Supplementary Information:
Recent selective pressure changes in human genes previously under balancing selection

Cesare de Filippo[1,*], Felix M. Key[1], Silvia Ghirotto[2], Andrea Benazzo[2], Juan R. Meneu[1], Antje Weihmann[1], NISC Comparative Sequence Program[3], Genís Parra[1], Eric D. Green[3], Aida M. Andrés[1,*]

1 Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

2 Department of Life Sciences and Biotechnology, University of Ferrara, Ferrara, Italy

3 National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland, USA

∗ **E-mail: Corresponding cesare_filippo(at)eva.mpg.de and aida.andres(at)eva.mpg.de**

Contents

1 Data

1.1 Sanger resequencing

Almost all exons of the genes were resequenced with Sanger technology. The exons were PCR-amplified and sequenced with unique primers in both directions. After removing the first and last 50 bp of the each electropherogram to minimize errors, SNPs were detected by means of Polyphred/Polyphrap (Bhangale, Stephens, and Nickerson, 2006). Additionally, we manually reviewed all genotypes associated with discordant results between overlapping amplimers, variants with a quality score lower than 99, singletons, and triallelic SNPs. The list of primers and other experimental details are available upon request.

1.2 Illumina resequencing

To increase the amount of sequence per gene we sequenced with Illumina a number of non-coding regions for each gene. Specifically, we targeted up to 2,000 bp of intronic regions around each exon, up to 2,000 bp of the 5'-UTRs and 3'-UTRs, 2,000 kb upstream (to include promoter regions) and 1,000 bp downstream. If there were individual UTR fragments shorter than 50 bp they were removed. For some intron-less genes (i.e. *PKDREJ* and *SDR39U1*) an extra of 1,000 bp was added to the downstream region. In total 138,634 bp were targeted with the array (Table S1). We designed an Agilent custom 244k array with probes that are 60 bp long and have a tiling of 1 bp. We implemented methods previously described to filter for repetitive probes in the human genome (Burbano, Hodges, Green et al., 2010).

Library preparation and multiplexing (50 individuals per pool) was done following the protocol described elsewhere (Meyer and Kircher, 2010), with the difference that individuals were double indexed in order to minimize multiplexing inaccuracies (Kircher, Sawyer, and Meyer, 2012). After capturing, the enriched fraction of the libraries was sequenced using the Illumina GAIIx platform with paired-end runs of 2x76 cycles and seven additional cycles for reading the index sequence. Base calling was performed with Ibis Kircher, Stenzel, and Kelso (2009) and reads were assigned to individuals according to their double-index sequence.

Reads were mapped to the human reference genome (hg19) using *BWA* software (Li and Durbin, 2009) with default parameters. We used Genome Analysis ToolKit (GATK) *IndelRealigner* v1.2-60 (McKenna, Hanna, Banks et al., 2010; DePristo, Banks, Poplin et al., 2011) to improve sequence alignment in and around insertion/deletions (indels). We excluded: (i) duplicated reads, (ii) reads that did not have a perfect pair-mate, and (iii) reads having Mapping Quality (MQ) lower than 25. Finally, we obtained approximately 20x coverage per individual. Figure S1 shows the distribution of the average coverage per individual for each population across the whole targeted region of the array; CHB have the highest values (23.8x) and GIH the lowest values (17.3x).

Given the properties of array capturing we were able to retrieve additional 21,457 bp (9.3% of the total) of sequences flanking the array's probes and that passed the coverage filter of 8x in 50% individuals (see later). Together with the sequences generated via Sanger technology, we analyzed a total of 230,452 bp (see Table S1)

Genotype calls were performed by means of the software GATK *UnifiedGenotyper* v1.2-60-g585a45b (McKenna et al., 2010; DePristo et al., 2011) on the complete dataset. Because the GATK *UnifiedGenotyper* v1.2 could not call multi-allelic sites, and therefore to detect them, we performed the genotype calls first for single individuals and subsequently we merged the single VCF files into one.

Because a modest amount (11%) of resequencing data, corresponding to 26,478 bp of the control regions (Table S1), was generated with two different technologies (Illumina and Sanger), we were able to ascertain the quality of our data. Given the lower error rate associated with Sanger sequencing we considered Sanger genotypes as the gold standard and compared it with the Illumina data. Without any filters specificity was 96.6%, that means on average 3.4% genotypes are erroneously called in the Illumina data (false positive). Applying a cutoff based on the quality of the SNPs (QUAL column in VCF file < 50) did not increase specificity. However, after applying strand bias filter (SB > 10) – i.e. by removing SNPs where one allele is mainly present on one DNA strand – specificity increased to 98.2%, the same value obtained if both QUAL and SB cutoffs were applied. In other words, the QUAL filter does not improve the quality of the data, but it reduces slightly the number of true positives. In summary, regardless of the filters applied the sensitivity was always higher then 99% and specificity ranged from 96.6% with no filters to 98.2% with our final set of filters (see Table S2 for more details. Therefore, to maximize the true positive rate and minimize the false positive rate we used the following thresholds:

1. Strand Bias (SB) ≤ 10;
2. SNP quality (QUAL) ≥ 20;
3. At least one individual with Genotype Quality (GQ) ≤ 10;
4. Coverage 8x in 50% individuals. We also disregarded genotypes with $\leq$ 3x coverage and those falling outside the 97.5% quartile of the coverage distribution specific for each individual. This filter removed only 12 sites, four and eight in genes and controls, respectively. This procedure increased the amount of missing genotypes from 2.2% to 3.3%.
5. We disregard multi-allelic sites, insertions and deletions (indels).

We took advantage of the software *FreeBayes* v0.9.6(Garrison and Marth, 2012) to reconstruct the chromosomal phase of SNPs that are on the same read (in our case within 170 bp considering paired-end reads). We use these first short haplotypes as known haplotypes to further infer the phase of variants with *fastPHASE* v1.4.0.(Scheet and Stephens, 2006). We observed on average of 90% concordance per individuals across three different runs of *fastPHASE*.

1.3 Other data: 1000Genomes

We used the 1000Genomes phase 1 data(1000 Genomes Project Consortium, Abecasis, Auton, Brooks, DePristo, Durbin, Handsaker, Kang, Marth, and McVean, 2012) applying the following filters:

1. We considered the same number of unrelated individuals (n=50) for all populations.

2. We considered SNPs that are ascertained in the low coverage data, having either the flag "LOWCOV" or "LOWCOV,EXOME" in the VCF file, respectively. This was done in order to avoid biases from coverage variation.

3. In order to avoid artifacts due to poor alignment, we considered only unique regions in the genome by disregarding positions outside the 50mer CRG Alignability track, which allows up to two mismatches (Derrien, Estellé, Marco Sola, Knowles, Raineri, Guigó, and Ribeca, 2012).

4. We removed positions that fall within simple units of repeat as detected by the Tandem Repeat Finder (TRF).

5. We disregarded regions that are segmental duplications (Cheng, Ventura, She et al., 2005; Alkan, Kidd, Marques-Bonet et al., 2009; Prüfer, Munch, Hellmann et al., 2012).

6. For some analyses, in particular those involving divergence, we considered only regions of the human genome that are orthologous to the chimpanzee reference genome (PanTro3).

2 Signatures of balancing selection are not due to mapping errors and duplications

Three (out of the four) target genes have more than 99% of the sequence mapping uniquely in the human genome when we consider 50 bp segments; in the fourth gene (*CLCNKB*), 60% of the sequence is unique with 50 bp segments and 80% is unique with segments of 100 bp (Figure S3), still a conservative criterion because our library insert sizes are on average much longer (i.e. 170 bp). Also, the distribution of coverage in these genes is not unusually high when compared to the remaining capture and sequencing array (making up to 63% of the total data, see Materials and Methods and Table S2), and coverage is similar across populations (Figure S4). Therefore mapping errors and undetected duplications are not responsible for the signatures of balancing selection we observe in African populations.

3 SNPs at intermediate allele frequencies in Africa and low or high frequency outside of Africa

We focused on SNPs with intermediate derived allele frequency ($0.20 \leq \text{DAF} \leq 0.80$) in Africa and low (DAF ≤ 0.05) or very high (DAF ≥ 0.95) frequency in non-Africans; we call these alleles *intermediate in Africa different Out-of-Africa alleles* (*iAdO-alleles*). Among these *iAdO-alleles* there are 22 non-synonymous (Figure 3B), 20 synonymous and 115 non-coding variants. We focused on the 22 non-synonymous *iAdO-alleles* as they have a higher probability of having functional consequences. To further quantify allele frequency differences we

calculated F_{ST} (Weir and Cockerham 1984) between two African (YRI and LWK) and two non-African (TSI and CHB) populations from the 1000 Genomes dataset, and compared it with the empirical genome-wide F_{ST} distribution of coding alleles conditioned on the frequency in African population. Although, only two *iAdO-alleles* fall in the top 5% of the F_{ST} empirical distribution in at least one Africa-Eurasia pairwise comparison (Table S5), 95% of all pairwise comparisons are among the most differentiated SNPs ($p < 0.50$) showing large allele frequency differences between these populations. Figure 4 shows the allele frequencies in the 1000 Genomes populations of the most differentiated SNPs, but the pattern is similar for all non-synonymous *iAdO-alleles* (Figure S9). We note that many *iAdO-alleles* are not independent because of high linkage disequilibrium.

4 Approximate Bayesian Computation

4.1 Simulations

We analyzed 160,000 simulations for each of the five evolutionary scenarios to model the changes in selective pressure after the out-of-Africa migration (Figure 5A). In all models a balanced polymorphism arises Tbs generations ago with selection coefficient Sbs and frequency equilibrium of approximatively 0.50, and is maintained until present day in African populations. The five models differ in the selective regime of the non-African populations, with changes straight after the out-of-Africa migration event as described in Results and Figure 5A. We used the same number of simulations in each model to avoid biases produced by different number of simulations (Bertorelle, Benazzo, and Mona, 2010). In the *B-Pdn* model we conditioned on the new advantageous mutation not to be lost in Europeans and Asians.

We fixed the dominance coefficient (h = 25.5) in order to have a frequency equilibrium of 0.51, which is not exactly 0.50 because the two homozygotes cannot have the same fitness (w) according to the model used in *SLiM* (Messer, 2013), where the fitness of the three genotypes are: $w_{AA} = 1$, $w_{Aa} = 1+hs$, $w_{aa}=1+s$ (Gillespie, 1978). The following parameters were drawn from uniform prior distributions: mutation rate, $\mu = U$(1e-8, 4e-8 per site per generation); recombination rate, $\rho = U$(0, 4e-8 per site per generation); time since balancing selection, Tbs = U(40000, 240000 generations); selection coefficient of the balanced polymorphism, Sbs = U(0.0001, 0.1); selection of the de novo advantageous mutation in model *B-Pdn*, Sps = $logU$(0.0001, 0.01).

All other parameters of the demographic model are the same as in the simulations for the neutrality tests (Gravel, Henn, Gutenkunst, Indap, Marth, Clark, Yu, Gibbs, 1000 Genomes Project, and Bustamante, 2011), with the exception that we did not allow migration. We set the divergence time between human and chimpanzee to 6.5 million years, i.e. 260,000 generations considering a generation time of 25 years. Furthermore, in order to increase the speed of the simulations we scaled by a factor of 5 the parameters of the forward simulations. This scaling did not affect the summary statistics we used. Figure S10 shows the schematic demographic models and its parameters considered in the simulations.

4.2 Choice of summary statistics

The choice of summary statistics (SuSt) is fundamental in an ABC framework (Beaumont, Zhang, and Balding, 2002; Bertorelle et al., 2010). It is necessary to maximize the amount of information retained and to minimize redundant information; SuSt are indeed often not "sufficient" to summarize the data and are correlated with each other. Table S6 lists 25 SuSt calculated for each African, European and Asian population (total = 75), and two SuSt between populations (i.e. F_{ST}). Orthogonal transformation of summary statistics such as Partial Least Squared, PLS (Wegmann, Leuenberger, and Excoffier, 2009; Wegmann, Leuenberger, Neuenschwander, and Excoffier, 2010) were developed for parameters estimation procedures (single model), but not for model comparison (multiple models). Therefore, in order to minimize both the loss of information and the correlation between SuSt, we detected those able to better discriminate the five evolutionary models and that exhibit relatively low correlation (Benazzo, Ghirotto, Vilaça, and Hoban, 2015; Veeramah, Wegmann, Woerner, Mendez, Watkins, Destro-Bisol, Soodyall, Louie, and Hammer, 2012), as follows.

First, we considered the distribution of each SuSt from 10,000 simulations for each model and performed a Mann-Whitney U (MWU) test to ascertain whether the distributions differed between models. We calculated the mean *p-value* across the 10 pairwise MWU tests performed among the five models. As expected and as a control for our simulations, no SuSt of African populations showed any significant MWU *p-value* because the five models for Africans are identical. Second, we calculated the correlation coefficient (as Pearson r^2) between all SuSt and considered as 'moderately correlated' those SuSt with $r^2 \geq 0.8$ (an arbitrary cutoff). For pairs of SuSt with an $r^2 \geq 0.8$, we kept the SuSt with the lowest mean *p-value* of MWU tests in the previous step, meaning that it better discerns between the different selection models. We did this for each model and retained 16 SuSt, which we call informative SuSt (see Table S6 for the complete list).

4.3 F_{ST} correction for migration

Migration among populations was absent in the simulations of the evolutionary models because it would create a mixture of different types of natural selection. In essence, the simulation software SLiM (Messer, 2013) did not perform as we desired with population-specific natural selection. For instance, when simulating any of the '*Balancing to Positive*' models (Figure 5), if migration carries the positively selected allele from Eurasia to Africa, there would be positive selection acting simultaneously with balancing selection in Africa; the same would happen if migration carries the balanced polymorphism into Eurasia.

We thus set migration to 0 and indirectly corrected the F_{ST} distribution for this difference between real data and the simulations. In essence, we aimed to quantify the influence of migration in F_{ST} and use this knowledge to correct the calculated F_{ST} in the five simulated evolutionary models. Specifically, we used 2,000 neutral simulations with and without migration (4,000 simulations in total), and for each type of simulations we generated the distribution of mF_{ST} (mean F_{ST} across sites) and sdF_{ST} (standard deviation of F_{ST} across sites). We found that all values have similar bias. Specifically, mF_{ST} and sdF_{ST} between Africa and Europe

with migration were 19.3% and 19.5% lower than those without migration, and a similar reduction for mF_{ST} and sdF_{ST} (both 16.7%) was between Africa and Asia. As expected, F_{ST} between Europe and Asia with migration exhibited less reduction (7.0%) compared to those without migration given that the two populations are more closely related. Considering the means or the modes of the distributions did not change the results. Therefore, we then corrected each mF_{ST} and sdF_{ST} in the simulations with selection between Africa and Europe reducing them by 19.3% and 19.5%, respectively, and the mF_{ST} and sdF_{ST} between Africa and Asia by 16.7%. We did not correct F_{ST} statistics between Asia and Europe because they resulted to be not informative in the ABC model selection (see later).

This correction did not affect the power of the ABC approach, which is the same when using the corrected and uncorrected F_{ST} (see next paragraphs). In addition, it did not alter the results of the model selection (Figure S11). The only differences are associated with an average increase of 2.2- and 1.4-fold in the posterior probability of models *B-B* and *B-N*, respectively after excluding one outlier (i.e. model *B-N* for *PKDREJ* in GIH). These models tend to be favored because they generate generally lower F_{ST} than the other models (Figure S12).

4.4 ABC model selection

Although the five models considered in this study are quite complex, they are only broad approximations to potential real evolutionary scenarios after the out-of-Africa. In order to incorporate some scenarios where the selective pressure is not common between non-African populations, we performed the model selection separately for European and for Asian populations. For instance, if in reality a new advantageous mutation arises in Europeans after their 'split' with Asians, while the balanced polymorphism is still selected in Asians, the model selection should favor model *B-P* in Europeans and model *B-B* in Asians. Two different ABC model selection approaches were performed.

4.4.1 Hierarchical ABC model selection

The first model selection is carried out using the logistic regression approach (Beaumont, 2008) in two hierarchical steps. The logistic regression is fitted where the model is the categorical dependent variable and the summary statistics are the predictive variables. The regression is local around the vector of observed SuSt, and the probability of each model is finally evaluated at the point corresponding to the observed vector of summary statistics. In the first step, we run the model selection considering the three models *B-Pcfe*, *B-Psv*, and *B-Pdn*; subsequently, we selected among these three models the one with the highest posterior probability and performed the model selection analysis by comparing it with models *B-B* and *B-N*. To estimate the posterior probabilities of each model we considered the 50,000 (out of 480,000) best simulations in every comparison.

The power of inferring the correct model was estimated by generating 1,000 pseudo-observed datasets (PODs) according to each model analyzed, with parameter values randomly chosen from the correspondent prior distri-

bution. We analyzed these PODs by means of the same ABC framework applied in the model selection (i.e. with logistic regression and 50,000 retained simulations). We considered all four different sets of models separately, hence considering all the possible comparisons that might appear in our hierarchical model selection procedure:
(i) *B-Psv*, *B-Pcfe* and *B-Pdn*;
(ii) *B-B*, *B-N* and *B-Psv*;
(iii) *B-B*, *B-N* and *B-Pcfe*;
(iv) *B-B*, *B-N* and *B-Pdn*.
In all comparisons, we evaluated for each model the proportion of cases where it is correctly recognized as the true model (i.e. true positives) and when it is not (i.e. false positives) considering a posterior probability threshold of 0.5 to assign the support. In other words, we considered as false positives situations where the POD is generated under one model, but the model selection procedure assigns the support to one of the other two tested models. For example, a false positive occurs when the PODs are generated under the model *B-B* but the ABC method assigns a posterior probability of only 0.10 to this model, and probabilities of 0.30 to *B-N* and 0.60 to *B-Psv*.

We performed this power analysis for each gene, which reflects the power for different sequence lengths (5,800 bp for *SDR39U1*, 8,800 bp for *PKDREJ*, and 10,000 bp for *CLCNKB* and *ZNF473*), separately for Europeans and Asians, using the 16 informative summary statistics.

Tables S7 and S8 show the results of the power analyses using the two triplets of models as explained in the previous paragraph. The true positive rate was good for model B-B, moderate for any of the B-P models but modest for the model *B-N*. The true positive rate for the three models *B-P* was small, with model *B-Pdn* having the higher rate (47%) and model *B-Psv* the lowest (19%). The power associated with the model choice can be seen in Figure S13, where we used principal component analyses of 500 random simulations to show how similar or different the models are. Figure S13A shows that *B-B* and *B-P* are fairly separated from each other and the model *B-N* is in between the two, as a result of lower true positive rate. Figure S13B further shows how poorly separated (and therefore with low power) the three models of *B-P* are, with only modest differentiation in the model *B-Pdn*. This is expected since positive selection on a de-novo mutation can be detected by classical population genetics tests (Przeworski, Coop, and Wall, 2005; Nielsen, Hellmann, Hubisz, Bustamante, and Clark, 2007; Pritchard, Pickrell, and Coop, 2010; Messer and Petrov, 2013).

4.4.2 All-together ABC model selection

The second model selection take into account potential biases that can be present because the similarities of the three models *B-P*. Therefore, we considered all five model together in one step and assigned to the models the same priors probabilities: 1/3 for *B-B*, 1/3 for *B-N*, 1/9 for each *B-P* model (1/3 in total). As in the hierarchical approach, we used the logistic regression (Beaumont, 2008) and retained 50,000 (out of 480,000) best simulations to estimate the posterior probabilities of each model. The results are shown in Figures S14C-D and are similar to the hierarchical model selection (Figure 5B-C, and Figures S14A-B). Because our aim is

to discriminate among *B-B*, *B-N* and *B-P models*, we present both the posterior probabilities for each model (Figure S14D), as well as the combined probabilities of the three *B-P* models (Figure S14C).

5 Biased gene conversion

GC-biased gene conversion could favor the segregation and fixation of weak-to-strong (from A/T to C/G) mutations and mimic the pattern of positive selection, i.e. excess of low-frequency variants and substitutions (Galtier, Piganeau, Mouchiroud, and Duret, 2001). We did not observe any evidence for enrichment of weak-to-strong mutations not only in non-African but also in African populations for all four genes. Actually, we do observe the opposite pattern of more strong-to-weak (from C/G to A/T) mutations although not significantly so for all populations.

6 Excess of diversity over divergence

Within the 400,000 bp region spanning *PKDREJ* the two highest peaks of *PtoD* are within the genes *PKDREJ* and *PPARA*. However, 83% of the sequence of *PPARA* does not pass our filtering criteria in the 1000 Genomes data (see SOM section 1.3 and Materials and Methods). For *SDR39U1* the highest peak is slightly downstream of the gene but *SDR39U1* is the closest gene (Figure 2C). Actually, this highest peak corresponds to an uncharacterized long RNA gene *LOC101927045*. Furthermore, 50% of the sequence of *SDR39U1* overlaps with the *KHNYN* gene, which also shows two peaks of high diversity. With regards to *ZNF473*, high peaks of diversity are located within 10,000 bp distance up- and down-stream of the gene, and these regions correspond to *VRK3F* and *LJ26850* genes, respectively.

7 Functional consequences of the SNPs within the candidate genes

Given that *CLCNKB* did not show strong evidence of changes in selective pressure in non-African populations (see Figure 5 and the main text for more details), we focused on the other three candidate genes in these analyses.

We consider the C-scores (Kircher, Witten, Jain, O'Roak, Cooper, and Shendure, 2014) as measurements of the levels of deleteriousness of the SNP: The higher the C-score, the higher is the probability of having functional consequences. First, we asked whether the non-synonymous SNPs in the three candidate genes show different C-scores than the rest of non-synonymous SNPs in the genome, and find that there are no significant differences at any frequency bins conditioning on allele frequency in Yoruba from the 1000 Genomes (Figure S15). However, when we repeat the same analyses with all SNPs (non-synonymous, synonymous and non-coding) there are significantly higher C-scores than those in the rest of the genome in almost all frequency bins (Figure S15B). This further suggests a potential regulatory role for the SNPs in the three candidate genes. It should be noted

that in the last bin of frequency (between 0.2 and 0.5) we report two distinct distributions of C-scores with and without conditioning the SNPs to be at low- or high-derived frequency in non-Africans. The latters are what we defined iA-alleles and the formers are *iAdO-alleles* (from *intermediate in Africa different Out-of-Africa*).

We further asked whether the SNPs in the three candidate genes have a potential role in regulation by using *RegulomeDB* (Boyle, Hong, Hariharan et al., 2012), which combines data sets from ENCODE and other sources, and classifies variants based on their likelihood to be located in a functional region. We divided the SNPs in two groups: *iAdO-alleles* (their category is reported in Table S9) and all the 'other alleles'. We found that both groups of SNPs are enriched for categories of high scores (Figure S16), i.e. those belonging to categories "1". In other words, these sites have a high-predicted likelihood to affect transcription factor binding and are mapped to an eQTL (Boyle et al., 2012). The enrichment was performed by 1,000 samplings of three random genes (i.e. the number of our candidate genes) within the genome and none of the samplings showed a proportion of SNPs belonging to categories "1" higher than that observed in the three candidate genes (i.e. $p < 0.001$). More interestingly we found that the proportion of SNPs belong to category "1" is significantly higher (exact binomial test $p = 4.2\text{e-}06$) for the *iAdO-alleles* (31%) than for the rest of alleles (13%).

8 Other mechanisms of balancing selection

There are several possible mechanisms of balancing selection, including overdominance (Allison, 1956), frequency-dependent selection (Wright, 1939), fluctuating selection (Gillespie, 1978), as well as pleiotropy (Gendzekhadze, Norman, Abi-Rached, Graef, Moesta, Layrisse, and Parham, 2009). The double signature of excess of polymorphisms (due to unusually old time to the most recent common ancestor, T_{MRCA} and excess of intermediate-frequency alleles observed in Africa are expected under balancing selection with frequency equilibrium around 0.5, such as overdominance (with similar fitness of both homozygotes), frequency-dependent selection (with favored frequency close to 0.5) and perhaps pleiotropy. These mechanisms could explain the concordant signatures we observe in the two African populations. Nevertheless, none of them would predict the very different signatures we observe in Eurasia (where alleles segregate at extremely high or low frequency) in the absence of changes in selection out of Africa. In fact, these mechanisms of balancing selection are expected to result in modest differences across all populations considered.

Other types of frequency-dependent selection such as rare allele advantage (negative frequency-dependent selection) might in theory increase the T_{MRCA} and the levels of genetic diversity as opposed to neutral expectations, but the extend of this increase is smaller than that of overdominance (Takahata and Nei, 1990). The SFS under rare allele advantage is though expected to be enriched in low-frequency alleles, rather than in intermediate-frequency ones (Tellier, Moreno-Gámez, and Stephan, 2014). This is opposite to what we observe in the two African populations and, therefore, this type of selection is extremely unlikely in the four genes.

Mildly fluctuating selection (where the range of allele frequencies of the selected allele is modest) could result in very similar patterns to the ones we observed in Africa: Old T_{MRCA} and excess of intermediate-frequency

alleles (if it happens to sample the population at a time when the selected allele is at about 0.5 frequency). This model would require the two African populations to be in synchrony, so that at time of sampling both populations have a similar frequency of the balanced allele. As before, though, in the absence of changes in selection mild fluctuating selection would not predict the very high and low allele frequencies at which these alleles are present in non-African populations (Table 2, Figures 3 and S7).

Strongly fluctuating selection (where the selected allele fluctuates strongly in time) is perhaps unlikely to maintain the balanced polymorphism for millions of years. If it did, we would not expect a strong excess of linked polymorphism at intermediate frequency, as alleles would eventually be lost or fixed by drift when they reached low or high frequency, respectively. Also, the four genes are enriched in intermediate-frequency alleles in the two separate African populations (i.e. between 20% and 80%), which is also unlikely under strong fluctuating selection unless the two African populations are somehow in synchrony. This is the only type of selection that would predict the differences we observe between African and non-African populations (if they were out of synchrony in the fluctuation), but it seems unlikely to explain the signatures in Yoruba and Luhya.

Since a mechanism that maintains alleles long-term at frequency close to 0.5 seems most likely given the data, we focused on this case. We note that we simulate balancing selection with overdominance. This has practical advantages, as this is a type of selection that we understand (and can simulate) well, and since it can simulate reasonably well the consequences of other types of selection that we consider likely under the patterns in Yoruba and Luhya: Frequency-dependent selection with frequency equilibrium at 0.5 and mild fluctuating selection that maintains alleles, long-term, close to 0.5 average allele frequency.

Web Resources

The URLs for data presented herein are as follows:

Agilent www.genomics.agilent.com

ENCODE https://www.encodeproject.org

UCSC Genome Browser http://genome.ucsc.edu

RegolomeDB http://regulomedb.org

References

1000 Genomes Project Consortium, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT, McVean GA. 2012. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 491:56–65.

Alkan C, Kidd JM, Marques-Bonet T, et al. (13 co-authors). 2009. Personalized copy number and segmental duplication maps using next-generation sequencing. *Nat Genet*. 41:1061–7.

Allison AC. 1956. The sickle-cell and haemoglobin C genes in some African populations. *Ann Hum Genet*. 21:67–89.

Beaumont MA. 2008. Joint determination of topology, divergence time, and immigration in population trees, Cambridge: Cambridge: McDonald Institute for Archaeological Research, pp. 134–1541.

Beaumont MA, Zhang W, Balding DJ. 2002. Approximate Bayesian Computation in population genetics. *Genetics*. 162:2025–35.

Benazzo A, Ghirotto S, Vilaça ST, Hoban S. 2015. Using ABC and microsatellite data to detect multiple introductions of invasive species from a single source. *Heredity (Edinb)*. 115:262–72.

Bertorelle G, Benazzo A, Mona S. 2010. ABC as a flexible framework to estimate demography over space and time: some cons, many pros. *Mol Ecol*. 19:2609–25.

Bhangale TR, Stephens M, Nickerson DA. 2006. Automating resequencing-based detection of insertion-deletion polymorphisms. *Nat Genet*. 38:1457–62.

Boyle AP, Hong EL, Hariharan M, et al. (12 co-authors). 2012. Annotation of functional variation in personal genomes using regulomedb. *Genome Res*. 22:1790–7.

Burbano HA, Hodges E, Green RE, et al. (20 co-authors). 2010. Targeted investigation of the Neandertal genome by array-based sequence capture. *Science*. 328:723–5.

Cheng Z, Ventura M, She X, et al. (12 co-authors). 2005. A genome-wide comparison of recent chimpanzee and human segmental duplications. *Nature*. 437:88–93.

DePristo MA, Banks E, Poplin R, et al. (18 co-authors). 2011. A framework for variation discovery and genotyping using next-generation dna sequencing data. *Nat Genet*. 43:491–8.

Derrien T, Estellé J, Marco Sola S, Knowles DG, Raineri E, Guigó R, Ribeca P. 2012. Fast computation and applications of genome mappability. *PLoS One*. 7:e30377.

Ewing G, Hermisson J. 2010. Msms: a coalescent simulation program including recombination, demographic structure and selection at a single locus. *Bioinformatics*. 26:2064–5.

Fay JC, Wu CI. 2000. Hitchhiking under positive darwinian selection. *Genetics*. 155:1405–13.

Fu YX, Li WH. 1993. Statistical tests of neutrality of mutations. *Genetics*. 133:693–709.

Galtier N, Piganeau G, Mouchiroud D, Duret L. 2001. Gc-content evolution in mammalian genomes: the biased gene conversion hypothesis. *Genetics*. 159:907–11.

Garrison E, Marth G. 2012. Haplotype-based variant detection from short-read sequencing. *arXiv preprint arXiv:1207.3907*. .

Gendzekhadze K, Norman PJ, Abi-Rached L, Graef T, Moesta AK, Layrisse Z, Parham P. 2009. Co-evolution of KIR2DL3 with HLA-C in a human population retaining minimal essential diversity of KIR and HLA class I ligands. *Proc Natl Acad Sci U S A*. 106:18692–7.

Gillespie JH. 1978. A general model to account for enzyme variation in natural populations. v. the sas–cff model. *Theor Popul Biol*. 14:1–45.

Gravel S, Henn BM, Gutenkunst RN, Indap AR, Marth GT, Clark AG, Yu F, Gibbs RA, 1000 Genomes Project, Bustamante CD. 2011. Demographic history and rare allele sharing among human populations. *Proc Natl Acad Sci U S A*. 108:11983–8.

Hudson RR, Kaplan NL. 1985. Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics*. 111:147–64.

Kelly JK. 1997. A test of neutrality based on interlocus associations. *Genetics*. 146:1197–206.

Kircher M, Sawyer S, Meyer M. 2012. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res*. 40:e3.

Kircher M, Stenzel U, Kelso J. 2009. Improved base calling for the illumina genome analyzer using machine learning strategies. *Genome Biol*. 10:R83.

Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. 2014. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet.* 46:310–5.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 25:1754–60.

McKenna A, Hanna M, Banks E, et al. (11 co-authors). 2010. The genome analysis toolkit: a mapreduce framework for analyzing next-generation dna sequencing data. *Genome Res.* 20:1297–303.

Messer PW. 2013. SLiM: simulating evolution with selection and linkage. *Genetics.* 194:1037–9.

Messer PW, Petrov DA. 2013. Population genomics of rapid adaptation by soft selective sweeps. *Trends Ecol Evol.* 28:659–69.

Meyer M, Kircher M. 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb Protoc.* 2010:pdb.prot5448.

Nei M. 1987. Molecular Evolutionary Genetics. Columbia University Press.

Nei M, Li WH. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc Natl Acad Sci U S A.* 76:5269–73.

Nielsen R, Hellmann I, Hubisz M, Bustamante C, Clark AG. 2007. Recent and ongoing selection in the human genome. *Nat Rev Genet.* 8:857–68.

Pritchard JK, Pickrell JK, Coop G. 2010. The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Curr Biol.* 20:R208–15.

Prüfer K, Munch K, Hellmann I, et al. (41 co-authors). 2012. The bonobo genome compared with the chimpanzee and human genomes. *Nature.* 486:527–31.

Przeworski M, Coop G, Wall JD. 2005. The signature of positive selection on standing genetic variation. *Evolution.* 59:2312–23.

R Core Team. 2013. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Ramos-Onsins SE, Rozas J. 2002. Statistical properties of new neutrality tests against population growth. *Mol Biol Evol.* 19:2092–100.

Scheet P, Stephens M. 2006. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *Am J Hum Genet.* 78:629–44.

Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics.* 123:585–95.

Takahata N, Nei M. 1990. Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of major histocompatibility complex loci. *Genetics*. 124:967–78.

Tellier A, Moreno-Gámez S, Stephan W. 2014. Speed of adaptation and genomic footprints of host-parasite coevolution under arms race and trench warfare dynamics. *Evolution*. 68:2211–24.

Thornton K. 2003. Libsequence: a C++ class library for evolutionary genetic analysis. *Bioinformatics*. 19:2325–7.

Thornton K, Andolfatto P. 2006. Approximate Bayesian inference reveals evidence for a recent, severe bottleneck in a Netherlands population of Drosophila melanogaster. *Genetics*. 172:1607–19.

Veeramah KR, Wegmann D, Woerner A, Mendez FL, Watkins JC, Destro-Bisol G, Soodyall H, Louie L, Hammer MF. 2012. An early divergence of KhoeSan ancestors from those of other modern humans is supported by an ABC-based analysis of autosomal resequencing data. *Mol Biol Evol*. 29:617–30.

Wall JD. 1999. Recombination and the power of statistical tests of neutrality. *Genetical Research*. 74:65–79.

Watterson GA. 1975. On the number of segregating sites in genetical models without recombination. *Theor Popul Biol*. 7:256–76.

Wegmann D, Leuenberger C, Excoffier L. 2009. Efficient approximate Bayesian computation coupled with Markov chain Monte Carlo without likelihood. *Genetics*. 182:1207–18.

Wegmann D, Leuenberger C, Neuenschwander S, Excoffier L. 2010. ABCtoolbox: a versatile toolkit for approximate Bayesian computations. *BMC Bioinformatics*. 11:116.

Weir BS, Cockerham CC. 1984. Estimating F-Statistics for the analysis of population structure. *Evolution*. 38:pp. 1358–1370.

Wright S. 1939. The distribution of self-sterility alleles in populations. *Genetics*. 24:538–52.

Table S1. Resequenced regions.

Genes	array	IN target ¶	OFF target †	Sanger §	Total Σ
AKAP13	29,293	24,863	3,255	6,287	34,405
BTNL2	4,998	4,367	423	3009	7,799
CLCNKB	5,917	5,523	1,067	6,572	13,162
FBLN2	5,918	5,220	1,117	10,779	17,116
GP6	3,225	2,482	425	3,645	6,552
HSD3B1	5,218	4,613	815	1938	7,366
MLPH	6,679	5,896	632	6,892	13,420
PKDREJ	2,844	2,509	189	6,102	8,800
SDR39U1	3,800	3,325	517	1,984	5,826
TARBP1	4,875	3,378	686	14,723	18,787
TRPV5	4,675	4,031	788	4,831	9,650
TRPV6	5,736	5,042	932	5,242	11,216
TXNRD2	5,082	5,058	814	9,298	15,170
ZNF473	6,389	5,893	1,284	3,130	10,307
controls	43,985	42,363	8,513	*X*	50,876
total	**138,634**	**124,563**	**21,457**	**84,432**	**230,452**

¶ number of base-pairs in the array that passed the coverage filter (i.e. at least 8x coverage in half of the individuals).
† sequences that have not been included in the array for capturing but that passed the coverage cutoffs.
the values correspond to (mainly exonic) regions that do not overlap with the Illumina sequences.
Σ sum of 'IN target', 'OFF target' and 'Sanger'.
X 26,478 base-pairs resequenced by Sanger, which are all included in the Illumina sequences.

Table S2. Genotype calls validation.

Filters	True SNPs ¶	FP †	FN §	Specificity $	Sensitivity Σ
none	229.19 (207-231)	6.04 (3-10)	0.21 (0-1)	96.62 (85.51-98.72)	99.91 (99.53-100)
QUAL<50	224.19 (202-226)	5.97 (2.4-10)	0.23 (0-1)	96.58 (85.16-98.69)	99.90 (99.52-100)
SB>10	229.19 (207-231)	2.37 (1-5)	0.21 (0-1)	98.15 (87.48-99.57)	99.91 (99.53-100)
QUAL<50 & SB>10	224.19 (202-226)	2.36 (1-4.6)	0.23 (0-1)	98.11 (87.17-99.56)	99.90 (99.52-100)

Comparison of sequences generated with Illumina and those produced by Sanger in 47 control regions (total of 26,478 bp). Different filters were considered in order to minimize the false discovery rate and to maximize the true discovery rate. The values report the median of the distributions across all resequenced individuals. In parentheses are the 2.5% and 97.5% quantiles of these distributions.
¶ number of true positive (TP)
† number of false positive (FP)
§ number of false negative (FN)
$ specificity in percentage (%) calculated as TN/(TN+TP)
Σ sensitivity in percentage (%) calculated as TP/(TP+TN)

Table S3. Segregating sites and fixed differences.

GENES	type	LWK		YRI		TSI		CHB		GIH	
		S	fd	S	fd	S	fd	S	fd	S	fd
AKAP13	all	194	391	163	323	133	321	103	411	133	317
	CDS	45	67	36	45	32	45	27	69	29	45
BTNL2	all	68	97	50	63	52	62	81	91	67	62
	CDS	16	13	11	14	11	13	14	15	7	16
CLCNKB	all	174	246	181	154	131	160	112	256	140	160
	CDS	23	23	26	19	16	20	11	22	15	20
FBLN2	all	126	234	120	205	75	206	54	259	89	208
	CDS	23	43	24	36	15	35	13	45	19	34
GP6	all	62	123	51	103	44	103	36	129	42	102
	CDS	19	39	18	29	18	29	15	38	15	29
HSD3B1	all	57	100	40	71	12	78	26	106	42	70
	CDS	16	13	7	8	3	9	7	15	20	7
MLPH	all	103	146	92	126	66	126	58	152	71	126
	CDS	10	30	13	20	10	20	11	29	9	20
PKDREJ	all	59	90	49	62	39	63	14	98	40	64
	CDS	34	58	27	35	20	35	10	64	25	35
SDR39U1	all	60	46	62	33	43	33	44	50	44	33
	CDS	8	6	8	2	8	2	6	8	8	2
TARBP1	all	121	217	109	203	82	206	73	227	83	203
	CDS	21	34	19	27	15	28	15	35	18	27
TRPV5	all	55	104	54	89	29	92	26	114	21	92
	CDS	9	19	9	13	8	13	9	19	5	13
TRPV6	all	69	175	71	127	45	127	21	188	23	139
	CDS	15	18	14	11	8	11	4	22	4	15
TXNRD2	all	115	188	115	165	57	172	39	205	68	168
	CDS	10	13	13	10	7	10	4	14	6	10
ZNF473	all	65	100	61	77	16	94	15	112	16	94
	CDS	16	24	17	17	2	23	4	30	2	23
controls		311	738	317	728	189	737	181	766	208	736

Number of segregating sites (S) and fixed differences (fd) to PanTro3 reference genome for each population using the entire sequence of the gene (all) or only the coding sequence (CDS).

Table S4. Allele frequency correlations between populations of 1000 Genomes.

pop1	**pop2**	**non-genic ¶**	**genic †**	**genes §**
Yoruba	Toscani	0.81	0.80	0.57
Yoruba	Han Chinese	0.78	0.77	0.51
Toscani	Han Chinese	0.85	0.85	0.92

The table shows the correlation as Pearson's r^2 for:
¶ the genic SNPs defined in the genes coordinates of *refGenes*;
† the non-genic SNPs, which are all the other SNPs outside the genes coordinates;
§ the SNPs in the four candidate genes (*CLCNKB*, *PKDREJ*, *SDR39U1*, and *ZNF473*).

Table S5. The non-synonymous *iAdO-alleles*.

CHR	POSITION	dbSNP ID	Anc	Der	Gene	YRI vs. TSI ¶	YRI vs. CHB ¶	LWK vs. TSI ¶	LWK vs. CHB ¶
1	16373062	rs5256	A	C	*CLCNKB*	0.203	0.240	0.142	0.160
1	16375063	rs1889789	G	C	*CLCNKB*	0.358	0.425	0.371	0.417
1	16375510	rs11588392	G	A	*CLCNKB*	0.474	0.532	0.445	0.513
1	16376191	NA	T	C	*CLCNKB*	NA	NA	NA	NA
1	16378000	rs6650119	A†	G†	*CLCNKB*	0.050	0.056	0.318	0.380
1	16380196	rs5253	T	C	*CLCNKB*	0.060	0.073	0.123	0.150
14	24909362	rs1043831	C	T	*SDR39U1*	0.061	0.083	0.076	0.115
14	24909475	rs3211056	C	G	*SDR39U1*	0.074	0.136	0.070	0.104
14	24910973	rs11625819	G	T	*SDR39U1*	0.210	0.281	0.231	0.281
19	50545025	rs10419876	G	A	*ZNF473*	0.275	0.312	0.265	0.298
19	50545070	rs10419911	G	A	*ZNF473*	0.189	0.215	0.248	0.272
19	50547945	rs73932407	A	G	*ZNF473*	0.189	0.215	0.248	0.272
19	50548191	rs16981705	C	T	*ZNF473*	0.156	0.195	0.225	0.287
19	50548367	rs61745068	C	T	*ZNF473*	0.156	0.195	0.231	0.293
19	50548443	rs61730172	A	G	*ZNF473*	0.156	0.195	0.225	0.287
19	50548626	rs16981706	A	G	*ZNF473*	0.189	0.215	0.248	0.272
19	50549661	rs10424809	C	T	*ZNF473*	0.156	0.195	0.225	0.287
19	50549684	rs10426374	T	G	*ZNF473*	0.156	0.195	0.225	0.287
22	46655948	rs6008384	T	C	*PKDREJ*	0.018	0.029	0.024	0.041
22	46656242	rs34798212	A	G	*PKDREJ*	0.402	0.439	0.385	0.421
22	46656479	rs6519993	G	A	*PKDREJ*	0.014	0.023	0.017	0.026
22	46659186	rs113101219	C	A	*PKDREJ*	0.860	0.883	0.841	0.858

¶ *P-values* assessed by comparing the SNP to the empirical distribution of pairwise F_{ST} values between African (YRI and LWK) and non-African (TSI and CHB) populations of all 1000 Genomes SNPs. The values have been calculated by conditioning on the frequencies in African populations. NA: Not ascertained because absent in the 1000 Genomes phase 1 data (1000 Genomes Project Consortium et al., 2012).

† Ancestry could not be determined; therefore, the reference and alternative alleles are considered as ancestral and derived states, respectively.

Table S6. Summary statistics used for the ABC framework.

SuSt	Description	tool ¶	Reference	Used †
PtoD	polymorphism-to-divergence	R-script		yes
ht	mean observed heterozygosity	R-script		
mMAF	mean of minor allele frequencies	R-script		
mDAF	mean of derived allele frequencies	R-script		yes
sdMAF	s.d. of minor allele frequencies	R-script		yes
sdDAF	s.d. of derived allele frequencies	R-script		yes
S	number of segregating sites	*msstats*		
n1	number of singletons	*msstats*		
nhaps	number of haplotypes	*msstats*		yes
hdiv	haplotype diversity	*msstats*	(Nei, 1987)	yes
θ_w	Watterson's θ	*msstats*	(Watterson, 1975)	
π	nuvcleotide diveristy	*msstats*	(Nei and Li, 1979)	yes
θ_H	θ calculated from homozygosity	*msstats*	(Fay and Wu, 2000)	
H'	standardized summary of the SFS related to θ_H	*msstats*	(Thornton and Andolfatto, 2006)	yes
D	Tajima's D	*msstats*	(Tajima, 1989)	yes
F	Fu & Li's F statistic	*msstats*	(Fu and Li, 1993)	yes
$D_{F\&L}$	Fu & Li's D statistic	*msstats*	(Fu and Li, 1993)	yes
F*	Fu & Li's F* statistic	*msstats*	(Fu and Li, 1993)	yes
D*	Fu & Li's D* statistic	*msstats*	(Fu and Li, 1993)	
rm	minimum number of recombination events	*msstats*	(Hudson and Kaplan, 1985)	yes
B_w	Wall's B statistic	*msstats*	(Wall, 1999)	
Q_w	Wall's Q statistic	*msstats*	(Wall, 1999)	yes
R_2	difference between the number of singletons and π	*msstats*	(Ramos-Onsins and Rozas, 2002)	
R_{2E}	R_2 but including mutations on external branches	*msstats*	(Ramos-Onsins and Rozas, 2002)	
ZnS	mean r^2 in the sample	*msstats*	(Kelly, 1997)	
mF_{ST}	mean of F_{ST} between populations across loci	R-script		yes
sdF_{ST}	s.d. of F_{ST} between populations across loci	R-script		yes

The summary statistics (SuSt) are calculated for each populations (African, European and Asian) and for all three populations pairwise comparisons [i.e. F_{ST} (Weir and Cockerham, 1984)].
¶ Using either in-house R-scripts (R Core Team, 2013) or *msstats* package from *Libsequence* (Thornton, 2003).
† These SuSt are informative (i.e. show only modest correlation and are different across the evolutionary scenarios) and used in the ABC model selection.

Table S7. ABC power analyses of the models *B-B*, *B-N* and *B-P*.

GENES	MODELS	ASIA		EUROPE	
		TP	FP	TP	FP
SDR39U1	*B-B*	0.818	0.123	0.814	0.142
	B-N	0.474	0.139	0.458	0.133
(5,800 bp)	*B-P*	0.644	0.118	0.654	0.113
PKDREJ	*B-B*	0.818	0.124	0.810	0.129
	B-N	0.469	0.143	0.473	0.142
(8.700 bp)	*B-P*	0.624	0.119	0.630	0.112
CLCNKB & ZNF473	*B-B*	0.815	0.123	0.802	0.132
	B-N	0.484	0.143	0.475	0.144
(10,000 bp)	*B-P*	0.609	0.115	0.626	0.097

The table reports the true positive (TP) and false positive (FP) rates when comparing the models *B-B*, *B-N* and *B-P*. The values are the average across three comparisons of *B-B* and *B-N* with each of the three *B-P* models.

Table S8. ABC power analyses of the three sub-models *B-P*.

GENES	MODELS	ASIA		EUROPE	
		TP	FP	TP	FP
SDR39U1	*B-Psv*	0.189	0.081	0.280	0.146
	B-Pcfe	0.258	0.086	0.299	0.095
(5,800 bp)	*B-Pdn*	0.462	0.068	0.490	0.074
PKDREJ	*B-Psv*	0.201	0.084	0.296	0.149
	B-Pcfe	0.256	0.078	0.332	0.088
(8,700 bp)	*B-Pdn*	0.458	0.061	0.492	0.067
CLCNKB & *ZNF473*	*B-Psv*	0.197	0.090	0.273	0.136
	B-Pcfe	0.252	0.082	0.307	0.087
(10,000 bp)	*B-Pdn*	0.467	0.062	0.499	0.065

Table S9. List of the *iAdO-alleles* in *RegulomeDB*.

Score	SNPs [chr_position]	Description
1b	14_24911168; 22_46652371; 19_50528695	eQTL + TFbinding + any motif + DNase footprint + DNase peak
1d	19_50551050; 19_50551200	eQTL + TF binding + any motif + DNase peak
1f	22_46652152; 14_24907287; 14_24908265; 14_24909071; 14_24909280; 14_24909361; 14_24909730; 14_24910209; 14_24912754; 22_46652302; 22_46656606; 22_46656804; 22_46657260; 19_50527912; 19_50528061; 19_50528232; 19_50528283; 19_50528593; 19_50545739; 19_50548190; 19_50548625; 19_50548686; 19_50549660; 19_50549683; 19_50550126; 19_50551664	eQTL + TF binding/DNase peak
2b	14_24911851; 14_24912168; 19_50528786	TF binding + any motif + DNase footprint + DNase peak
3a	14_24906848; 14_24907645; 14_24908210; 14_24912678; 14_24912696; 19_50529355; 19_50541608	TF binding + any motif + DNase peak
4	14_24907512; 14_24907551; 14_24908116; 14_24908183; 14_24908730; 14_24909474; 14_24911115; 14_24911763; 19_50545968; 19_50548366; 19_50548442; 19_50548917	TF binding + DNase peak

The table lists the *iAdO-alleles* (i.e. those at intermediate frequency in Africa and low/high derived frequency in non-Africans for *PKDREJ*, *SDR39U1*, and *ZNF473*) that are in the top six categories of *RegulomeDB*.Boyle et al. (2012) Only the first three categories "1" exhibit a significant enrichment after 1,000 random sampling of three random genes (see Figure S16).

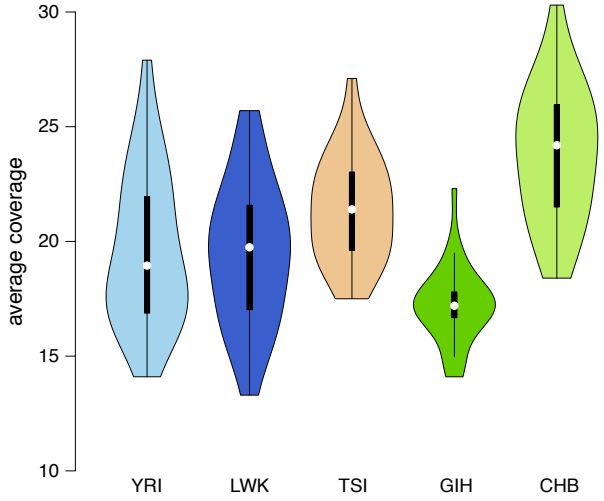

Figure S1. Average coverage distribution by population. The violinplot represents the average coverage in the targeted regions of each individual in the five populations. Positions without coverage are disregarded for the calculation. CHB have a higher coverage because they were resequenced in an extra lane due to poor coverage in the previous re-sequencing.

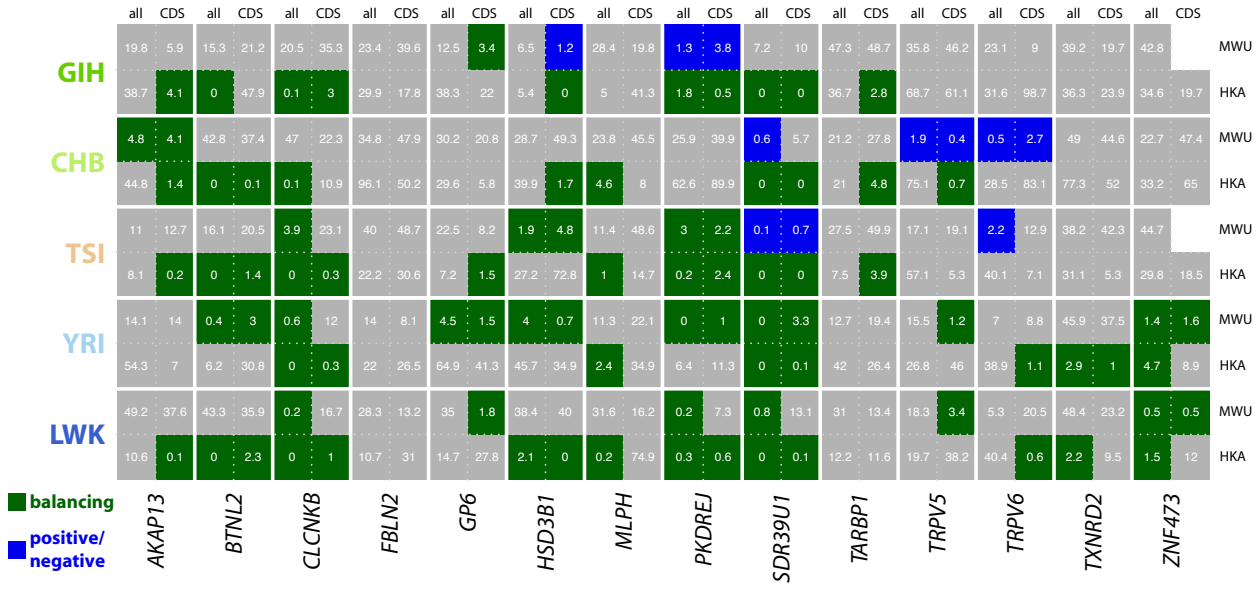

Figure S2. Neutrality tests: MWU and HKA. The tests were carried for the entire sequence ('all') and the coding part ('CDS') of the genes. The numbers in the cells are the *p-values* (in percentage). The cells are colored according to 5% significance threshold indicating balancing and positive/negative selection. MWU of *ZNF473* using the coding region in TSI and GIH was not performed because there are only two SNPs (see Table S3). Only two genes show no significant evidence of balancing selection, and eight genes display partial signatures of balancing selection in at least one African population.

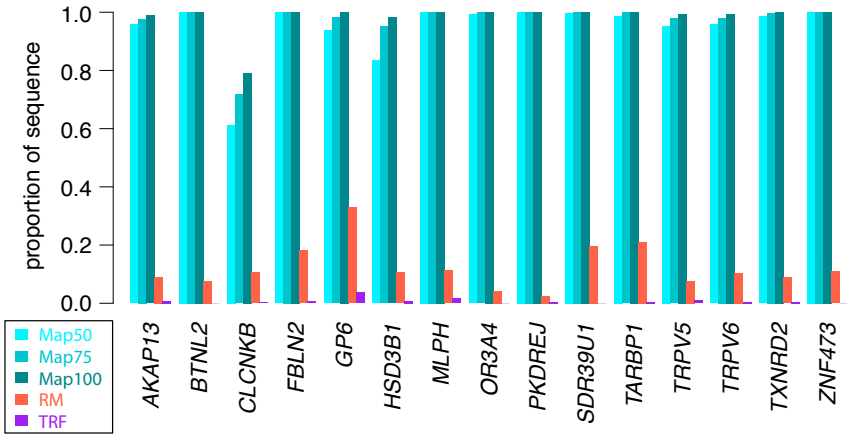

Figure S3. Genomic features of the genes. The y-axis is the proportion of each gene (x-axis) in the mappability tracks of 50mer 'Map50', of 75mer 'Map75', of 100mer 'Map100', in RepeatMasker 'RM', and in simple units of repeats 'TRF'.

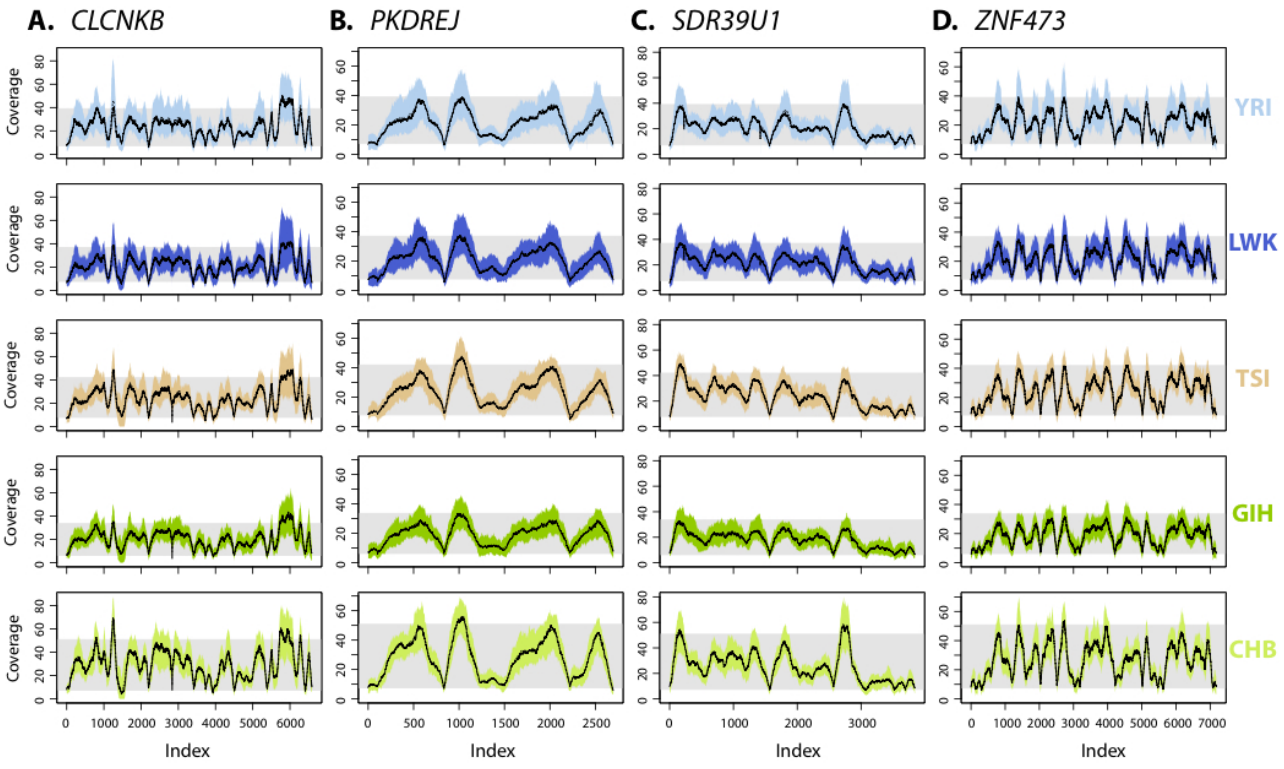

Figure S4. Coverage along each gene by population. The gray areas represent the 90% quantile of the coverage distribution of the entire targeted position of the array. The black lines and the colored areas correspond to mean coverage across individuals and the 90% quantile for a given position for the gene (x-axis).

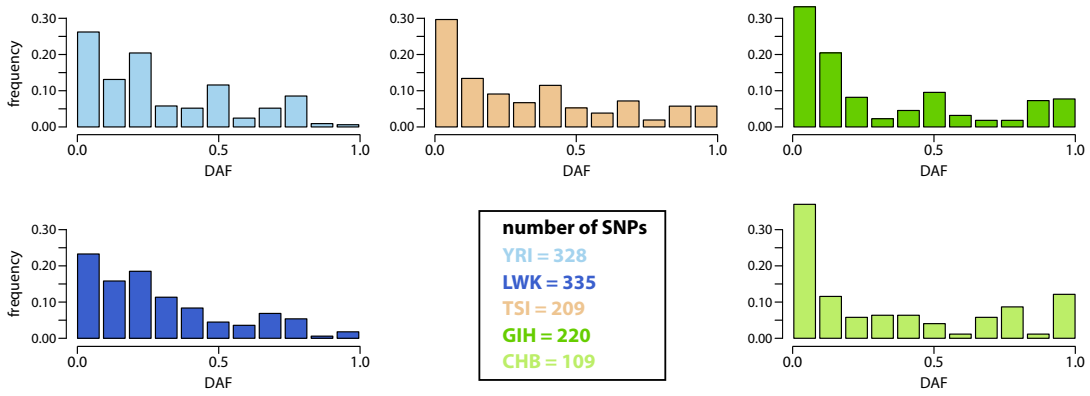

Figure S5. Site frequency spectra of all candidate genes. The SFS includes all SNPs of the candidate genes *CLCNKB*, *PKDREJ*, *SDR39U1*, and *ZNF473*. Notice that low frequency variants are higher in non-Africans than Africans.

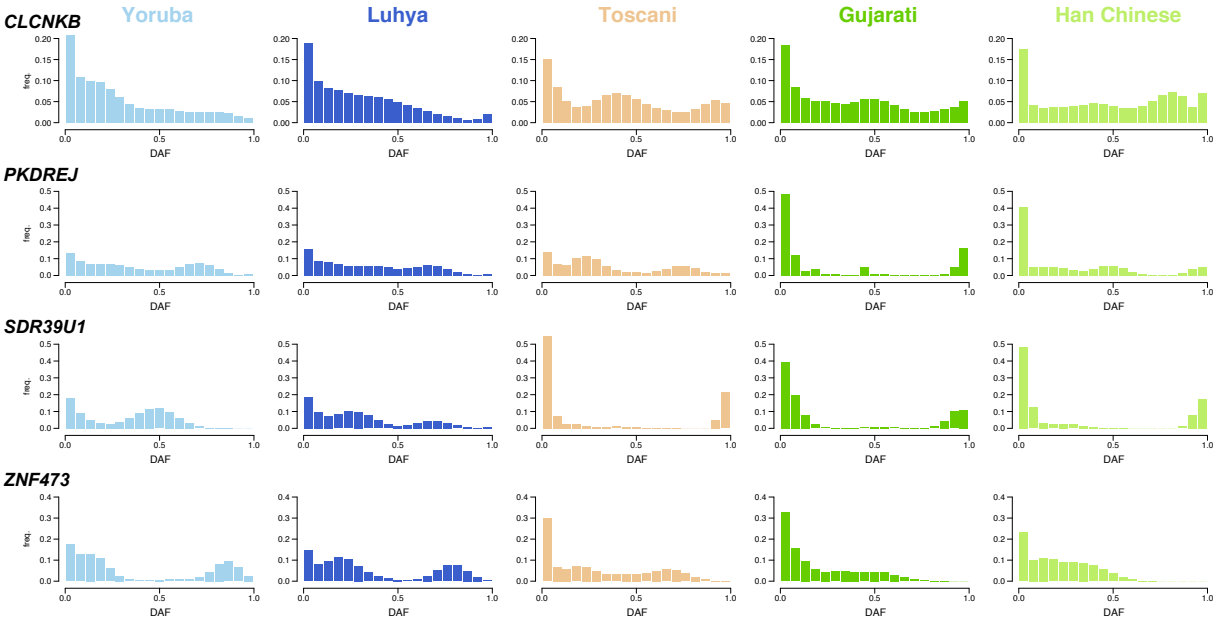

Figure S6. Site frequency spectra for each population and each candidate genes.

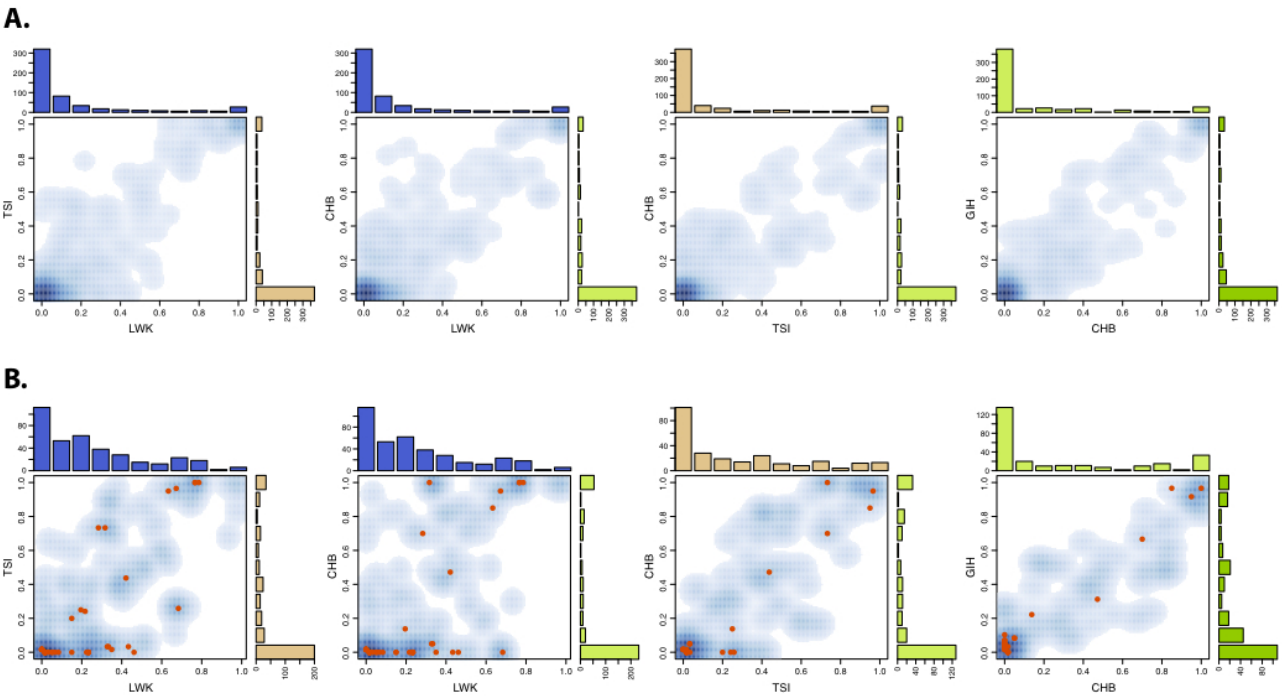

Figure S7. Two-dimensional SFS. (A) the control regions and **(B)** the four candidate genes combined. The red dots represent non-synonymous SNPs. The histograms on the top and right side of the scatterplot are the SFS for the x and y population. The representation of the scatter plot is colored according to the SNPs density. The figure is the same as Figure 3 from the main text, but it shows other pairwise comparisons and the histograms on the top and bottom sides are not necessary the same due to monomorphic sites, which number varies when compared with a different population. Nevertheless, the comparisons of Africans vs. non-Africans in these plots and those of Figure 3 are similar.

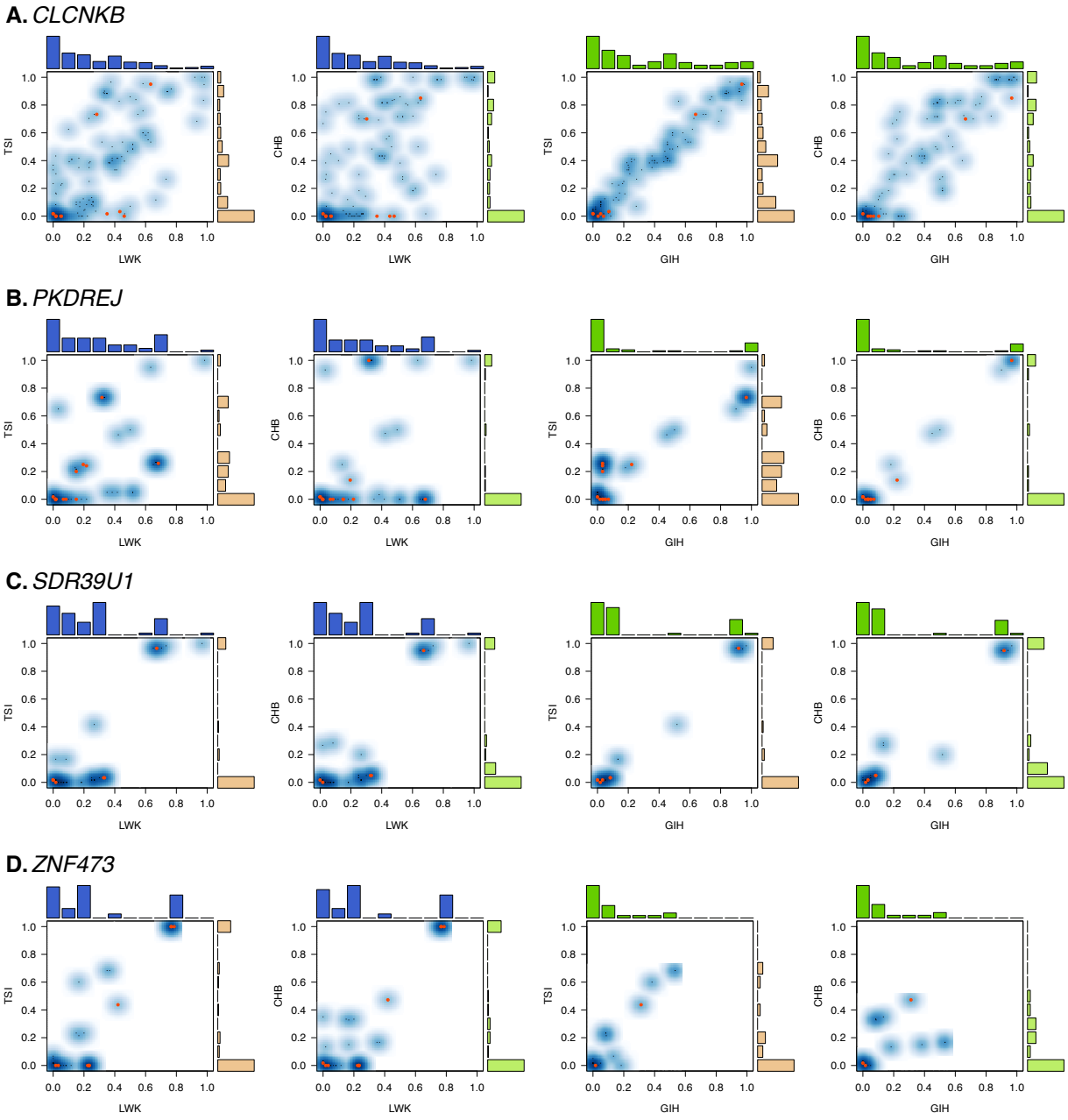

Figure S8. Two-dimensional SFS. SNPs in each of the four candidate genes as in Figure 5 and Figure S7.

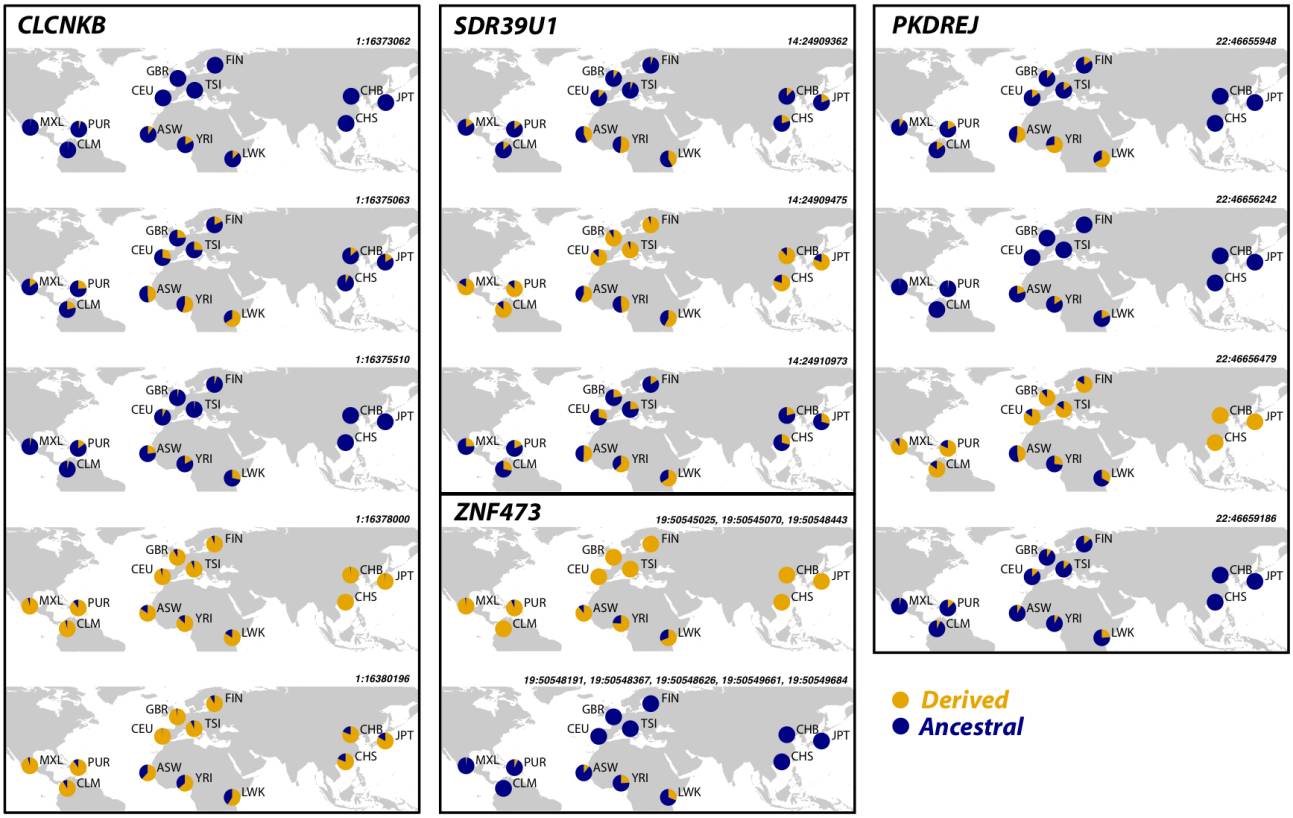

Figure S9. Worldwide allele frequencies of 21 non-synonymous *iAdO-alleles*. The pies are for all populations from the 1000 Genomes phase 1 data (1000 Genomes Project Consortium et al., 2012). For *ZNF473*, the allele frequencies are identical for all *iAdO-alleles*, but here we grouped them into two according to ancestry. One of the 22 non-synonymous *iAdO-alleles* is missing because it is not present in 1000 Genomes phase 1 data (Table 3).

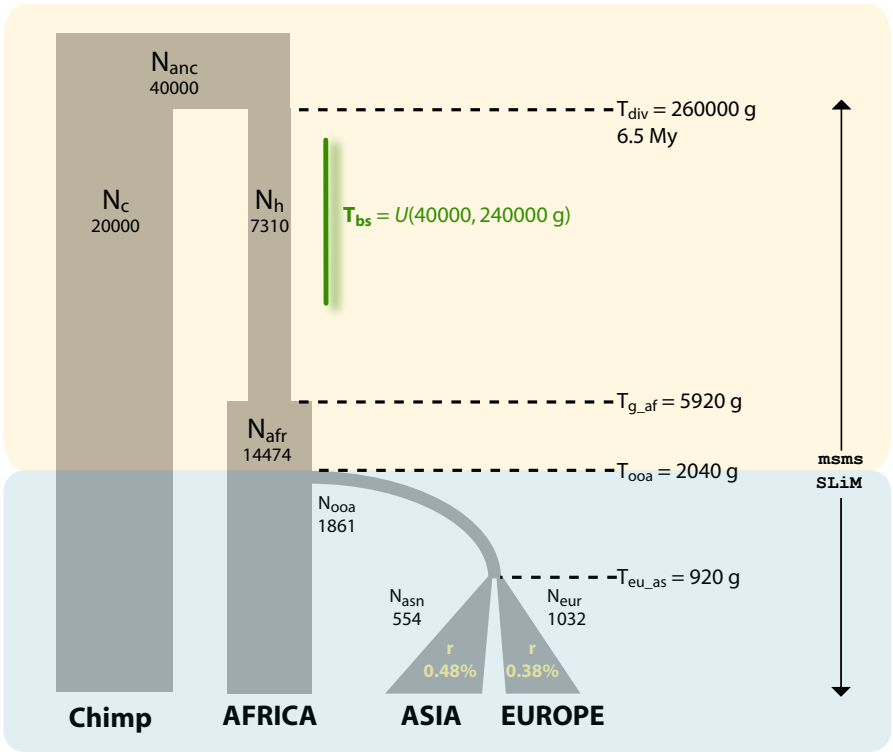

Figure S10. Simulation outline. The figure shows the schematic view of the simulations' procedure and the parameters used in the ABC. While the time since balancing selection Tbs is drawn from a uniform distribution (as Sbs, Sps, μ, and ρ, see main text for their ranges), all other demographic parameters are fixed to the values reported in the picture. The two different background colors indicate which software is used in that lapse of time: *msms* (Ewing and Hermisson, 2010) in yellow and *SLiM* (Messer, 2013) in blue with the direction of the arrows pointing to coalescence and forward simulator, respectively.

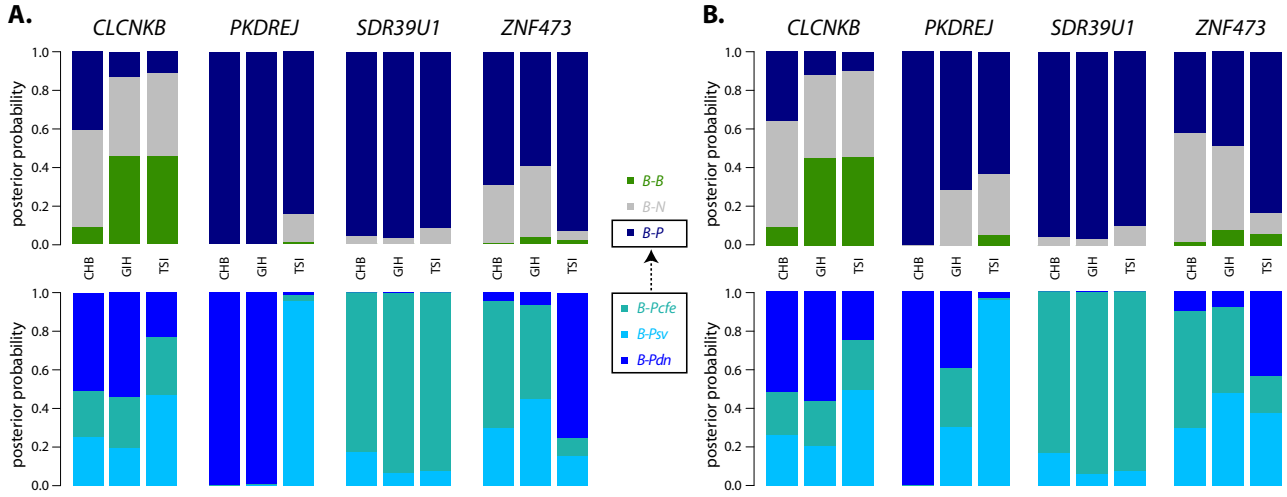

Figure S11. ABC model selection. **(A)** with and **(B)** without F_{ST} correction.

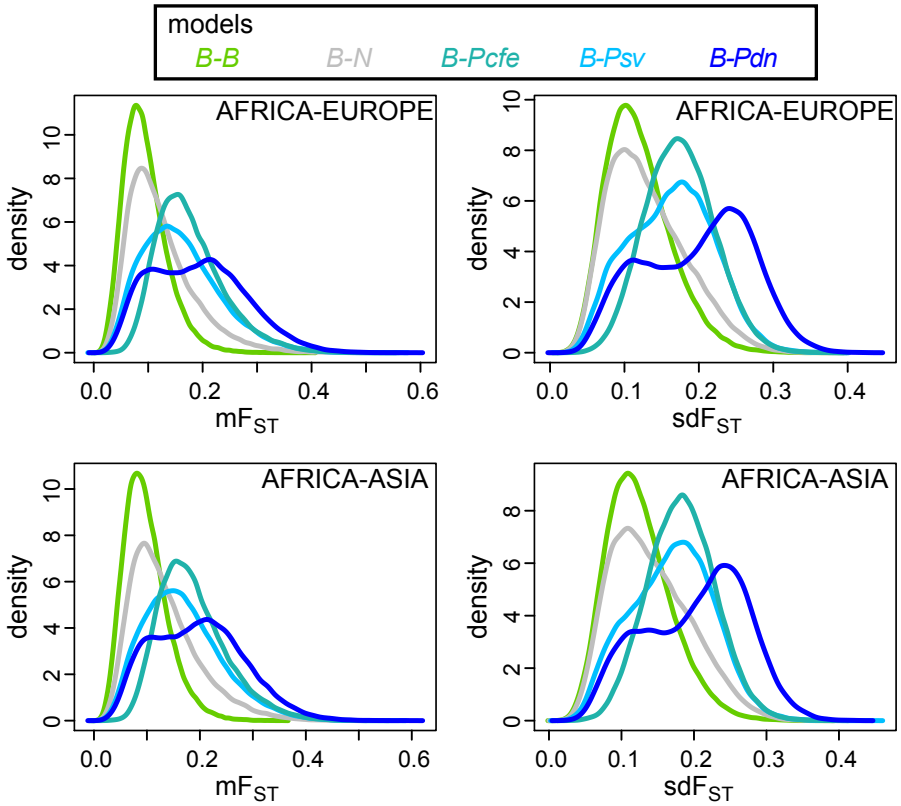

Figure S12. Distributions of mF_{ST} and sdF_{ST} statistics.. The curves for each of the five simulated model in two pairwise comparisons, Africa-Europe and Africa-Asia.

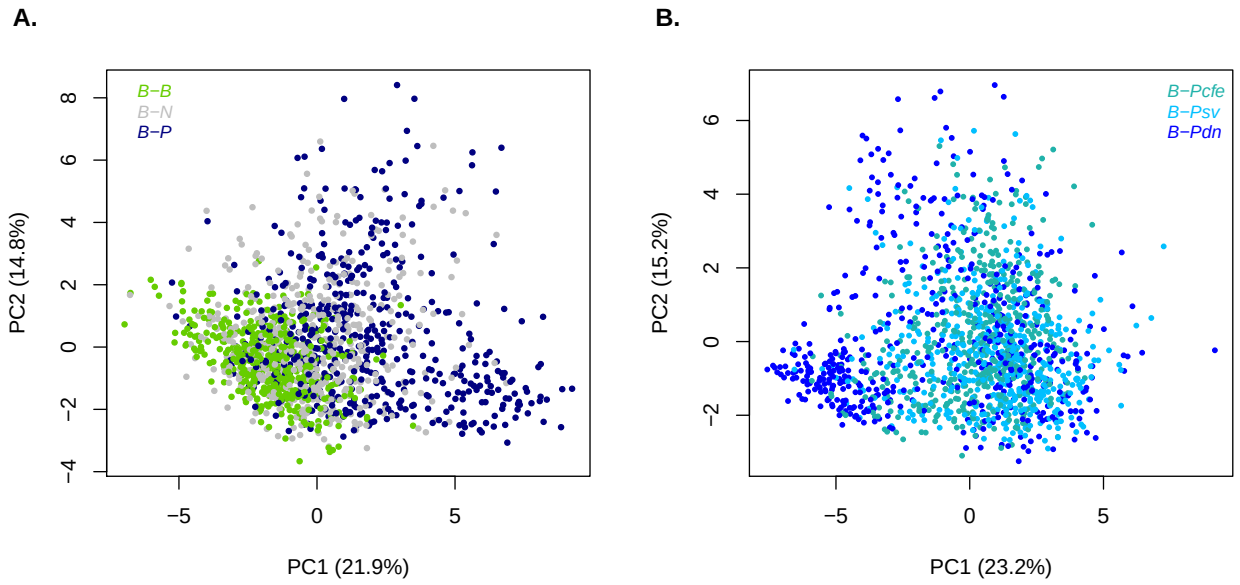

Figure S13. Comparisons of the simulated models. Principal Component Analyses (PCA) of 500 random simulations for each models using 16 informative summary statistics. In (**A**) the model *B-P* includes 500 simulations sampled among all the three sub-models *B-Pcfe*, *B-Psv*, and *B-Pdn*, which are then represented in (**B**).

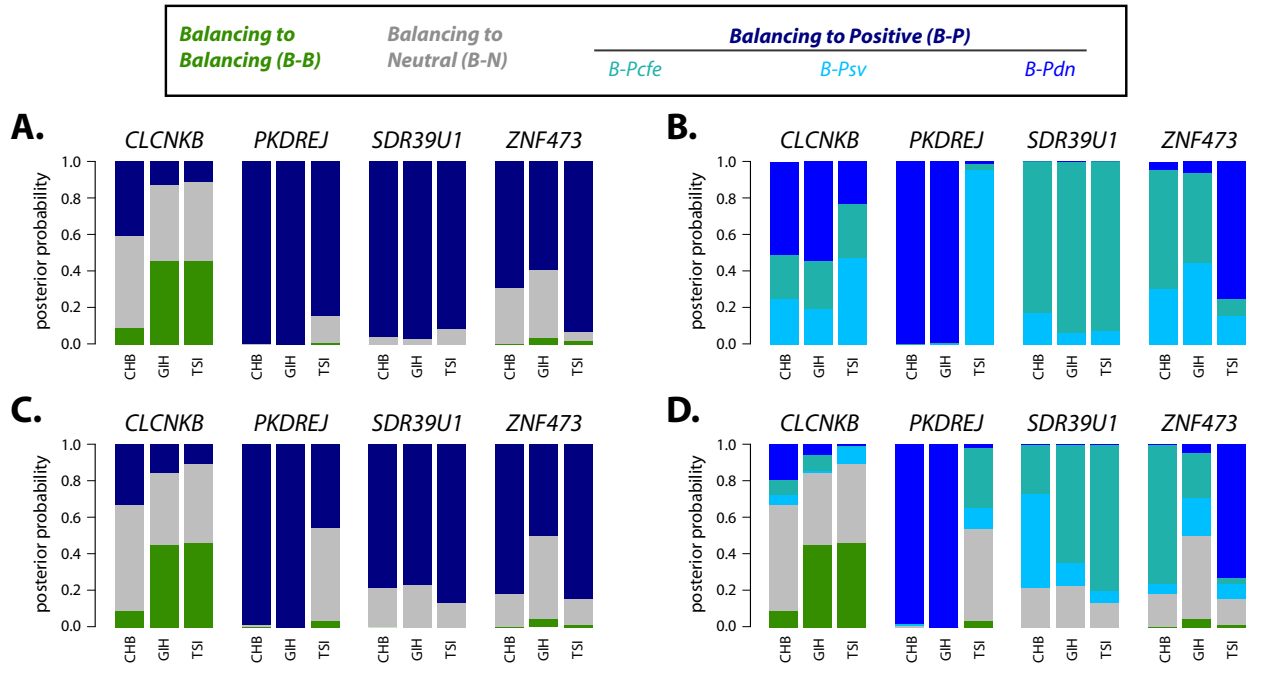

Figure S14. Two ABC model selections. (**A-B**) Posterior probabilities using a hierarchical approach as described in the main text and in Fig. 5; this two plots are the same as in Fig. 5. (**C-D**) Posterior probabilities using all five models together, where the three *B-P* models together have the same prior as each of *B-B* and *B-N* models. For visual representation and comparison with the hierarchical approach we colored in (**C**) the three *B-P* models with the same dark blue color as in (*A*).

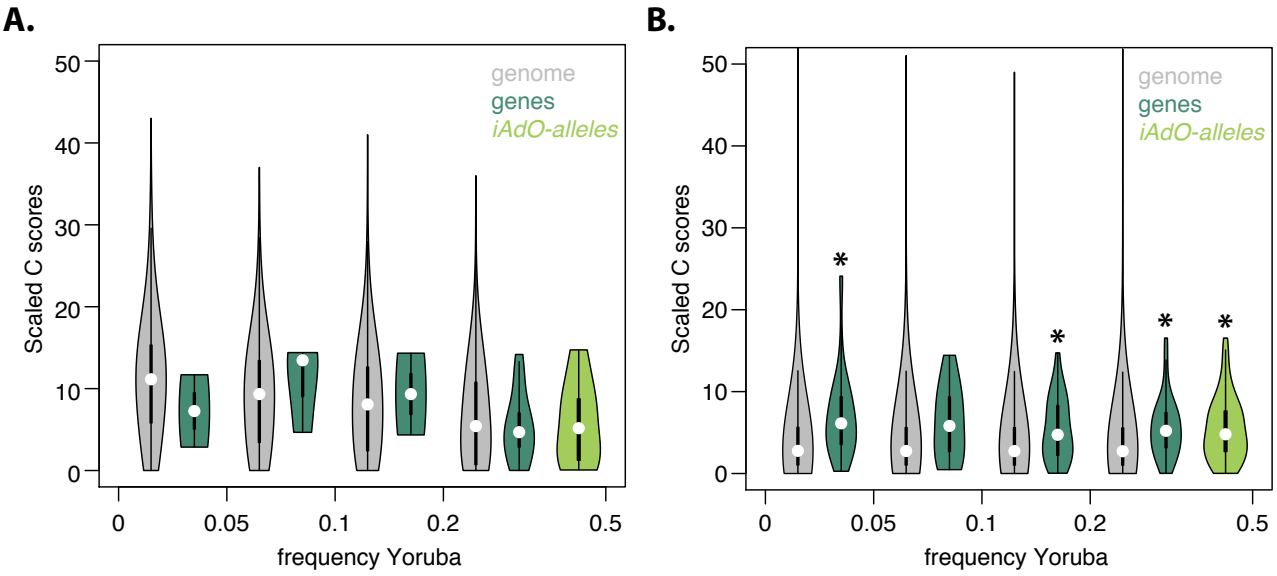

Figure S15. Distributions C-scores. C-scores are a combined measurement for the relative pathogenicity of mutations (Kircher et al., 2014). The non-synonymous SNPs (**A**) and all (i.e. non-synonymous, synonymous, and non-coding) SNPs (**B**). The distributions are conditioned on the minor frequency of the alleles in 1000 Genomes Yoruba in four bins (0-0.05, 0.06-0.10, 0.11-0.25, 0.26-0.50) on the x-axis. The different colors represent all SNPs in the 'genome' (gray), the SNPs in the three candidate 'genes' (green), and the *iAdO-alleles* (light green). We note that *CLCNKB*, which does not show evidence of changes in selective pressures, was excluded from this analysis. The asterisk * above the distributions indicate a difference between that distribution and the entire genome with a Mann-Whitney U $p < 0.01$.

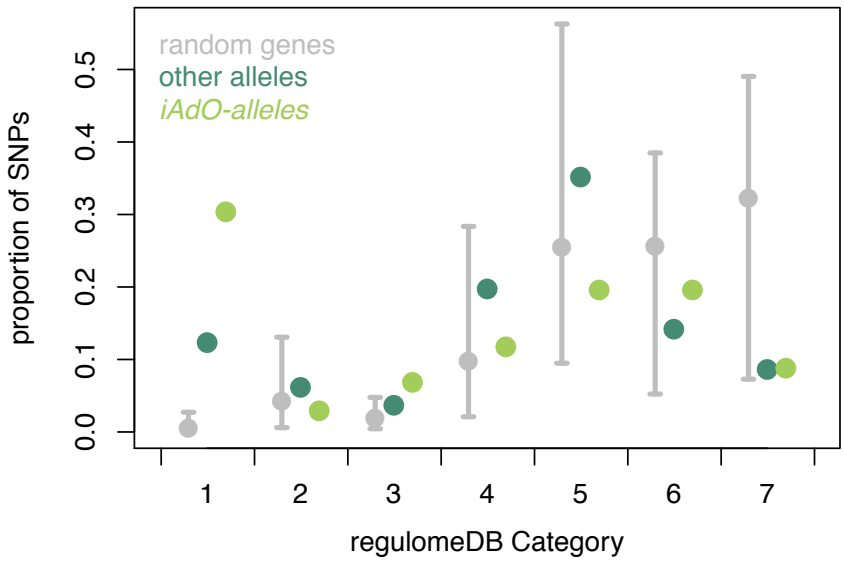

Figure S16. SNPs and regulatory functions. The proportion of SNPs (y-axis) belonging to a given category of *RegulomeDB* (Boyle et al., 2012) (x-axis) are shown. The gray dots and bars correspond to the mean and the 95% CI given by 1,000 samplings of three 'random genes' from the entire genome. The light and dark green dots are *iAdO-alleles* and all the other alleles in the three candidate genes, respectively.

.