No evidence that natural selection has been less effective at removing deleterious mutations in Europeans than in West Africans

Ron Do[1,2], Daniel Balick[1,3], Heng Li[1,2], Ivan Adzhubei[3], Shamil Sunyaev[1,3] and David Reich[1,2,4]

[1]Broad Institute of Harvard and MIT, Cambridge, MA, USA, 02142; [2]Department of Genetics, Harvard Medical School, Boston, MA 02115; [3]Division of Genetics, Brigham & Women's Hospital, Harvard Medical School, Boston, MA, USA 02115; [4] Howard Hughes Medical Institute, Harvard Medical School, Boston, MA, USA 02115

To whom correspondence should be addressed:
ssunyaev@rics.bwh.harvard.edu or reich@genetics.med.harvard.edu

Non-African populations have experienced major bottlenecks in the time since their split from West Africans, which has led to the hypothesis that natural selection to remove weakly deleterious mutations may have been less effective in non-Africans. To directly test this hypothesis, we measure the per-genome accumulation of deleterious mutations across diverse humans. We fail to detect any significant differences, but find that archaic Denisovans accumulated non-synonymous mutations at a higher rate than modern humans, consistent with the longer separation time of modern and archaic humans. We also revisit the empirical patterns that have been interpreted as evidence for less effective removal of deleterious mutations in non-Africans than in West Africans, and show they are not driven by differences in selection after population separation, but by neutral evolution.

The effectiveness with which natural selection removes deleterious mutations from a population depends not only on the selection coefficient (s) of a mutation, but on the product of this and the population size (N) (Ns, reviewed in 1). Thus, the rate at which deleterious alleles are removed from a population depends on demographic history. Profound demographic differences across humans are well documented. Founder events in the last hundred thousand years have reduced the nucleotide diversity (the number of differences per base pair between an individual's two chromosomes) in non-Africans by at least 20% relative to West Africans[2-4], reflecting times

when the ancestors of non-Africans had relatively smaller population sizes. Similarly, the advent of agriculture in the last ten thousand years has led to rapid population expansions.

To investigate the empirical evidence for natural selection having differed in its effectiveness across human populations, past studies have contrasted mutation classes thought to be subject to little selection (synonymous mutations in genes) to those potentially subject to purifying selection (non-synonymous mutations)[5-7]. The most important study on this topic measured the proportion of polymorphic positions in genes that are non-synonymous in 20 European and 15 African American samples, and showed that while all such segregating sites have a reduced rate in non-Africans, the reduction is proportionally less for non-synonymous sites.[5] Based on simulations, this observation was interpreted as being consistent with reduced effectiveness of selection against weakly deleterious alleles in European than in West African populations due to their smaller size since separation[5]. Subsequent studies have confirmed the empirical observation and favored a similar interpretation[6,8,9]. These studies have clearly documented that there has been an interaction between the forces of natural selection and demographic history with regard to their effects on densities of non-synonymous sites[5,6,8,9]. However, such observations do not necessarily imply that there have been differences in the effectiveness of selection between the two populations after the split. An alternative explanation is that in the common ancestral population of Europeans and West Africans, the average derived frequency for non-synonymous alleles was lower than for synonymous alleles, as negative selection places downward pressure on the frequency of derived alleles[5]. The different initial distributions for the two classes of alleles would have responded differently to the bottleneck that then occurred in European populations, simply because they started out with different shapes, an effect that would be expected to occur even in the absence of any differences in the effectiveness of selection between the two populations since they split.

The most direct way to contrast the effectiveness of selection between two populations is to sample a single haploid genome from each population, count all the differences, and measure which of the two lineages carries an excess. Any two lineages that are compared in this manner must, by definition, be separated by the same amount of time since their common ancestor, and are thus expected to harbor the same number of lineage-specific mutations in the absence of

selection or differences in mutation rate. In the presence of selection, however, mutations are removed from each lineage at a rate that depends on the product Ns, such that differences in the effectiveness of selection can be inferred from a detected asymmetry in the number of mutations in the two lineages. Here, we test for differences in the accumulation of mutations between two lineages by sampling one allele from population X and one from Y, determining the ancestral state based on the chimpanzee genome (*PanTro2*), and recording all the differences. We count the number of derived mutations in population X but not Y ($L_{X\,not\,Y}$) and in Y but not X ($L_{Y\,not\,X}$), and define a statistic $R_{X/Y} = L_{X\,not\,Y} / L_{Y\,not\,X}$. We average over all possible pairs of samples, and compute a standard error using a Weighted Block Jackknife to correct for correlation among neighboring sites (Methods)[10]. If selection has been equally effective since the split and mutation rates have been the same on the two lineages, $R_{X/Y}$ should be within a few standard errors from 1.

We measured $R_{WestAfrica/Europe}$ in four sequencing datasets: (1) coding regions of genes (exomes) from 15 African Americans and 20 European Americans[5]; (2) exomes from 1,089 individuals in the 1000 Genomes Project (1KG)[11]; (3) exomes from 1,088 African Americans and 1,351 European Americans[6], and (4) 24 whole genomes sequenced to high coverage[12,13] (Table S1). As expected for sites unaffected by selection and for mutation rates being indistinguishable on the two lineages, $R_{WestAfrica/Europe}$(*synonymous)* is always within 2 standard errors of 1 (Table 1 and Table S1). However, $R_{WestAfrica/Europe}$(*non-synonymous)* is also indistinguishable from 1 (Table 1, Table 2, and Table S2). Thus, our empirical data provide no evidence that purging of weakly deleterious mutations has been less effective in Europeans than in West-Africans. This null result is consistent with the recently reported finding of ref. [14] for similar population comparisons, which used a somewhat different approach. To generalize these results, we extended the analysis to a diverse set of human populations by computing $R_{X/Y}$ between all possible pairs of 5 diverse sub-Saharan African and 6 non-African populations[13], and fourteen 1000 Genomes Project populations. We observe no significant differences for any pair despite profound differences in demographic history (Table 2 and Table S3).

To interpret these null findings, we carried out simulations using fitted models of the demographic histories of West African and European populations[5,6,15]. (For most of our simulations, we assumed that mutations act additively although in Figure S1 we show that

qualitatively different patterns are expected for mutations that act recessively, a phenomenon that we explore in a separate study[16].) The simulations predict that $R_{WestAfrica/Europe}$ will be below 0.95 when s is between -0.004 and -0.0004 (Figure 1). However, if many mutations have selection coefficients outside this range, the signal could be diluted to the point of not being detectable. Indeed, when we compute the expected value of $R_{WestAfrica/Europe}$ integrating over a previously fitted distribution of selection coefficients[17], we find that $R_{WestAfrica/Europe}(non-synonymous)$ is expected to be 0.987, too close to 1 to be easily detectable given the standard errors of our empirical measurements (Table 1)[18]. We next attempted to boost our power by stratifying non-synonymous sites based on their predicted functional effects. The PolyPhen-2[19] and SIFT[20] algorithms both predict function in a way that is dependent on the ancestral/derived status of allelic variants compared with the human reference genome, which has a particular ancestry at every segment that can bias measurements. We thus implemented a version of PolyPhen-2 that is independent of the allelic status of the human reference (Methods). We continue to observe no significant differences[18] (Table 1, Table 2, Table S2, Table S3), although this null result may also reflect low power as when we infer the distribution of newly arising mutations for different PolyPhen-2 classes (Note S1), $R_{WestAfrica/Europe}$ is predicted to be 0.984-0.993 (Table S4).

To increase statistical power to detect any real differences in the net effectiveness of selection, we leveraged the fact that the population split between African and non-African populations occurred only in the last roughly one hundred thousand years. Mutations that arose prior to the population divergence are expected to dilute any true signal. We therefore computed a time-stratified $R_{African/non-African}$-statistic, taking advantage of data from 4 experimentally phased African and 6 experimentally phased non-African genomes, all processed similarly[13]. We created pseudo-diploid genomes by merging 1 African and 1 non-African phased genome, for a total of $96 = (4 \times 2) \times (6 \times 2)$ comparisons, and then used the Pairwise Sequential Markovian Coalescent method (PSMC) to infer the local time since the most recent common ancestor (TMRCA) at each location, masking out the exome to avoid circularly using the same sites for inferring TMRCA and computing $R_{African/non-African}$. Table S5 shows that when we restrict to the subset of the genome with the lowest inferred TMRCA in this analysis, we still detect no differences.

4

As a third way to increase statistical power, we analyzed a class of sites that is much larger than the class of coding substitutions so that we can make measurements with much smaller standard errors; this is the class of sites affected by biased gene conversion (BGC). BGC is a selection-like process in which DNA repair acting on heterozygous sites in gene conversion tracts favors GC over AT alleles[21]. Since BGC only acts on heterozygous sites, it occurs at a rate proportional to heterozygosity, or $2p(1$-$p)$ for a mutation of frequency p, exactly mimicking additive selection[21]. We find that $R'^{GC \to AT}_{West\ Africans/Europeans} = 0.995 \pm 0.002$ and $R'^{AT \to GC}_{West\ Africans/Europeans} = 1.000 \pm 0.002$ by using an $R'^{class1 \to class2}_{X/Y} = R^{class1 \to class2}_{X/Y}/(R^{A \leftrightarrow T}_{X/Y} + R^{C \leftrightarrow G}_{X/Y})$ statistic that normalizes by the rate of A↔T and C↔G substitutions not expected to be affected by BGC. A further advantage of this normalization is that it also corrects for possible differences in mutation rates across lineages. For diverse population comparisons, we detect no differences significant at |Z|>3 standard errors from zero with the exception of San Bushmen who have about 1% more GC→AT mutations than other humans (significant at up to 8 standard errors) (Table S6). This is the first direct detection of less effective removal of mutations in some present-day humans than in others, and is consistent with the San being amongst the most deeply diverged human populations, which may have provided more opportunity for slight differences in the effectiveness of removal of mutations across populations to have a cumulatively measurable effect[22].

To demonstrate empirically that differences in the accumulation of non-synonymous sites can be empirically detected given a sufficiently ancient population divergence time and sufficiently different subsequent demographic histories, we also analyzed two deeply sequenced genomes from archaic humans: Denisovan and Neanderthal. The ancestors of both are inferred to have maintained relatively small effective population sizes for on the order of a half million years since their main separation from present-day humans, consistent their levels of genetic diversity being 3-6 times smaller[12]. A challenge in comparing the accumulation of mutations in present-day human samples to ancient samples is that fewer mutations are expected to have occurred on the lineage of ancient samples because they are from individuals who lived closer in time to the common ancestor. To correct for this, we divide the accumulation of non-synonymous mutations on each lineage by the synonymous sites: $R'_{X/Y}(non$-$synonymous\ class) = R_{X/Y}\ (non$-$synonymous\ class)/R_{X/Y}(synonymous)$. After removing C→T and G→A mutations in the archaic genomes that

have evidence of degradation leading to ancient DNA errors, we find that present-day humans have accumulated deleterious mutations at a significantly lower rate than Denisovans since separation: $R'_{Modern/Denisova}$(non-synonymous) = 0.872 ± 0.034 (P=0.0002) (Table S7)[12]. In contrast, $R'_{Modern/Neanderthal}$(non-synonymous) = 1.037 ± 0.037 is consistent with 1, suggesting that deleterious mutations have not been removed as effectively on the Neanderthal as on the Denisovan lineage (Table S7). The higher resolution BGC analysis further detects that the effectiveness of removal of mutations on the Neanderthal lineage was intermediate between that of Denisovans and modern humans: $R'^{GC\rightarrow AT}_{Yoruba/Denisova} = 0.925 \pm 0.004$, $R'^{GC\rightarrow AT}_{Yoruba/Neandertal} = 0.961 \pm 0.003$, and $R'^{GC\rightarrow AT}_{Denisova/Neandertal} = 1.041 \pm 0.005$. The different rates of accumulation in Neanderthals and Denisovans despite similar inferred demographic histories (Figure 1) suggests that fitted models of demographic history, e.g. for West Africans and Europeans, may not always provide accurate predictions of the relative effectiveness of removal of mutations.

In light of these results, is there any reason to believe that weakly deleterious mutations have been removed less effectively in Europeans than in West Africans? The strongest previous evidence for such an effect was based on an alternative statistic: the proportion of polymorphic sites in the exome that are non-synonymous[5]. We therefore carried out simulations of demographic history that allowed us to study the change in this statistic over time; our simulations agree with previous simulations of the same statistic[5], but allow for additional insights into the evolutionary forces responsible for the dynamics (Figure 2). In the first set of simulations, we adjusted the selection coefficient *s* in Europeans in each generation so that the quantity governing the effectiveness of selection *Ns* was by construction always the same in Europeans and West Africans. The simulations show that the proportion of non-synonymous sites in Europeans in fact rises more, not less, when we adjust our simulations in this way to eliminate any differences in the effectiveness of selection (blue curves in Figure 2B). Thus, the observed rise in the proportion of non-synonymous sites in Europeans is not due to differences in the effectiveness of selection after the population split, but occurs in spite of such differences. In the second set of simulations, we partitioned the change in the proportion of non-synonymous sites over time into selection and neutral effects. The simulations show that the dynamics of the previously studied statistic are driven by neutral forces (correlation coefficient of ρ = 0.96 to 0.99). In contrast, changes in the effectiveness with which selection removes deleterious alleles

6

in Europeans compared with West Africans has an effect that is opposite to the observed change (ρ = -0.45 to -0.27) (Figure 2C and Figure 2D). Intuitively, prior to the West African / European split, allele frequencies of non-synonymous polymorphisms would, on average, have been much lower due to the depletion of non-synonymous sites by selection, and the per-site density of non-synonymous segregating sites would also have been lower. The population entering the bottleneck primarily loses rare alleles, so the non-synonymous site allele frequency distribution would be expected to adjust faster than that for synonymous sites. Once the population re-expands, the allele frequency distribution for non-synonymous sites also adjusts faster, in this case because the same flux of new mutations into both classes causes a faster rate of replenishment of non-synonymous sites than synonymous sites due to an initially lower density. These findings illustrate the complexity of the interactions between selection and demographic history in their effects on genetic variation, and highlight an opportunity first identified by ref.[5], which is that joint analyses of demographic history and natural selection can provide more insight into the nature of both evolutionary forces that either type of analysis alone.

It is tempting to interpret the indistinguishable accumulations of deleterious mutations across present-day humans as implying that the overall genetic burden of disease should be similar for diverse humans. To the extent that mutations act additively this is correct, and it implies that the complex demographic events of the past are not expected to lead to substantial population differences in prevalence rates of complex disease that have an additive genetic architecture[18]. However, recessively or epistatically acting mutations work in combination to contribute to disease risk, and since demographic history affects allele frequencies, it affects the rate of co-occurrence of alleles. For example, Table 1 and Table S8 show that the absolute count of alleles occurring in homozygous form is empirically higher in non-Africans than in Africans for all functional site classes, confirming previous findings[5]. Thus, the relative risk for diseases that are contributed to by recessively acting mutations could be expected to be influenced by demography, as indeed is known to be the case for populations that experienced recent founder events like Ashkenazi Jews and Finns. An important direction for future work is to determine the extent to which mutations contributing to phenotypes act non-additively, as this will determine the extent to which demographic history is important in affecting human disease risk.

Methods

Data

The datasets we analyzed were published previously and are summarized here.

"*Celera*": PCR amplification and Sanger sequencing was performed on 15 African American (AA) and 20 European American (EA) samples over the coding sequences of 10,150 genes. We downloaded ancestral and derived allele counts for 39,440 autosomal SNPs from the supplementary materials of the original study, restricting to sites with genotypes available from both AA and EA[5].

"*1000 Genomes (1KG)*": A total of 1,089 samples from 14 populations were analyzed in Phase 1 of the 1000 Genomes project. Illumina-based exome sequencing[11] was performed to ~100× average coverage after solution hybrid capture of the exome[23].

"*ESP*": A total of 1,088 African Americans and 1,351 European Americans were sequenced as part of the National Heart, Lung and Blood Institute Exome Sequencing Project. Illumina-based exome sequencing was performed to ~100× average coverage after solution hybrid capture of the exome [6].

"*24 Genomes*": This dataset includes 2 samples each from 6 non-African and 5 sub-Saharan African genomes, an archaic human from Denisova Cave in Siberia sequenced to 30× coverage, and an archaic Neanderthal from Denisova Cave in Siberia sequenced to 52× coverage. All sequencing data is based on Illumina technology. We used the version of this dataset reported in ref. [13]. We only analyzed sites with genotype quality scores (GQ) of ≥45[24].

Mutation annotation

We annotated coding mutations using ANNOVAR[25], which classifies sites as "non-synonymous", "synonymous", "stop-gain" or "stop-loss". We sub-classified variants using a simplified version of PolyPhen-2 that is independent of the ancestral/derived status of the human genome reference sequence ("human-free Polyphen-2"). To guarantee independence of PolyPhen-2 predictions from the human genome reference sequence, we created a simplified version that relies solely on the multi-species conservation score used in this method[26]. This score reflects the likelihood of observing a given amino acid at a site conditional on the observed pattern of amino acid changes in the phylogeny, and is the most informative feature of PolyPhen-

8

2. The predictions in our simplified PolyPhen-2 method are based on the absolute values of the difference of the scores for the two alleles. By construction, this is symmetric with respect to reference/non-reference (and also ancestral/derived, major/minor) allele status. This procedure is similar to the original version of PolyPhen, but relies on the PolyPhen-2 homology search and alignment pipeline.

Statistics

We are interested in the expected number of mutations in a randomly sampled haploid exome from one population that are not seen in a randomly sampled comparison exome from another population. To compute this in a situation where we have many exomes available from each population, we do not wish to literally randomly choose an exome from each population as this would throw away data decreasing the precision of our estimates. Instead, we obtain the expected value if we performed an infinite number of random samplings. To compute this, at each variable site i in the genome we define d_A^i as the count of the mutant allele at that site in a sample of n_A^i exomes from population A and d_B^i as the count of the mutant allele in a sample of n_B^i exomes from population B. The expectations are obtained by summing over all sites in the genome.

$$L_{A\ not\ B} = \sum_i \left(d_A^i / n_A^i \right) \left(1 - d_B^i / n_B^i \right)$$

For some analyses, we also wished to compute the relative probability that a population is homozygous for a derived allele whereas the other is not. Thus, we defined two additional statistics, now imposing a correction for limited sample size (since we need to sample two alleles from each population, we need to sample without replacement):

$$L_{A\ not\ B}^2 = \sum_i \frac{d_A^i(1-d_A^i)}{n_A^i(n_A^i-1)} \left(1 - \frac{d_B^i(1-d_B^i)}{n_B^i(n_B^i-1)} \right)$$

We then defined the ratio statistics

$$R_{X/Y} = L_{X\ not\ Y} / L_{Y\ not\ X}$$
$$R_{X/Y}^2 = L_{X\ not\ Y}^2 / L_{Y\ not\ X}^2$$

9

Weighted block jackknife to estimate standard errors

We obtained standard errors using a weighted block jackknife[10]. We divided the SNP datasets into 100 contiguous blocks and then recomputed the statistic on all the data except for the data from that block. The variation can be converted to a standard error using jackknife theory. We assess significance based on the number of standard errors from the null expectation of R=1, and compute a P-value using a Z-score assuming a normal distribution.

Time-stratified computation of relative accumulation of deleterious mutation

We began with data from 10 experimentally phased genomes, all processed in a nearly identical way[13]. These genomes consisted of one each from the populations in Table 2 except for the Dinka. We then combined haploid genomes from one of 4 African and one of 6 non-African individual to make $96 = (2{\times}4){\times}(2{\times}6)$ pseudo-diploid individuals. We masked the data from the exome, and ran the Pairwise Sequential Markovian Analysis (PSMC)[2] on the data to estimate the time since the most recent common ancestor of the two phased genomes at each location in the genome. We stratified the data into three subsets of inferred time depth, and then computed the $R_{African/non\text{-}African}$-statistic within each time-stratified subset (using exomic sites that had been masked from the PSMC analysis so we could independently use them for analysis).

Analysis of sites susceptible to biased gene conversion (BGC)

We computed the accumulation of mutations susceptible to BGC for three different classes: GC→AT mimicking negative additive selection, AT→GC mimicking positive additive selection, and A/T or G/C polymorphisms which we treat as neutral (and use as the denominator of $R'_{X/Y}$). For BGC analyses we use the entire genome, after excluding sites in the exome. For analyses involving ancient samples, we exclude C→T and G→A sites from the analysis of GC→AT substitutions (we restrict to C→A and G→T substitutions), to avoid the degradation errors typical of ancient DNA that we are concerned may affect even the consensus genome sequences.

$R'_{X/Y}$-statistic: Correcting for branch shortening and differences in mutation rates

For analyses involving the archaic Denisovan and Neanderthal samples, which are many tens of thousands of years old and thus have experienced less evolution from the common ancestor than

present-day humans to which they are compared, we do not expect that $L_{archaic\ not\ modern} = L_{modern\ non\ archaic}$ even for neutral sites. For all analyses involving ancient samples, we instead compute normalized statistics $L_{A\ not\ B}'$ and $L_{B\ not\ A}'$, where we divide both $L_{A\ not\ B}$ and $L_{B\ not\ A}$ by the accumulation of mutations at sites that are expected to act neutrally (synonymous sites for coding sequences and A/T + C/G for BGC). Thus, $L'^{class}_{A\ not\ B} = L^{class}_{A\ not\ B} / L^{normalization}_{A\ not\ B}$. We then define:

$$R'^{class}_{X/Y} = (L^{class}_{X\ not\ Y} / L^{class}_{Y\ not\ X}) / (L^{normalization}_{X\ not\ Y} / L^{normalization}_{Y\ not\ X}) = R^{class}_{X/Y} / R^{normalization}_{X/Y}$$

This $R'_{X/Y}$-statistic not only corrects for branch shortening in the ancient samples, but also has the benefit of correcting for any differences in mutation rate that might have arisen on one lineage or another since population separation.

Simulations

We wrote a forward simulation in C that implements an infinite sites model. Each mutation is assumed to occur at an unlinked site. To the extent that linkage affects the expected values of the statistics we compute, our simulations are not capturing these effects.

There is an initial burn-in period of 250,000 generations to generate an equilibrium allele frequency spectrum. The simulator samples the allele counts in the current generation based on the frequencies in the previous generation, the selection coefficient s, the dominance coefficient h (usually set to additive or $h=0.5$), and the current population size.

For modeling West African and European history in the simulations reported in the main text, we use a demographic model previously fitted to genetic data[6]. (Figure S1 reports results for other four different demographic histories, shown in Table S9). For comparisons of West African and archaic population history we also use a previously fitted demographic model[13]. We use a mutation rate of 2×10^{-8}/base pair/generation.

For analyses of the proportion of sites that are expected to be non-synonymous in a sample size of 40 chromosomes, we use a hypergeometric distribution to obtain the expected value of this statistic. If the population size in a generation is N and K_i is the number of copies of the derived allele at site i, then we can compute the probability that a sample of 40 chromosomes is polymorphic at a site as 1 minus the hypergeometric probability of 0 or 40 derived alleles:

11

$$Probability\ that\ site\ i\ is\ segregating = 1 - \frac{\binom{K_i}{0}\binom{N-K_i}{40}}{\binom{N}{40}} - \frac{\binom{K_i}{0}\binom{N-K_i}{0}}{\binom{N}{40}}$$

We average this probability over all simulated positions to obtain the density of segregating sites.

Integrating over distributions of selection coefficients

For some statistics, we wished to obtain an expected value integrating over distributions of selection coefficients. To achieve this, we carried out simulation series for different selection coefficients; for example, for Figure 2, each of 19 values: $s = \{$ -1×10^6, -2×10^6, -5×10^6, -1×10^5, -2×10^5, -5×10^5, -1×10^4, -2×10^4, -5×10^4, -1×10^3, -2×10^3, -5×10^3, -0.01, -0.02, -0.05, -0.1, -0.2, -0.5, -1$\}$. To compute expected values for $L_{X\,not\,Y}$, $L_{Y\,not\,X}$, and the density of segregating sites per base pair in a fixed sample size of 40 chromosomes, we use a weighted average of the values of the simulated single selection coefficient statistics. For some analyses, we use the distribution of human selection coefficients for non-synonymous sites from ref. [17], where the probability of a given value of -s is drawn from a gamma distribution fitted to European genetic data with α=0.206 and β=15400. For analyses of the expected value of $R_{WestAfrica/Europe}$ stratified by PolyPhen-2 functional class, we use the values inferred in Note S1.

Simulations forcing the effectiveness of selection in Europeans and West Africans to be the same

To evaluate whether differences in the effectiveness of selection in the Europeans and West African populations since their split could be explaining the observed rise in the proportion of non-synonymous sites in Europeans above baseline, we modified the simulator so that in every generation i, the selection coefficient in Europeans $s_{e,i}$ is determined dynamically. Define $N_{e,i}$ and $N_{a,i}$ as the diploid population sizes in Europeans and Africans, respectively, and define the selection coefficient in Africans (held constant) as $s_{a,i}$. We then set the selection coefficient in Europeans in each generation to be $s_{e,i} = s_{a,i}(N_{i,a}/N_{i,e})$. This procedure has the consequence that $N_{e,i}s_{e,i} = N_{e,i}s_{a,i}(N_{a,i}/N_{e,i}) = N_{a,i}s_{a,i}$. Since the quantity Ns governs the effectiveness of selection, selection is guaranteed to be equally effective in both populations at all times.

Partitioning the evolutionary dynamics into the effects of selection and neutral effects

We modified the simulation to sample two counts of derived alleles in each generation for a given selection coefficient s and nucleotide position i. The first count is $A_{s,i}$, which reflects the forces of selection, mutation, and stochastic sampling. The second is $N_{s,i}$, which reflects the effects of mutation and stochastic sampling but not selection. The counts in each generation are only updated based on $A_{s,i-1}$ in the previous generation, so the long-term evolutionary trajectory properly incorporates the effects of natural selection as they have accumulated over generations.

We average the values of $N_{s,i}$ and $A_{s,i}$ over simulation replicates, and compute a weighted average of the results based on the distribution of selection coefficients from ref. [17]. We define the proportion of sites that are expected to be non-synonymous in a given generation as $PropAll_i = A_{s,i}/(A_{s,i} + A_{0,i})$, and the proportion that would be non-synonymous if selection had been switched off in that generation as $PropNeu_i = N_{s,i}/(N_{s,i} + A_{0,i})$.

With this notation, the expected change in the proportion of nonsynonymous sites due to all evolutionary forces in generation i is $\delta PropAll_i = PropAll_i - PropAll_i$. The proportion due to neutral forces only is $\delta PropNeu_i = PropNeu_i - PropAll_{i-1}$, and the proportion due to selective forces only is $\delta PropSel_i = PropAll_i - PropNeu_i$.

To compare the effect of an evolutionary force in a given generation to what it was in the ancestral population prior to the split (>2,040 generations ago in the simulations of Figure 2), we defined the baseline-corrected statistics $\Delta PropSel_i = \delta PropSel_i - \delta PropSel_{baseline}$, $\Delta PropNeu_i = \delta PropNeu_i - \delta PropNeu_{baseline}$, and $\Delta PropAll_i = \delta PropAll_i - \delta PropAll_{baseline}$. These statistics are positive if the effectiveness of removal of mutations due to an evolutionary force is less than in the ancestral population, and negative if the effectiveness is greater than in the ancestral population. These are the statistics plotted in Figure 2C and integrated to obtain cumulative effects in Figure 2D.

13

Figure Legends

Figure 1: The effect of demographic history on the accumulation of deleterious mutations..
To study the expected value of $R_{WestAfrica/Europe}$ stratified by selection coefficient, we simulated a previously published model of the joint history of West Africans and Europeans[6], for a range of selection coefficients, assuming an additive model of nature selection (Figure S1 shows similar results for other demographic models). The simulations show that $R_{WestAfrica/Europe}$ dips below a potentially detectable ratio of 0.95 for $s \in (-0.0004, -0.004)$. We also simulated a published model of the history of archaic Denisovans, archaic Neanderthals, and West Africans [13]. The simulations predicts similar curves for $R'_{WestAfrica/Denisova}$ and $R'_{WestAfrica/Neanderthal}$ reflecting their similar inferred demographic histories (we use a normalized R' statistic to correct for the effects of branch shortening in these ancient lineages). The simulations show that $R'_{WestAfrica/Denisova}$ is expected to below a detectable ratio of 0.95 for $s \in (-0.00002, -0.03)$ and that $R'_{WestAfrica/Neanderthal}$ is expected to be below 0.95 for $s \in (-0.00002, -0.09)$.

Figure 2: The rise in the proportion of non-synonymous sites in Europeans compared with West Africans is not due to a reduced effectiveness of selection in Europeans since the split.
(A) The West African and European diploid population sizes for the two simulated models, both of which specify a population split 2,040 generations ago. The plots of the temporal dynamics of the statistics in subsequent panels are restricted to the European population, as the West African population size does not fluctuate enough to result in appreciable changes from the baseline. (B) We show the values of key statistics as a fraction of the baseline. The present-day proportion of non-synonymous sites in Europeans is higher than in the ancestral populations (black curves). This cannot be attributed to less effective removal of deleterious mutations in Europeans than in Africans since the population split, as we can see from the fact that when we carry out simulations in which the selection coefficient in Europeans per generation is set so that the effectiveness of selection Ns in the two populations is the same, the rise actually becomes greater (blue curves). (C) Partitioning of the change in the proportion of non-synonymous sites per generation into neutral and selective forces shows that for the West African / European

comparison, the temporal dynamics are driven by neutral forces (strong positive correlation in the dynamics) and not by the selective forces (negative correlation). (D) Plots of the cumulative effect of each evolutionary force compared to baseline show that differences in the effectiveness of selection between Europeans and West Africans since they separated in fact have a net negative, not a net positive effect on the proportion of non-synonymous sites in Europeans.

Table 1: Accumulation of different classes of mutation in exomes of West African compared to exomes of European ancestry

Data set	West Africans	Europeans	R: Relative rate of lineage specific mutations					R^2: Relative rate of homozygosity for lineage specific mutations				
			$R(synonymous)$	$R(All\ non\text{-}synonymous)$	$R(Benign)$	$R(Possibly\ damaging)$	$R(Probably\ damaging)$	$R^2(synonymous)$	$R^2(All\ non\text{-}synonymous)$	$R^2(Benign)$	$R^2(Possibly\ damaging)$	$R^2(Probably\ damaging)$
24 deep genomes	4	4	1.022 (0.012)	1.008 (0.015)	1.002 (0.018)	1.007 (0.040)	1.031 (0.038)	0.652 (0.014)	0.629 (0.019)	0.628 (0.021)	0.602 (0.043)	0.660 (0.047)
Celera exomes	15	20	0.982 (0.011)	1.010 (0.019)	1.011 (0.022)	1.019 (0.043)	0.992 (0.039)	0.610 (0.011)	0.586 (0.047)	0.583 (0.021)	0.616 (0.053)	0.586 (0.047)
1KG exomes	88	85	1.019 (0.010)	0.994 (0.012)	0.999 (0.015)	0.955 (0.028)	1.011 (0.026)	0.655 (0.012)	0.656 (0.014)	0.639 (0.018)	0.599 (0.032)	0.642 (0.032)
ESP exomes	1,088	1,351	1.004 (0.009)	1.001 (0.011)	0.993 (0.013)	1.001 (0.021)	1.037 (0.029)	0.605 (0.010)	0.608 (0.011)	0.598 (0.015)	0.578 (0.025)	0.630 (0.034)

Notes: ±1 standard errors are obtained from a Block Jackknife with 100 equally sized blocks. For the whole genomes, Yoruba+Mandenka represent West Africans, and French+Sardinian represent Europeans. For the 1000 Genomes Data (1KG), YRI represent West Africans and CEU Europeans. The results for the Celera and ESP analyses represent people of West African ancestry using African Americans.

Table 2: $R_{X/Y}$(probably damaging) for all pairs of deep genome populations (bottom left) and 1000 Genomes populations (top right)

1000 Genomes populations (top-right triangle). Ratios of row population divided by column population. Sample sizes in parentheses after labels.

	IBS (Spanish)	GBR (British)	FIN (Finns)	CEU (European)	JPT (Japanese)	CHS (Chinese)	CHB (Chinese)	PUR (Puerto Rican)	MXL (Mexican)	CLM (Colomb.)	YRI (Nigerian)	LWK (Kenyan)	ASW (Afr. Am.)
TSI (98)	1.025 (0.013)	1.004 (0.007)	1.030 (0.009)	1.012 (0.007)	1.014 (0.025)	1.029 (0.025)	1.024 (0.025)	1.026 (0.009)	1.042 (0.018)	1.022 (0.012)	0.998 (0.026)	0.981 (0.024)	1.008 (0.021)
IBS (14)		0.979 (0.013)	1.005 (0.014)	0.987 (0.013)	0.991 (0.028)	1.006 (0.028)	1.001 (0.028)	1.002 (0.014)	1.018 (0.022)	0.998 (0.016)	0.978 (0.026)	0.962 (0.025)	0.987 (0.021)
GBR (89)			1.026 (0.008)	1.008 (0.007)	1.010 (0.026)	1.026 (0.026)	1.021 (0.026)	1.023 (0.01)	1.039 (0.019)	1.018 (0.013)	0.995 (0.027)	0.978 (0.025)	1.005 (0.022)
FIN (93)				0.983 (0.008)	0.986 (0.024)	1.001 (0.025)	0.996 (0.025)	0.998 (0.011)	1.014 (0.017)	0.993 (0.012)	0.975 (0.026)	0.958 (0.024)	0.984 (0.021)
CEU (85)					1.002 (0.025)	1.018 (0.025)	1.013 (0.025)	1.015 (0.01)	1.030 (0.018)	1.010 (0.013)	0.988 (0.026)	0.971 (0.024)	0.998 (0.021)
JPT (89)						1.017 (0.009)	1.012 (0.009)	1.011 (0.023)	1.028 (0.02)	1.007 (0.022)	0.986 (0.028)	0.969 (0.026)	0.995 (0.024)
CHS (100)							0.994 (0.007)	0.996 (0.022)	1.012 (0.02)	0.991 (0.021)	0.973 (0.028)	0.957 (0.026)	0.982 (0.024)
CHB (97)								1.001 (0.022)	1.017 (0.021)	0.996 (0.021)	0.978 (0.027)	0.961 (0.026)	0.986 (0.024)
PUR (55)									1.015 (0.014)	0.995 (0.008)	0.977 (0.022)	0.96 (0.021)	0.986 (0.017)
MXL (64)										0.980 (0.012)	0.964 (0.027)	0.948 (0.026)	0.972 (0.022)
CLM (60)											0.980 (0.024)	0.963 (0.023)	0.989 (0.019)
YRI (88)												0.983 (0.009)	1.010 (0.009)
LWK (96)													1.027 (0.011)
ASW (61)													

Deep genomes (bottom-left triangle). Highlighted numbers (Denisova column) are >4 standard errors from expectation.

	Denisova (1)	Neanderthal (1)	Dinka (2)	Mandenka (2)	Mbuti (2)	San (2)	Yoruba (2)	Dai (2)	French (2)	Han (2)	Karitiana (2)	Papuan (2)
Neanderthal	**0.695 (0.046)**											
Dinka	**0.724 (0.037)**	0.979 (0.058)										
Mandenka	**0.734 (0.036)**	1.004 (0.055)	1.029 (0.0442)									
Mbuti	**0.734 (0.034)**	1.014 (0.048)	1.041 (0.041)	1.026 (0.04)								
San	**0.759 (0.038)**	1.026 (0.056)	1.024 (0.039)	1.005 (0.04)	0.98 (0.038)							
Yoruba	**0.738 (0.036)**	1.004 (0.057)	1.007 (0.035)	0.985 (0.039)	0.961 (0.036)	0.975 (0.036)						
Dai	**0.765 (0.038)**	1.066 (0.063)	1.075 (0.046)	1.047 (0.046)	1.013 (0.044)	1.051 (0.04)	1.085 (0.041)					
French	**0.732 (0.037)**	0.993 (0.060)	0.994 (0.045)	0.959 (0.044)	0.937 (0.044)	0.957 (0.04)	0.972 (0.039)	0.908 (0.039)				
Han	**0.762 (0.036)**	1.078 (0.057)	1.061 (0.047)	1.048 (0.048)	1.013 (0.039)	1.043 (0.041)	1.080 (0.043)	0.984 (0.041)	1.110 (0.053)			
Karitiana	**0.712 (0.041)**	0.988 (0.062)	0.967 (0.047)	0.938 (0.047)	0.929 (0.045)	0.938 (0.044)	0.96 (0.044)	0.861 (0.041)	0.966 (0.046)	0.875 (0.046)		
Papuan	**0.726 (0.038)**	1.008 (0.063)	1.006 (0.052)	0.974 (0.046)	0.964 (0.044)	0.979 (0.044)	0.995 (0.048)	0.924 (0.041)	1.020 (0.048)	0.911 (0.042)	1.046 (0.053)	
Sardinian	**0.750 (0.039)**	1.001 (0.057)	1.020 (0.046)	0.969 (0.043)	0.958 (0.038)	0.969 (0.043)	0.988 (0.044)	0.919 (0.041)	1.014 (0.045)	0.917 (0.040)	1.052 (0.055)	1.008 (0.048)

Notes: ±1 standard errors (parentheses) are based on a Block Jackknife with 100 equally sized blocks. We highlight numbers >4 standard errors from expectation. Ratios for Neanderthal and Denisova are normalized by the number of synonymous sites on each lineage, to adjust for the expectation of fewer mutations in the ancient samples than in the present-day human samples due to less time elapsed since divergence (all other comparisons are un-normalized). Ratios are based on the accumulation observed in the population in the row divided by that in the population in the column. Numbers in parentheses are sample sizes.

Acknowledgments

References

1.   Charlesworth, B. Fundamental concepts in genetics: effective population size and patterns of molecular evolution and variation. *Nat Rev Genet* 10, 195-205 (2009).

2.   Li, H. & Durbin, R. Inference of human population history from individual whole-genome sequences. *Nature* 475, 493-6 (2011).

3.   Keinan, A., Mullikin, J.C., Patterson, N. & Reich, D. Measurement of the human allele frequency spectrum demonstrates greater genetic drift in East Asians than in Europeans. *Nat Genet* 39, 1251-5 (2007).

4.   Gronau, I., Hubisz, M.J., Gulko, B., Danko, C.G. & Siepel, A. Bayesian inference of ancient human demography from individual genome sequences. *Nat Genet* 43, 1031-4 (2011).

5.   Lohmueller, K.E. *et al.* Proportionally more deleterious genetic variation in European than in African populations. *Nature* 451, 994-7 (2008).

6.   Tennessen, J.A. *et al.* Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* 337, 64-9 (2012).

7.   Fu, W. *et al.* Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nature* 493, 216-20 (2013).

8.   Casals, F. *et al.* Whole-exome sequencing reveals a rapid change in the frequency of rare functional variants in a founding population of humans. *PLoS Genet* 9, e1003815 (2013).

9.   Kidd, J.M. *et al.* Population genetic inference from personal genome data: impact of ancestry and admixture on human genomic variation. *Am J Hum Genet* 91, 660-71 (2012).

10.   Kunsch, H.R. The Jackknife and the Bootstrap for General Stationary Observations. *Annals of Statistics* 17, 1217-1241 (1989).

11.   Abecasis, G.R. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* 491, 56-65 (2012).

12.   Meyer, M. *et al.* A high-coverage genome sequence from an archaic Denisovan individual. *Science* 338, 222-6 (2012).

13.   Prufer, K. *et al.* The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* 505, 43-9 (2014).

14.   Turchin, M.C. *et al.* Evidence of widespread selection on standing variation in Europe at height-associated SNPs. *Nat Genet* 44, 1015-9 (2012).

15.   Gravel, S. *et al.* Demographic history and rare allele sharing among human populations. *Proc Natl Acad Sci U S A* 108, 11983-8 (2011).

16.   Balick, D.J., Do, R., Reich, D. & Sunyaev, S.R. Response to a population bottleneck can be used to infer recessive selection. *Submitted. arXiv. 1312.3028.*

17.   Boyko, A.R. *et al.* Assessing the evolutionary impact of amino acid mutations in the human genome. *PLoS Genet* 4, e1000083 (2008).

18.   Simons, Y.B., Turchin, M.C., Pritchard, J.K. & Sella, G. The deleterious mutation load is insensitive to recent population history. *Nat Genet* (2014).

19.	Adzhubei, I., Jordan, D.M. & Sunyaev, S.R. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet* Chapter 7, Unit7 20 (2013).

20.	Ng, P.C. & Henikoff, S. Predicting deleterious amino acid substitutions. *Genome Res* 11, 863-74 (2001).

21.	Duret, L. & Galtier, N. Biased gene conversion and the evolution of mammalian genomic landscapes. *Annu Rev Genomics Hum Genet* 10, 285-311 (2009).

22.	Schuster, S.C. *et al.* Complete Khoisan and Bantu genomes from southern Africa. *Nature* 463, 943-7 (2010).

23.	Gnirke, A. *et al.* Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat Biotechnol* 27, 182-9 (2009).

24.	Reich, D. Insights into human population history from a high coverage Neandertal genome. *Society for Molecular Biology and Evolution 2013 Meeting.* Abstract 37(2013).

25.	Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 38, e164 (2010).

26.	Sunyaev, S.R. *et al.* PSIC: profile extraction from sequence alignments with position-specific counts of independent observations. *Protein Eng* 12, 387-94 (1999).

Tennessen et al.   Bottleneck and growth