

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

Ron Do^{1,2}, Daniel Balick^{1,3}, Heng Li^{1,2}, Ivan Adzhubei³, Shamil Sunyaev^{1,3} & David Reich^{1,2,4}

Non-African populations have experienced size reductions in the time since their split from West Africans, leading to the hypothesis that natural selection to remove weakly deleterious mutations has been less effective in the history of non-Africans. To test this hypothesis, we measured the per-genome accumulation of nonsynonymous substitutions across diverse pairs of populations. We find no evidence for a higher load of deleterious mutations in non-Africans. However, we detect significant differences among more divergent populations, as archaic Denisovans have accumulated nonsynonymous mutations faster than either modern humans or Neanderthals. To reconcile these findings with patterns that have been interpreted as evidence of the less effective removal of deleterious mutations in non-Africans than in West Africans, we use simulations to show that the observed patterns are not likely to reflect changes in the effectiveness of selection after the populations split but are instead likely to be driven by other population genetic factors.

The effectiveness with which natural selection removes deleterious mutations from a population depends not only on the selection coefficient (s) of a mutation but also on the population size (N), which determines the magnitude of the stochastic force of genetic drift. For a constant-size population in equilibrium, effectiveness is fully determined by the product Ns (ref. 1). Because of the dependence on N , the rate at which deleterious alleles are removed from a population depends, in theory, on demographic history. Demographic differences across human populations are well documented. For example, founder events in the last 100,000 years have reduced nucleotide diversity (the number of differences per base pair between paired chromosomes in an individual) 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 10,000 years has led to rapid population expansions.

To investigate whether selection has differed in its effectiveness across populations, some studies have contrasted mutation classes thought to be subject to little selection (synonymous mutations in genes) to those potentially subject to purifying selection (nonsynonymous mutations)^{5–9}. A key study measured the proportion of polymorphic positions that were nonsynonymous in 20 Europeans and 15 African Americans and showed that, whereas both classes of sites had a reduced rate in Europeans, the reduction was proportionally less for nonsynonymous sites⁵. This pattern was interpreted as being due, in part, to a reduced effectiveness of natural selection against weakly deleterious alleles in Europeans in comparison to West Africans due to the smaller effective population size in Europeans since the separation of these populations⁵. Subsequent studies have confirmed the observation^{6,8,10} and have often given a similar interpretation^{7–9,11}. What these studies have shown is that there has been an interaction between the forces of natural selection and demographic history that has affected the total number of nonsynonymous polymorphisms^{5–11}. However, it does not follow that there have been differences in the effectiveness of selection after the population split. In the common ancestral population of Europeans and West Africans, the average derived allele frequency for nonsynonymous sites would have been lower than for synonymous sites, as negative selection places downward pressure on derived allele frequencies⁵. The different frequency distributions at nonsynonymous and synonymous sites would have responded differently to the bottleneck that then occurred in European populations, simply because the different initial shapes of the distributions would be distorted in different ways by the bottleneck. Thus, the empirically observed differences between Europeans and West Africans could arise independently of differences in the effectiveness of selection after the population split.

RESULTS

No significant differences in the load of deleterious mutations across human populations

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 carries an excess. Any genomes that are compared in this manner are by definition separated by the same amount of time since their most recent common ancestor at each chromosomal location. In the absence of selection and assuming no differences in the mutation rate between the two populations, both genomes are expected to harbor the same number of genome-specific mutations. In the presence of selection, however, mutations are removed from the ancestral populations of

¹Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA.

²Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA.

³Division of Genetics, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA.

⁴Howard Hughes Medical Institute, Harvard Medical School, Boston, Massachusetts, USA. Correspondence should be addressed to S.S. (ssunyaev@rics.bwh.harvard.edu) or D.R. (reich@genetics.med.harvard.edu).

Received 19 February 2014; accepted 9 December 2014; published online 12 January 2015; doi:10.1038/ng.3186

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

Data set	R (relative accumulation of mutations)				R ² (relative accumulation of homozygous mutations)				
	R (all nonsynonymous)	R (synonymous)	R (benign)	R (possibly damaging)	R ² (all nonsynonymous)	R ² (synonymous)	R ² (benign)	R ² (possibly damaging)	R ² (probably damaging)
24 deep genomes	4	1.022 (0.012)	1.002 (0.018)	1.007 (0.040)	1.031 (0.038)	0.652 (0.014)	0.628 (0.021)	0.602 (0.043)	0.660 (0.047)
Celera exomes	15	0.982 (0.011)	1.011 (0.022)	1.019 (0.043)	0.992 (0.039)	0.610 (0.011)	0.583 (0.021)	0.616 (0.053)	0.586 (0.047)
1000 Genomes Project exomes	88	1.019 (0.010)	0.999 (0.015)	0.955 (0.028)	1.011 (0.026)	0.655 (0.012)	0.639 (0.018)	0.599 (0.032)	0.642 (0.032)
ESP exomes	1,088	1.004 (0.009)	0.993 (0.013)	1.001 (0.021)	1.037 (0.029)	0.605 (0.010)	0.598 (0.015)	0.578 (0.025)	0.630 (0.034)

We obtained ± 1 standard errors from a weighted-block jackknife with 100 equally sized blocks. For the whole genomes, Yoruba and Mandenka samples represent West Africans and French and Sardinian samples represent Europeans. For the 1000 Genomes Project data, YRI represents West Africans and CEU represents Europeans. The Celera and Exome Sequencing Project (ESP) analyses represent people of West African ancestry using data from African Americans.

each of the two genomes at a rate that depends on the product Ns . Thus, differences in the effectiveness with which selection removes mutations from the two populations can be inferred from a detected asymmetry in the mutation counts between the two genomes. Here we test for differences in the accumulation of mutations between two genomes by sampling one genome from population X and one genome from population Y , determining the ancestral state on the basis of the chimpanzee genome (PanTro2) and recording all the differences. We count the number of derived mutations in genome X that are not seen in genome Y and vice versa, and we define a statistic $R_{X/Y}$ that is a ratio of the two counts. We average over all possible pairs of genomes from the two populations and compute a standard error using a weighted-block jackknife to correct for correlation among neighboring sites (Online Methods)¹². If selection has been equally effective and mutation rates have been the same since the population split, $R_{X/Y}$ is expected to equal 1. This statistic is monotonically related to the difference in mutation load between the two populations.

We measured $R_{\text{WestAfrican/European}}$ in four sequencing data sets: (i) the coding regions of genes (exomes) in 15 African Americans and 20 European Americans⁵; (ii) exomes from 1,089 individuals in the 1000 Genomes Project¹³; (iii) exomes from 1,088 African Americans and 1,351 European Americans⁶; and (iv) 24 whole genomes sequenced to high coverage^{14,15} (**Supplementary Table 1**). We inferred the ancestral allele on the basis of comparison to chimpanzee. As expected for sites unaffected by selection and for indistinguishable differences in mutation rate in the history of the two populations, $R_{\text{WestAfrican/European}}$ (synonymous) was within two standard errors of 1 (**Table 1** and **Supplementary Table 2**). However, $R_{\text{WestAfrican/European}}$ (nonsynonymous) was also indistinguishable from 1. Thus, our data provide no evidence that purging of weakly deleterious mutations has been less effective in Europeans than in West Africans, similar to the finding in ref. 16 for similar population comparisons. To extend these results to a more diverse set of populations, we computed $R_{X/Y}$ for all possible pairs of 11 populations, each represented by 2 deep genome sequences¹⁵, and all pairwise comparisons of 14 populations from the 1000 Genomes Project¹³. We observed no significant differences for any population pair, despite there sometimes being profound differences in demographic history (**Fig. 1** and **Supplementary Table 3**).

To contextualize these null findings, we carried out simulations using fitted models of the histories of West African and European populations^{5,6,17} (**Supplementary Table 4**). The simulations showed that, if selection acts additively and coefficients are in the range $s \in [-0.004, -0.0004]$, $R_{\text{WestAfrican/European}}$ is expected to be below 0.95 and detectable given the standard errors of our measurements (**Fig. 2**). 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 computed the expected value of $R_{\text{WestAfrican/European}}$ integrating over a previously fitted distribution of selection coefficients¹⁸, we found that $R_{\text{WestAfrican/European}}$ (nonsynonymous) was expected to be 0.987, too close to 1 to be reliably detected given the standard errors of our measurements (**Table 1**). This is consistent with other studies that have concluded that, assuming additively acting mutations, the mutational load in West Africans and Europeans is expected to be indistinguishable when measured on a per-genome basis^{16,19–21}. We also simulated recessively acting mutations and in this case predict a stronger difference across populations. The direction of the difference was opposite to that for additively acting mutations, however, reflecting the fact that recessively acting mutations that drift up in frequency owing to a bottleneck can be efficiently purged through the action of selection, as they are exposed in homozygous form²² (**Fig. 2** and **Supplementary Fig. 1**). The difference in direction for

			IBS (Spanish)	GBR (British)	FIN (Finn)	CEU (Euro- pean)	JPT (Japa- nese)	CHS (Chin- ese)	CHB (Chin- ese)	PUR (Puerto Rican)	MXL (Mex- ican)	CLM (Co- lombian)	YRI (Niger- ian)	LWK (Ken- yan)	ASW (Afr. Am.)	1KG	
	TSI (98)		1.026 (0.005)	1.003 (0.003)	1.003 (0.004)	1 (0.003)	0.998 (0.01)	1.005 (0.011)	1.001 (0.011)	1.017 (0.004)	1.014 (0.006)	1.004 (0.005)	1.005 (0.012)	0.992 (0.011)	1.013 (0.01)	TSI	
		IBS (14)		0.978 (0.005)	0.977 (0.005)	0.974 (0.005)	0.974 (0.011)	0.981 (0.011)	0.978 (0.011)	0.993 (0.006)	0.989 (0.008)	0.979 (0.006)	0.986 (0.012)	0.972 (0.012)	0.992 (0.01)	IBS	
	Deni sova (1)			GBR (89)		0.999 (0.003)	0.996 (0.002)	0.995 (0.01)	1.002 (0.011)	0.998 (0.01)	1.014 (0.005)	1.011 (0.006)	1.001 (0.005)	1.003 (0.012)	0.989 (0.011)	1.01 (0.01)	GBR
Nean derthal	0.857 (0.043)	Nean derthal (1)			FIN (93)	0.997 (0.003)	0.995 (0.01)	1.003 (0.011)	0.999 (0.011)	1.015 (0.005)	1.011 (0.007)	1.001 (0.005)	1.003 (0.013)	0.99 (0.012)	1.011 (0.01)	FIN	
Dinka	0.849 (0.035)	0.98 (0.042)	Dinka (2)			CEU (85)	0.998 (0.011)	1.005 (0.011)	1.002 (0.011)	1.018 (0.005)	1.014 (0.007)	1.004 (0.005)	1.006 (0.013)	0.992 (0.012)	1.013 (0.01)	CEU	
Man denka	0.853 (0.038)	0.99 (0.04)	1.014 (0.017)	Man denka (2)			JPT (89)	1.008 (0.004)	1.004 (0.003)	1.019 (0.009)	1.016 (0.008)	1.006 (0.009)	1.007 (0.013)	0.993 (0.012)	1.014 (0.011)	JPT	
Mbuti	0.88 (0.035)	1.03 (0.041)	1.03 (0.016)	1.017 (0.016)	Mbuti (2)			CHS (100)	0.996 (0.002)	1.012 (0.009)	1.009 (0.008)	0.999 (0.009)	1.001 (0.013)	0.988 (0.012)	1.008 (0.011)	CHS	
San	0.899 (0.037)	1.05 (0.04)	1.014 (0.016)	1.004 (0.016)	0.99 (0.017)	San (2)			CHB (97)	1.015 (0.01)	1.012 (0.008)	1.002 (0.009)	1.004 (0.013)	0.99 (0.012)	1.011 (0.011)	CHB	
Yoruba	0.85 (0.037)	0.993 (0.039)	0.99 (0.018)	0.982 (0.016)	0.965 (0.016)	0.975 (0.017)	Yoruba (2)			PUR (55)	0.996 (0.005)	0.987 (0.003)	0.992 (0.011)	0.978 (0.01)	0.998 (0.008)	PUR	
Dai	0.86 (0.036)	1.017 (0.041)	1.012 (0.019)	1.001 (0.021)	0.983 (0.018)	0.999 (0.019)	1.024 (0.019)	Dai (2)			MXL (64)	0.99 (0.004)	0.995 (0.011)	0.981 (0.011)	1.001 (0.009)	MXL	
French	0.832 (0.036)	0.971 (0.037)	0.992 (0.019)	0.978 (0.02)	0.961 (0.019)	0.973 (0.019)	0.997 (0.019)	0.972 (0.018)	French (2)			CLM (60)	1.002 (0.011)	0.989 (0.01)	1.01 (0.008)	CLM	
Han	0.868 (0.036)	1.01 (0.04)	1.027 (0.018)	1.016 (0.021)	1.00 (0.019)	1.018 (0.019)	1.044 (0.02)	1.021 (0.017)	1.051 (0.019)	Han (2)			YRI (88)	0.986 (0.004)	1.007 (0.004)	YRI	
Karitia na	0.833 (0.037)	0.967 (0.034)	0.97 (0.018)	0.961 (0.019)	0.943 (0.018)	0.954 (0.018)	0.975 (0.019)	0.941 (0.019)	0.97 (0.019)	0.923 (0.018)	Kari tiana (2)			LWK (96)	1.27 (0.011)	LWK	
Papuan	0.85 (0.037)	0.991 (0.039)	1.012 (0.021)	0.999 (0.019)	0.982 (0.02)	0.997 (0.018)	1.021 (0.02)	1.002 (0.02)	1.022 (0.018)	0.976 (0.02)	1.057 (0.021)	Papuan (2)		ASW (61)			
Sardinian	0.859 (0.038)	0.988 (0.038)	0.997 (0.018)	0.985 (0.019)	0.967 (0.018)	0.977 (0.018)	1.006 (0.018)	0.978 (0.017)	1.004 (0.017)	0.958 (0.017)	1.035 (0.022)	0.987 (0.018)	Sar dinian (2)				
Deep genomes	Deni sova	Nean derthal	Dinka	Man denka	Mbuti	San	Yoruba	Dai	French	Han	Kari tiana	Papuan					

Figure 1 Relative load of nonsynonymous mutations $R_{X/Y}$ for diverse pairs of populations. Results for the deep genomes are given at the bottom left, and results for 1000 Genomes Project populations are given at the top right. Ratios are based on the accumulation of mutations observed in the population in the row divided by the accumulation of mutations for the population in the column. Standard errors (± 1 ; in parentheses) are based on a weighted-block jackknife. We highlight numbers >4 standard errors from expectation. Ratios for Neanderthal and Denisovan samples are normalized by the number of synonymous sites specific to each genome, to adjust for the expectation of fewer mutations in the ancient samples than in the present-day human samples owing to less time elapsed since divergence (all other comparisons are not normalized). Ratios involving Neanderthal and Denisovan samples also remove C→T and G→A substitutions to avoid high error rates due to ancient DNA degradation.

additively and recessively acting mutations suggests that, until there is a reliable model of demographic history and a joint distribution of dominance and selection coefficients in humans, it will be impossible to make a reliable theoretical prediction about whether West Africans or Europeans carry a higher per-genome load.

To boost statistical power to detect differences in the load of nonsynonymous mutations, we stratified nonsynonymous sites in two ways. First, we stratified by predicted functional effect. The PolyPhen-2 (ref. 23) and SIFT²⁴ algorithms both predict function in a way that is dependent on the ancestral/derived status of allelic variants relative to the human reference genome, which has a particular ancestry at every segment that can bias measurements. We therefore implemented a version of PolyPhen-2 that is independent of the allelic status of the human reference genome (Online Methods) and that fully corrects for the bias (Supplementary Fig. 2). We detected no significant differences in the accumulation of deleterious mutations¹⁶ (Table 1, Figure 1 and Supplementary Tables 2 and 3). We caution, however, that these tests might also have limited power; for example, when we fit a distribution of additively acting selection coefficients to each of the three PolyPhen functional classes (Supplementary Note) and then simulated under these distributions, we predicted ratios of $R_{\text{WestAfrican}/\text{European}}$ of 0.984–0.993 (Supplementary Table 5), close enough to 1 to not be able to distinguish a significant difference given the standard errors of our measurements. The second way we

attempted to boost power was by restricting analysis to locations where pairs of African and non-African individuals share relatively recent common ancestors. We reasoned that this approach might enhance power, as the population split between African and non-African populations occurred only in the last roughly 100,000 years and the mutations that arose before population divergence would be expected to contribute equally to the descendant populations and thus dilute any true signal. We carried out this analysis using four experimentally phased African and six experimentally phased non-African genomes¹⁵. For each pair of populations, we used the pairwise sequential Markovian coalescent (PSMC) to infer the local time since the most recent common ancestor at each location in the genome (masking the exome to avoid circularly using the same sites for our statistic). When we pooled over sample pairs to increase statistical power and computed $R_{\text{African}/\text{non-African}}$ only for the subset of the genome with the shortest times to the most recent common ancestor, we still detected no significant differences between the populations (Supplementary Table 6).

A greater load of deleterious mutations in archaic Denisovans than in present-day humans

Tests for differences in the load of nonsynonymous mutations are not always null, as we found when we analyzed the deeply sequenced genomes from an archaic Denisovan¹⁴ and Neanderthal¹⁵ and compared

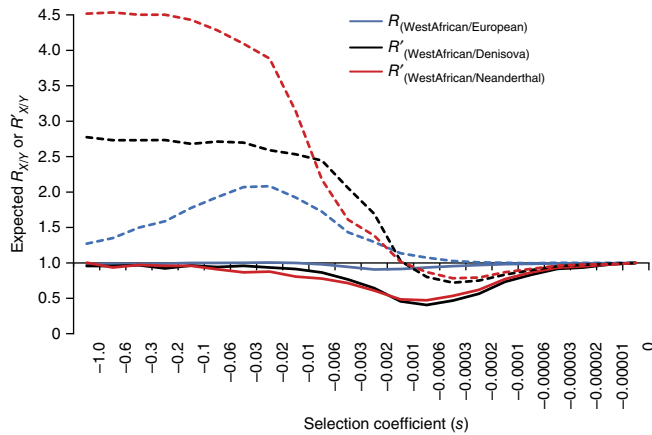


Figure 2 The effect of demographic history on the accumulation of deleterious mutations. To study the expected value of $R_{\text{WestAfrican/European}}$ stratified by selection coefficient, we simulated a previously published model of the joint history of West Africans and Europeans⁶, for a range of selection coefficients, assuming both additive ($h = 0.5$) and recessive ($h = 0$) models of selection. For the additive case (solid lines), $R_{\text{WestAfrican/European}}$ dips below a confidently detectable ratio of 0.95 (given the standard errors of our empirical measurements) for $s \in (-0.0004, -0.004)$. Real distributions of selection coefficients may include a large fraction of their density outside this range, and a true signal may thus be difficult to detect. We also simulated a published model of the history of Denisovans and Neanderthals¹⁵. The simulations predict similar curves for $R'_{\text{WestAfrican/Denisovan}}$ and $R'_{\text{WestAfrican/Neanderthal}}$, reflecting their similar inferred demographic histories (we use a normalized R' statistic to correct for the effects of branch shortening in these ancient genomes). The simulations show that $R'_{\text{WestAfrican/Denisovan}}$ is expected to be below a detectable ratio of 0.95 for $s \in (-0.00002, -0.03)$ and that $R'_{\text{WestAfrican/Neanderthal}}$ is expected to be below 0.95 for $s \in (-0.00002, -0.09)$. For recessively acting alleles (dashed lines), the directionalities of the effects are often opposite.

them to genomes from present-day humans. The ancestors of both archaic individuals are inferred to have maintained relatively small effective population sizes since their main separation from present-day humans, consistent with their levels of genetic diversity having been three to six times smaller¹⁴. A challenge in comparing the accumulation of mutations in present-day samples to that in ancient samples is that fewer mutations are expected to have occurred in the ancient samples, as they are closer in time to the common ancestor. We therefore divided the accumulation of nonsynonymous mutations specific to each genome by the accumulation of synonymous sites: $R'_{X/Y}$ (nonsynonymous) = $R_{X/Y}$ (nonsynonymous)/ $R_{X/Y}$ (synonymous). After removing C→T and G→A mutations in the archaic genomes that might be susceptible to degradation leading to errors in ancient DNA sequences, we inferred that present-day humans had accumulated deleterious mutations at a significantly lower rate than Denisovans since their separation: $R'_{\text{modern/Denisovan}}$ (nonsynonymous) = 0.872 ± 0.034 ($P = 0.0002$) (Supplementary Table 7)¹⁴. However, $R'_{\text{modern/Neanderthal}}$ (nonsynonymous) = 1.037 ± 0.037 is consistent with 1, indicating that deleterious mutations accumulated faster in Denisovan than in Neanderthal ancestors since they separated (Fig. 1). The different patterns—despite similar inferred demographic histories—suggest that fitting models to patterns of neutral genetic variation and then simulating under these models might not produce an accurate prediction of the relative effectiveness of selection in pairs of populations (Fig. 2). A corollary is that current models for the joint history of West Africans and non-Africans might not produce accurate predictions. Little is known about the duration

of the out-of-Africa bottleneck: a short, sharp bottleneck and a long, drawn-out bottleneck are both consistent with most analyses. The primary influence on the cumulative effectiveness of selection is the duration of the bottleneck, and so the uncertainty about its duration is important.

Differences in the effectiveness of biased gene conversion across populations

As a final way to boost power to detect any differences in the effectiveness of the removal of mutations across populations, we analyzed a class of sites that is much larger than the class of nonsynonymous substitutions so that we could make measurements with far smaller standard errors. Specifically, we focused on the class of sites affected by biased gene conversion (BGC), a process in which DNA repair acting on heterozygous sites in gene conversion tracts favors GC over AT alleles²⁵. Because 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 selection²⁵. We found that $R'_{\text{WestAfrican/European}}^{\text{GC} \rightarrow \text{AT}} = 0.995 \pm 0.002$ and $R'_{\text{WestAfrican/European}}^{\text{AT} \rightarrow \text{GC}} = 1.00 \pm 0.002$ by using a statistic

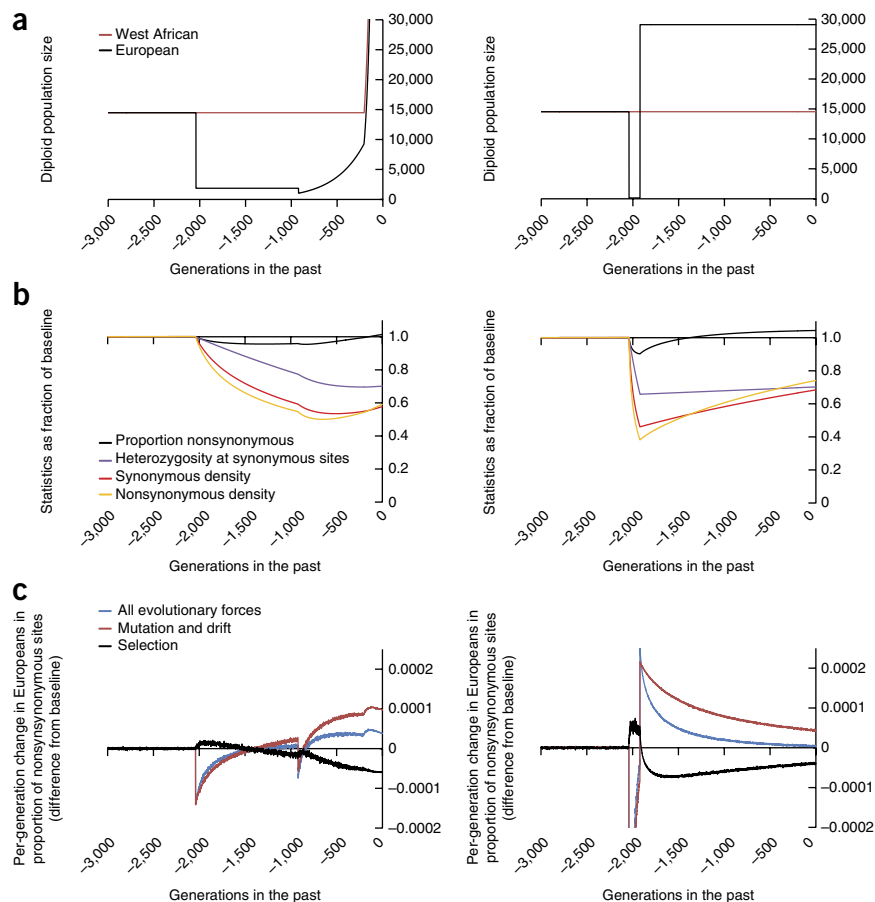
$$R'_{X/Y}^{\text{class1} \rightarrow \text{class2}} = \frac{R_{X/Y}^{\text{class1} \rightarrow \text{class2}}}{(R_{X/Y}^{\text{A} \leftrightarrow \text{T}} + R_{X/Y}^{\text{C} \leftrightarrow \text{G}})}$$

that normalizes by the rates of A↔T and C↔G substitutions that are not expected to be affected by BGC. This approach also corrects for possible differences in mutation rate across populations. For the comparison of diverse West Africans to diverse non-Africans, we detected no significant differences (multiple-hypothesis testing—adjusted $P > 0.05$) (Supplementary Table 8). The very small standard errors allow us to state with confidence that any differences in the effectiveness of selection in the history of non-Africans and West Africans since the split is likely to have been extremely small. However, San Bushmen from southern Africa had about 1% fewer G/C→A/T mutations than all other humans (significant at up to eight standard errors). To our knowledge, this is the first direct detection of a different rate of accumulation of a class of mutations in one present-day human population in comparison to others. The San are among the most deeply diverged present-day humans, and a possible explanation for our observation is that the longer isolation time for the San would have provided an opportunity for differences in the effectiveness of mutation removal across populations to accumulate to the point of having a detectable effect in the high-sensitivity BGC analysis²⁶.

Reinterpretation of previous evidence for less effective selection in Europeans

Previous suggestions that weakly deleterious mutations have been removed less effectively in Europeans than in West Africans were largely based on the study of an alternative statistic: the proportion of polymorphic sites in the exome that are nonsynonymous. This statistic is significantly higher in Europeans than in West Africans⁵. We investigated the population genetic forces shaping this statistic by carrying out simulations that allowed us to study the dynamics of this statistic over time. While our simulations showed qualitative patterns that were consistent with those reported previously⁵, they also provided new insight owing to a modification to the program that allowed us, in every generation, to dissect how selection (versus mutation and genetic drift) contributed to the expected change in the proportion of nonsynonymous sites in that generation. The simulations showed that, during and after a population bottleneck, the per-generation change in the proportion of segregating sites that were nonsynonymous was not driven by selection being less effective at reducing this ratio than

Figure 3 The rise in the proportion of nonsynonymous sites in Europeans in comparison to West Africans is not due to reduced effectiveness of selection in Europeans since the population split. (a) The West African and European diploid population sizes for the two simulated models (left (ref. 6) and right, a bottleneck followed by expansion), both of which specify a population split 2,040 generations ago. Subsequent panels are restricted to Europeans, as the West African population size does not fluctuate enough to cause statistics to deviate substantially from the baseline. (b) Key statistics as a fraction of the baseline. The present proportion of nonsynonymous sites in Europeans is higher than in the ancestral population (black). We also show heterozygosity at unselected sites (blue), synonymous site density (red) and nonsynonymous site density (yellow). (c) Partitioning of the change in the proportion of nonsynonymous sites per generation into selective and other forces. For both models, the temporal dynamics are driven by the forces of mutation and stochastic changes in allele frequency (the curves are positively correlated) and not by negative selective forces (negatively correlated). We plot the per-generation change in the proportion of nonsynonymous mutations due to selection minus its value before the West African–European population split, used as a baseline. A positive value does not mean that selection is working to increase the proportion of nonsynonymous mutations, just that the decrease per generation due to this quantity is less than in the past.



it was in the ancestral population. Instead, after a short period at the start of the bottleneck when the effectiveness of selection in changing this statistic was reduced, selection began to be more effective at reducing the proportion of nonsynonymous sites per generation than it was before the bottleneck (Fig. 3). Thus, the rate at which selection reduced the value of this statistic per generation was enhanced rather than diminished by the bottleneck, which for much of the history means that selection pushed the statistic in the direction opposite to that in which it actually moved. We can conclude from this observation that it is primarily non-selective forces that drove the dynamics of this statistic since the separation of West African and European populations.

Intuitively, what explains these simulation results? Before the West African–European split, allele frequencies of nonsynonymous polymorphisms would, on average, have been much lower owing to the depletion of nonsynonymous sites by selection, and the per-site density of nonsynonymous segregating sites would also have been lower. A population entering a bottleneck primarily loses rare alleles, so the nonsynonymous distribution is predicted to be affected more strongly in each generation by the constant flux of new mutations than the synonymous site distribution, as our simulations show. Once the population expands again, the allele frequencies for nonsynonymous sites also adjust faster because the same flux of new mutations into both classes causes a faster rate of replenishment of nonsynonymous sites than synonymous sites, owing to an initially lower density for this class. It is the greater proportional impact of new mutations on nonsynonymous sites per generation that occurs after a bottleneck—because the class has been depleted by the bottleneck—that is driving the observed effects (Fig. 3). Putting this another way, we agree with previous reports that have suggested that interactions

between the effects of demographic history and natural selection are responsible for the empirically observed differences in the proportion of nonsynonymous segregating sites across human populations^{5–11}. However, we differ in the interpretation. Our simulations show that the observed patterns are not driven by a reduced effectiveness of selection at removing slightly deleterious alleles in some human populations in comparison to others since they separated, as has been hypothesized to explain the patterns observed in comparisons of West Africans to non-Africans^{5,7}, as well as in comparisons of French Canadians⁸, Finns⁹ and Ashkenazi Jews¹¹ to European populations that have not experienced recent bottlenecks. Instead, the patterns are driven primarily by new mutation and drift, acting on the different distributions that existed at nonsynonymous and synonymous sites before the population split.

DISCUSSION

It is tempting to interpret the indistinguishable accumulations of deleterious mutations across present-day human populations as implying that the overall genetic burden of disease should be similar for diverse populations. To the extent that mutations act additively, this is correct, as it implies that the complex demographic events of the past are not expected to lead to substantial population differences in the prevalence rates of complex diseases that have an additive genetic architecture^{16,19}. However, recessively or epistatically acting mutations work in combination to contribute to disease risk, and, because demography affects allele frequencies, it affects the rate of co-occurrence of alleles. For example, the absolute count of alleles occurring in homozygous form is higher in non-Africans than in Africans for all functional site classes (Table 1 and Supplementary Table 9)⁵. Thus,

the relative risk for diseases with non-additive architectures might be influenced by demography. It will be important to determine the extent to which mutations contributing to phenotypes act non-additively, which will largely determine the extent to which demographic differences among human population affect disease risk.

METHODS

Methods and any associated references are available in the [online version of the paper](#).

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank J. Akey, D. Altshuler, C. Bustamante, S. Castellano, C. de Filippo, A. Keinan, A. Kondrashov, E. Lander, K. Lohmueller, S. Mallick, S. Pääbo, N. Patterson, J. Pritchard, M. Przeworski, J. Schaiber, G. Sella and M. Slatkin for valuable discussions. R.D. was supported by a Banting fellowship from the Canadian Institutes of Health Research. S.S. was supported by US National Institutes of Health grants R01GM078598 and R01MH101244. D.R. was supported by US National Institutes of Health grants GM100233 and HG006399 and US National Science Foundation grant 1032255 and is an investigator of the Howard Hughes Medical Institute.

AUTHOR CONTRIBUTIONS

R.D., D.B., H.L., I.A., S.S. and D.R. performed analyses. S.S. and D.R. supervised the research. R.D., D.B., S.S. and D.R. wrote the manuscript with the assistance of all coauthors.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

- Charlesworth, B. Fundamental concepts in genetics: effective population size and patterns of molecular evolution and variation. *Nat. Rev. Genet.* **10**, 195–205 (2009).
- Li, H. & Durbin, R. Inference of human population history from individual whole-genome sequences. *Nature* **475**, 493–496 (2011).
- 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**, 1251–1255 (2007).
- 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**, 1031–1034 (2011).
- Lohmueller, K.E. *et al.* Proportionally more deleterious genetic variation in European than in African populations. *Nature* **451**, 994–997 (2008).
- Tennessen, J.A. *et al.* Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* **337**, 64–69 (2012).
- Fu, W. *et al.* Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nature* **493**, 216–220 (2013).
- 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).
- Lim, E.T. *et al.* Distribution and medical impact of loss-of-function variants in the Finnish founder population. *PLoS Genet.* **10**, e1004494 (2014).
- 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**, 660–671 (2012).
- Carmi, S. *et al.* Sequencing an Ashkenazi reference panel supports population-targeted personal genomics and illuminates Jewish and European origins. *Nat. Commun.* **5**, 4835 (2014).
- Kunsch, H.R. The jackknife and the bootstrap for general stationary observations. *Ann. Stat.* **17**, 1217–1241 (1989).
- Abecasis, G.R. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
- Meyer, M. *et al.* A high-coverage genome sequence from an archaic Denisovan individual. *Science* **338**, 222–226 (2012).
- Prüfer, K. *et al.* The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43–49 (2014).
- Simons, Y.B., Turchin, M.C., Pritchard, J.K. & Sella, G. The deleterious mutation load is insensitive to recent population history. *Nat. Genet.* **46**, 220–224 (2014).
- Gravel, S. *et al.* Demographic history and rare allele sharing among human populations. *Proc. Natl. Acad. Sci. USA* **108**, 11983–11988 (2011).
- Boyko, A.R. *et al.* Assessing the evolutionary impact of amino acid mutations in the human genome. *PLoS Genet.* **4**, e1000083 (2008).
- Gazave, E., Chang, D., Clark, A.G. & Keinan, A. Population growth inflates the per-individual number of deleterious mutations and reduces their mean effect. *Genetics* **195**, 969–978 (2013).
- Lohmueller, K.E. The impact of population demography and selection on the genetic architecture of complex traits. *PLoS Genet.* **10**, e1004379 (2014).
- Lohmueller, K.E. The distribution of deleterious genetic variation in human populations. *bioRxiv* doi:10.1101/005330 (2014).
- Balick, D.J., Do, R., Reich, D. & Sunyaev, S.R. Response to a population bottleneck can be used to infer recessive selection. *bioRxiv* doi:10.1101/003491 (2014).
- Adzhubei, I., Jordan, D.M. & Sunyaev, S.R. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr. Protoc. Hum. Genet.* Chapter 7 Unit 7.20 (2013).
- Ng, P.C. & Henikoff, S. Predicting deleterious amino acid substitutions. *Genome Res.* **11**, 863–874 (2001).
- Duret, L. & Galtier, N. Biased gene conversion and the evolution of mammalian genomic landscapes. *Annu. Rev. Genomics Hum. Genet.* **10**, 285–311 (2009).
- Schuster, S.C. *et al.* Complete Khoisan and Bantu genomes from southern Africa. *Nature* **463**, 943–947 (2010).

ONLINE METHODS

Data. The data sets we analyzed were published previously and are summarized here. We determined the ancestral allele at each position on the basis of comparison to the chimpanzee genome (PanTro2), except in the case of the Celera data set where we used the previously reported determination⁵.

Celera. PCR amplification and Sanger sequencing were performed on 15 African-American and 20 European-American 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 for both African Americans and European Americans⁵.

1000 Genomes Project. A total of 1,089 samples from 14 populations were analyzed in Phase 1 of the 1000 Genomes Project. Illumina-based exome sequencing¹³ was performed to ~100× average coverage after solution hybrid capture of the exome²⁷.

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 exome⁶.

24 Genomes. This data set included two samples each from six non-African and five sub-Saharan African genomes, an archaic human from Denisova Cave in Siberia sequenced to 31× coverage and an archaic Neanderthal from Denisova Cave in Siberia sequenced to 52× coverage. All sequencing data are based on Illumina technology. We used the version of this data set reported in ref. 15. We only analyzed sites with genotype quality (GQ) scores of ≥45.

Mutation annotation. We annotated coding mutations using ANNOVAR²⁸, which classifies sites as 'nonsynonymous', 'synonymous', 'stop gain' or 'stop loss'. We subclassified variants using a version of PolyPhen-2 that was created specifically for this study where annotation is independent of the ancestral/derived status of the human genome reference sequence (human-free PolyPhen-2). To guarantee the independence of the PolyPhen-2 predictions from the human genome reference sequence, we modified PolyPhen-2 to rely solely on the multi-species conservation score used in this method²⁹. 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-2. The predictions in our simplified PolyPhen-2 method are based on the absolute value of the difference in the scores for the two alleles. By construction, this is symmetric with respect to reference/non-reference (and also ancestral/derived and 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 were interested in the expected number of mutations in a randomly sampled haploid exome from one population that were not seen in a randomly sampled comparison exome from another population. To compute this in a situation where we had many exomes available from each population, we did not wish to literally randomly choose a single exome from each population, as this would reduce the sample size in our analysis, resulting in decreased precision of our estimates. Instead, we obtained the expected value if we were to perform an infinite number of random samplings. To compute this value, at each variable site i , we defined d_X^i as the count of the mutant allele at that site in a sample of n_X^i exomes from population X . Similarly, d_Y^i was the count of the mutant allele in a sample of n_Y^i exomes from population Y . The expectation values were obtained by summing over all sites:

$$L_{X, \text{not } Y} = \sum_i (d_X^i / n_X^i) (1 - d_Y^i / n_Y^i)$$

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

$$L_{X, \text{not } Y}^2 = \sum_i \frac{2d_X^i (n_X^i - d_X^i)}{n_X^i (n_X^i - 1)} \left(1 - \frac{2d_Y^i (n_Y^i - d_Y^i)}{n_Y^i (n_Y^i - 1)} \right)$$

We then defined the ratio statistics as follows:

$$R_{X/Y} = L_{X, \text{not } Y} / L_{Y, \text{not } X}$$

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

Weighted-block jackknife to estimate standard errors. We obtained standard errors using a weighted-block jackknife¹². We divided the SNP data sets into 100 contiguous blocks and then recomputed the statistic on all of the data except for the data from that block. The variation could be converted to a standard error using jackknife theory. We assessed significance on the basis of the number of standard errors from the null expectation of $R_{X/Y} = 1$ and computed a P value using a z score assuming a normal distribution.

Time-stratified computation of the relative accumulation of deleterious mutation. We began with data from ten experimentally phased genomes, all processed identically¹⁵. These genomes consisted of one each from the populations in **Figure 1** except for the Dinka. We then combined the haploid genomes from 4 African and 6 non-African individuals in all possible pairs to make $96 = (2 \times 4) \times (2 \times 6)$ pseudo-diploid individuals. We masked the data from the exome and ran PSMC² 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 $R_{\text{African/non-African}}$ statistic within each time-stratified subset (using exomic sites that had been masked from the PSMC analysis so we could independently use these data for downstream analysis).

Analysis of sites susceptible to biased gene conversion. We computed the accumulation of mutations susceptible to BGC for three different substitution classes: $G/C \rightarrow A/T$ ($G \rightarrow A$, $G \rightarrow T$, $C \rightarrow A$ or $C \rightarrow T$) mimicking negative selection, $A/T \rightarrow G/C$ ($A \rightarrow G$, $T \rightarrow G$, $A \rightarrow C$ or $T \rightarrow C$) mimicking positive selection and $A \leftrightarrow T$ or $G \leftrightarrow C$ ($A \rightarrow T$, $T \rightarrow A$, $C \rightarrow G$ or $G \rightarrow C$), which we treated as neutral (and used as the denominator of $R_{X/Y}$). For BGC analyses, we used the entire genome, excluding sites in the exome.

The $R'_{X/Y}$ statistic: correcting for branch shortening and differences in mutation rate. For analyses involving samples from the archaic Denisovan and Neanderthal populations, which are many tens of thousands of years old and thus have experienced less evolution from the common ancestor than the present-day humans to whom they were compared, we did not expect $L_{\text{archaic, not-modern}}$ to equal $L_{\text{modern, not-archaic}}$, even for neutral sites. For all analyses involving ancient samples, we normalize both $L_{X, \text{not } Y}$ and $L_{Y, \text{not } X}$ by the accumulation of mutations at sites that were expected to act neutrally (synonymous sites for coding sequences and sites with $A/T + C/G$ for BGC). We then defined:

$$R'_{X/Y}{}^{\text{class}} = \left(L_{X, \text{not } Y}{}^{\text{class}} / L_{Y, \text{not } X}{}^{\text{class}} \right) / \left(L_{X, \text{not } Y}{}^{\text{normalization}} / L_{Y, \text{not } X}{}^{\text{normalization}} \right) = R_{X/Y}{}^{\text{class}} / R_{X/Y}{}^{\text{normalization}}$$

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 in one population or the other since they separated.

Avoiding the confounder of ancient DNA degradation. Ancient DNA data are known to have a high rate of $C \rightarrow T$ and $G \rightarrow A$ errors, which persist at a measurable rate even in high-coverage genomes such as those from Denisovan or Neanderthal individuals¹³. In **Supplementary Table 7**, we document that this error process is substantial enough to bias statistics involving Denisovans. We therefore restricted the computation of $R_{X/Y}$ involving the ancient samples to sites that were not $C \rightarrow T$ and $G \rightarrow A$ substitutions (for the sake of comparability, we also did not analyze these classes of sites for non-ancient samples in analyses that also involved archaic samples).

Simulations. We wrote a forward simulation program in C that implemented an infinite-sites model. Each mutation was assumed to occur at an unlinked site.

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 on the basis of 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 main text, we used a demographic model previously fitted to genetic data⁶, as well as a simple bottleneck and expansion model (**Supplementary Fig. 1** reports the results for four histories using the simulation parameters shown in **Supplementary Table 4**). For comparisons of West African and archaic population histories, we also used a previously fitted demographic model¹⁵. We used a mutation rate of 2×10^{-8} mutations per base pair per generation.

At each simulated site, we computed the probability of it being discovered as polymorphic in a sample of size 40, assuming that K_i is the total number of derived alleles for a total of N individuals in the population. We computed the probability that 40 chromosomes were polymorphic at a site i as 1 minus the hypergeometric probability of 0 or 40 derived alleles:

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

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

Code availability. The code in C and Perl that was used for the simulations is available on request from D.R.

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 simulations for different coefficients. For **Figure 3**, we simulated each of 19 values: $s = \{-1 \times 10^{-6}, -2 \times 10^{-6}, -5 \times 10^{-6}, -1 \times 10^{-5}, -2 \times 10^{-5}, -5 \times 10^{-5}, -1 \times 10^{-4}, -2 \times 10^{-4}, -5 \times 10^{-4}, -1 \times 10^{-3}, -2 \times 10^{-3}, -5 \times 10^{-3}, -0.01, -0.02, -0.05, -0.1, -0.2, -0.5, -1\}$. To compute expected values for $L_{X, \text{not } Y}$ and the density of segregating sites per base pair in a fixed sample size of 40 chromosomes, we used a weighted average of the values of the simulated single selection coefficient statistics. For most analyses, we used weighting based on the distribution of human selection coefficients for the nonsynonymous sites inferred in ref. 18, where the probability of a given value of $-s$ was drawn from a gamma distribution fitted to European genetic data with $\alpha = 0.206$ and $\beta = 15,400$. For analyses of the expected value of $R_{\text{WestAfrican/European}}$ stratified by PolyPhen-2 functional class, we used the values inferred in the **Supplementary Note**. Further details of the integration over selection coefficients are given in the **Supplementary Note**.

Partitioning evolutionary dynamics into effects due to selection, mutation and drift. We modified the simulations to sample derived alleles in the next generation at each simulated nucleotide under two alternative assumptions: assuming that all evolutionary forces are operating, and assuming that only non-selective forces are operating.

Let $All_{s,i,j-1,k}$ be the count of derived alleles that have selection coefficient s at nucleotide position i in the $j-1$ generation in simulation replicate k . We used our simulation machinery to sample the count of derived alleles in the subsequent generation $All_{s,i,j,k}$, assuming that the selection coefficient in the next generation was the same. We also independently sampled the count of derived alleles in that generation, $NonSel_{s,i,j,k}$, assuming that selection stopped in that generation ($s = 0$). Because the count of derived alleles was always sampled on the basis of $All_{s,i,j-1,k}$ in the previous generation (not $NonSel_{s,i,j-1,k}$), this procedure ensured that the accumulation of derived alleles at each position corresponded to what is expected for an unchanging selection coefficient over time.

To compute the expected value of counts of segregating sites in generation j assuming that all analyzed nucleotides when mutated produce an allele of

selection coefficient s , we averaged over A simulation replicates and B simulated nucleotides per replicate:

$$E[All_{s,j}] = \frac{1}{A \times B} \sum_{i=1}^A \sum_{k=1}^B All_{s,i,j,k}$$

$$E[NonSel_{s,j}] = \frac{1}{A \times B} \sum_{i=1}^A \sum_{k=1}^B NonSel_{s,i,j,k}$$

In practice, we also wanted to integrate over a distribution of selection coefficients. Let z_{nonsyn} be the fraction of nucleotides in the genome than when mutated result in a nonsynonymous substitution, which we empirically adjusted to obtain a ratio that matched the data in West Africans (**Supplementary Note**). Let $f(s)$ be the distribution of selection coefficients for *de novo* substitutions (in many of our simulations, we used a distribution fitted to data by ref. 18). We then obtained the expected density of nonsynonymous sites by integrating over the distribution of selection coefficients, which we did in practice by performing a large number of simulations for each of a range of selection coefficients and then grid averaging:

$$E[All_{\text{nonsyn},j}] = z_{\text{nonsyn}} \int E[All_{s,j}] f(s) ds$$

$$E[NonSel_{\text{nonsyn},j}] = z_{\text{nonsyn}} \int E[NonSel_{s,j}] f(s) ds$$

We also defined the expectation for synonymous sites:

$$E[All_{\text{syn},j}] = E[NonSel_{\text{syn},j}] = (1 - z_{\text{nonsyn}}) E[All_{0,j}]$$

We defined the proportion of sites that were nonsynonymous in a given generation as follows:

$$\text{PropAll}_j = \frac{All_{\text{nonsyn},j}}{All_{\text{nonsyn},j} + All_{\text{syn},j}}$$

$$\text{PropNonSel}_j = \frac{NonSel_{\text{nonsyn},j}}{NonSel_{\text{nonsyn},j} + NonSel_{\text{syn},j}}$$

The expected change in the proportion of nonsynonymous sites in generation j is:

$$\delta \text{PropAll}_j = \text{PropAll}_j - \text{PropAll}_{j-1} \text{ (all evolutionary forces)}$$

$$\delta \text{PropNonSel}_j = \text{PropNonSel}_j - \text{PropNonSel}_{j-1} \text{ (mutation and drift only)}$$

$$\delta \text{PropSel}_j = \text{PropAll}_j - \text{PropNonSel}_j \text{ (selective forces only)}$$

We defined the effectiveness of an evolutionary force in a generation—measured as the magnitude of its effect in that generation on a statistic of interest—by comparing it to the baseline when the population was constant in size (we call this 2,500 generations ago for convenience, as for both the demographic histories we simulated, the 2 populations had not yet split 2,500 generations ago and were in mutation-selection-drift equilibrium):

$$\Delta \text{PropSel}_j = \delta \text{PropSel}_j - \delta \text{PropSel}_{-2,500}$$

$$\Delta \text{PropNonSel}_j = \delta \text{PropNonSel}_j - \delta \text{PropNonSel}_{-2,500}$$

$$\Delta \text{PropAll}_j = \delta \text{PropAll}_j - \delta \text{PropAll}_{-2,500}$$

These statistics are positive if the effectiveness of the 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.

27. Gnirke, A. *et al.* Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat. Biotechnol.* **27**, 182–189 (2009).
28. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
29. Sunyaev, S.R. *et al.* PSIC: profile extraction from sequence alignments with position-specific counts of independent observations. *Protein Eng.* **12**, 387–394 (1999).