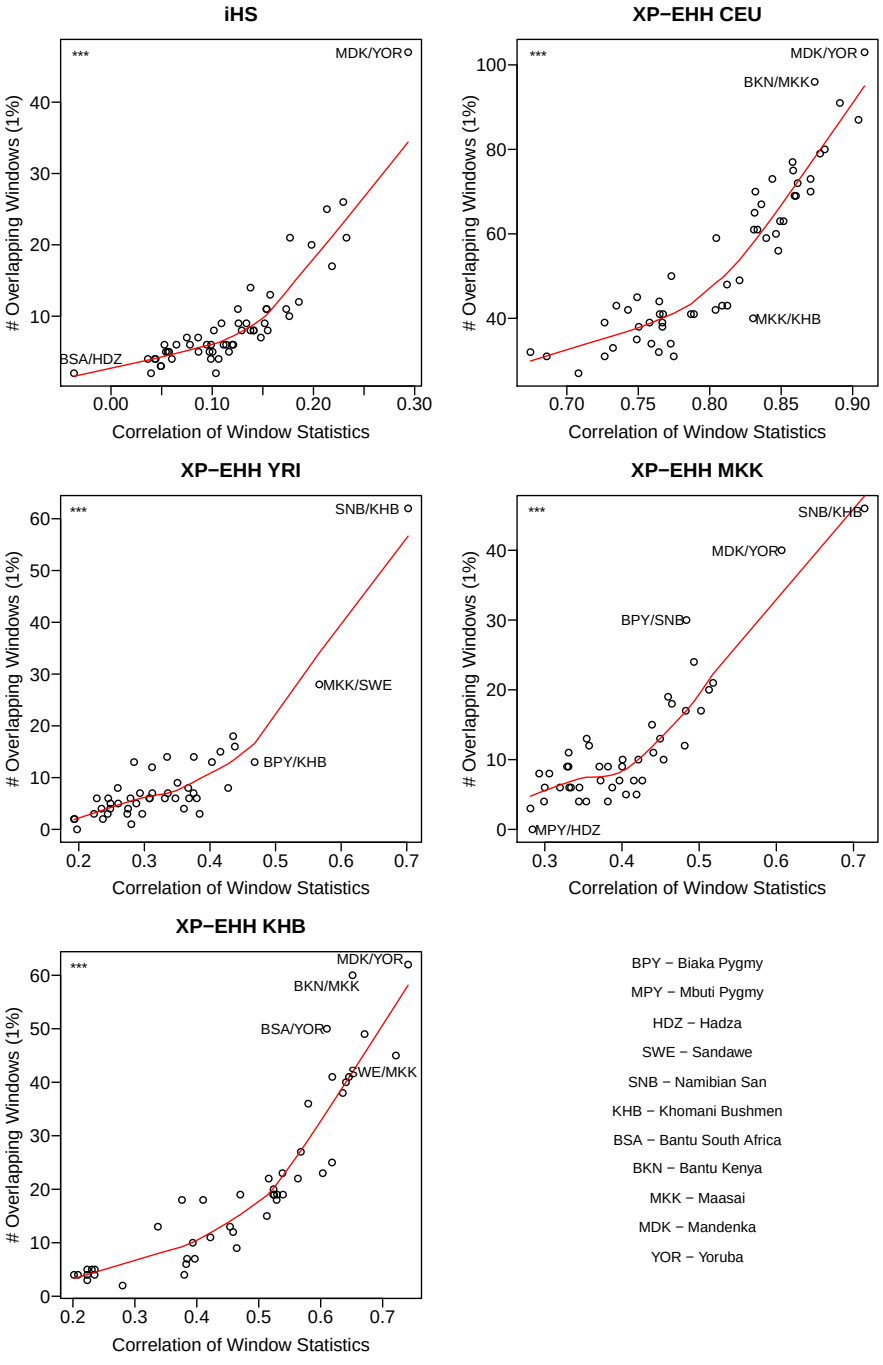

Figure S10. Number of shared 100 kb genomic windows in the top empirical 1% vs. correlations of window statistics over all genomic windows for pairs of populations. Each point is a different population pair. Titles indicate the haplotype statistics for which the shared windows occur. Line is a best-fit *lowess* curve through the points. Significance of the p-value for the Mantel correlation between the x and y variables is indicated in the upper left corner: ``NS'': not significant, ``*'': < 0.05, ``**'': $< 0.01$, ``***'': < 0.001.

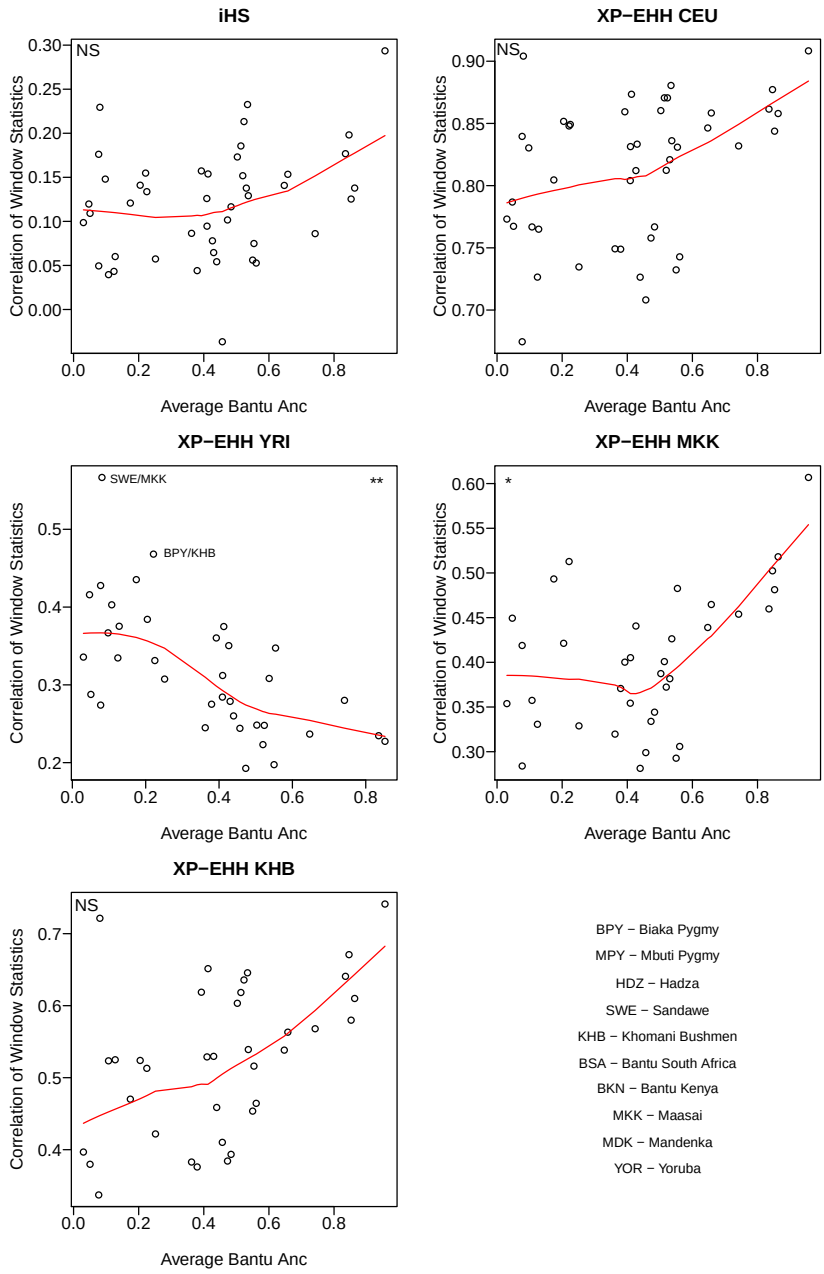

Figure S11. Correlations of window statistics over all 100 kb genomic windows vs. average Bantu ancestry for pairs of populations. Average proportion of Bantu ancestry is inferred from HENN *et al.* (2011), see Methods. Each point is a different population pair. Titles indicate the haplotype statistics for which the correlations are calculated. Line is a best-fit *lowess* curve through the points. Significance of the p-value for the Mantel correlation between the x and y variables is indicated in the upper right or left corner: ``NS'': not significant, ``*'': < 0.05, ``**'': $< 0.01$, ``***'': < 0.001. Note that the Namibian San are not included, as Bantu ancestral proportions for all individuals are not available from HENN *et al.* (2011).

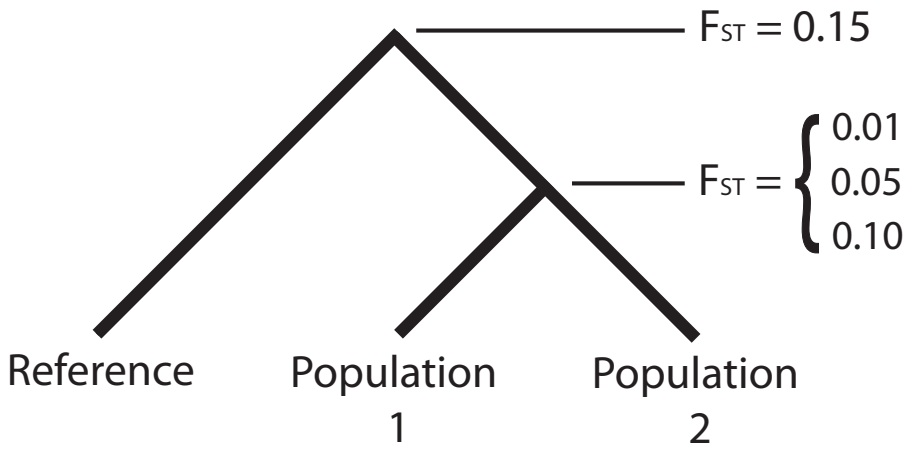

Figure S12. Demographic model simulated using *msms*. F_{ST} values determine the divergence times in the model; see main text.

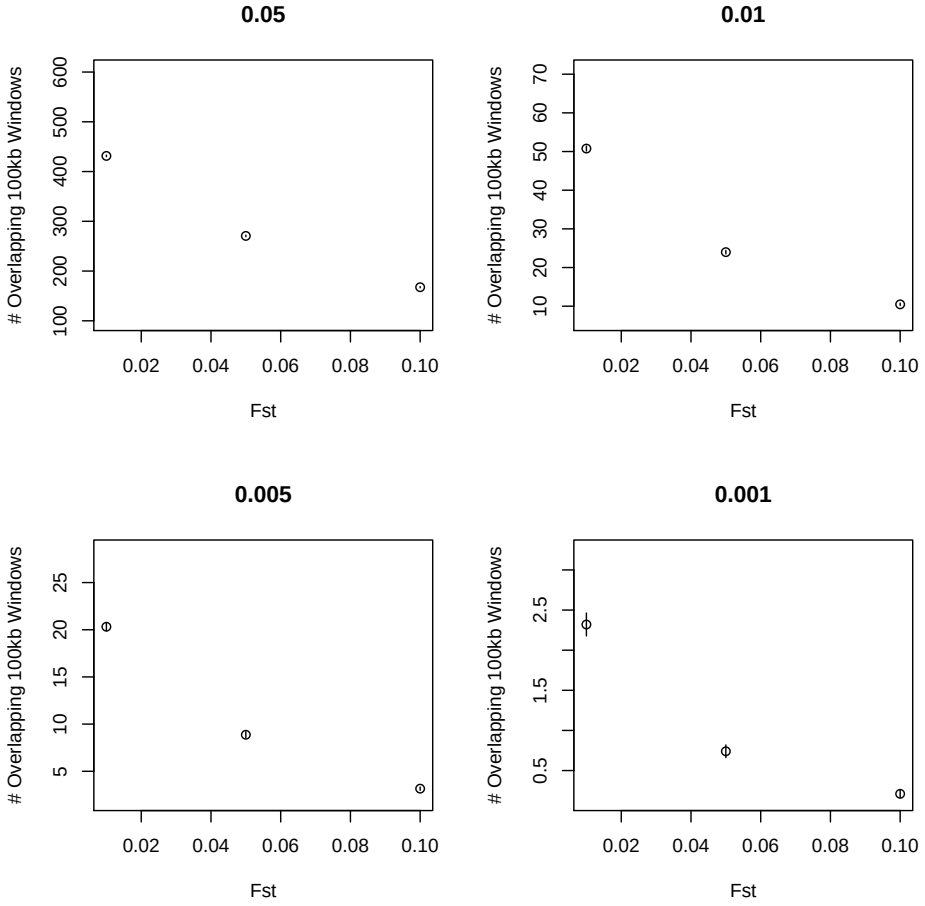

Figure S13. Number of overlapping 100 kb windows in the empirical tails of XP-EHH for pairs of populations simulated under neutrality. Means +/- standard errors over 100 simulations are shown for each value of F_{ST} simulated between Populations 1 and 2 (see Figure S12). The empirical tail examined in Populations 1 and 2 is shown above each plot.

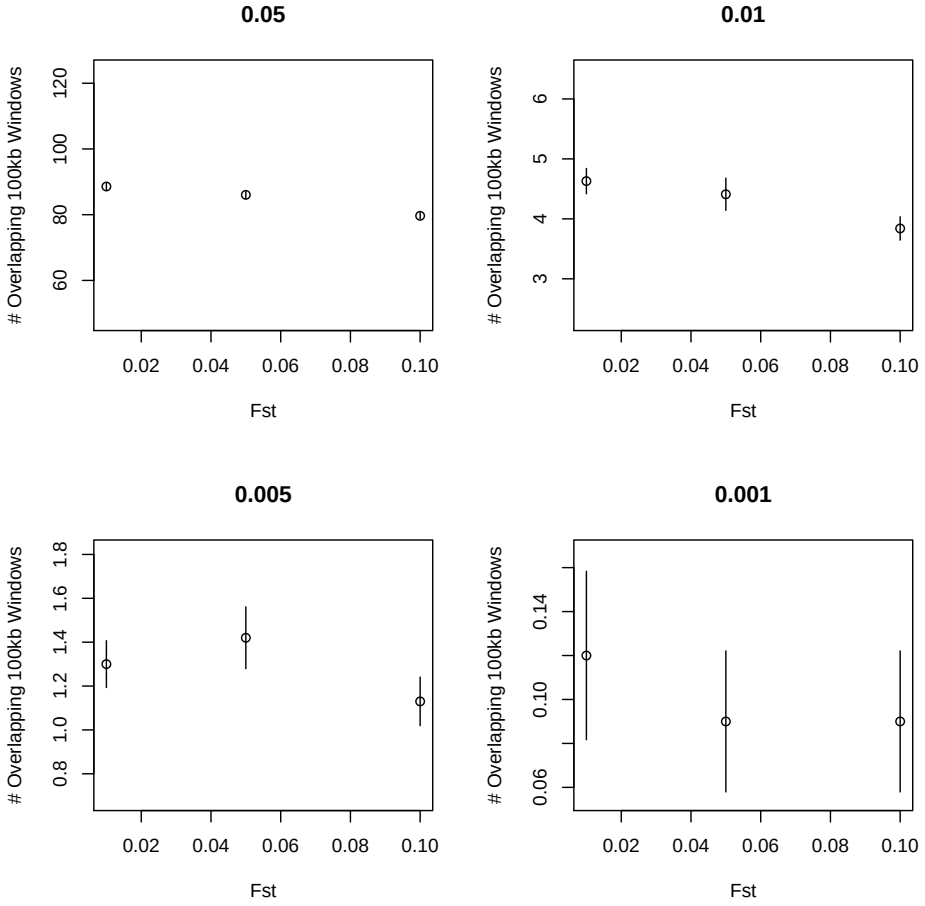

Figure S14. Number of overlapping 100 kb windows in the empirical tails of iHS for pairs of populations simulated under neutrality. Means +/- standard errors over 100 simulations are shown for each value of F_{ST} simulated between Populations 1 and 2 (see Figure S12). The empirical tail examined in Populations 1 and 2 is shown above each plot.

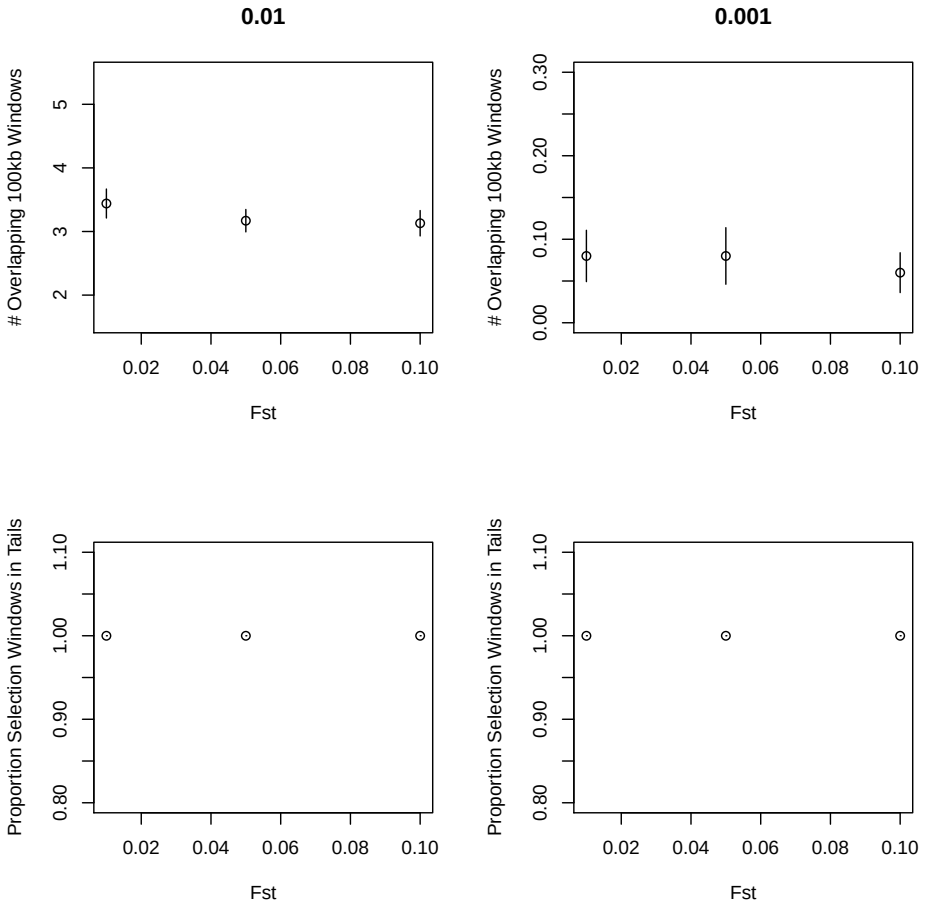

Figure S15. Number of overlapping 100 kb windows and true positive rate in the empirical tails of XP-EHH for pairs of populations simulated under recent selection. Strength of selection s = 20%, time of selection equal to 62 generations ago. The empirical tail examined is shown above each column, and the x-axes indicate the value of F_{ST} simulated between Populations 1 and 2 (see Figure S12). Top row y-axis: number of windows overlapping between the empirical tails of populations 1 and 2 (means +/- standard errors over 100 simulations). Bottom row y-axis: as an estimate of true positive rate, the proportion of 100 kb windows in the empirical tails of population 1 that fell within one of the 5 Mb regions in which a selective sweep was simulated (+/- standard deviation over 100 simulations).

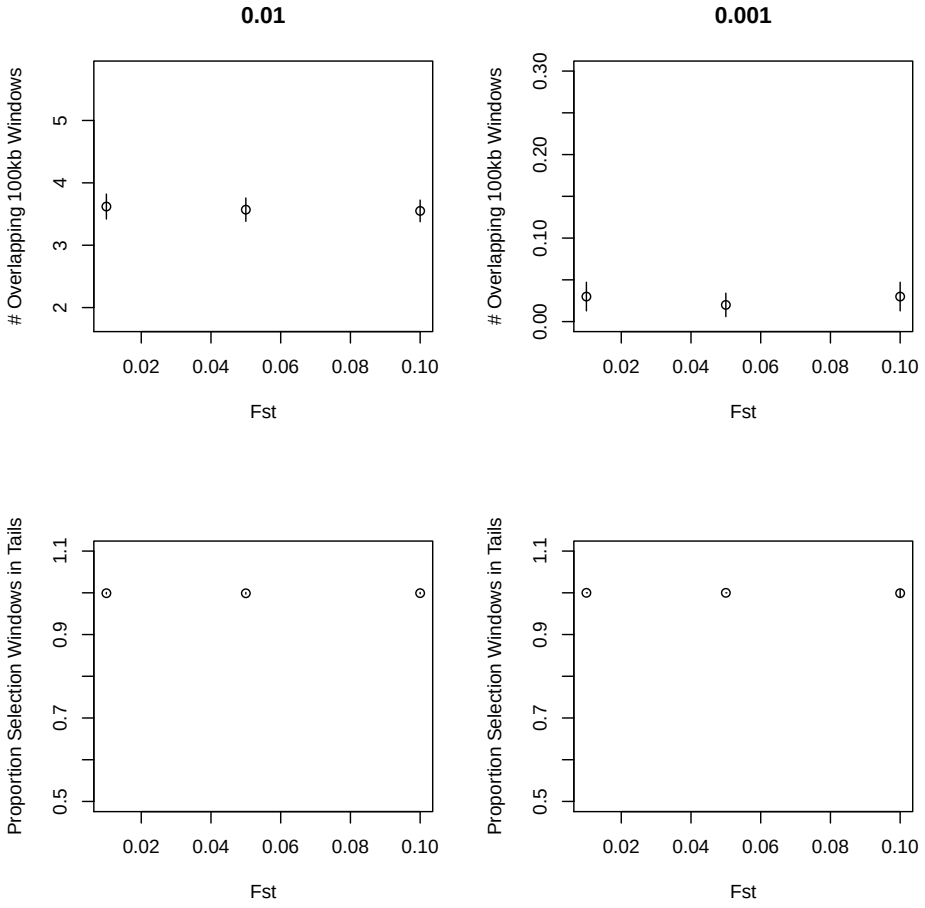

Figure S16. Number of overlapping 100 kb windows and true positive rate in the empirical tails of iHS for pairs of populations simulated under recent selection. Strength of selection s = 20%, time of selection equal to 62 generations ago. The empirical tail examined is shown above each column, and the x-axes indicate the value of F_{ST} simulated between Populations 1 and 2 (see Figure S12). Top row y-axis: number of overlapping 100 kb windows; bottom row y-axis: true positive rate. See caption for Figure S15.

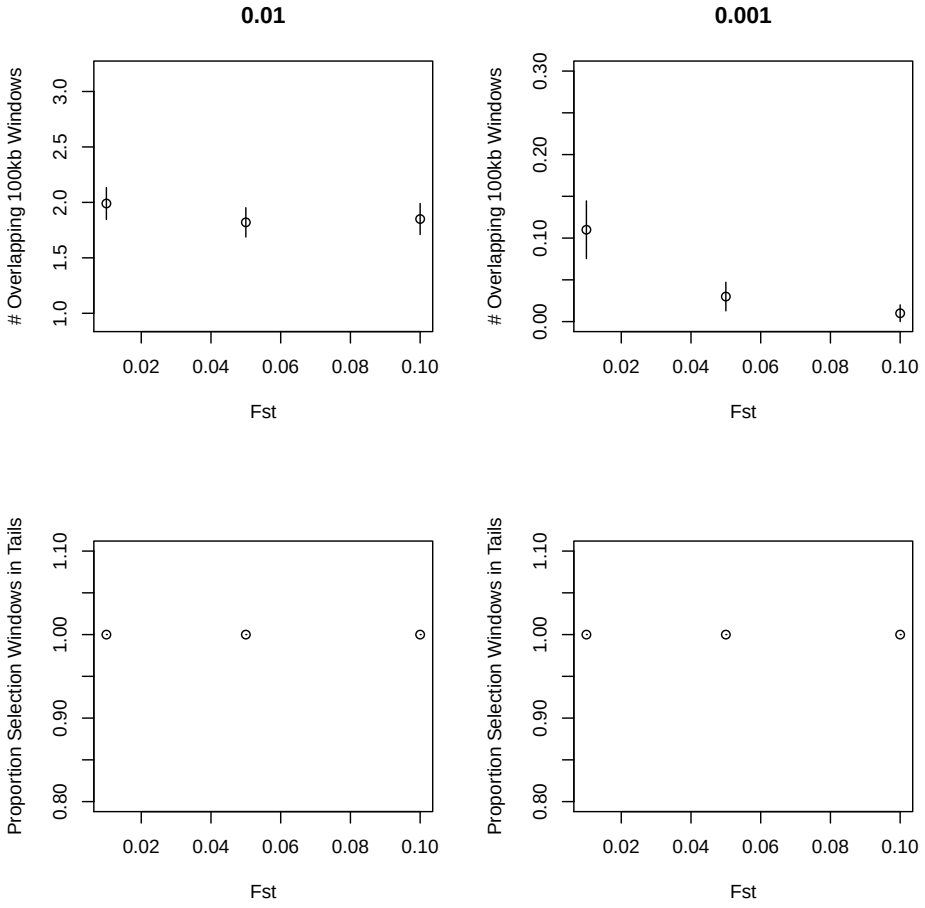

Figure S17. Number of overlapping 100 kb windows and true positive rate in the empirical tails of XP-EHH for pairs of populations simulated under less recent selection. Strength of selection s = 20%, time of selection equal to 144 generations ago. The empirical tail examined is shown above each column, and the x-axes indicate the value of F_{ST} simulated between Populations 1 and 2 (see Figure S12). Top row y-axis: number of overlapping 100 kb windows; bottom row y-axis: true positive rate. See caption for Figure S15.

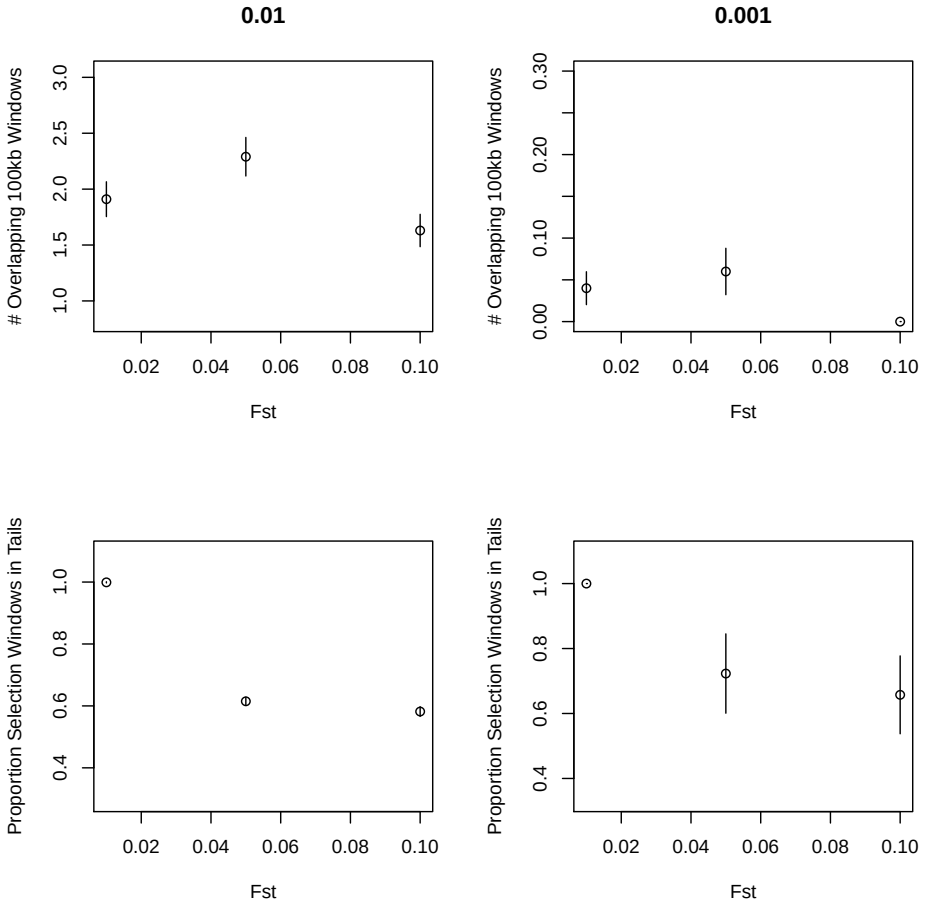

Figure S18. Number of overlapping 100 kb windows and true positive rate in the empirical tails of iHS for pairs of populations simulated under less recent selection. Strength of selection s = 20%, time of selection equal to 144 generations ago. The empirical tail examined is shown above each column, and the x-axes indicate the value of F_{ST} simulated between Populations 1 and 2 (see Figure S12). Top row y-axis: number of overlapping 100 kb windows; bottom row y-axis: true positive rate. See caption for Figure S15.

Tables

List of Tables

Table S1. Populations Analyzed.

Population	Source	Phased?	F_{ST}	iHS	XP-EHH CEU	XP-EHH YRI	XP-EHH MKK	XP-EHH KHB
Kenyan Bantu	HGDP	Yes	X	X	X	X	X	X
S. African Bantu	HGDP	Yes	X	X	X	X	X	X
Biaka	HGDP	Yes	X	X	X	X	X	X
Hadza	HENN *et al.* (2011)	Yes	X	X	X	X	X	X
Maasai	HapMap	Yes	X	X	X	X	REF	X
Mandenka	HGDP	Yes	X	X	X	X	X	X
Mbuti	HGDP	Yes	X	X	X	X	X	X
Namibian Bushmen	HGDP and SCHUSTER *et al.* (2010)	Yes	X	X	X	X	X	X
$\neq$ Khomani Bushmen	HENN *et al.* (2011)	Yes	X	X	X	X	X	REF
Sandawe	HENN *et al.* (2011)	Yes	X	X	X	X	X	X
Yoruba	HGDP	Yes	X	X	X	REF (test)	X	X
YRI	HapMap	Yes		X	X	REF	X	X
CEU	HapMap	Yes			REF			
Tuscans	HapMap	No	X					
All Europeans	HGDP	Yes			REF (test)			

``Phased?'' column indicates whether phased data were used (for haplotype analyses). An ``X'' indicates whether the population was used in the analyses of F_{ST} (and allele frequency differentiation), iHS, XP-EHH CEU, XP-EHH YRI, XP-EHH MKK, or XP-EHH KHB. ``REF'' indicates that population was used as a reference population. HGDP Europeans and HGDP Yorubans were only used as test reference populations for XP-EHH (see Supplemental Methods).

Table S2. Mean F_{ST} values between pairs of populations.

	Kenyan Bantu	S. African Bantu	Biaka Pygmy	Hadza	Maasai	Mandenka	Mbuti Pygmy	Namibian San	≠Khomani Bushmen	Sandawe	Yoruba
S. African Bantu	0.008										
Biaka	0.039	0.032									
Hadza	0.100	0.105	0.120								
Maasai	0.016	0.025	0.053	0.085							
Mandenka	0.015	0.015	0.046	0.108	0.028						
Mbuti	0.073	0.067	0.056	0.148	0.078	0.081					
Namibian Bushmen	0.091	0.073	0.074	0.153	0.090	0.097	0.092				
≠Khomani Bushmen	0.073	0.055	0.064	0.133	0.071	0.080	0.081	0.018			
Sandawe	0.021	0.025	0.050	0.085	0.015	0.033	0.075	0.082	0.065		
Yoruba	0.009	0.008	0.040	0.106	0.026	0.009	0.076	0.093	0.077	0.030	
Tuscan	0.135	0.147	0.172	0.180	0.095	0.146	0.193	0.198	0.172	0.108	0.148

Calculated over all autosomal SNPs. All populations are included except for the HapMap YRI, since the HGDP Yorubans are examined instead.

Table S3. Mantel correlations between matrices of the indicated variables for each haplotype statistic.

Haplotype Statistic	F_{ST} and		Correlation of Window Statistics and			Average Bantu Ancestry and	
	Overlap Top 1%	Overlap Top 0.1%	Overlap Top 1%	Overlap Top 0.1%	F_{ST}	Correlation of Window Statistics	Overlap Top 1%
iHS	-0.626 **(0.0002)**	-0.116 (0.212)	0.833 **(0.002)**	0.223 (0.057)	-0.669 **(0.003)**	0.300 (0.175)	0.468 **(0.029)**
XP-EHH CEU	-0.885 **(0)**	-0.664 **(0)**	0.894 **(0.0002)**	0.637 **(0.0004)**	-0.913 **(0)**	0.400 (0.123)	0.490 **(0.043)**
XP-EHH YRI	-0.200 (0.137)	-0.096 (0.346)	0.834 **(0.0002)**	0.310 (0.066)	-0.243 (0.164)	-0.643 **(0.008)**	-0.531 **(0.0004)**
XP-EHH MKK	-0.500 **(0)**	-0.341 **(0.001)**	0.866 **(0.0002)**	0.514 **(0.007)**	-0.747 **(0)**	0.455 **(0.043)**	0.443 **(0.023)**
XP-EHH KHB	-0.500 **(0)**	-0.647 **(0)**	0.869 **(0.0002)**	0.655 **(0.0002)**	-0.784 **(0.003)**	0.495 (0.063)	0.527 **(0.021)**

The indicated variables are calculated between all pairs of populations (creating matrices for which Mantel correlations can be calculated). P-values for the Mantel correlations, assessed by 5,000 randomizations, are indicated in parentheses and in bold if significant (p-value < 0.05). F_{ST}: mean F_{ST} over all SNPs between pairs of populations; Correlations of Window Statistics: correlations of window statistics across all genomic windows between pairs of populations; Overlap Top 1%: number of shared windows in the empirical top 1% between pairs of populations; Overlap Top 0.1%: number of shared windows in the empirical top 0.1% between pairs of populations.

Table S4. Significantly enriched Gene Ontology (GO) terms in the top empirical 1% and 5% of windows in each population.

Population	Statistic	GO term	Name of GO term	p value	Windows	Genes
Biaka Pygmy	XP-EHH CEU (0.05)	GO:0016571	histone methylation	0	6	*MEN1, EED, PRMT8, SUV420H2, NSD1, EHMT1*
	XP-EHH CEU (0.05)	GO:0007089	traversing start control point of mitotic cell cycle	$5x10^{-5}$	5	*CDK2, MDM2, CDC6, CDC25C, MTBP*
Namibian San	XP-EHH CEU (0.01)	GO:0006898 (P)	receptor-mediated endocytosis	0.00045	5	*PICALM, ASGR1, LRP1B, DAB2, IGF2R*
≠Khomani Bushmen	XP-EHH CEU (0.05)	GO:0006355	regulation of transcription, DNA-dependent	0.0	180	
Bantu SA	XP-EHH YRI (0.01)	GO:0006977	DNA damage response, signal transduction by p53 class mediator, cell cycle arrest	$5x10^{-5}$	3	*PML, GTSE1, GML*
Bantu Kenya	XP-EHH CEU (0.05)	GO:0030178	negative regulation of Wnt signaling pathway	$5x10^{-5}$	7	*DKK1, LRP1, GSC, TAX1BP3, FRZB, TLE1, BARX1*
Mandenka	iHS (0.01)	GO:002504 (also P)	antigen processing and presentation of peptide or polysaccharide via MHC Class II	0	3	*HLA-DRB5, HLA-DQB2, HLA-DPA1*
	iHS (0.01)	GO:0019882 (also P)	antigen processing	$5x10^{-5}$	5	*CD1A, HFE, HLA-DRB5, HLA-DQB2, HLA-DPA1*
	iHS (0.01)	GO:0015031 (P)	protein transport	0.00145	17	
	XP-EHH CEU (0.01)	GO:0006807 (P)	nitrogen compound metabolic process	0.00015	4	*NIT1, GLUL, NIT2, VNN1, VNN3*
	XP-EHH CEU (0.01)	GO:0006366 (P)	transcription from RNA polymerase II promoter	0.0009	8	*USF1, TRIM29, YBX2, HLF, SNAPC2, BAPX1, CREB5, POU6F2*
Yoruba	XP-EHH CEU (0.01)	GO:0007417	central nervous system development	$5x10^{-5}$	9	*ACCN1, NPTX1, NPAS1, ADAM23, DNER, CHRD, PARK2, POU6F2, GLI3*
HapMap YRI	XP-EHH CEU (0.01)	GO:0045786	negative regulation of cell cycle	0.0186	2	*HRASLS3, MOV10L1*
	XP-EHH CEU (0.01)	GO:0006366 (P)	transcription from RNA polymerase II promoter	0.00015	9	*PRDM4, MAF, MNT, YBX2, NFIX, BAPX1, CREB5, POU6F2, GATA4*
	XP-EHH MKK (0.01)	GO:0055088	lipid homeostasis	$5x10^{-5}$	3	*USF1, USF2, PPARG*
	XP-EHH MKK (0.01)	GO:0000432	positive regulation of transcription from RNA polymerase II promoter by glucose	0.0001	2	*USF1, USF2*
	XP-EHH MKK (0.01)	GO:0009991	response to extracellular stimulus	0.00025	2	*RASGRP4, RPS19*
	XP-EHH KHB (0.01)	GO:0009991	response to extracellular stimulus	0.0	2	*RASGRP4, RPS19*

Significance determined as described in Methods, using an FDR cutoff of 0.05, within each population. The ``Statistic'' column indicates the statistic resulting in significance; the numbers in parentheses after the statistic indicate the empirical tail for which an enrichment was tested (top 0.001 are shown in the main text). Significant GO terms are indicated in the ``GO Term'' column. Next to a GO term, ``(P)'' indicates that the term is only significant when examining the PANTHER subset of terms; ``(also P)'' indicates that the term is significant both when examining PANTHER terms and when examining all GO terms; if nothing is next to the term, it indicates significance only when all GO terms are examined. We count the number of windows in the indicated empirical tail to which each term was associated (``Windows'' column); rather than reporting those windows, we report the genes within those windows that were associated with the given GO term (in the ``Genes'' column).