SUPPLEMENTARY MATERIAL

Brenna M. Henn[1*§], Laura R. Botigué[1*], Stephan Peischl[2,6,7*], Isabelle Dupanloup[2], Mikhail Lipatov[1], Brian K. Maples[3], Alicia R. Martin[3], Shaila Musharoff[3], Howard Cann[4], Michael Snyder[3], Laurent Excoffier[2,6^], Jeffrey M. Kidd[5^], Carlos D. Bustamante[3^§]

[1] Stony Brook University, SUNY, Department of Ecology and Evolution, Stony Brook, NY 11794
[2] Institute of Ecology and Evolution, University of Berne, Berne, Switzerland 3012.
[3] Stanford University School of Medicine, Department of Genetics, Stanford, CA 94305
[4] Foundation Jean Dausset, Centre d'Etude du Polymorphisme Humain, Paris, France 75010
[5] University of Michigan Medical School, Department of Human Genetics and Department of Computational Medicine and Bioinformatics, Ann Arbor, MI 48109
[6] Swiss Institute of Bioinformatics, Lausanne, Switzerland, 1015
[7] Interfaculty Bioinformatics Unit, University of Berne, Berne, Switzerland 3012

Table of Contents

Supplementary Methods

Mapping and Variant Calling

Read-pairs were mapped onto the human genome reference (GRCh37, with the pseudo-autosomal regions of the Y chromosome masked). Briefly, reads are mapped using the *bwa* mapper (version 0.5.9), an aligner that reports a confidence metric associated with each aligned read. The resulting alignments are then processed to identify PCR duplicates (using Picard, http://picard.sourceforge.net/), empirically recalibrate the quality values associated with each base call based on observed rates of differences from the reference, and realignment of all samples together around candidate small insertion-deletion variants (using the Genome Analysis Tool Kit [GATK] version 1.2-65) (DePristo et al., 2011). The output of this process is a set of cleaned, calibrated, and mapped reads from each individual, suitable for subsequent analysis.

Contamination and Data Quality Control

Sample contamination and data quality issues can compromise the results of large-scale genome sequencing efforts. Contamination was assessed for each individual by comparing the genotypes from Illumina Human660K array SNP data (Li et al., 2008) and the Illumina HiSeq data from an initial per-sample call set using *samtools*. A concordance rate was calculated from the number of HiSeq homozygous non-reference calls (HNR) that were also homozygous non-reference on the Illumina Human660K array, divided by the total number of HNR calls from the 660K array. If the concordance dropped below 90%, a new library was made and contamination assessed in a second run.

Base pair composition plots were examined visually to identify reads with a skewed composition. In cases where the average quality score dropped below 15, all reads for a given lane were trimmed from base pair 101 backwards until the score became elevated above 15. Additionally, only reads with a minimum of 50 base pairs exceeding Q=15 were retained. This trimming procedure resulted in an increase in the percent of reads mapping to the human reference sequence. However, trimming did not appear to noticeably improve the concordance with the Illumina SNP array at homozygous non-reference sites.

Identifying Single Nucleotide Variants

Candidate single nucleotide variants were identified based on joint calling across all samples using the Unified Genotyper in the GATK. We applied the Variant Quality Score Recalibration (VQSR) procedure to retain a set of variants such that 99% of variant positions that overlap with HapMap3 SNPs were retained. Refined genotypes for the resulting set of positions were obtained using Beagle v3 (Browning and Yu, 2009). Sites were called on the autosomes and the pseudo-autosomal portions of the X chromosome, but only variants on the autosomes were utilized in subsequent analysis.

Callable Genome Mask for WGS analysis

To aid comparisons between exome and WGS calls, we created a mask file to identify regions of the genome that can be confidentially called based on the WGS data. We utilized metrics reported in the GATK UnifiedGenotyper 'Emitall Sites' file. We set cutoffs for DP, the total read depth at each site, MQ, the average mapping quality at a site, and the fraction of MQ0 reads at a site. We determined cutoffs based on comparison of putatively variable sites that pass or fail the VQSR selection criteria (**Figure S2**). We found that DP cutoffs of >= 192 and <=547 capture 98% of the VQSR pass sites, that 99% of VQSR pass sites have MQ >= 48 and 99.5 of VQSR sites have a MQ0 fraction <= 1%. Applying these cutoffs to the non-variable sites (variable site mask is determined by the VQSR procedure), identified 89.79% of the non-gap autosomes as being callable. We further refined this by removing sites within 5bp of a candidate indels,

removing annotated segmental duplications, and intersecting with the target regions of the exome capture array (**Figure S3**).

Exome Sequence Data Analysis

Exome capture data was processed as described above. Variants were selected based on the VQSR criteria implemented in the GATK. We restricted analysis to the 44 Mb target set for the Agilent Sure Select Exon Enrichment platform.

Variant Annotation

The putative ancestral state of each variant was annotated following the 1000 Genomes Project (1000 Genomes Project Consortium et al., 2012) based on ancestral sequences determined by Ortheus using multi-species alignments from Ensembl Compara release 59 (http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/phase1/analysis_results/supporting/ancestral_alignments/). Only variants for which the ancestral state was known were kept for downstream analysis.

Recent studies have shown that establishing the damaging potential of a variant is extremely difficult. As an example, one of the commonly used predictive algorithms, Polyphen (Sunyaev et al., 2001), has been shown to have a strong reference bias, annotating as neutral variants that are represented in the reference genome, regardless of their ancestral state (Simons et al., 2014). We therefore used two algorithms, one that is a measure of conservation across species (GERP scores) (Cooper et al., 2005), and another that is based on the biological effect of the variant (ANNOVAR) (**Table S2**). Positive GERP (RS) scores reflect a site showing high degree of conservation, based on the inferred number of "rejected substitutions" across the phylogeny. GERP scores were obtained from the UCSC genome browser (http://hgdownload.cse.ucsc.edu/ gbdb/hg19/bbi/All_hg19_RS.bw) based on an alignment of 35 mammals to human. The allele represented in the human hg19 sequence was not included in the calculation of GERP RS scores. GERP scores from the exome dataset range from -12 to 6.17, though only variants with a GERP score greater than -2 were selected for subsequent analysis, as negative values may be indicative of poor sequence alignments across the phylogeny. Most analysis focus on variants with positive GERP RS scores > 2. Since the range of RS scores is dependent on the depth of the multi-species phylogeny used, we re-annotated GERP scores for the 1000 Genomes data using this procedure. We first examined the distribution of GERP RS scores for both all exome single nucleotide changes and for only nonsynonymous changes (**Figure S8)**. We observe an approximately normal distribution for all exome variants between -2 and 6.5, but for nonsynonymous variants, there is sharp decrease in the number of variants greater than GERP score of 4. This difference in the distributions is consistent with the prediction that more conserved nonsynonymous sites are more likely to be functionally important and therefore subject to purifying selection when mutations occur at highly conserved sites. We focus on all exome variants in the analyses that follow. In order to explore a possible ancestral bias we examined the site frequency spectra across effects for the different populations. **Figure S4** shows results for moderate effect variants. Though no evidence of an ancestral bias was detected, we note that excluding variants where the ancestral allele is the alternate allele according to the reference sequence has a very strong effect in the site frequency spectrum overall shape. While the true fitness of these mutations cannot be measured directly, GERP scores are indicative of long-term selection in many species and the severity of mutation effect should be similar in human populations.

Local Ancestry Assignment

Local ancestry segments in the 8 Mayan samples were inferred using RFMix (Maples et al., 2013). Two reference panels were constructed, one for Native American ancestry and one for European ancestry. The Native American reference panel was constructed by including all samples from the Maya, Pima, Columbian, Karitiana and Surui populations in HGDP(Li et al., 2008). The European reference panel was constructed by using all samples from the Sardinian and French populations. One Mayan sample at a time from the Native American reference panel was removed to form the admixed panel for the initial inference step. RFMix was run in

PopPhased mode with the "Generations After Admixture" parameter set to 12. Expectation-maximization (EM) was performed and the results from the first iteration were used for analysis. All other RFMix parameters were left as their default values.

Individual-based simulations

To simulate changes in heterozygosity across human populations during a range expansion with founder effects, we kept track of allele frequencies at a set of 100 loci. All loci are diallelic and unlinked. At selected loci, the ancestral allele is assumed selectively neutral and mutants reduce an individual's fitness by a factor 1-*s* only if it is present in homozygous state, that is, deleterious mutations are completely recessive. Because we are modeling mutations at single nucleotides, we assume the frequency of back-mutation to be sufficiently rare that it can be neglected, and that each mutation occurs at a unique locus. We modeled human range expansion across an array of 10x100 demes, with an ancestral population restricted to the first 10x10 demes at one edge of the habitat. After reaching migration-selection-drift equilibrium, populations expand into the empty territory, which is separated from the ancestral population by a geographical barrier, through a spatial bottleneck (to mimic the bottleneck out of Africa, see **Figure S19** for an illustration of the model). After 3,000 generations, we computed the average expected heterozygosity for all populations. To compare the simulation results with the data, the spacing of demes was chosen such that distance between two neighboring demes is 250 km. Since computational limitations of individual-based simulations prohibit a complete exploration of the parameter space for this model, we focused on a set of reasonable demographic and mutations parameters (K = 100 diploid individuals per deme, mutation rate of $u = 10^{-5}$ per locus per generation), and the migration rate and selection coefficient were adjusted to generate heterozygosity consistent with the observed data, without formally maximizing the fit.

Models of dominance

Several models of dominance were considered in the calculation of mutational load. Formally, h=0 if the mutation is completely recessive (ancestral homozygotes and heterozygotes have the same fitness), h=0.5 indicates that the mutation effect is additive (the fitness is exactly intermediate between the reference homozygote and the alternate homozygote) and when h=1 the mutation is dominant (heterozygotes and derived homozygotes have the same fitness). We also consider a dominance model developed from mutation-accumulation results where the dominance coefficient is inversely related to the selection coefficient by an asymptotic distribution. Specifically, h decreases from additive to recessive as the selection coefficient becomes stronger.

Testing for a recessive model of dominance

Hardy-Weinberg: If deleterious variants are completely recessive, we would expect a deficit of derived homozygous mutations (or conversely, an excess of heterozygotes) as purifying selection would tend to remove recessive homozygotes. One might test for this hypothesis by considering the ratio of heterozygotes to derived homozygotes for different function effect classes; the het/hom_{Der} ratio increases as variants are predicted to be of greater effect. However, this pattern could also be due to the enrichment of low frequency variants (namely singletons) by purifying selection alone without a significant number of recessive variants. We thus considered the het/hom_{Der} ratio in the Luhya population, removing singletons from the dataset and calculating the ratio for different frequency bins (**Figure S15**). Our results show that even after removing singletons, extreme variants are enriched for heterozygotes, in the low frequency bins. This is consistent with a recessive model of purifying selection, whereby recessive homozygotes are more likely to be removed.

We also investigated deviations from Hardy-Weinberg using the polymorphic exome sites in the Luhya population from the 1000 Genome project (**Figure S16A**) by plotting the observed number of derived homozygotes versus heterozygotes. Color indicates the number of observations found in each bin (i.e. the number of sites that have x homozygotes and y

heterozygotes.) We used the derived allele frequency, q, to calculate the number of heterozygotes (het^{HW}) and derived homozygotes (hom^{HW}_{Der}) under the Hardy Weinberg expectation. Variants that do not follow the neutral pattern with a p-value of 0.01 are shaded.

$$\hom_{Anc}^{HW} = p^2 \times N = (1-q)^2 \times N$$
$$het^{HW} = 2pq \times N = 2(1-q)q \times N$$
$$\hom_{Der}^{HW} = q^2 \times N$$

where *N* is the population sample size. To calculate the significance we used the Chi-Square statistic to test whether the observed genotype frequencies where significantly different from the ones expected under Hardy-Weinberg Equilibrium, with a p-value of 0.01 and 1 degree of freedom.

$$\chi^2 = \frac{\left(\frac{\hom_{Anc}^{Obs}}{N} - \frac{\hom_{Anc}^{HW}}{N}\right)^2}{\frac{\hom_{Anc}^{HW}}{N}} + \frac{\left(\frac{het^{Obs}}{N} - \frac{het^{HW}}{N}\right)^2}{\frac{het^{HW}}{N}} + \frac{\left(\frac{\hom_{Der}^{Obs}}{N} - \frac{\hom_{Der}^{HW}}{N}\right)^2}{\frac{\hom_{Der}^{HW}}{N}}$$

Proportion of deleterious variants in dominant and recessive genes

We additionally tested for a recessive model of dominance by examining the average proportion of neutral, moderate, large and extreme effect variants in known recessive and dominant genes. With this purpose, we used the OMIM database (ftp.omim.org) to obtain a list of genes and physical positions of autosomal genes related with a recessive or dominant disease, and classified with a *Confirmed* status. Genes associated with both dominant and recessive diseases were excluded from the dataset. In this way we had a list of regions in the genome related with recessive and dominant diseases, respectively.

We next examined those regions in our HGDP exome dataset, as well as in 1000G Agilent exome dataset. For each gene we calculated the proportion of variants within each effect, and weighted the proportions according to the length of the gene. Specifically, genes further away from the median gene length distribution were down weighted. We then averaged the proportion of the number of variants within each effect category (**Figure S16B**) and performed a Wilcoxon test to determine if the distribution of the proportion of LARGE effect variants were different between dominant and recessive genes. Results for HGDP were not significant with a p-value of 0.06, but results for 1000G reached significance with p-value of 0.03. In both cases the proportion of LARGE effect variants in dominant genes was lower than in recessive genes, suggesting that the distribution of high effect variants varies with the degree of dominance of the gene or the genotype.

Model for the underlying distribution of dominance

We aimed to relate the dominance coefficient, *h,* and the absolute value of the selection coefficient, *s,* for deleterious single-nucleotide mutations segregating in human populations.

1. Boundary conditions

To begin, we make use of a relationship between *h* and *s* that was previously obtained for mutations in yeast (Agrawal and Whitlock, 2011):

$$h(s) = \frac{\beta_1}{1+\beta_2 s} - d \qquad (1)$$

where β_1, β_2, and d are some constants. As described in Agrawal and Whitlock and others (Agrawal and Whitlock, 2011), as selection strength increases, the dominance coefficient tends to zero. In other words, we assume that strongly deleterious mutations are fully recessive:

$$\lim_{s\to\infty} h(s) = -d = 0$$

We also make use of a frequent assumption that the dominance coefficient for neutral mutations (i.e. those for which s = 0) is equal to ½:

$$h(0) = \beta_1 = \frac{1}{2}$$

When we specify d = 0 and β_1 = ½ in equation (1), the dependence of the dominance coefficient on the selection coefficient becomes

$$h(s) = \frac{1/2}{1 + \beta_2 s}$$

2. Least-squares fit

In order to fins the best value for parameter β_2 in h(s) above, we start by defining h(s) as a function of both s and β_2:

$$h(\beta_2, s) = \frac{1/2}{1+\beta_2 s} \qquad (2)$$

We now make use of the selection coefficients we have obtained independently for four GERP categories of single nucleotide polymorphisms segregating in human populations. The absolute values of these selection coefficients are – in order of increasing selection strength – $s_0 = 0$, $s_1 = 10^{-4}$, $s_2 = 10^{-3}$, $s_3 = 2 \times 10^{-3}$.

We assume that, of the four classes of mutations mentioned above, the one with the smallest selection coefficient has a dominance coefficient that is very close to ½ and that the class of mutations with the largest selection coefficient has a dominance coefficient that is very close to zero.

One can show that the former requirement, that $|h(\beta_2, s = s_0) - ½|$ is minimized, tends to decrease β_2. At the same time, the latter requirement, that $|h(\beta_2, s = s_3) - 0|$ is minimized tends to increase β_2. If we require that the sum of $|h(\beta_2, s = s_0) - ½|$ and $|h(\beta_2, s = s_3) - 0|$ is minimized – or, similarly, that the sum of the squares of these two expressions is minimized – one obtains an intermediate value of β_2 that corresponds to a balance between the two requirements. In other words, we are looking for

$$\arg\min_{\beta_2} f(\beta_2),$$

the value of β_2 that results in a minimum of function $f(\beta_2)$, defined below:

$$f(\beta_2) = \left[\left(h(\beta_2, s = s_0) - \frac{1}{2}\right)^2 + (h(\beta_2, s = s_3) - 0)^2\right] \qquad (3)$$

In order to find β_2 that minimized $f(\beta_2)$, we take derivative of that function with respect to β_2 and set it to zero:

$$\frac{df}{d\beta_2} = \frac{\frac{1}{2}s_0^2\beta_2}{(s_0\beta_2+1)^3} - \frac{\frac{1}{2}s_3}{(s_3\beta_2+1)^3} = 0 \qquad (4)$$

When we use $s_0 = 0$ and $s_3 = 2 \times 10^{-3}$ in equation (4) and look for positive, real roots of that equation, we find that the only such root is

$$\beta_2 = \frac{1}{\sqrt{s_0 \times s_3}} = 7071.07$$

and that $f(\beta_2)$ is at its lowest value at this root if β_2 is restricted to be greater than 0.

3. Dominance coefficient function.

We can now use the resulting dependence of the dominance coefficient on the selection coefficient,

$$h(s) = \frac{1/2}{1 + 7071.07 \times s},$$

to obtain *h* for various values of *s*:

$$h(s_0) = h(0) = 0,$$
$$h(s_1) = h(10^{-4}) = 0.292893,$$
$$h(s_2) = h(10^{-3}) = 0.0619497 \; and$$
$$h(s_3) = h(2 \times 10^{-3}) = 0.0330204.$$

4. Variance of the dominance coefficient

We also make use of a previously described function, namely, a displaced gamma distribution, which has been shown to be a best fit to the data in previous studies (Agrawal Whitlock, 2011). In summary, the dominance coefficient for a given variant follows the equation:

$$h_{del,k}[s_j, \beta_1, \beta_2, \sigma_h^2, \delta] = \mu_{h(del),j} + Q_G\left(\delta^2/\sigma_h^2, \sigma_h^2/\delta, \frac{1}{25}\left(k - \frac{1}{2}\right)\right) - d,$$

where s_j, β_1, β_2 have already been estimated, $\mu_{h(del),j}$ is the dominance coefficient for each selection coefficient that has also been calculated, $d \approx \delta$, and $\sigma_h^2 \; and \; \delta$ are the variance and the mean of Q_G, which is a gamma distribution to introduce variance to the dominance coefficient.

Values for $\sigma_h^2 \; and \; \delta$ have been taken from (Agrawal and Whitlock, 2011) and are 0.010 and 0.038, respectively.

Testing for significance in differences in Load

In order to test whether differences in Load under the different models of dominance (**Fig. 5**) are significant we performed 1,000 iterations under each model where the 54 individuals in the dataset were randomly re-assigned to the 7 populations. For each iteration we would recalculate Load accordingly to the model of dominance assumed and then calculate the maximum difference in Load (Δ_{Load}) obtained in the simulated mosaic dataset. After the 1,000 iterations we would compare the real Δ_{Load} and the mosaic Δ_{Load}, and determine if the real observation was a statistical outlier (**Figure S22**). Under the recessive and intermediate model there were virtually no scenarios in which simulated Δ_{Load} was larger or equal to the observed one. For the additive model, the observed Δ_{Load} was still statistically significant, with only 1.6% of the mosaic populations having a greater value than the real one (p-value < 0.05).

Supplementary Results

PSMC Simulations and demography

We constructed profiles of effective population size through time using PSMC method (Li and Durbin, 2011). Since this model relies on heterozygous sites within an individual it is not applicable to low or moderate coverage whole genome sequencing. However, if the rate of 'missing' heterozygotes is known and uniform, the PSMC curves can be corrected through a rescaling of the mutation rate to an effective rate that incorporates heterozygote false negative rates. We applied this rescaling idea, utilizing Pathan sample HGDP00222, which has 22x coverage, as a test case. We subsampled reads from this sample to lower coverage levels, ran the PSMC calling procedure on the sub sampled read sets, and compare the proportion of heterozygous sites identified at each coverage level. Based on this, we constructed a correction curve relating coverage level with to heterozygote SNP false negative rates. We found that reasonable concordance between down-sampled and original PSMC curves could be obtained for coverage levels >10x. Since all of the samples were sequenced and processing in the same manner, we reasoned that the correction curve constructed for HGDP00222 would be applicable to other samples in this data set. We verified this through comparison of our corrected PSMC curves with PSMC curves constructed from a high coverage San individual and a high coverage Mbuti Pygmy sample obtained from (Prüfer et al., 2013) (**Figure S1)**.

Effect of sample size on mean number of homozygotes

We observe approximately equal numbers of extreme homozygotes per individual, unlike other effect ranges. The pattern may be the result of strong purifying selection equally efficient in different populations in removing homozygotes. However, these results could also be due to lack of power to find differences across populations due to the small number of variants we observe in the extreme effect category. One way to test this hypothesis is to sub-sample the same number of extreme homozygotes for the other effect categories and test whether there is a difference among populations. We took a random individual from the San population and counted the number of extreme homozygotes, n=24. We then randomly sampled 24 variants in the neutral, moderate, and large categories and calculated the homozygotes per individual within each population. We iterated over 1,000 bootstraps. Results can be found in **Figure S12A-D**, and demonstrate that the number of homozygotes increases with distance from Africa for each effect even for a small sample of variants. This result lends support to the interpretation that the pattern **Figure 2F** is due to strong purifying selection, rather than low power to detect a cline.

Effect of sample size on A_i for each functional category

In order to find out whether the observed pattern for moderate, large and extreme variants is actually a consequence of differences in variant sample size across categories we opted for following strategy. For each effect category, we randomly selected an increasing number of variants, and calculated individual load for the selected set. If the pattern of mean individual load across populations is random and a consequence of the variant sample size, one would expect a certain stochasticity in the individual counts, independent from the observations in **Figure 2D-F**. Alternatively, if the minimum informative sample size is reached, the pattern is expected to remain constant from that point on. Results are shown in **Figure S2** and show that the pattern we see in the individual load boxplots is already visible with fewer variants. This is especially true for the large effect variants, where the increase in derived counts with distance from Africa is can be detected with only 7,000 variants (vs. the more than 25,000 variants in the full exome dataset).

Extreme alleles (GERP >= 6) across populations

We were interested in looking at the pattern of extreme alleles across populations. Population theory predicts that extreme alleles will be held at low frequencies if their effect is deleterious, and eventually be eliminated. The SFS of extreme variants shows an excess of low frequency variants (namely singletons), compared to the neutral SFS (**Figure S13**). For a given

population, no less than 45% of the extreme variants are singletons. We next asked how variants were distributed across the complete dataset (**Figure S20A**). Surprisingly when we consider all the populations 60% of the variants are singletons (514 out of 854). If we focus on variants private to a specific population the percentage increases to 76%. Thus, the vast majority of variants with extreme effect are either new or kept at very low frequencies, being private to a population. Interestingly, few variants (a dozen) are almost fixed in the dataset. This could be due to errors in the assignment of the ancestral allele or evidences of positive selection.

When we focus on variants found in homozygosity (**Figure 20B**) we observe as expected an increase in the number of homozygotes, with distance from Africa. Sub-Saharan African populations have more variants in homozygosity that are found only once in the dataset (like "homozygous – singletons"), whereas Out of Africa populations have more homozygous singletons at higher frequencies. Some variants are found at high frequencies in African populations, and are found at lower frequencies elsewhere.

Hardy-Weinberg Equilibrium Test

We tested for deviations from Hardy-Weinberg equilibrium in a sample of 72 Luhya individuals from 1000 Genomes Nimblgen exome capture (**Figure S16A**). We show that there is an excess of heterozygotes compared to Hardy-Weinberg expectations, particularly when the homozygotes are at low frequency. However, no extreme effect alleles were found to have significantly more heterozygotes than predicted. The bulk of the heterozygotes with a paucity of corresponding derived homozygotes occurs in the neutral and moderate effect categories. We conclude that alleles are either generally additive or moderately recessive such that incomplete penetrance does not cause them to significantly violate Hardy-Weinberg at $p<0.01$. Alternatively, we note that the HW model has low power for rare allele frequencies, so if most selection occurs against deleterious recessive variants less than 25% in frequency than this test does not have sufficient power to detect deviations from an additive model. For example, if there is one derived homozygotes in the population then there would need to be more than 37 heterozygotes to deviate from Hardy-Weinberg at $p<0.01$, an allele frequency of at least 28%. Interestingly, we also observe many variants that have a deficient number of heterozygotes / excess of homozygotes. This pattern can occur due to haploinsufficiency (Huang et al., 2010) or false negatives in the next-generation sequencing data (i.e. heterozygotes are more error prone for variant calling software).

Inference based on the site frequency spectrum:

Although we classify extreme effect mutations as being potentially deleterious, there is also a possibility that these mutations are functionally adaptive, large effect mutations that are under positive selection. We test this hypothesis by considering the site frequency spectrum (SFS) of predicted extreme and neutral effect mutations. For each population, we considered the number of extreme and neutral effect variants in each allele frequency bin, proportional to the total number of mutations in the extreme and neutral category such that the spectra are directly comparable. While the two spectra generally demonstrate an exponential decay, as expected under constant size or low population growth, there is an enrichment of extreme effect mutations in low frequency bins for all populations. This observation is consistent with other studies that have shown an enrichment of deleterious alleles at low frequencies (Fu et al., 2013; Nelson et al., 2012). Some populations also display an enrichment of extreme effect variants at intermediate frequencies (e.g. Pathan), potentially indicative of adaptive alleles under balancing selection; such inference would require additional modeling (Andres et al., 2009). No populations display an enrichment of extreme effect alleles at fixation, suggesting that overall, selective sweeps have not played a dominant role in shaping the frequencies of extreme effect alleles (Hernandez et al., 2011). No such pattern is present for large effect alleles either (**Fig. 3B**).

Mutation Load

It has also been argued that the relationship between effective population size and load is non-linear for a model with partially, but not completely, recessive mutations (i.e. h=0.05) (Kimura et al., 1963). This is because in a population with larger effective size, mutations of equal s are less likely to be lost by drift and thus recessive deleterious alleles can float to higher frequencies, impacting more individuals when exposed as homozygotes. We do not observe this effect within African populations, which carry fewer weakly deleterious alleles per individual than non-African populations (**Figure 2A**).

It is also interesting to note that there are negligible differences in additive load between western Africans and Europeans. This is in keeping with the fact that western African populations have experienced dramatic population growth over the past 5,000 years (Tennessen et al., 2012), which alters the distribution of deleterious alleles within a population (Gazave et al., 2013). There are sharp differences in demography among African populations, and populations with western African ancestry should not be taken to be representative of all of Africa.

Table S1: Genome and Exome Variant Statistics by Population after Imputation

	San	Mbuti	Mozabite	Pathan	Cambodian	Yakut	Maya
Sample Size	6	7-8[5]	8	8	8	8	8
Genome Statistics							
Coverage[1]	10.57x	6.67x	6.32x	8.93x	7.41x	5.96x	7.86x
NR alleles[2]	3976209	3826512	3240806	3121928	3100036	3072826	3008568
Heterozygotes[3]	2424664	2316159	2002220	1870784	1762812	1715462	1609374
Singletons[4]	223066	151579	75293	66821	59120	36385	37099
T_i/T_v	2.166	2.17	2.176	2.175	2.17	2.168	2.167
NR alleles ≥ 2 reads	3948479	3774409	3198236	3089167	3066118	3024982	2969049
Homozygous NR concord.	0.992	0.979	0.981	0.987	0.990	0.987	0.991
Heterozygous concordance	0.978	0.964	0.981	0.986	0.990	0.986	0.988
Exome Statistics							
Coverage[1]	82.3	77	85	75.5	77	78	75.5
NR alleles[2]	34918	34148	28486	27380	27048	26889	26233
Heterozygotes[3]	21366	20994	17914	16645	15652	15232	14218
Singletons[4]	2936	2392	1513	1424	1328	1061	980

[1] Mean population coverage for genomes assessed by sampling each individual for ~650,000 sites on the Illumina Human660K BeadChip SNP platform and counting read depth after quality filtering. Median population coverage for the exomes encompassing all mapped, on target reads.
[2] Mean number of non-reference alleles for an individual in the population (i.e., a non-reference homozygous genotype is counted twice.)
[3] Mean number of heterozygotes for an individual in the population.
[4] Mean number of singletons for an individual in the population.
[5] Eight individuals were included for exome and genome sequencing; one sample did not pass genome quality control and was excluded from the full genome dataset.

Table S2: ANNOVAR functional annotations as compared to GERP

RS Score	Inter-genic	Intronic	UTR-5	UTR-3	Mis-sense	Non-sense	Synon	Total
Neutral: -2,2	5566	28592	1167	1751	14614	187	16883	68760
Moderate: 2,4	1676	8956	540	721	13648	160	9913	35614
Large: 4,6	682	2789	248	314	19645	183	4935	28796
Extreme: >6	11	64	4	8	741	7	88	923
Total	7935	40401	1959	2794	48648	537	31819	134093

Figure S1: Assessment of PSMC Coverage Correction

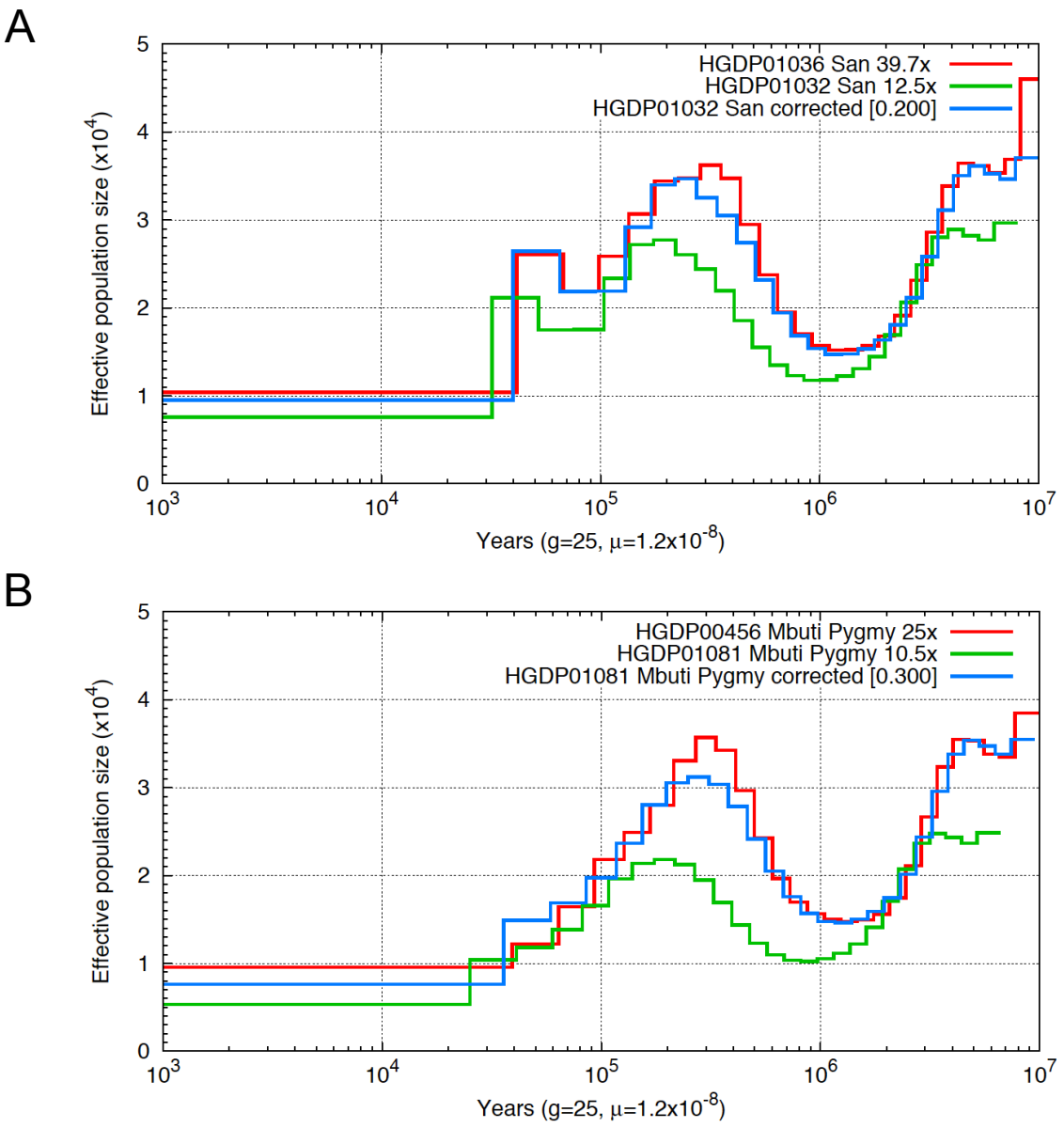

Figure S1 ***Assessment of PSMC coverage correction*****. A** PSMC curves from original moderate coverage data before and after coverage correction are compared with **B** PSMC curves constructed from high-coverage sequences from the same populations. Strong concordance is observed, with discrepancies mostly restricted to the point of maximum population size inferred by PSMC.

Figure S2: Determination of whole genome masks

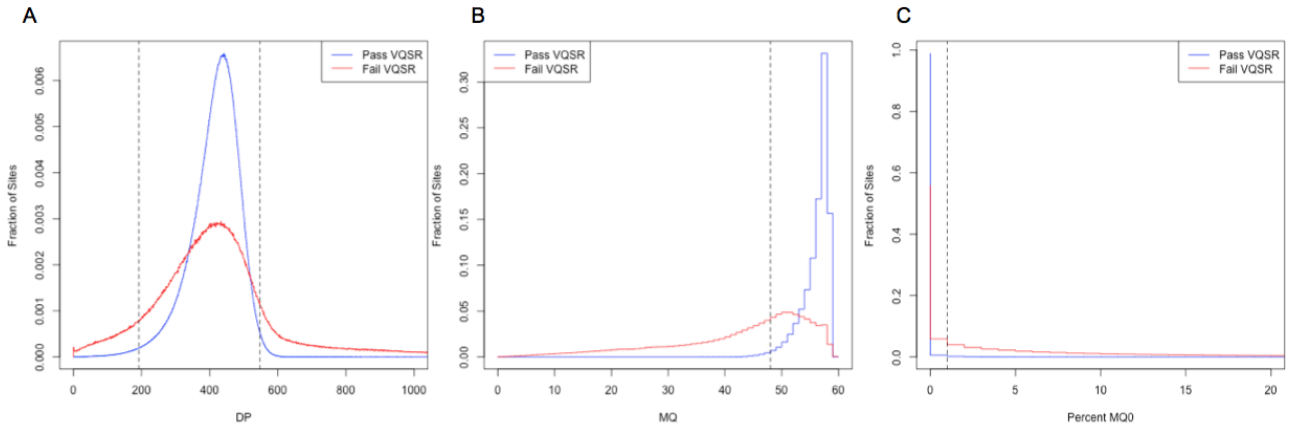

Figure S2: ***Determination of whole genome masks***. Distribution of DP, MQ, and MQ0 fraction values for genomic sites that pass (blue) and fail (red) the VQSR procedure are shown. Cutoffs correspond to 192 <= DP <= 547, MQ >= 48 and MQ0 fraction <= 0.01.

Figure S3: Genotype Concordance for Full Genome Data

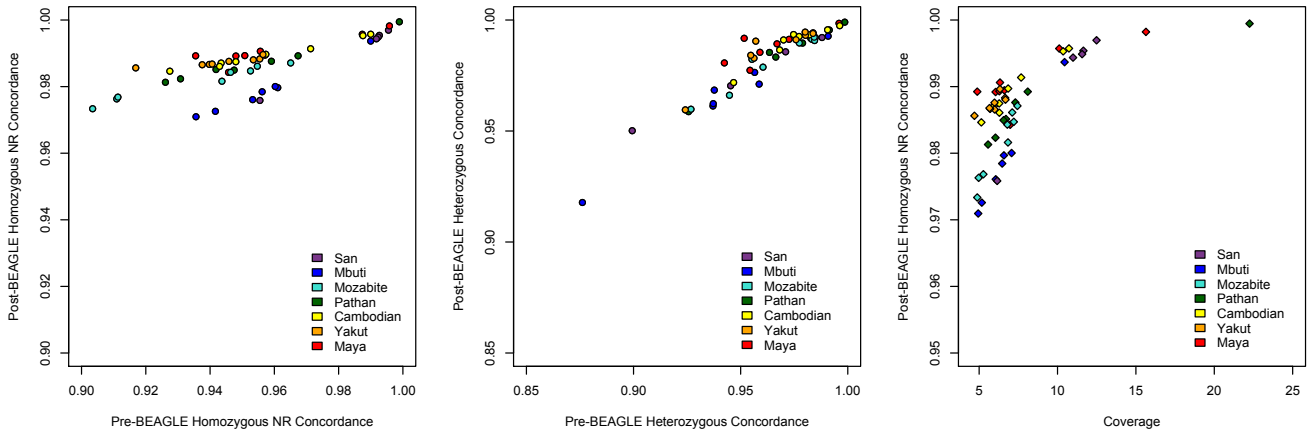

Figure S3: ***Genotype Concordance for Full Genome Data***. Phasing and imputation for the full genomes was performed using BEAGLE v3.2. We assessed genotype concordance for SNP calls pre- and post-BEAGLE by comparing genotypes to the Illumina 660K SNP array data for each individual. A) Concordance between homozygous non-reference genotypes for each of 53 individuals. B) Concordance between heterozygous genotypes. C) Relationship between concordance at homozygous non-reference genotypes for the post-BEAGLE imputed genome data and overall genome coverage.

Figure S4: Contrasting the SFS for Ancestral and Derived Alleles.

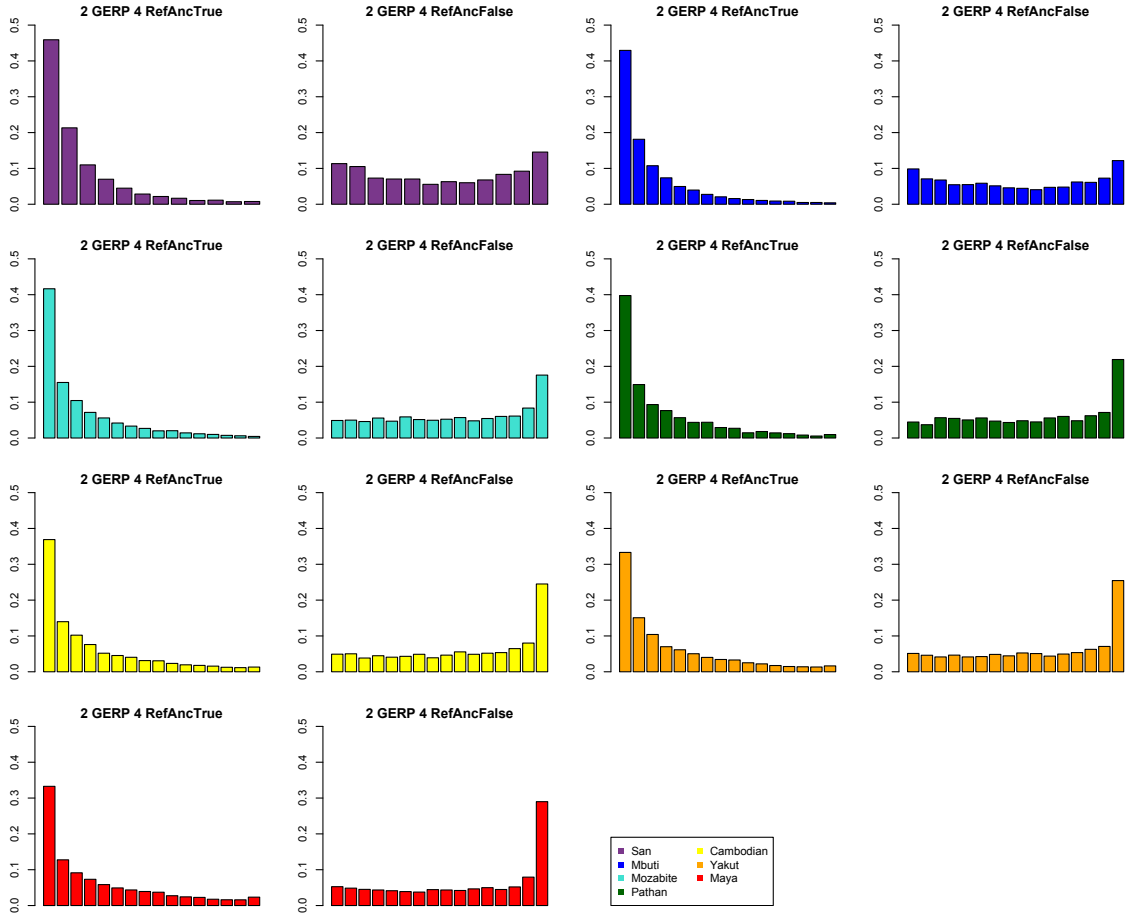

Figure S4: ***Contrasting the SFS for Ancestral and Derived Alleles***. We asked whether there were systematic differences in the SFS for ancestral and derived variants, relative to the human reference genome. Shown are moderate effect variants, GERP >2 and <4. The left plot for each population shows the SFS for which the reference allele is ancestral, and thus the non-reference allele is derived. The right hand SFS shows the opposite pattern, where the reference allele is derived and the non-reference allele is ancestral. This pattern has been observed elsewhere (Chen et al., 2007), and is even expected because alleles that have already been observed once, in the human reference genome, have a higher probability of being observed again when sampling a new population. OOA populations, being more closely related to the human reference genome, have more alleles that have been previously observed in the single human reference sample. African populations have a higher proportion of novel, derived alleles (or conversely fewer derived alleles shared with the reference).

Figure S5: Number of heterozygotes per individual genome for 7 populations

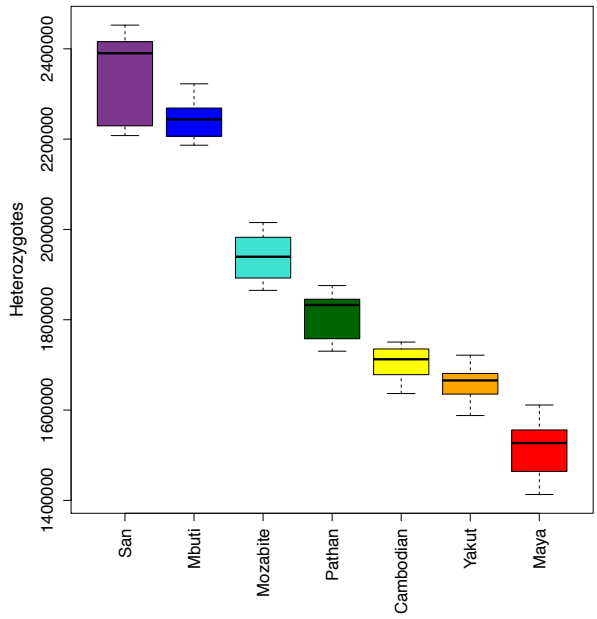

Figure S5: ***Number of heterozygotes per individual genome for 7 populations.*** Boxplots of number of heterozygotes per individual from the 2.48Gb callable region of the human genome for all 7 seven populations.

Figure S6: Karyograms of the Maya individuals reflecting the inferred ancestry

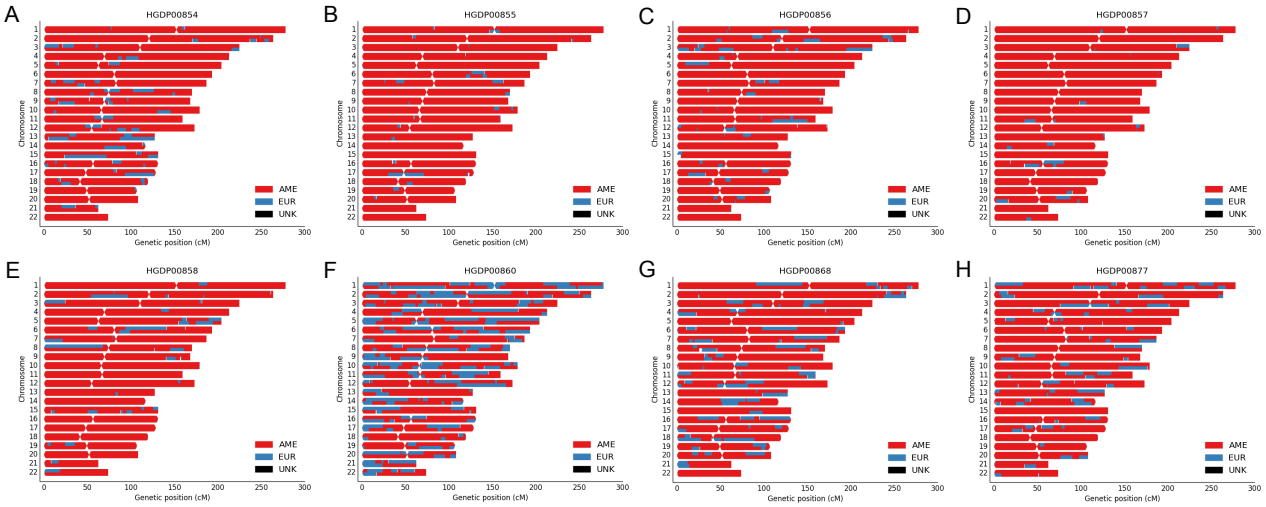

Figure S6: Estimates of European (blue) and Native American (red) ancestry at the chromosome level were plotted for every individual (**A-G**). Every pair of chromosomes is depicted along the Y-axis and the genetic position is reflected on the X-axis. Note that two out of eight individuals (**F,G**) showed more than 20% of European ancestry and were thus removed from analysis based on deleterious variants.

Figure S7: Simulations of Bottleneck Length and Magnitude as Inferred from PSMC.

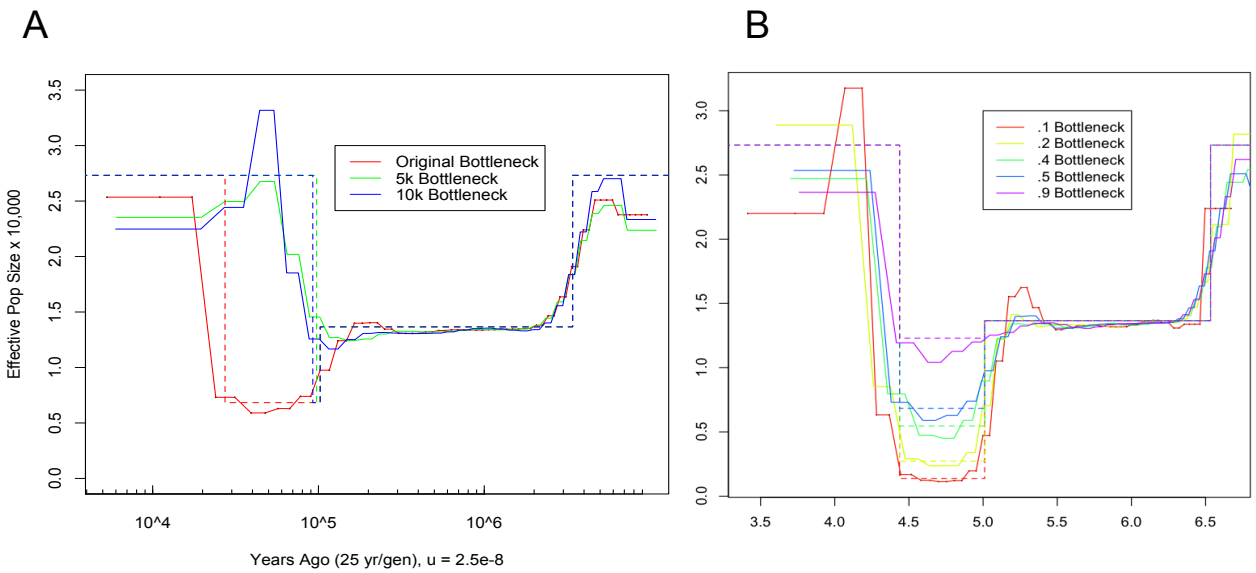

Figure S7: ***Simulations of Bottleneck Length and Magnitude as Inferred from PSMC.*** We tested whether PSMC was robust to changes in either the duration of a bottleneck or the magnitude of a population bottleneck. **A)** Using a simulation of population history parameters similar to the original paper (Li and Durbin, 2011), we varied the duration of a bottleneck to reflect more realistic 5,000 or 10,000 year periods. The inferred time of the bottleneck is substantially overestimated for briefer bottleneck periods (by approximate 25% to 75% for the tested scenarios). Additionally, when the bottleneck is of brief duration the magnitude of the bottleneck is underestimated. **B)** Using the original 70,000y bottleneck, we varied the magnitude of the reduction in effective population size. The magnitude of shallower bottlenecks maybe somewhat overestimated, but approaches accuracy for severe (e.g. 90%) reductions in effective population size.

Figure S8: Distribution of derived variants with conservation scores -2 ≤ GERP ≤ 6.5

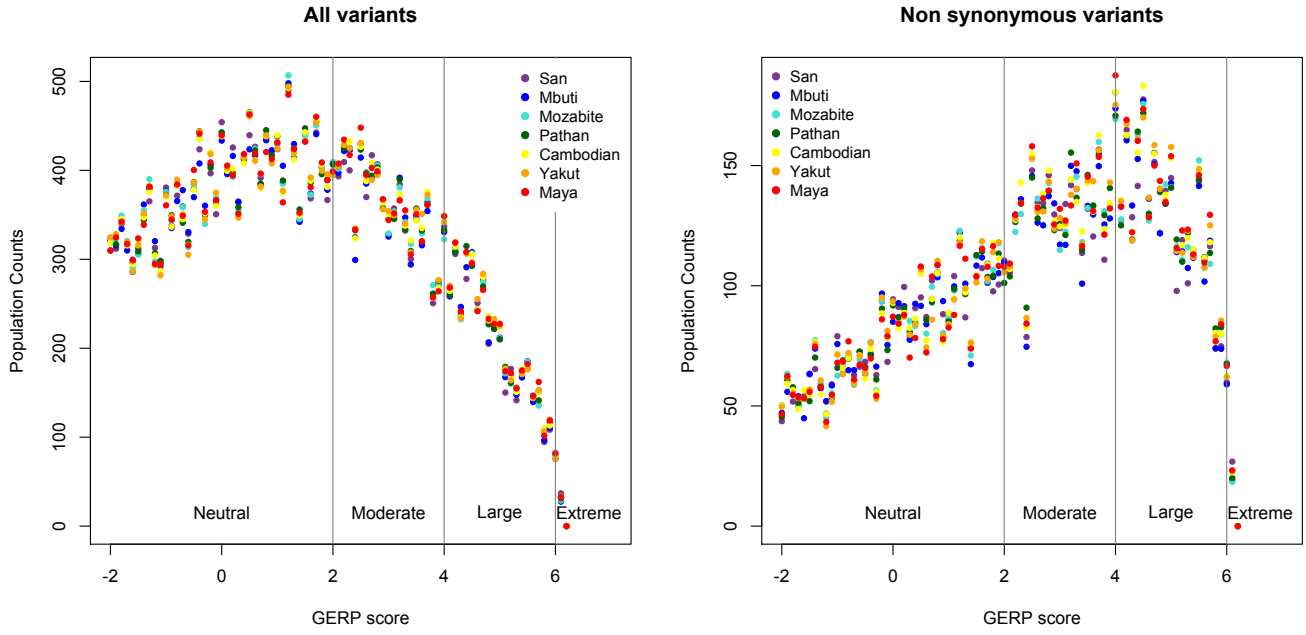

Figure S8: ***Distribution of derived variants with conservation scores -2 ≤ GERP ≤ 6.5.*** For bin sizes of 0.2, the number of derived variants within each population are plotted according to the prior population color scheme (Figure 1A). Binned counts were standardized by the number of samples per population. GERP scores were divided into four functional categories: neutral (-2 to 2), moderate (2 to 4), large (4 to 6), extreme (>6). A) Nonsynonymous variants are not normally distributed. B) All exome variants conform to a normal distribution. No population had a significant excess or deficit of variants within a particular GERP score range.

Figure S9: Median number of derived variants per individual

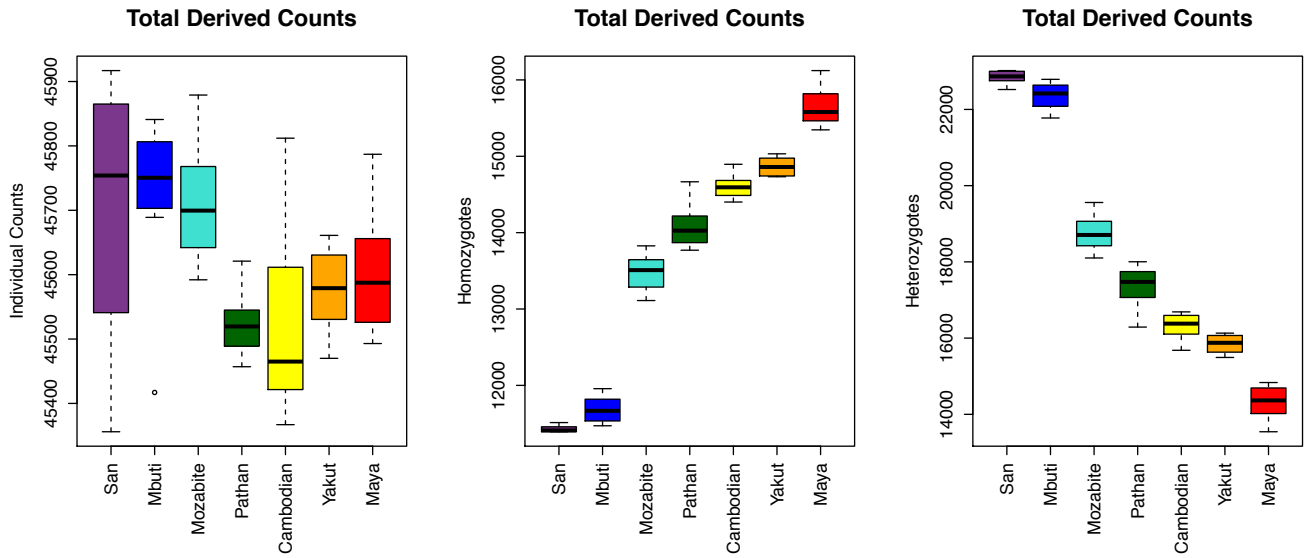

Figure S9: ***Median number of derived variants per individual***. For all variants, regardless of GERP score annotation, we tabulated the number of derived variants per individual, heterozygotes and derived homozygotes. Out of Africa populations have roughly equivalent numbers of derived variants per individual. African populations have ~1% fewer derived variants per individual.

Figure S10: Individual counts of Neutral derived variants

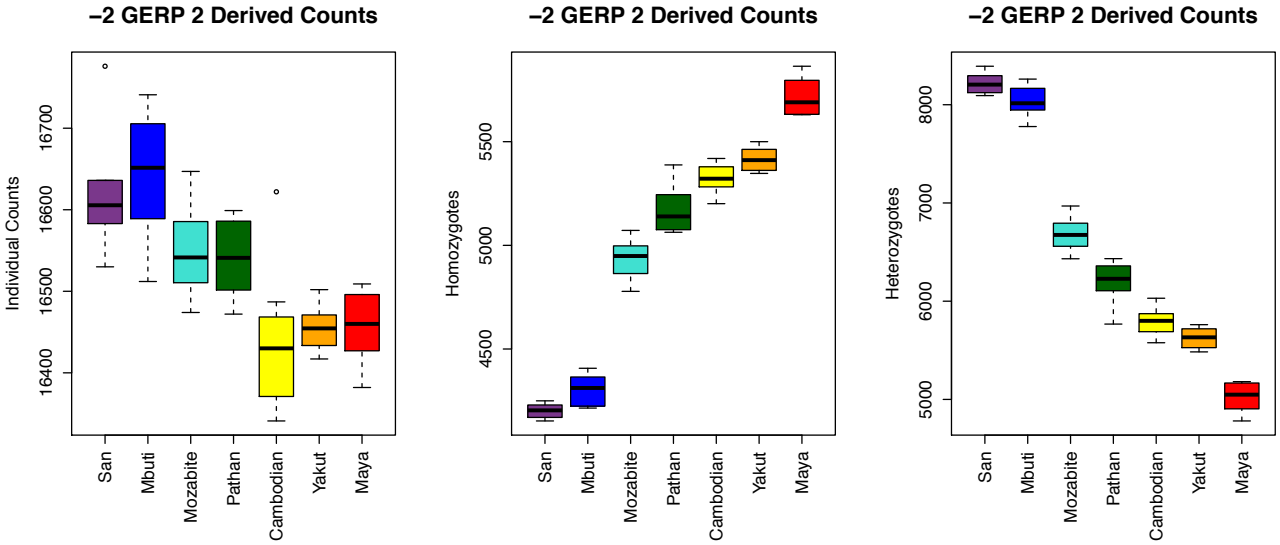

Figure S10: ***Individual counts of Neutral derived variants***. For exome variants with GERP score in the -2 to 2 range, we evaluated the average number of A) The total number of derived variants (equivalent to number of heterozygotes + twice the number of homozygotes) B) derived homozygotes and C) heterozygotes by population.

Figure S11: Number of Common and Rare variants per individual's genome by predicted effect

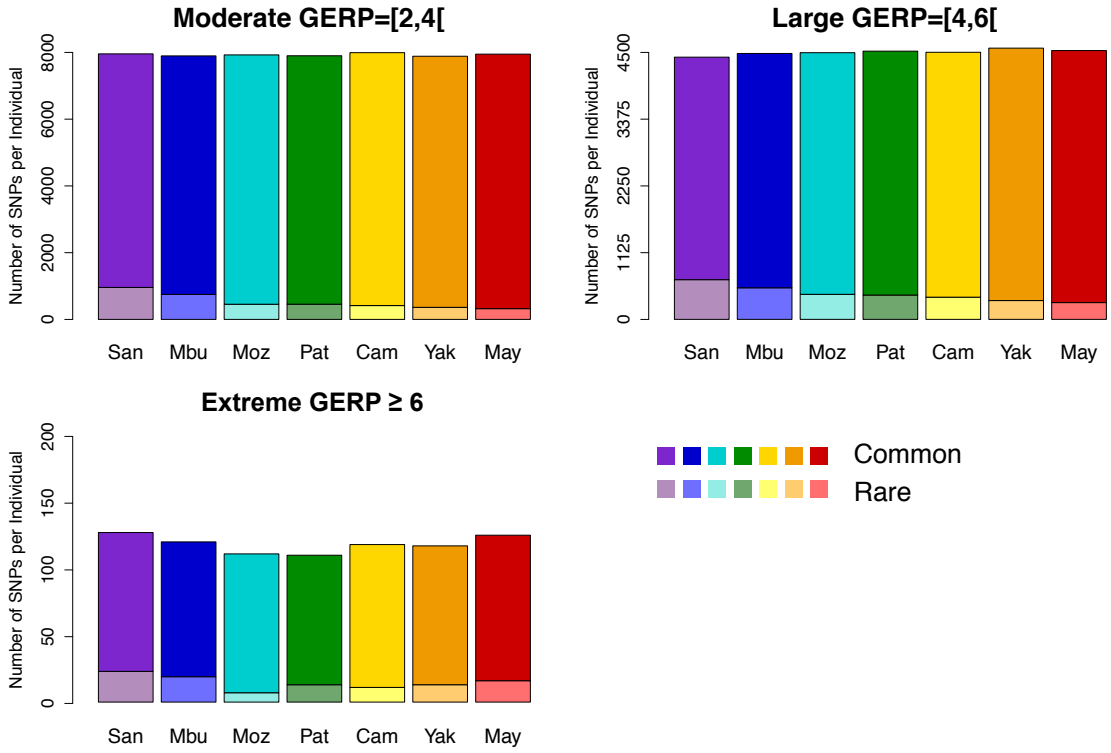

Figure S11: ***Number of common and rare variants per individual's genome by predicted effect***. For a given individual, deleterious variants within each predicted effect category were divided into common (>10%, solid colors) and rare (<10%, shaded colors). The contribution of common deleterious variants to an individual's burden is much greater than rare variants.A) Moderate B) Large and C) Extreme.

Figure S12: Number of homozygotes per population with subsampling

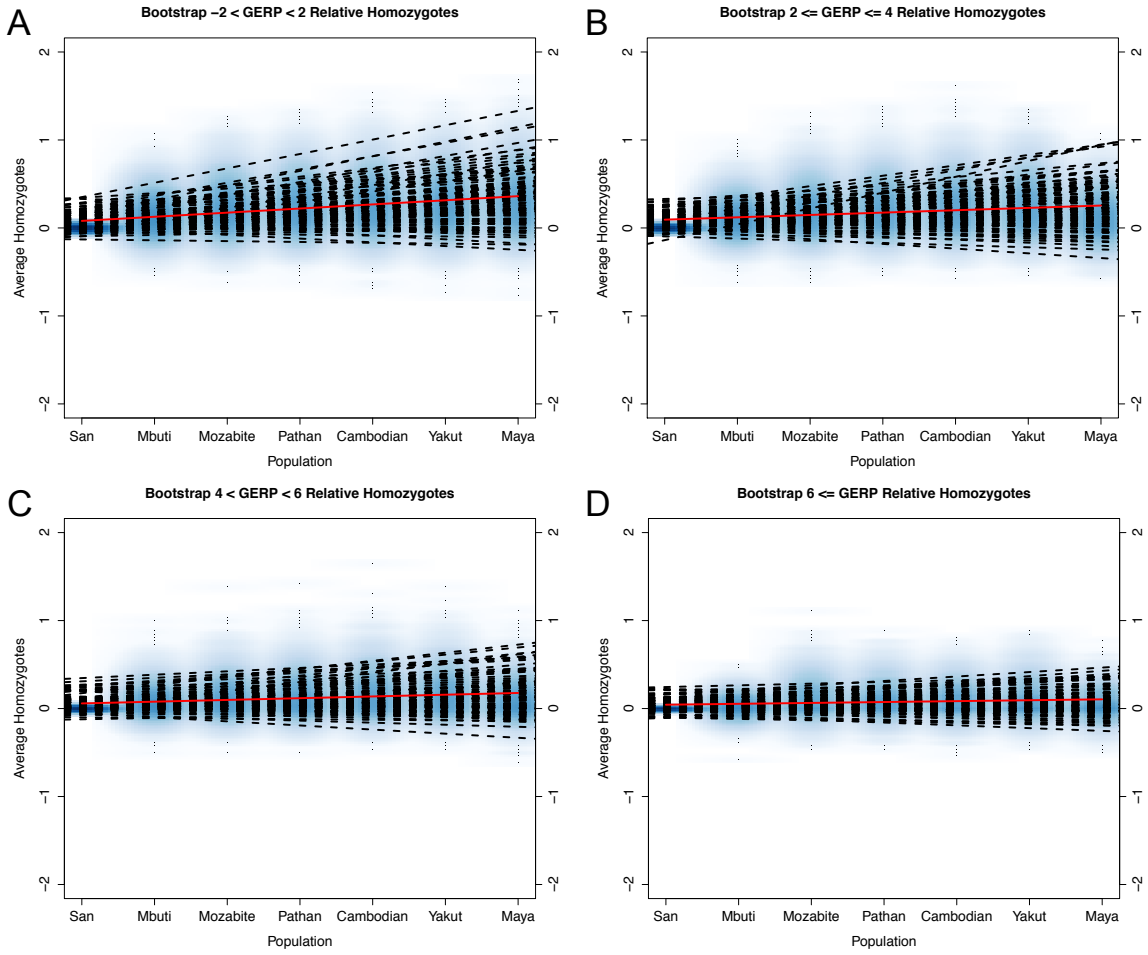

Figure S12: ***Number of homozygotes per population with subsampling***. The minimum number extreme homozygotes in a San individual, 26, was used to sub-sample the variants of neutral, moderate, large and extreme effect, and calculate the average number of homozygotes per population. The red line indicates the average number across 10,000 bootstraps, the dashed lines indicate the average number per population for every bootstrap and blue background indicated the individual ranges for every bootstrap.

Figure S13: Site Frequency Spectra (SFS) of Neutral and Extreme effect Variants

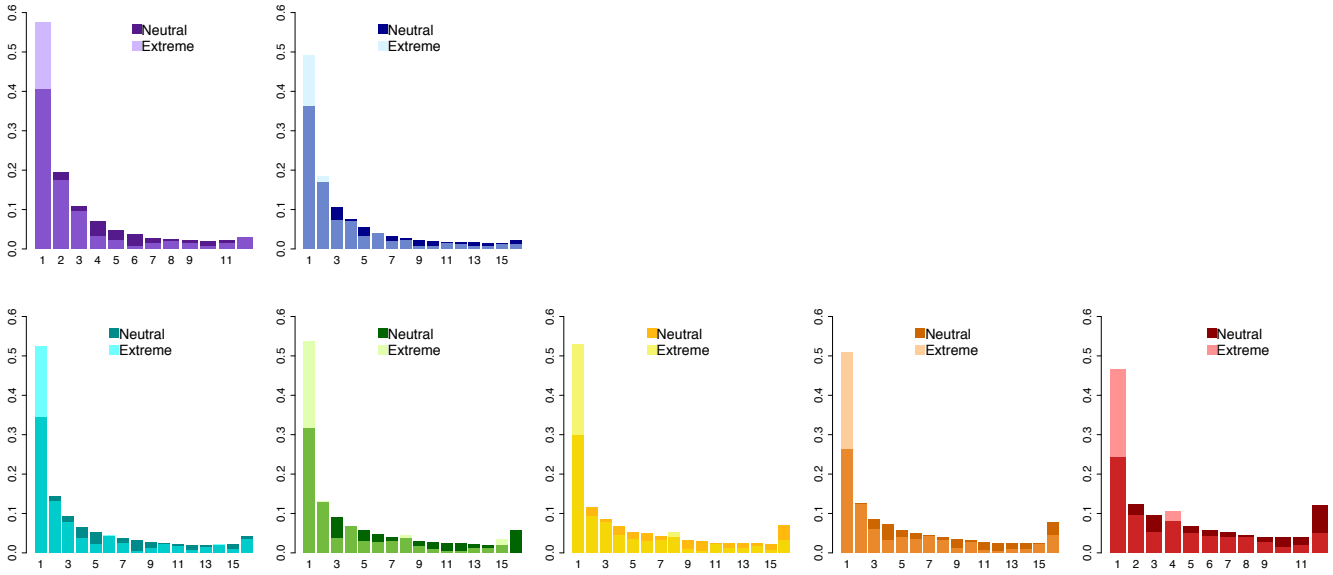

Figure S13: *Site Frequency Spectra of Neutral and Extreme Effect Variants*. We compared the proportion of neutral variants by their frequency class to extreme effect variants by frequency class. The proportion of variants is shown along the Y-axis and each frequency bin is shown along the X-axis. Extreme effect variants are colored translucent. Neutral variants are shaded grey. Overlap between the two categories is opaque.

Figure S14: Relative reduction in heterozygosity (RH)

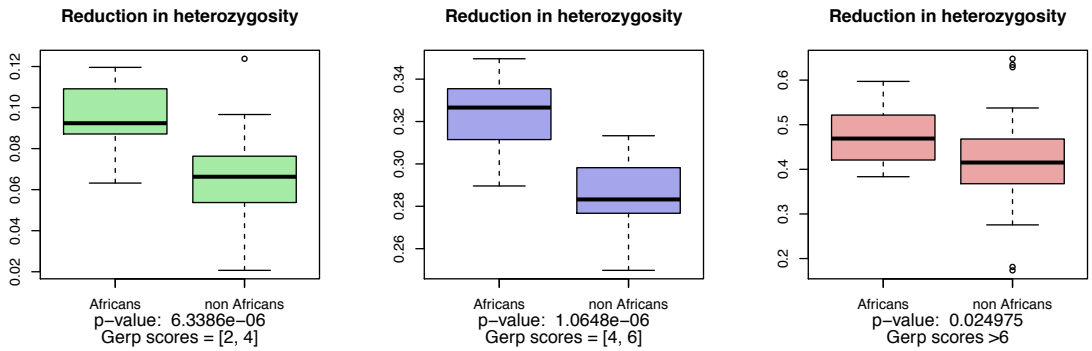

Figure S14: ***Relative reduction in heterozygosity (RH) at selected sites as compared to neutral sites***. Comparison of the distribution of RH between African and non-African individuals for different GERP categories, tested with a two-tailed Student t-test.

Figure S15: Luhya het/hom$_{der}$ ratio by effect category

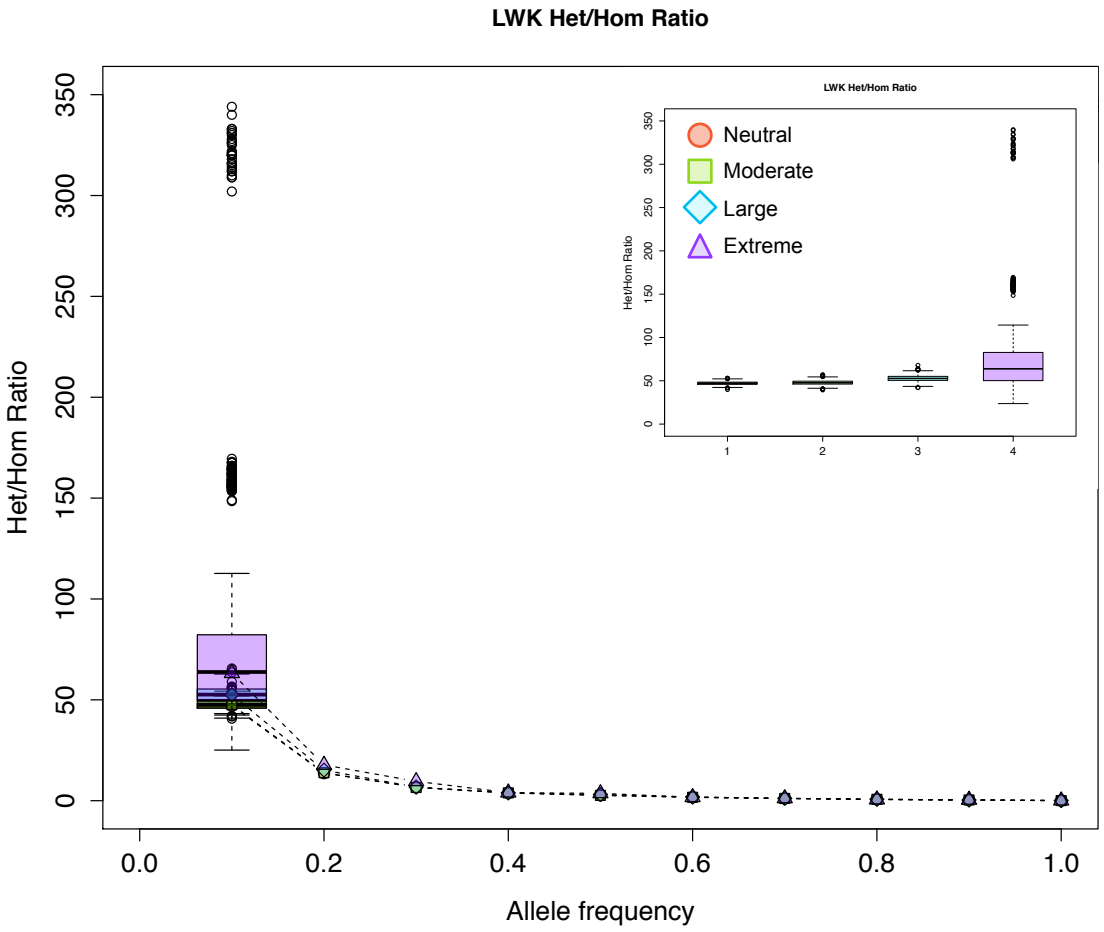

Figure S15: ***Luhya (LWK) Het/Hom Ratio by effect catgory****:* Under a recessive model, it is expected that EXTREME effect variants will have an excess of heterozygotes, compared to homozygotes, because of the effect of purifying selection on homozygotes. However, this pattern could also be biased by an excess of low frequency variants with extreme effect, compared to other categories. In order to distinguish between the two processes, we removed singletons for the dataset and calculated the ratio of heterozygotes / homozygotes in the 1000G LWK for all variants within each effect, and plotted the results according to the variants frequency. Results show an excess of heterozygotes in variants of extreme effect for low frequency bins (≤ 30%), being particularly evident for variants between 10% derived allele frequency. The inset shows boxplots for the 10% allele frequency bin along the x-axis.

Figure S16: Testing a recessive model

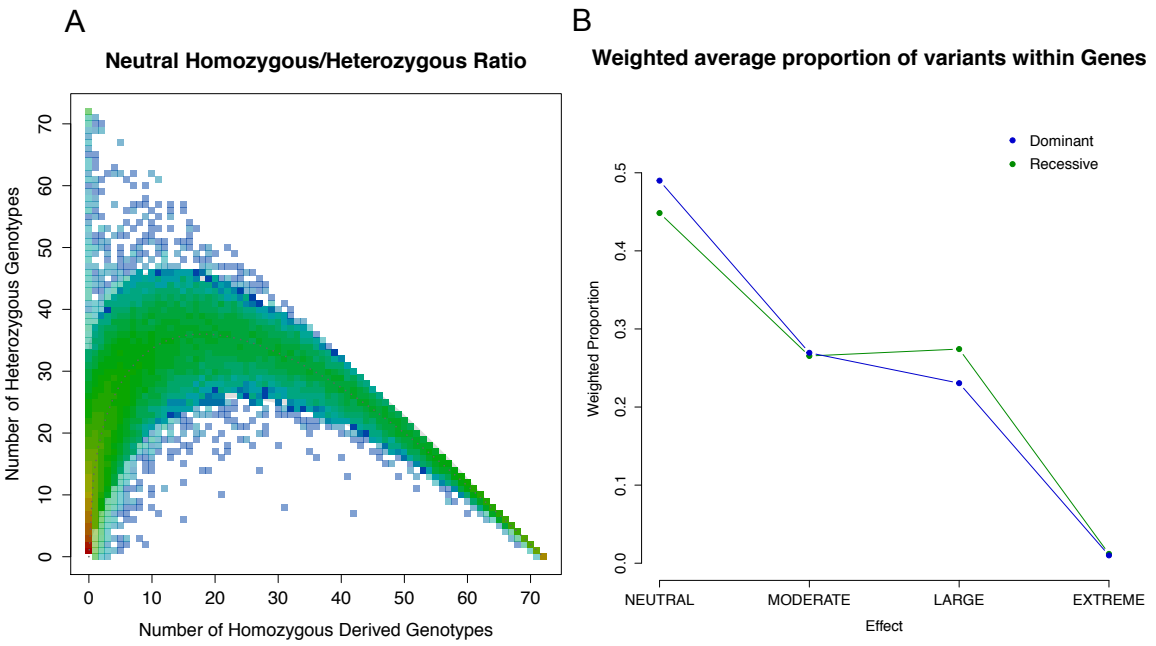

Figure S16: ***Testing a recessive model.*** **A** A non-additive model of dominance could potentially lead to deviations from HW Equilibrium, if the derived variant selection coefficient is high enough. We compared the observed genotype frequencies with the expected ones, considering the known allele frequencies. Variants are plotted according to the observed number of homozygotes (x-axis) and heterozygotes (y-axis) in the LWK population. Heat colors reflect a higher number of variants. The grey dashed line reflects the HW expectation. Colors are shaded when variants significantly deviate from HW expectation (p-value < 0.01). Specifically, variants on the upper left corner represent an excess of heterozygotes compared to what would be expected, compatible with a recessive model. **B** *Weighted average proportion of variants grouped by effect in recessive (green) vs. dominant (blue) genes.* LARGE effect variants are found on average at lower proportions in dominant genes, compared to recessive genes, consistent with purifying selection being more efficient in dominant genes, where LARGE effect variants are always expressed.

Figure S17: Mutational Load in 1000 Genomes Exome Data

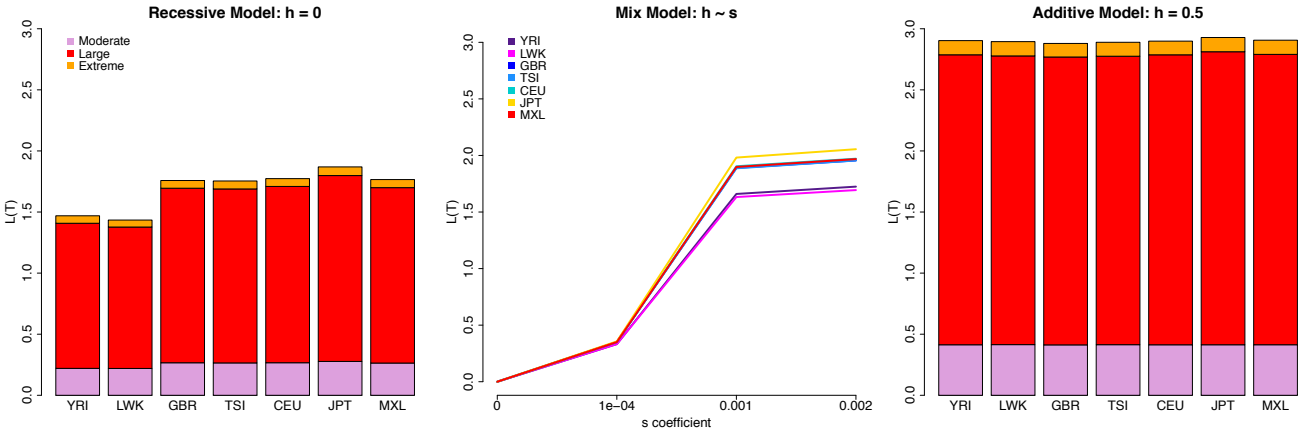

Figure S17: ***Differences in Load - 1000 Genomes Dataset***. For each population, load is calculated under a recessive, intermediate and dominant model (as in **Figure 4**), reflecting contributions from variants with moderate, large and extreme effect.

Figure S18: Distribution of highly differentiated variants vs. the genome

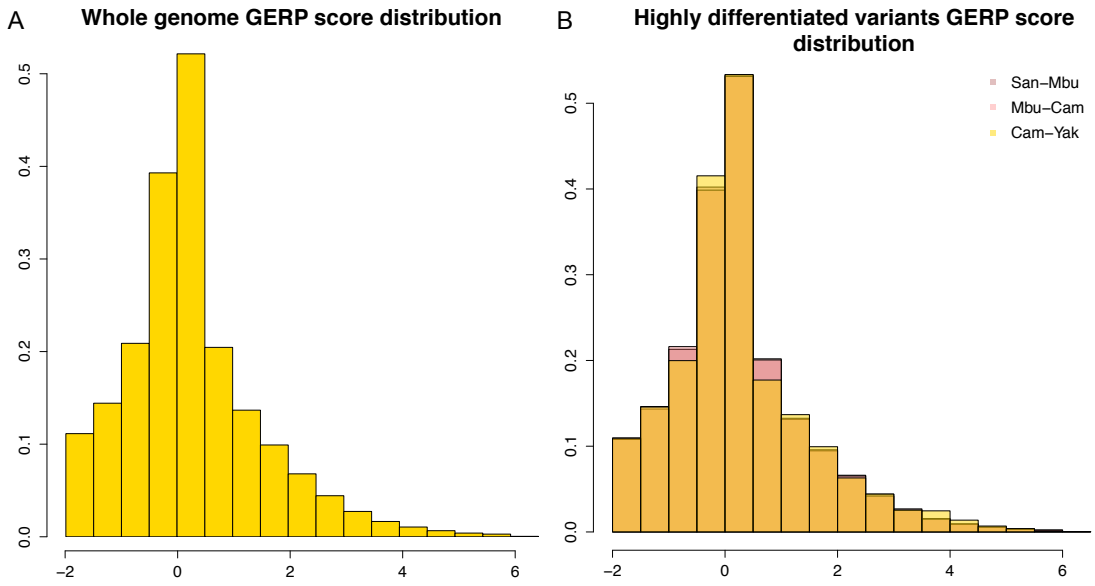

Figure S18: ***Distribution of functional alleles in highly differentiated variants vs. the whole Genome.*** **A** Distribution of GERP scores across the Genome **B** Distribution of GERP scores in highly differentiated variants for different demographically relevant population comparisons: (Afr-Afr), (Afr-OoA), (OoA,OoA). Results show now apparent differences in the distribution of functional variants in those two datasets.

Figure S19: Schematic of the range expansion model

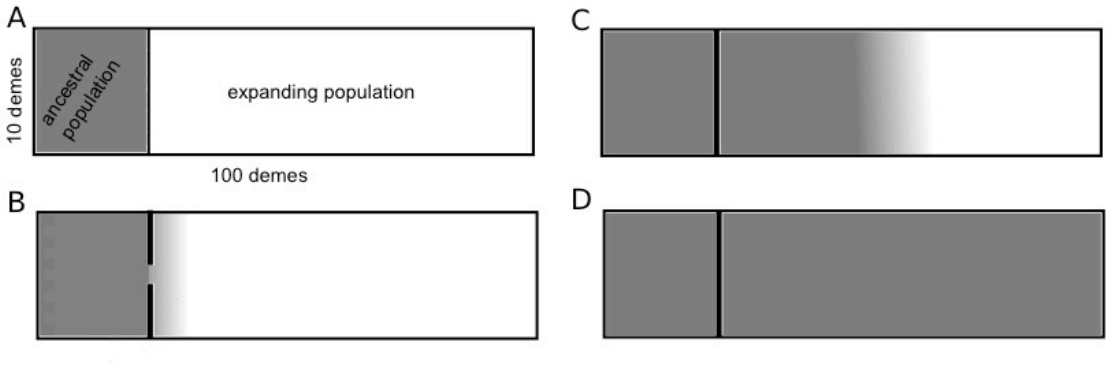

Figure S19: ***Schematic of the range expansion model.*** The model includes a spatial bottleneck we used to simulate the evolution of heterozygosity during a linear 2D expansion. Panel **A** shows the ancestral population (gray) separated from the empty habitat by a migration barrier (black line). After a burn-in phase of 20,000 generations, a single deme in the middle of the migration barrier is removed for 5 generations, during which individuals from the ancestral population can migrate into the empty habitat. Panel **B** shows the onset of the expansion and panel **C** the colonization of the empty habitat by the expanding population (gray). Panel **D** shows the whole metapopulation after the colonization is complete. Migration is bidirectional among demes in the simulation. For a similar simulation model, see (Peischl et al., 2013).

Figure S20: Sharing of GERP ≥ 6 Variants Across Populations

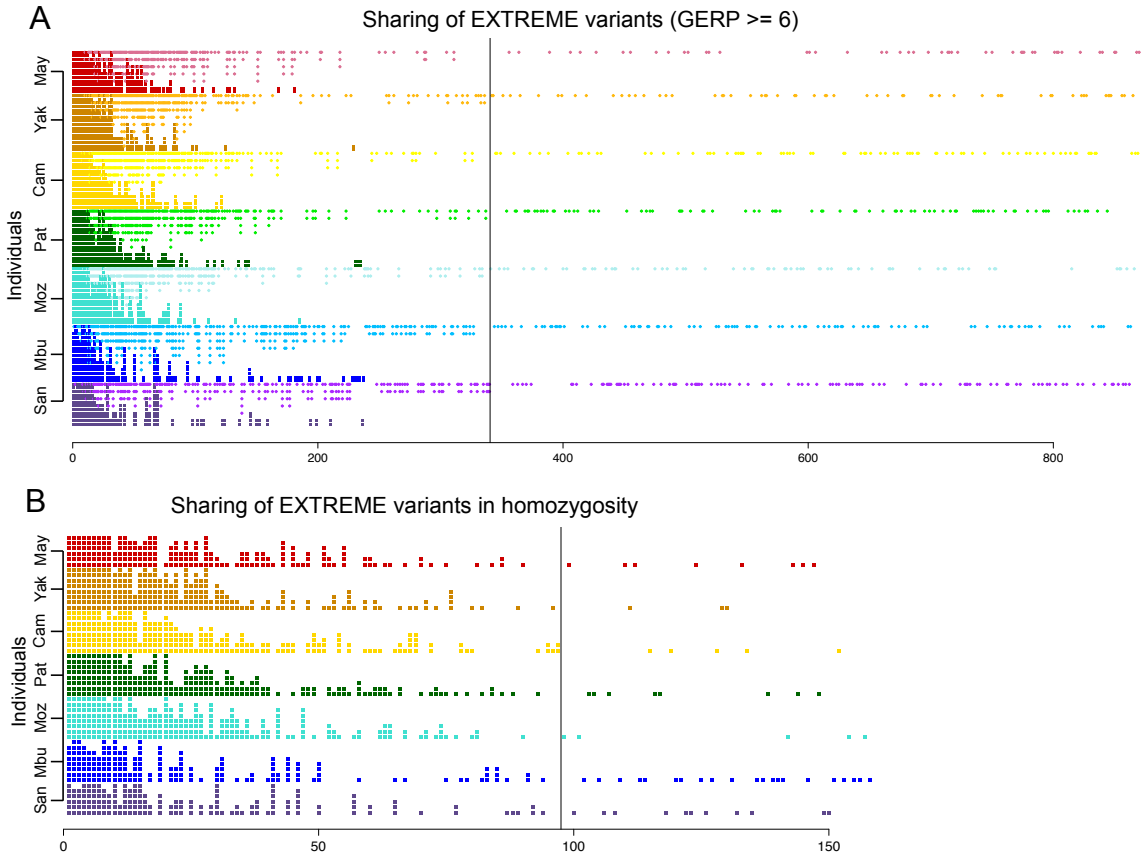

Figure S20*: Sharing of GERP >6 Variants Across Populations*. A For the 800 extreme variants we sorted alleles into homozygote and heterozygote states. Variants are sorted along the X-axis according to their global frequency in the dataset, with common variants on the left and rare variants on the right. In each population, the counts of heterozygotes are ordered in decreasing frequency from top to bottom. Homozygotes are ordered in the opposite fashion, with frequent counts on the bottom row and increasing toward the top within each population. The majority of variants are singletons, indicated to the right of grey line. Out of Africa populations carry more EXTREME variants at higher frequencies and share more EXTREME variants with each other than they share with African populations. Only a small number of GERP >6 variants are fixed in African populations. **B** A version of the homozygous GERP>6 variants is shown in the bottom panel.

Figure S21: Site Frequency Spectrum under different selection regimes and locations of the range expansion.

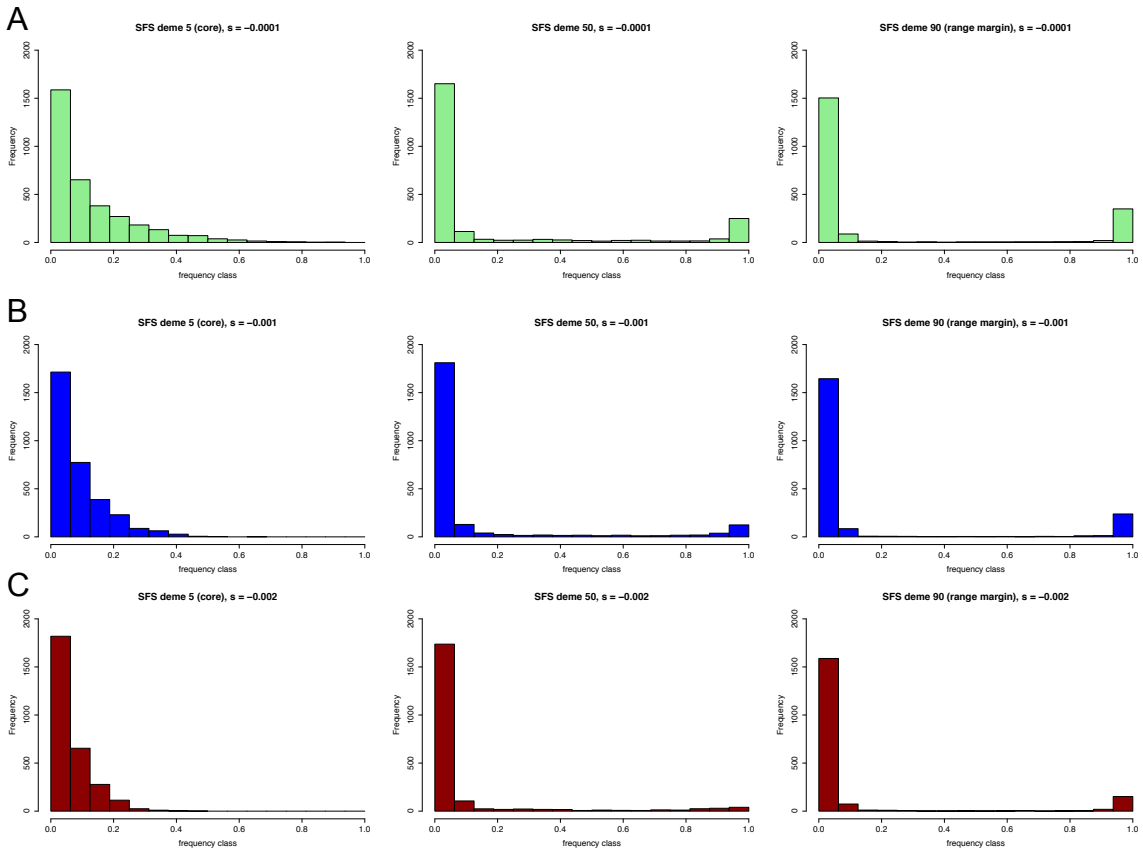

Figure S21: ***Site frequency spectrum under different selection regimes and locations of the range expansion***. The site frequency spectrum was plotted for simulated demes from different locations under a range expansion model. Each row represents a different simulated selection coefficient, corresponding to **A)** moderate **B)** large **C)** extreme estimated effect. As the negative selection coefficient increases, the proportion of low frequency variants increases, and as geographic distance between the deme and the ancestral population increases, a greater amount of variants reaches fixation, even for highly deleterious variants.

Figure S22: Testing significance in observed differences in Load under the assumed models of dominance.

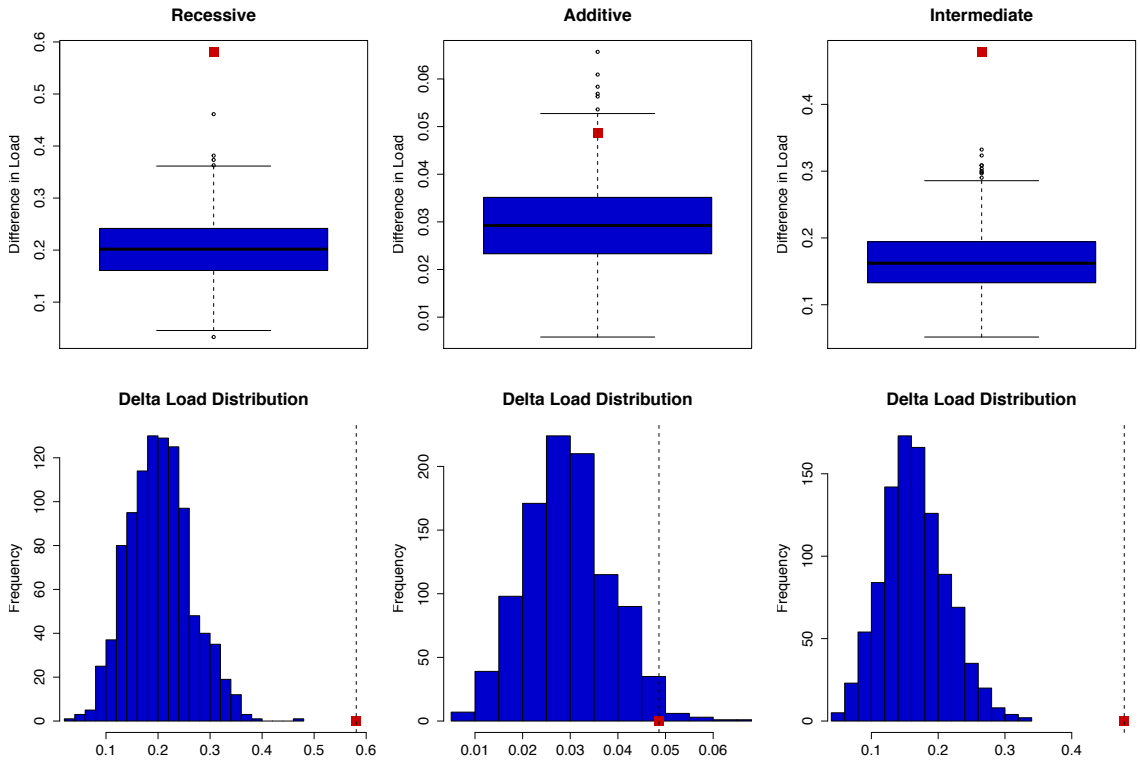

Figure S22: ***Testing significance in observed differences in Load under the assumed models of dominance***. A) Under each model, 1,000 iterations were performed where individuals were randomly re-assigned to populations and the maximum difference in mutation load was calculated. The observe difference in load is represented by a red square and the simulated differences are represented via boxplots. Under all three models the observed difference in Load is statistically significant with a p-value < 0.05 (See *SI Methods*). B) Distribution of the simulated differences in mutation load (blue) and the observed difference in load (red square).

Supplementary References:

1000 Genomes Project Consortium, Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T., and McVean, G.A. (2012). An integrated map of genetic variation from 1,092 human genomes. Nature *491*, 56–65.

Agrawal, A.F., and Whitlock, M.C. (2011). Inferences About the Distribution of Dominance Drawn From Yeast Gene Knockout Data. Genetics *187*, 553–566.

Andres, A.M., Hubisz, M.J., Indap, A., Torgerson, D.G., Degenhardt, J.D., Boyko, A.R., Gutenkunst, R.N., White, T.J., Green, E.D., Bustamante, C.D., et al. (2009). Targets of Balancing Selection in the Human Genome. Mol Bio Evol *26*, 2755–2764.

Browning, B.L., and Yu, Z. (2009). Simultaneous Genotype Calling and Haplotype Phasing Improves Genotype Accuracy and Reduces False-Positive Associations for Genome-wide Association Studies. The American Journal of Human Genetics *85*, 847–861.

Chen, H., Green, R.E., Paabo, S., and Slatkin, M. (2007). The Joint Allele-Frequency Spectrum in Closely Related Species. Genetics *177*, 387–398.

Cooper, G.M., Stone, E.A., Asimenos, G., NISC Comparative Sequencing Program, Green, E.D., Batzoglou, S., and Sidow, A. (2005). Distribution and intensity of constraint in mammalian genomic sequence. Genome Research *15*, 901–913.

DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nature Genetics *43*, 491–498.

Fu, W., O'Connor, T.D., Jun, G., Kang, H.M., Abecasis, G., Leal, S.M., Gabriel, S., Rieder, M.J., Altshuler, D., Shendure, J., et al. (2013). Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. Nature *493*, 216–220.

Gazave, E., Chang, D., Clark, A.G., and Keinan, A. (2013). Population growth inflates the per-individual number of deleterious mutations and reduces their mean effect. Genetics.

Hernandez, R.D., Kelley, J.L., Elyashiv, E., Melton, S.C., Auton, A., McVean, G., 1000 Genomes Project, Sella, G., and Przeworski, M. (2011). Classic Selective Sweeps Were Rare in Recent Human Evolution. Science *331*, 920–924.

Huang, N., Lee, I., Marcotte, E.M., and Hurles, M.E. (2010). Characterising and Predicting Haploinsufficiency in the Human Genome. PLoS Genet *6*, e1001154.

Kimura, M., Maruyama, T., and Crow, J.F. (1963). The Mutation Load In Small Populations. Genetics *48*, 1303–1312.

Li, H., and Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. Nature *475*, 493–496.

Li, J.Z., Absher, D.M., Tang, H., Southwick, A.M., Casto, A.M., Ramachandran, S., Cann, H.M., Barsh, G.S., Feldman, M., Cavalli-Sforza, L.L., et al. (2008). Worldwide Human Relationships Inferred from Genome-Wide Patterns of Variation. Science *319*, 1100–1104.

Maples, B.K., Gravel, S., Kenny, E.E., and Bustamante, C.D. (2013). RFMix: A Discriminative Modeling Approach for Rapid and Robust Local-Ancestry Inference. The American Journal of Human Genetics 1–11.

Nelson, M.R., Wegmann, D., Ehm, M.G., Kessner, D., St Jean, P., Verzilli, C., Shen, J., Tang, Z., Bacanu, S.A., Fraser, D., et al. (2012). An Abundance of Rare Functional Variants in 202 Drug Target Genes Sequenced in 14,002 People. Science *337*, 100–104.

Peischl, S., Dupanloup, I., Kirkpatrick, M., and Excoffier, L. (2013). On the accumulation of deleterious mutations during range expansions. Molecular Ecology *22*, 5972–5982.

Prüfer, K., Racimo, F., Patterson, N., Jay, F., Sankararaman, S., Sawyer, S., Heinze, A., Renaud, G., Sudmant, P.H., de Filippo, C., et al. (2013). The complete genome sequence of a Neanderthal from the Altai Mountains. Nature *505*, 43–49.

Simons, Y.B., Turchin, M.C., Pritchard, J.K., and Sella, G. (2014). The deleterious mutation load is insensitive to recent population history. Nature Genetics *46*, 220–224.

Sunyaev, S., Ramensky, V., Koch, I., Lathe, W., Kondrashov, A.S., and Bork, P. (2001). Prediction of deleterious human alleles. Human Molecular Genetics *10*, 591–597.

Tennessen, J.A., Bigham, A.W., O'Connor, T.D., Fu, W., Kenny, E.E., Gravel, S., McGee, S., Do, R., Liu, X., Jun, G., et al. (2012). Evolution and Functional Impact of Rare Coding Variation from Deep Sequencing of Human Exomes. Science *337*, 64–69.