

# Distance from sub-Saharan Africa predicts mutational load in diverse human genomes

Brenna M. Henn<sup>a,1,2</sup>, Laura R. Botigué<sup>a,1</sup>, Stephan Peischl<sup>b,c,d,1</sup>, Isabelle Dupanloup<sup>b</sup>, Mikhail Lipatov<sup>a</sup>, Brian K. Maples<sup>e</sup>, Alicia R. Martin<sup>e</sup>, Shaila Musharoff<sup>e</sup>, Howard Cann<sup>f,3</sup>, Michael P. Snyder<sup>e</sup>, Laurent Excoffier<sup>b,c,4</sup>, Jeffrey M. Kidd<sup>g,4</sup>, and Carlos D. Bustamante<sup>e,2,4</sup>

<sup>a</sup>Department of Ecology and Evolution, Stony Brook University, The State University of New York, Stony Brook, NY 11794; <sup>b</sup>Institute of Ecology and Evolution, University of Berne, 3012 Berne, Switzerland; <sup>c</sup>Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland; <sup>d</sup>Interfaculty Bioinformatics Unit, University of Berne, 3012 Berne, Switzerland; <sup>e</sup>Department of Genetics, Stanford University School of Medicine, Stanford, CA 94305; <sup>f</sup>Centre d'Etude du Polymorphisme Humain, Foundation Jean Dausset, 75010 Paris, France; and <sup>g</sup>Department of Human Genetics and Department of Computational Medicine and Bioinformatics, University of Michigan Medical School, Ann Arbor, MI 48109

Edited by Charles F. Aquadro, Cornell University, Ithaca, NY, and accepted by the Editorial Board November 13, 2015 (received for review June 9, 2015)

The Out-of-Africa (OOA) dispersal ~50,000 y ago is characterized by a series of founder events as modern humans expanded into multiple continents. Population genetics theory predicts an increase of mutational load in populations undergoing serial founder effects during range expansions. To test this hypothesis, we have sequenced full genomes and high-coverage exomes from seven geographically divergent human populations from Namibia, Congo, Algeria, Pakistan, Cambodia, Siberia, and Mexico. We find that individual genomes vary modestly in the overall number of predicted deleterious alleles. We show via spatially explicit simulations that the observed distribution of deleterious allele frequencies is consistent with the OOA dispersal, particularly under a model where deleterious mutations are recessive. We conclude that there is a strong signal of purifying selection at conserved genomic positions within Africa, but that many predicted deleterious mutations have evolved as if they were neutral during the expansion out of Africa. Under a model where selection is inversely related to dominance, we show that OOA populations are likely to have a higher mutation load due to increased allele frequencies of nearly neutral variants that are recessive or partially recessive.

mutation | founder effect | range expansion | expansion load | purifying selection

It has long been recognized that a human genome may carry many strongly deleterious mutations; Morton et al. (1) estimated that each human carries on average four or five mutations that would have a “conspicuous effect on fitness” if expressed in a homozygous state. Empirically estimating the deleterious mutation burden is now feasible through next-generation sequencing (NGS) technology, which can assay the complete breadth of variants in a human genome. For example, recent sequencing of over 6,000 exomes revealed that nearly half of all surveyed individuals carried a likely pathogenic allele in a known Mendelian disease gene (i.e., from a disease panel used for newborn screening) (2). Although there is some variation across individuals in the number of deleterious alleles per genome, we still do not know whether there are significant differences in deleterious variation among populations. Human populations vary dramatically in their levels of neutral genetic diversity, which suggests variation in the effective population size,  $N_e$ . Theory suggests that the efficacy of natural selection is reduced in populations with lower  $N_e$  because they experience greater genetic drift (3, 4). In an idealized population of constant size, the efficacy of purifying selection depends on the relationship between  $N_e$  and the selection coefficient  $s$  against deleterious mutations. If  $4N_e s << 1$ , deleterious alleles evolve as if they were neutral and can, thus, reach appreciable frequencies. This theory raises the question of whether human populations carry differential burdens of deleterious alleles due to differences in demographic history.

Several recent papers have tested for differences in the burden of deleterious alleles among populations; these papers have

focused on primarily comparing populations of western European and western African ancestry. Despite similar genomic datasets, these papers have reached a variety of contradictory conclusions (4–9). Initially, Lohmueller et al. (10) found that a panel of European Americans carried proportionally more derived, deleterious alleles than a panel of African Americans, potentially as the result of the Out-of-Africa (OOA) bottleneck. More recently, analyses using NGS exome datasets from samples of analogous continental ancestry found small or no differences in the average number of deleterious alleles per genome between African Americans and European Americans—depending on which prediction algorithm was used (11–13). Simulations by Fu et al. (11) found strong bottlenecks with recovery could recapitulate patterns of differences in the number of deleterious alleles between African and

## Significance

Human genomes carry hundreds of mutations that are predicted to be deleterious in some environments, potentially affecting the health or fitness of an individual. We characterize the distribution of deleterious mutations among diverse human populations, modeled under different selection coefficients and dominance parameters. Using a new dataset of diverse human genomes from seven different populations, we use spatially explicit simulations to reveal that classes of deleterious alleles have very different patterns across populations, reflecting the interaction between genetic drift and purifying selection. We show that there is a strong signal of purifying selection at conserved genomic positions within African populations, but most predicted deleterious mutations have evolved as if they were neutral during the expansion out of Africa.

Author contributions: B.M.H., M.P.S., L.E., J.M.K., and C.D.B. designed research; B.M.H., L.R.B., S.P., and J.M.K. performed research; S.P., H.C., and L.E. contributed new reagents/analytic tools; B.M.H., L.R.B., S.P., I.D., M.L., B.K.M., A.R.M., S.M., and J.M.K. analyzed data; and B.M.H., L.R.B., S.P., L.E., J.M.K., and C.D.B. wrote the paper.

Conflict of interest statement: C.D.B. is the founder of IdentifyGenomics, LLC, and is on the scientific advisory boards of Personalis, Inc. and Ancestry.com as well as the medical advisory board InVitae. None of this played a role in the design, execution, or interpretation of experiments and results presented here.

This article is a PNAS Direct Submission. C.F.A. is a guest editor invited by the Editorial Board.

Freely available online through the PNAS open access option.

Data deposition: The sequence reported in this paper has been deposited in the NCBI Sequence Read Archive (accession no. SRP036155).

See Commentary on page 809.

<sup>1</sup>B.M.H., L.R.B., and S.P. contributed equally to this work.

<sup>2</sup>To whom correspondence may be addressed. Email: brenna.henn@stonybrook.edu or cdbustam@stanford.edu.

<sup>3</sup>Deceased May 3, 2014.

<sup>4</sup>L.E., J.M.K., and C.D.B. contributed equally to this work.

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1510805112/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1510805112/-DCSupplemental).

non-African populations, supporting Lohmueller et al. (10), but in contrast to work by Simons et al. (12).

It is important to note two facts about these contradictory observations. First, these papers tend to use different statistics, which differ in power to detect changes across populations, as well as the impact of recent demographic history (6, 11). Lohmueller et al. (10) compared the relative number of nonsynonymous to synonymous (or “probably damaging” to “benign”) SNPs per population in a sample of  $n$  chromosomes, whereas Simons et al. (12) examined the special case of  $n = 2$  chromosomes, namely, the average number of predicted deleterious alleles per genome (i.e., heterozygous + 2 \* homozygous derived variants per genome). One way to think about these statistics is that the total number of variants,  $S$ , gives equal weight,  $w = 1$ , to an SNP regardless of its frequency,  $p$ . The average number of deleterious variants statistic gives weights proportional to the expected heterozygous and homozygous frequencies or  $w = 2p(1 - p) + p^2 = 2p - p^2$ . The average number of deleterious alleles per genome is fairly insensitive to differences in demographic history because heterozygosity is biased toward common variants. In contrast, the proportion of deleterious alleles has greater power to detect the impact of recent demographic history for large  $n$  across the populations because it is sensitive to rare variants that tend to be more numerous, younger, and enriched for functionally important mutations (14–16). Second, empirical comparisons between two populations have focused primarily on an additive model for deleterious mutations, even though there is evidence for pathogenic mutations exhibiting a recessive or dominant effect (17, 18), and possibly an inverse relationship between the strength of selection  $s$  and the dominance parameter  $h$  (19).

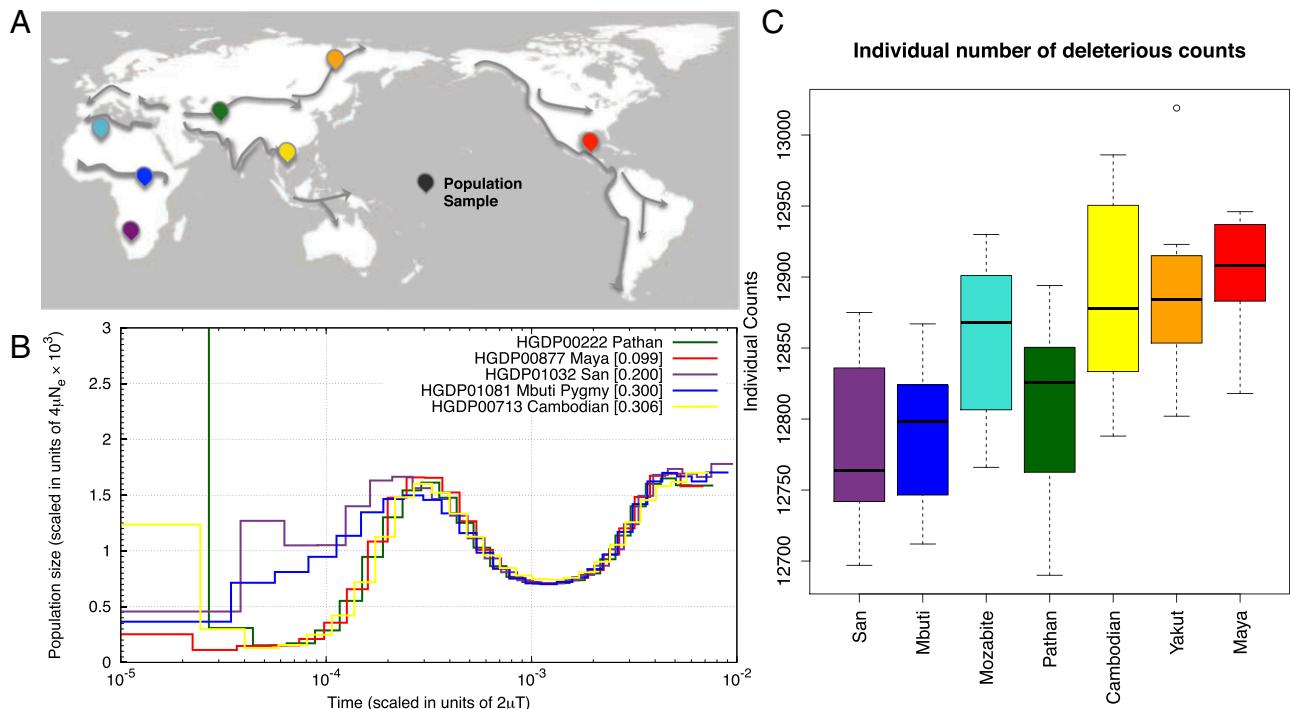
There remains substantial conceptual and empirical uncertainty surrounding the processes that shape the distribution of

deleterious variation across human populations. We aim here to clarify three aspects underlying this controversy: (i) Are there empirical differences in the total number of deleterious alleles among multiple human populations? (ii) Which model of dominance is appropriate for deleterious alleles (i.e., should zygosity be considered in load calculations)? (iii) Are the observed patterns consistent with predictions from models of range expansions accompanied by founder effects? We address these questions with a new genomic dataset of seven globally distributed human populations.

## Results

**Population History and Global Patterns of Genetic Diversity.** We obtained moderate coverage whole-genome sequence (median depth 7 $\times$ ) and high coverage exome sequence data (median depth 78 $\times$ ) from individuals from seven populations from the Human Genome Diversity Panel (HGDP) (20). Unrelated individuals (no relationship closer than first cousin) were selected from seven populations chosen to represent the spectrum of human genetic variation from throughout Africa and the OOA expansion, including individuals from the Namibian San, Mbuti Pygmy (Democratic Republic of Congo), Algerian Mozabite, Pakistani Pathan, Cambodian, Siberian Yakut, and Mexican Mayan populations (Fig. 1A). The 2.48-Gb full genome callset consisted of 14,776,723 single nucleotide autosomal variants, for which we could orient 97% to ancestral/derived allele status (*SI Appendix*).

Heterozygosity among the seven populations decreases with distance from southern Africa, consistent with an expansion of humans from that region (21). The Namibian San population carried the highest number of derived heterozygotes, ~2.39 million per sample, followed closely by the Mbuti Pygmies (*SI Appendix*, Table S1 and Fig. S5). The North African Mozabites



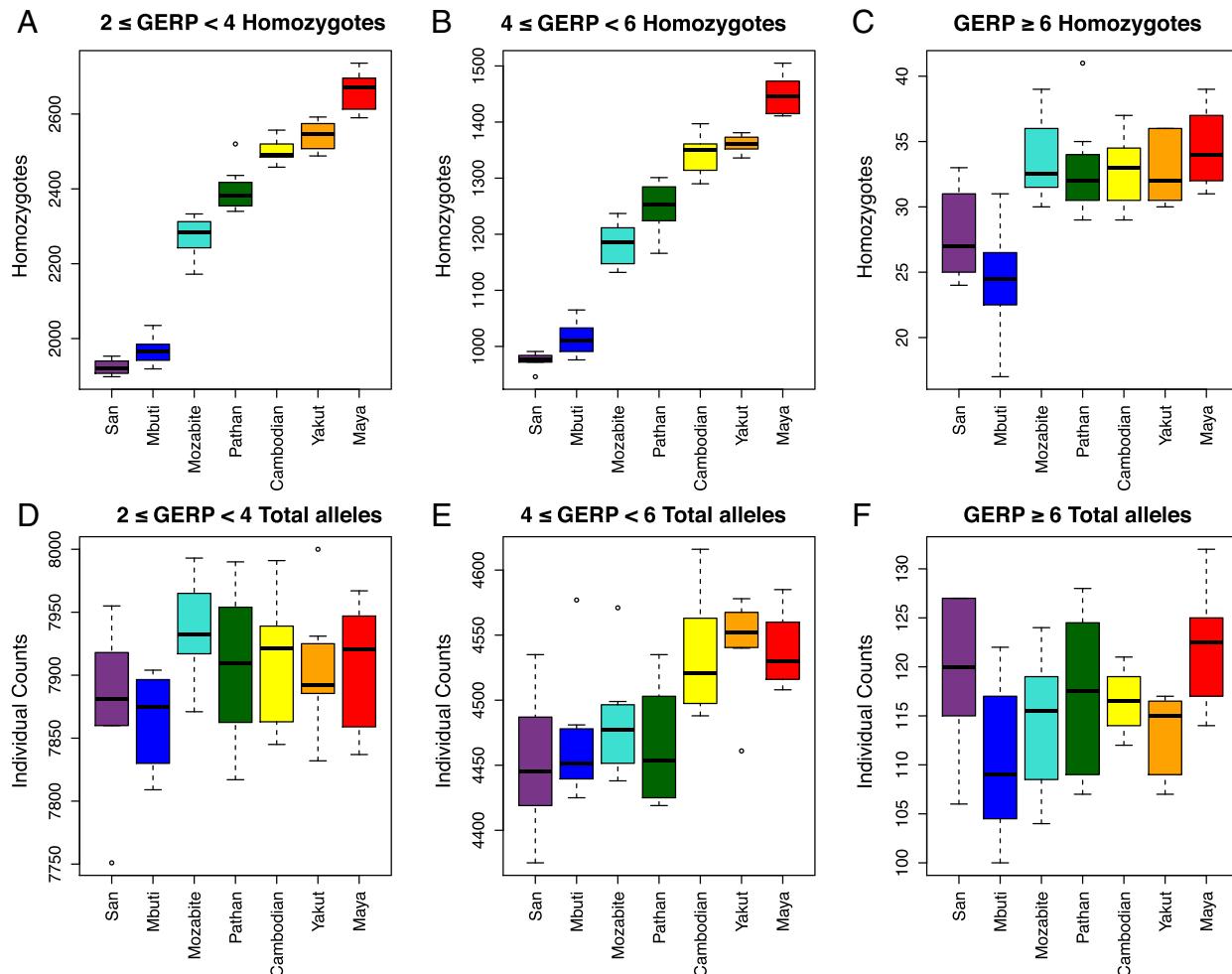
**Fig. 1.** Decrease in heterozygosity and estimated  $N_e$  with distance from southern Africa. (A) Locations of HGDP populations sampled for genome and exome sequencing are indicated on the map. Putative migration paths after the origin of modern humans are indicated with arrows (adapted from ref. 46). (B) PSMC curves for individual genomes, corrected for differences in coverage. Whereas populations experiencing an OOA bottleneck have substantially reduced  $N_e$ , African populations also display a reduction in  $N_e$  between ~100 kya and 30 kya (see *SI Appendix* for simulations of population history with resulting PSMC curves). (C) For each individual's exome, the number of putatively deleterious variants (equivalent to number of heterozygotes + twice the number of derived homozygotes) is shown by population.

PNAS

carry more heterozygotes than the OOA populations in our dataset (2 million) but substantially fewer than the sub-Saharan samples, likely reflecting a complex history of an OOA migration, followed by reentry into North Africa and subsequent recent gene flow with neighboring African populations (22). The Maya have the lowest median number of heterozygotes in our sample, ~1.5 million, which may be inflated due to recent European admixture (23). Two Mayan individuals displayed substantial recent European admixture (>20%) as assessed with local ancestry assignment (24) (*SI Appendix*, Fig. S6); these individuals were removed from analyses of deleterious variants. When we recalculated heterozygosity in the Maya, it was reduced by 3.5%. The decline in heterozygosity in OOA populations with distance from Africa strongly supports earlier results based on SNP array and microsatellite data for a serial founder effect model for the OOA dispersal (25, 26). We analyzed population history for individuals having sufficient coverage from five of the studied populations using the pairwise sequential Markovian coalescent software (PSMC) to estimate changes in  $N_e$  (11, 12, 27). Because dating demographic events with PSMC is dependent on both the assumed mutation rate and the precision with which a given event can be inferred, we compare relative bottleneck magnitudes and timing among the seven HGDP populations. Consistent with previous analyses (27), the OOA populations show a sharp reduction in  $N_e$ , with virtually identical population histories (Fig. 1B

and *SI Appendix*). Simulations indicate that the magnitude of the 12-fold bottleneck is accurately estimated (*SI Appendix*, Fig. S7), even if the time of the presumed bottleneck is difficult to estimate precisely using PSMC. Interestingly, both the Mbuti and the Namibian San show a moderate reduction in  $N_e$  relative to the ancestral maximum, with the San experiencing an almost twofold reduction in  $N_e$  and the Mbuti displaying a reduction intermediate between the San and OOA populations (see also refs. 20, 28, and 29). These patterns are consistent with multiple population histories (e.g., both short and long bottlenecks) and multiple demographic events, including a reduction in substructure from the ancestral human population rather than a bottleneck per se (27).

**Differences in Deleterious Alleles per Individual Genomes.** Owing to differences in coverage among the whole genome sequences, our subsequent analyses focus on the high-coverage exome dataset (78 $\times$  median coverage) to minimize any bias in comparing populations (*Materials and Methods*). We classified all mutations discovered in the exome dataset into categories based on Genomic Evolutionary Rate Profiling (GERP) Rejected Substitution (RS) scores. These conservation scores reflect various levels of constraint within a mammalian phylogeny (*Materials and Methods*) and are used to categorize mutations by their predicted deleterious effect (30, 31). Importantly, the allele present in the human reference genome was not used in the GERP RS calculation, avoiding the



**Fig. 2.** Individual counts of deleterious variants. (A–C) For each individual's exome, the number of derived homozygotes is plotted by population for moderate-, large-, and extreme-effect GERP categories. (D–F) For each individual's exome, the number of derived variants (equivalent to number of heterozygotes + twice the number of homozygotes) is plotted by population for moderate-, large-, and extreme-effect GERP categories.

reference-bias effect previously observed in other algorithms (11, 12) (*SI Appendix*, Fig. S8A). Variants were sorted into four groups reflecting the likely severity of mutational effects: “neutral” ( $-2 < \text{GERP} < 2$ ), “moderate” ( $2 \leq \text{GERP} < 4$ ), “large” ( $4 \leq \text{GERP} < 6$ ), and “extreme” ( $\text{GERP} \geq 6$ ) (*SI Appendix*, Fig. S9). GERP categories were concordant with ANNOVAR functional annotations (*SI Appendix*, Table S2 and Fig. S8B).

When considering the total number of derived alleles per individual, defined here as  $A_I = (1 \times \text{HET}) + (2 \times \text{HOM}_{\text{der}})$ , we observe an increase of predicted deleterious alleles with distance from Africa (Fig. 1C). The number of predicted deleterious alleles per individual increases along the range expansion axis (from San to Maya), consistent with theoretical predictions for expansion load (32). The maximal difference in the number of deleterious alleles between African and OOA individuals is  $\sim 150$  alleles. This result is consistent with theoretical predictions; the rate at which deleterious mutations accumulate in wave-front populations is limited by the total number of mutations occurring during the expansion (32). Assuming an exomic mutation rate of  $u = 0.5$  per haploid exome and an expansion that lasted for  $t = 1,000$  generations, a very conservative upper limit for the excess of deleterious alleles in OOA individuals would be  $2^*u^*t = 1,000$ . The cline in  $A_I$  is most pronounced for large-effect alleles ( $4 \leq \text{GERP} < 6$ , Fig. 2E), whereby the San individuals carry  $A_I = 4,450$  large-effect alleles on average, increasing gradually to 4,550 in Yakut. The Mayans carry slightly fewer large-effect mutations per individual than the Yakut, which may be influenced by the residual European ancestry (between 5–20%) in our sample. For extreme alleles ( $\text{GERP} \geq 6$ ), each individual in the dataset carries on average 110–120 predicted highly deleterious alleles with no significant differences among populations (Fig. 2F). The average additive GERP score—obtained by counting the GERP scores at homozygous sites twice—for all predicted deleterious variants per individual is lowest in the San ( $\sim 3.3$ ) and highest in the Maya ( $\sim 3.8$ ).

Similar patterns are found when we consider the number of derived homozygous sites per individual. We find that individuals from OOA populations exhibit significantly more homozygotes for moderate, large, and extreme variants than African populations (Fig. 2A–C). In addition, we observe a clear increase in the number of derived homozygotes with distance from Africa for moderate ( $2 \leq \text{GERP} < 4$ ) and large ( $4 \leq \text{GERP} < 6$ ) mutation effects categories, whereas the number of derived “extreme” homozygotes ( $\text{GERP} \geq 6$ ) is similar among OOA populations: All OOA genomes possess 30–40 extremely deleterious alleles in homozygous state (Fig. 2C). These patterns are in excellent agreement with theoretical predictions for the evolution of genetic variation during range expansions (7). The average GERP score per individual for derived homozygous variants is less differentiated than the additive model (above), varying between 2.43–2.49.

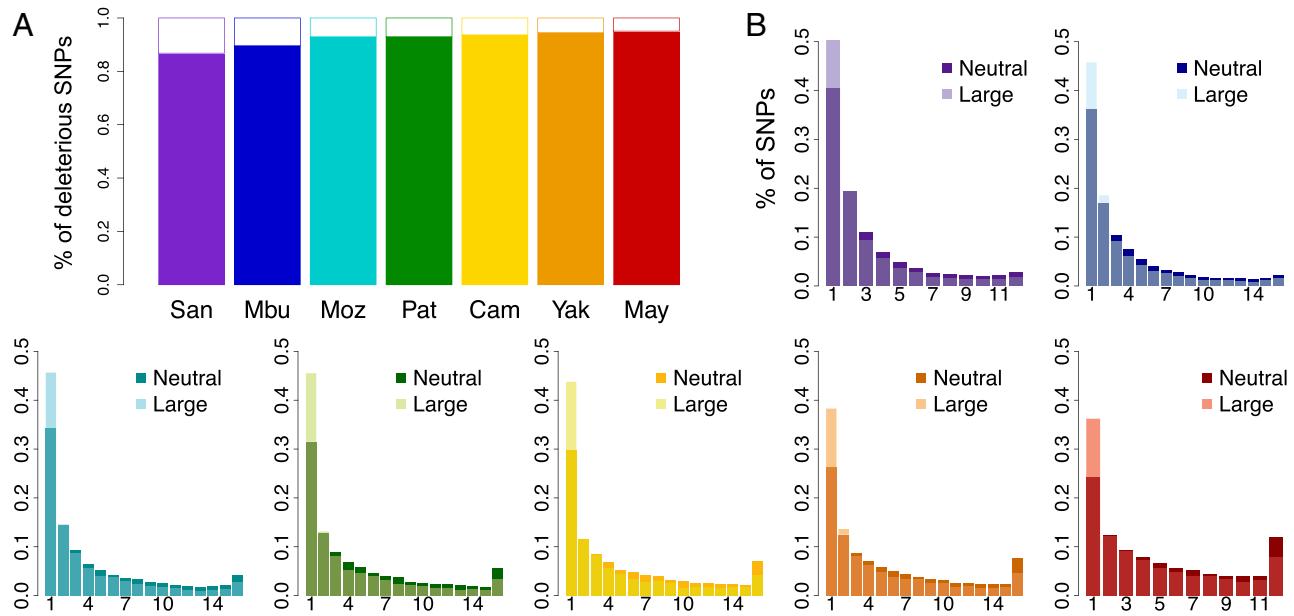
It is important to note that  $A_I$  is strongly influenced by common variants. Goode et al. (33) observed that as much as 90% of deleterious alleles in a single genome have a derived allele frequency greater than 5%, suggesting that the bulk of mutational burden using this metric will come from common variants. To explore this idea, we randomly chose an individual in each population and calculated the proportion of deleterious variants that are rare (<10%, i.e., a singleton within our population samples) and common (>10%), for each GERP category (Fig. 3A). Common deleterious alleles contribute to more than 90% of an individual’s  $A_I$ , and the proportion of common deleterious variants increases with distance from Africa, as can be seen by the decrease of rare deleterious variants. This includes common large-effect variants, which make up proportionally more of  $A_I$  for an OOA individual than for an African individual. For example, in a Mayan individual, 93% of large-effect variants are common compared with a San individual, where only 85% of large-effect variants are common (*SI Appendix*, Fig. S12). Given

the small number of chromosomes in each population ( $n = 14$ –16), estimates of allele frequencies are subject to sampling effects. We recently performed the same analysis on exome data from the 1000 Genome Phase 1 Project (34). We find a similar pattern as in our HGDP data: On a per-genome basis, common variants represent a majority of the alleles predicted to be deleterious (5).

**Differences in Deleterious Alleles at the Population Level.** To further elucidate the relationship between predicted mutation effect and allele frequencies, we compared the site frequency spectrum (SFS) for neutral and large- ( $4 \leq \text{GERP} < 6$ ) effect variants (Fig. 3B; see *SI Appendix*, Fig. S14 for a comparison between neutral and extreme variants). For all populations, singletons are enriched for deleterious variants (compared with neutral variants), consistent with the effect of purifying selection against deleterious variants (15, 35). However, the SFSs of OOA and African populations show marked differences. The neutral and deleterious SFSs of OOA populations show a global shift toward higher frequencies, consistent with the effects of serial bottlenecks/founder effects. It follows that OOA populations have fewer rare deleterious variants than Africans, as well as a larger proportion of fixed deleterious alleles; almost 7.9% of large-effect variants are fixed in the Maya, whereas the San have only 1.8% of deleterious variants fixed (Fig. 3B).

**Simulations of Purifying Selection Under a Range Expansion.** We sought to interpret the population-specific patterns of genetic diversity for each GERP category under a model including serial founder effects across geographic space and purifying selection. We simulated the evolution of both neutral and deleterious mutations under a simple model of range expansion in a 2D habitat (*SI Appendix*, Fig. S21). At selected loci, the ancestral allele was assumed selectively neutral and mutants reduced an individual’s fitness by a factor  $1 - s$  only if it was present in homozygous state, that is, deleterious mutations were assumed to be completely recessive. Three thousand generations (corresponding to about 75 kya) after the onset of the range expansion, we computed the average expected heterozygosity for all populations. Computational limitations of individual-based simulations prohibit a complete exploration of the parameter space for this model, but, by varying migration rates and selection coefficients, we identified parameter values that fit the observed clines in heterozygosity reasonably well (Fig. 4B). Specifically, we first identified selection coefficients that yield the same relative differences between observed neutral and selected heterozygosities (Fig. 4A). Then, the migration rate was adjusted to fit the observed clines in heterozygosities, assuming that the distance between two demes is 250 km (Fig. 4B). The fit selection coefficients were  $0, 1.25 \times 10^{-4}, 1 \times 10^{-3}$ , and  $2 \times 10^{-3}$  for neutral, moderate, large, and extreme GERP scores categories, respectively; the  $\text{GERP} \geq 6$  category showed the worst fit and observed counts indicate that even stronger selection coefficients should be considered for these extreme mutations (16). We performed the same analysis using a model in which mutations are codominant and, as expected, we found that the fit selection coefficients are smaller than those obtained a recessive model. These coefficients are estimated as  $s = 0, 0.5 \times 10^{-4}, 1.2 \times 10^{-4}$ , and  $2 \times 10^{-4}$ , respectively (*SI Appendix*, Fig. S16) (16).

**Evolutionary Forces Acting on Heterozygosity.** To better understand which evolutionary forces have acted in different populations to shape their levels of genetic diversity, we define a new statistic,  $RH$ .  $RH$  measures the reduction in heterozygosity at conserved sites relative to neutral heterozygosity,  $RH = (H_{\text{neu}} - H_{\text{del}})/H_{\text{neu}}$ , where  $H_{\text{neu}}$  indicates heterozygosity at neutral sites and  $H_{\text{del}}$  at GERP score categories  $> 2$ .  $RH$  can be seen as a way to quantify changes of functional diversity across populations relative to neutral expectations. For instance, a constant  $RH$  value across



**Fig. 3.** Differences in the proportion of deleterious alleles by frequency class. (A) The proportion of rare versus common deleterious variants per individual. For a given individual, deleterious variants were divided into common (>10%, solid colors) and rare (<10%, white space). The contribution of common deleterious variants to an individual's burden is much greater than rare variants. (B) For each population, we calculated the proportional site frequency spectrum by plotting the proportion of deleterious large-effect alleles in each frequency class (translucent coloring) along with the proportion of neutral alleles for each frequency class (opaque coloring). African populations have proportionally fewer rare deleterious alleles than expected from neutrality. Populations with OOA ancestry have proportionally more fixed deleterious mutations.

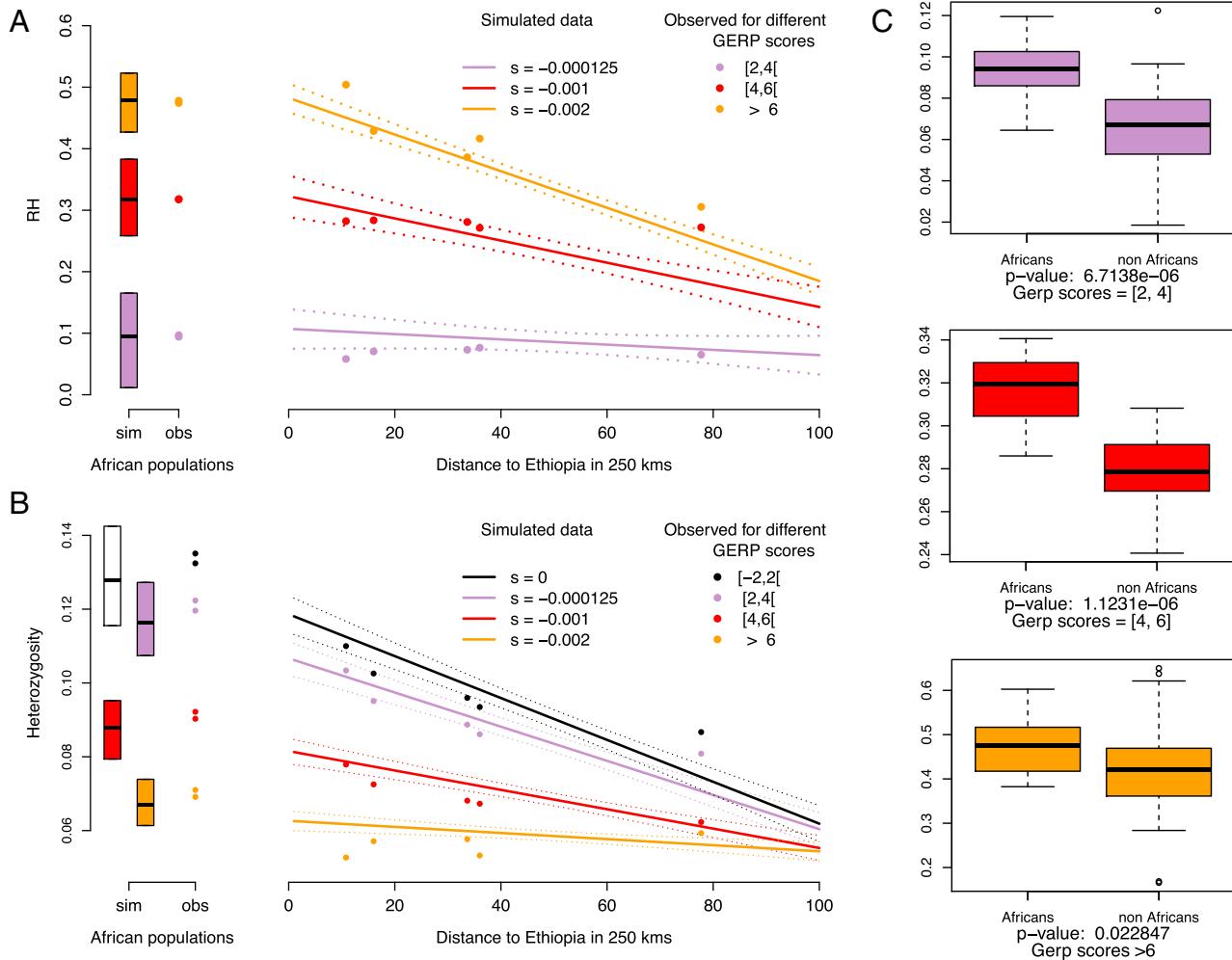
populations would suggest that average functional diversity is determined by the same evolutionary force(s) as neutral diversity, that is, genetic drift and migration. In contrast, if *RH* changes across populations, it suggests that different evolutionary forces have shaped neutral and functional diversity, that is, selection has changed functional allele frequencies.

In our dataset, *RH* is significantly larger in sub-Saharan Africans than in OOA populations across all functional GERP categories (Fig. 4C), indicating that selection has acted differently relative to drift between the two groups. The correlation between *RH* value and predicted mutation effect observed in Africa (Fig. 4A) confirms that purifying selection has kept strongly deleterious alleles at lower frequencies than in OOA populations. We then asked whether there were significant differences across OOA population, as oriented by their distance from eastern Africa. Interestingly, we see that the OOA *RH* values do not depend on their distance from Africa for predicted moderate-effect alleles ( $P = 0.82$ ; SI Appendix, Fig. S15), suggesting that the frequencies of moderate mutations have evolved mainly according to neutral demographic processes during the range expansion out of Africa. In contrast, for strongly deleterious variants (large and extreme GERP categories) we see a significant cline in *RH* ( $P = 0.01$  and  $P = 1.12 \times 10^{-6}$ , respectively; SI Appendix, Fig. S15), which implies that purifying selection has also contributed to their evolution relative to demographic processes.

**Models of Dominance.** We next considered whether there is empirical evidence for nonadditive effects for deleterious variants. Prior studies generally calculated “mutation load” by assuming an additive model, summing the number of deleterious alleles per individual, without factoring in whether an SNP occurs in a homozygous or heterozygous state. Determining an individual’s mutation load is, however, highly dependent on the underlying model of dominance (36) (a formal definition of mutation load is given below). For humans, Mendelian diseases tend to be overrepresented in endogamous populations or consanguineous pairings, indicating that many of these mutations are recessive

(37); Gao et al. (38) estimate 0.58 lethal recessive mutations per diploid genome in the Hutterite population. Gene conversion can also lead to differential burden of derived, recessive disease alleles among populations (39). Even height, a largely quantitative trait, seems to be affected by the architecture of recessive homozygous alleles in different populations (40).

To further clarify the impact of dominance, we compared the distribution of deleterious variants across genes associated with dominant or recessive disease as reported in Online Mendelian Inheritance in Man (OMIM) (41). We expect to see a lower proportion of large- and extreme-effect variants in genes with dominant OMIM mutation annotations, compared with genes with recessive OMIM mutation annotations. We tested this hypothesis with the HGDP as well as the much larger 1000 Genomes Phase 1 dataset (SI Appendix, Fig. S18B). We averaged the proportion of variants within each effect category and performed a Wilcoxon test to determine whether the distribution of the proportion of large-effect variants was different between dominant and recessive genes. In the HGDP dataset, we observed  $P = 0.06$ , and for the larger 1000 Genomes dataset,  $P = 0.03$ . Our results indeed show a significantly higher proportion of large-effect variants in genes with recessive annotations, compared with genes with dominant annotations, suggesting that deleterious variants in the genome may tend to be recessive. However, we caution that OMIM genes are here annotated as dominant or recessive, whereas dominance is a property of specific mutations, and therefore all deleterious variants in a gene will not necessarily have the same dominance coefficient. Nonetheless, our results are consistent with an interpretation that genes may have certain properties, for example negative selection against dominant mutations in crucial housekeeping or developmental genes, that influence the tolerable distribution of dominance among variants. We consider the effect of dominance (summarized by  $h$ , which measures the effect of selected mutations in heterozygotes relative to homozygotes) on mutation load in the HGDP population samples given the observed differences in heterozygosity.



**Fig. 4.** Heterozygosity under range expansion simulations with different selection coefficients. (A) Observed and simulated patterns of the reduction of heterozygosity (RH). Selection coefficients used in the simulations are  $s = 0$  (black),  $s = -0.000125$  (lavender),  $s = -0.001$  (red), and  $s = -0.002$  (orange). (B) Colored circles show average expected heterozygosity for populations with ancestry from the OOA bottleneck. Solid lines show the regression lines obtained from simulations and dashed lines indicate 95% confidence intervals for the regression. The boxplots and colored circles on the left show the simulated heterozygosities in ancestral (i.e., African) populations, and the observed heterozygosities in our African dataset (San/Mbuti), respectively. (C) Comparison of the distribution of RH between African and non-African individuals for different GERP categories, tested with a two-tailed Student *t* test (SI Appendix, Fig. S15).

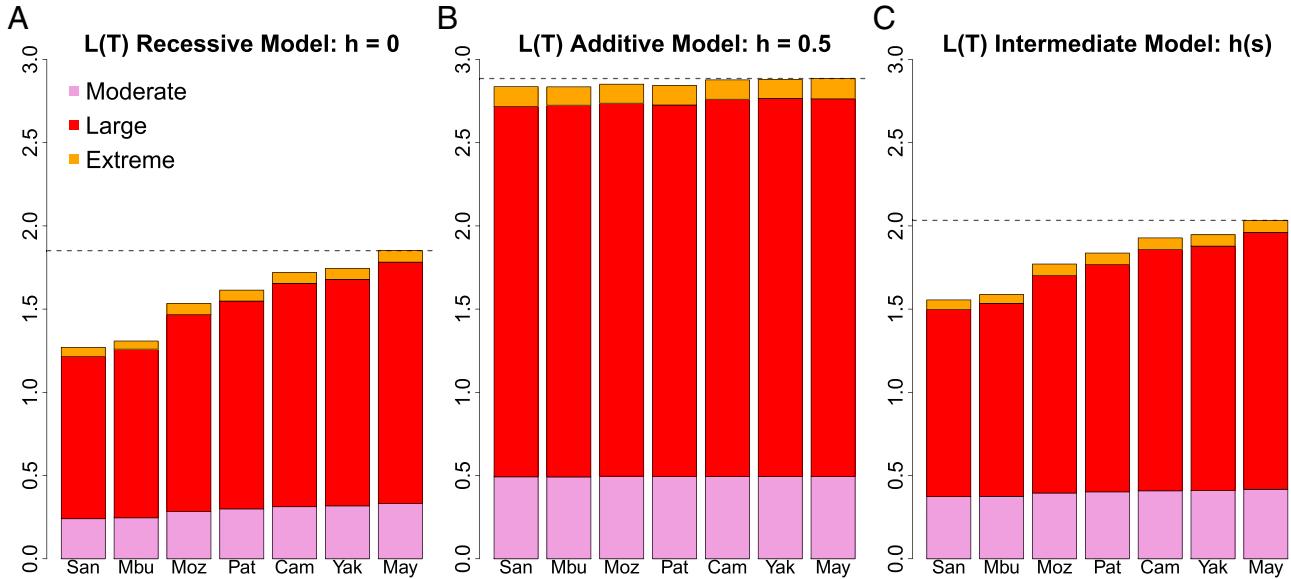
**Modeling the Burden of Deleterious Alleles.** We modeled three different scenarios to estimate the burden of deleterious alleles across populations. The relationship between fitness  $W$  and load for a given locus  $v$  is classically defined (36) as

$$L_v = 1 - W = 1 - (1 - W_{het} - W_{hom}).$$

$W_{het} = g_{Aa} \times (1 - hs)$  and  $W_{hom} = g_{aa} \times (1 - s)$ , where  $g_{Aa}$  and  $g_{aa}$  are the observed genotype frequencies of the heterozygotes and derived homozygotes, respectively. The estimated population load (ignoring epistasis) is the sum of the load for all variants:  $L_T = \sum_v L_v$ . For each variant we assigned the selection coefficient inferred by the range expansion simulations according to its GERP score [see also Henn et al. (5)]. Given that we do not know the distribution of dominance effects in human variation, we started by estimating the bounds for the mutation load for each population by considering two extreme scenarios: completely recessive and complete additive models for deleterious variants. We calculated  $L_T$  for each HGDP population (Fig. 5). When all mutations are considered strictly additive ( $h = 0.5$ ), values for mutation load are very similar across populations, with sub-Saharan African populations having the lowest mutational

load ( $L_T = 2.83$ ), followed by the Pathan and Mozabites, and finally the Asian and Native American populations showing the highest load ( $L_T = 2.89$ ) (Fig. 5B). We consider this model, as adopted in earlier studies, to demonstrate that even under an additive assumption there is a statistically significant 1.7% difference in the spectrum of load between populations (SI Appendix, Fig. S24). When all mutations are considered recessive ( $h = 0$ ), this model yields a much larger 45% difference in load ( $L_T$  ranges between 1.27 and 1.85) between the San and the Maya (Fig. 5A). Although this is surely an overestimate, it illustrates the broad range of potential values and consistent signal in the data for differences among populations in estimated load. The mutation load under a recessive model is not explained by inbreeding, as measured by the cumulative amount of the genome in runs of homozygosity (cROH) greater than 1 Mb ( $r = 0.27$ ,  $P = 0.55$ ) (SI Appendix, Fig. S25); this is because the African hunter-gatherers have relatively high cROH compared with other global populations, as is commonly observed in small endogamous populations (21, 42).

For the third scenario we used a model based on studies of dominance in yeast and *Drosophila* (19, 43, 44), in which there is an inverse relationship between selection and dominance (highly deleterious mutations tend to be recessive), and where  $h$  is



**Fig. 5.** Estimates of mutational load in seven populations as a function of dominance assumptions. Total mutation load was summed over all annotated mutations in the exome dataset for the observed heterozygote and derived homozygote genotype frequencies in each population. The cumulative mutational load is shown in increased order from neutral to extremely deleterious mutations. Strongly deleterious mutations contribute the most to mutational load. Mutations were assigned an  $s$ , selection coefficient, based on their GERP score. (A)  $h = 0$ , recessive model; (B)  $h = 0.05$ , additive model; (C)  $h(s)$ , intermediate dominance model. For each selection coefficient, an  $h$  dominance coefficient was assigned based on the inverse relationship between  $s$  and  $h$ .

sampled from a distribution following Agrawal and Whitlock (19). The maximal difference in load under this model was 30.8% (Fig. 5C), again between the San and Maya, and the minimum difference in load was 1%, between the Cambodians and Yakut. We note that the difference in relative fitness [ $e^{-L(T)}$ ] is much less than the difference in mutation load (i.e., a relative reduction of 79% in the San versus 87% in the Maya translates to a 8% difference between the two populations under the  $h(s)$  model; see also *Discussion*). As in the other modeled dominance scenarios, the majority of calculated mutational load is contributed by the large-effect mutational category, because this category has a relatively strong selection coefficient and thousands of mutations (>4,000 on average per individual). Thus, this category contributes proportionally more to the total load, even though the extreme-effect mutations have a higher selection coefficient. We note, however, that our assumed selection coefficients, particularly for the extreme effect, are somewhat lower than those obtained by other distribution of fitness effect studies (16, 45) and simulations under an additive model results in even smaller selection coefficients (discussed above). Because selection coefficients are the same across populations in our calculations,  $s$  will affect the absolute value of load but not relative differences across populations.

## Discussion

Two primary demographic signals are reflected in human genetic data from non-African populations. First, a major 5- to 10-fold population bottleneck is associated with the OOA dispersal(s) (46–48). Second, the distribution of genetic diversity among non-African populations is characterized by a decrease in heterozygosity proportional to geographic distance from northeastern Africa. A model of serial founder effects in the ancestral populations of Eurasia, Oceania, and the Americas has been posited as the most likely model for explaining the systematic variation in genetic diversity across this geographic range for humans (25, 26), as well as commensal human species (49, 50). By directly ascertaining genomic variation in over 50 individuals from seven populations, we observe a clear cline of genetic diversity as a function of distance from Africa, supporting evidence for a serial founder effect model. We also observe differences in the amount

of predicted deleterious variation across populations. These differences seem to result from the genetic drift of existing deleterious variants to higher frequencies during the sequential range expansion after the OOA exit (Fig. 3B). Clines in heterozygosity for the different mutational effect categories can be reproduced by spatially explicit simulations with negative selection and recessive mutations (Fig. 4; see also codominant simulations in *SI Appendix*, Fig. S16). Although both moderate- and large-effect deleterious mutations have evolved under negative selection in Africa (Fig. 4C and *SI Appendix*, Fig. S15), many predicted moderate variants have evolved as if they were neutral in non-African populations. However, selection has remained a major force during the OOA expansion for strongly deleterious variants.

**Impact of the OOA Bottleneck.** There is an ongoing debate on whether selection has been equally or more efficient in African versus non-African populations due to the major bottleneck that occurred in the ancestors of OOA populations (10, 12, 13, 35). Two studies found no significant differences in mutation load between European Americans and African Americans under an additive model with two classes of alleles: deleterious and neutral (12, 13, 33). Fu et al. (11) identified small but significant differences in the average number of alleles and the SFS, potentially due to a different algorithm for predicting mutation effect than earlier studies. We argue that estimates of the efficacy of selection should take into account not only the number of mutations per individual but also the predicted severity of mutational effect. Here, we classify mutations into four categories and find differences across populations in some, but not all, mutational categories. For variants that have putatively moderate ( $2 \leq \text{GERP} < 4$ ) or extreme deleterious effect ( $\text{GERP} \geq 6$ ), we do not see a significant difference between African and non-African populations in the number of mutations per individual. Significant per-individual differences are only observed for the intermediate large-effect category. We used PhyloP scores (51) as an alternative measure of conservation to verify our main results (*SI Appendix*, Fig. S26). We found qualitatively very similar patterns for both the spatial distribution of the number of

derived homozygous sites per individual (*SI Appendix*, Fig. S26A) as well as the number of derived alleles per individual, suggesting that our results are robust to the choice of prediction algorithm that is used to estimate deleteriousness of mutations.

We note that the observed differences between populations are relatively small compared with the within-population variance (Fig. 2). Nonetheless, a novel measure of the efficacy of selection,  $RH$ , is significantly different across all three mutational categories (Fig. 4C and *SI Appendix*, Fig. S15) between sub-Saharan Africans and non-Africans in our dataset. That is, the observed heterozygosity at deleterious loci is greater in non-Africans than in Africans—after correcting for neutral genetic diversity in each group. This is particularly significant for moderate- and large-effect mutations, in agreement with theory that would suggest that differences in purifying selection will primarily emerge for variants at the  $N_s$  boundary.

**Serial Founder Effects/Range Expansion.** Several simulation studies have attempted to characterize the distribution of deleterious alleles under OOA demographic scenarios. Some simulations focused on differences in the cumulative number of deleterious alleles per individual; others focused on differences in the proportion of segregating alleles within a population that are deleterious. Lohmueller et al. (10) found that a long bottleneck lasting more than 7,500 generations (>150,000 y) could produce the excess proportion of deleterious mutations observed in European Americans. A bottleneck model with subsequent explosive growth has also been proposed to explain the proportionally greater number of nonsynonymous or deleterious mutations in Eurasian populations (52, 53). As a consequence, deleterious mutations accumulate in populations during the expansion process. Simons et al. (12) tested a long bottleneck and subsequent population expansion model contrasting African and non-African populations and found no evidence that human demography played a role in the differential accumulation of deleterious alleles per individual.

A recent theoretical study of spatial range expansions (i.e., a model similar to geographic serial founder effects) showed that strong genetic drift at the wave front of expanding populations decreases the efficiency of selection (32). Under a spatial range expansion model, deleterious variants, unless they have a large selection coefficient, should evolve as if they are neutral on the wave front (32), and their overall frequency should therefore not change much during the range expansion (7). The loss of deleterious variants at some loci should be compensated by an increase of their frequencies at other loci. The frequency of deleterious homozygotes should therefore increase with distance from Africa, which is observed here in the rightward shift of the SFS in OOA populations (Fig. 3), except for the most evolutionarily constrained sites. We can address the question of whether this increased frequency is driven entirely by drift and gene surfing or by differential selection in non-African populations by considering the spatial distribution of the  $RH$  statistic (Fig. 4C). The fact that  $RH$  does not change among OOA populations for moderately deleterious alleles suggests that they have evolved as if they were neutral alleles during the expansion and that selection has not yet purged the deleterious mutations that increased in frequency. In contrast, extremely deleterious alleles ( $GERP \geq 6$ ) exhibit similar heterozygosity in all OOA populations, suggesting that they are subject to similar levels of purifying selection in these populations. The remaining deleterious alleles ( $4 \leq GERP < 6$ ) present an intermediate pattern, implying that both drift and selection have acted on this category of sites.

A recent controversy concerns whether there are differences in the efficacy of purifying selection between African and non-African populations (6, 12, 13). It is difficult to discuss our results in the context of this controversy because there is no generally accepted definition of “efficacy of selection,” and different

definitions will lead to different interpretations (4). We therefore prefer to interpret our results in the context of our spatially explicit model of range expansions, and the relative roles of drift and selection in this model. Recurrent founder events should contribute to a decrease in the effective population sizes with distance from Africa, and it is commonly assumed that selection will become weaker with smaller effective population sizes. However, reducing the impact of a range expansion to a simple gradient in effective size, and thus to a decrease of the efficacy of selection, can be misleading. Diversity-based estimates of  $N_e$  are not necessarily informative about the strength of selection in nonequilibrium scenarios because estimates of  $N_e$  may lag behind recent demographic changes (e.g., ref. 54). Rather, if one considers that deleterious alleles were kept at low frequencies by purifying selection in ancestral African populations, those that increased in frequency by gene surfing during the OOA expansion also became more accessible to subsequent selection, especially for those alleles that were recessive. The observed cline in  $RH$  for large-effect mutations is more compatible with an unequal purging of deleterious variants by selection. Indeed, selection will have had less time to act on newly formed populations that are further away from Africa, and it will also operate more slowly on populations that have less diversity and therefore lower interindividual differences in fitness. Furthermore, the fact that our simulations can reproduce the observed pattern with spatially uniform population sizes and strength of selection against deleterious mutations implies that the simulated gradients in  $RH$  in Fig. 4A, as well as the increased number of deleterious homozygous sites, is not the consequence of reduced strength of selection away from Africa. Rather, it is caused by increased drift during the expansion, as well as by differential purging of deleterious mutations after the expansion.

**The Importance of Dominance.** Multiple modeling assumptions are crucial when considering the burden of deleterious alleles across populations. In addition to the selection coefficients, the assumed dominance terms are critical. An estimated 16% of Mendelian diseases are known to be autosomal recessive (estimated from the OMIM) and many contribute significantly to infant mortality. Owing to the difficulty of detecting recessive diseases, unless they are extremely damaging, there are potentially many more disease mutations that have an  $h$  coefficient less than 0.5. Autosomal recessive diseases seem to be more frequent than autosomal dominant diseases (55), and even mildly deleterious mutations are predicted to have a mean  $h$  of 0.25 (56). Although formal calculations of genetic load require multiple assumptions, we demonstrate that differences in calculated load across human populations are primarily sensitive to assumptions about dominance, as expected given the increased extent of homozygosity in OOA populations.

We have modeled deleterious mutations as having variable  $h$  coefficients. Whereas strongly deleterious mutations are likely recessive, dominance for weakly deleterious mutations is particularly problematic to estimate because there is less power to measure weak effects and  $h$  may be upwardly biased in model organism competition experiments (19). When sampling  $h$  coefficients under our model, we allowed weakly deleterious mutations to be assigned a coefficient  $h > 0.5$ , but this had little effect on mutational load because the bulk of the load was contributed by large-effect variants. However, a fraction of strongly deleterious mutations are clearly dominant, as ascertained from disease studies, and future work may need to model different mixtures distributions on  $h$ . We also note that the absolute mutational load is twofold higher under an additive model than under a recessive model (Fig. 5), as expected from theory (36).

**Estimates of Mutational Load.** We estimate that there are differences in mutational burden calculated using a formal load model, among extant human populations, particularly if we

depart from a simple additive assumption. We found that the change in mutation load between sub-Saharan African populations versus Native American populations (the two ends of the range) were significantly different at  $P < 0.05$  under recessive, partially recessive, and additive models (*SI Appendix*, Fig. S24). Mutational load under a fully or partially recessive model is 10 to 30% greater in non-African populations (Fig. 5A), as the result of higher homozygosity from the legacy of the OOA bottleneck across all (deleterious) mutation categories [e.g.,  $L_{T(Mbuti)} = 1.59$  and  $L_{T(Yakut)} = 1.95$  under the  $h(s)$  model]. All populations carry significant load, relative to a population with the alternate, ancestral allele genotype. Under a model where fitness differences are determined only by genotype and environments are equal across individuals, the relative fitness [ $e^{-L(T)}$ ] of 0.204 for the Mbuti indicates a reduction in fitness of 79.6%, whereas a relative fitness of 0.142 for the Yakut indicates an 85.7% reduction. These fitness differences are relatively small, even under a partially recessive model.

Although illustrative, such models of load have important limitations. The mutations identified in this dataset have not been functionally characterized and are predicted to be deleterious based on degree of sequence conservation. The assumed selective coefficients across GERP categories are fit based on a recessive model, which is not applicable to all sites. However, although different selection coefficients will change the values of load in our calculation, it will not change the relative difference among populations because the same set of coefficients were applied to all populations (5). If mutations have different fitness effects across heterogeneous global environments, then the values of mutation load will change. Indeed, a proportion of the alleles may be locally adaptive, or neutral, and hence the sign of the selection coefficient for the mutation would be miscalculated in our analysis. For example, the Duffy null allele is classified as a large-effect mutation using GERP (RS = 4.27) and is found at high frequency in western Africa; however, it has likely increased in frequency due to positive selection as a response to malaria (57). Recent genome-wide studies have stressed the paucity of selective sweeps in the human genome (35, 58, 59); only 0.5% of nonsynonymous mutations in 1000 Genomes Pilot Project were identified as having undergone positive selection. Others have emphasized evidence for pervasive adaptive selection (60, 61) and a variety of studies have identified specific beneficial alleles locally adapted to high altitude, immune response, and pigmentation (62–64). We considered local adaptive evolution by examining highly differentiated alleles in our dataset, that is, alleles that differ by 80% in frequency between a pair of populations, indicative of a strong local adaptation. We find that highly differentiated alleles have the same GERP score distribution as nondifferentiated alleles, indicating there is little reason to believe that most large- and extreme-effect mutations have been subjected to strong local adaptation (*SI Appendix*, Fig. S20; also see ref. 65). We conclude that the raw, calculated mutational burden may differ across human populations, although the effects of positive selection, varying environments, and epistasis have yet to be explored and remain a significant challenge to fully understanding mutational burden.

**Conclusions.** A major difference between our work and previous results is the interpretative framework we present, which underlines the role of range expansions out of Africa to explain patterns of neutral and functional diversity. Whereas previous comparisons between African and non-African diversity attributed the observed increased proportion of deleterious variants in non-Africans to the OOA bottleneck (10), our study shows that a single bottleneck is not sufficient to reproduce the gradient we observe in the number of deleterious alleles per individual with distance from Africa (Fig. 2). Taking into account the range expansion of modern humans (66) sheds new light on this apparent controversy. Finally, we note that recent simulation work

(4) suggests that the impact of a bottleneck on the efficacy of natural selection depends critically on the distribution of fitness and dominance effects as well as postbottleneck demographic history. Although these models and parameter choices clearly affect the interpretation of the pattern of deleterious alleles across populations, we find empirical evidence for significant differences in deleterious alleles as tabulated by a variety of statistics across the spectrum of human genetic diversity.

## Materials and Methods

**Samples and Data.** Aliquots of DNA isolated from cultured lymphoblastoid cell lines were obtained from Centre d'Étude du Polymorphisme and prepared for both full genome sequencing on Illumina HiSeq technology and exome capture with an Agilent SureSelect 44Mb array. One hundred one base pair read-pairs were mapped onto the human genome reference (GRCh37) using a mapping and variant calling pipeline designed to effectively manage massive amounts of short-read data. This pipeline followed many of the best practices developed by the 1000 Genomes Project Consortium (34).

**Variant Annotation.** Ancestral state was inferred based on orthologous regions in a great ape and rhesus macaque phylogeny as reported by Ensembl Compara and used by the 1000 Genomes Project. To determine the biological impact of a variant we used GERP score (30) as a measure of conservation across a phylogeny. Positive scores reflect a site showing a high degree of conservation, based on the inferred number of "rejected substitutions" across the phylogeny. GERP scores were obtained from the University of California, Santa Cruz genome browser ([hgdownload.cse.ucsc.edu/gbdb/hg19/bbi/All\\_hg19\\_RS.bw](http://hgdownload.cse.ucsc.edu/gbdb/hg19/bbi/All_hg19_RS.bw)) based on an alignment of 35 mammals to human. The allele represented in the human hg19 sequence was not included in the calculation of GERP RS scores. The human reference sequence was excluded from the alignment for the calculation of both the neutral rate and site-specific "observed" rate for the RS score to prevent any bias in the estimates. In addition to GERP, we also used PhyloP scores (51) as measures of genomic constraint during the evolution of mammals. We used the PhyloP<sub>NH</sub> scores computed in Fu et al. (11) from the 36 eutherian-mammal EPO alignments [available in Ensembl release 70 (67)], which is also computed without using the human reference sequence.

**Classification of Mutation Effects by GERP Scores.** Variants were classified as being neutral, moderate, large, or extreme for GERP scores with ranges [-2,2], [2,4], [4,6], and [6,max], respectively. The use of four "bins" of GERP scores simplifies the range expansion simulations performed for distinct selection coefficients. For every individual the total number of derived deleterious counts found in homozygosity (i.e., 2 × HOM), and the total number of deleterious counts [i.e., HET + (2 × HOM)] within each category was recorded.

**Individual-Based Simulations.** To simulate changes in heterozygosity, we modeled human range expansion across an array of  $10 \times 100$  demes (32). After reaching migration-selection-drift equilibrium, populations expand into the empty territory, which is separated from the ancestral population by a geographical barrier, through a spatial bottleneck (*SI Appendix*, Fig. S21). After 3,000 generations, we computed the average expected heterozygosity for all populations. The migration rate and selection coefficients were adjusted to generate heterozygosity consistent with the observed data, without formally maximizing the fit. The code used for simulations can be downloaded from <https://github.com/CMPG/ADMRE>.

**Calculating Load.** Mutational load was calculated following Kimura et al. (36), but using observed genotype frequencies instead of inferring them from Hardy-Weinberg based on the allele frequencies. In this way, the fitness of the heterozygotes and the homozygotes will be  $W_{het} = Aa \times (1 - hs)$  and  $W_{hom} = aa \times (1 - s)$ , where  $Aa$  and  $aa$  are the genotype frequencies of the heterozygotes and derived homozygotes, respectively. The fitness for a given variant will be relative to that of the ancestral variant, which for numerical convenience is set to 1. The relationship between fitness and load is  $L_v = 1 - W_v = 1 - (1 - W_{het} - W_{hom})$ , and the total population load is the sum of the load for all variants,  $L_T = \sum_v L_v$ .

**ACKNOWLEDGMENTS.** We thank Chris Tyler-Smith, David Reich, Yuval Simons, Spencer Koury, and Simon Gravel for helpful discussion. L.R.B. was supported by a Beatriu de Pinós Programme Fellowship. This work was supported by NIH Grants 3R01HG003229 (to C.D.B. and B.M.H.) and DP5OD009154 (to J.M.K.). S.P. and I.D. were supported by Swiss SNSF Grant 31003A-143393 (to L.E.).

- Morton NE, Crow JF, Muller HJ (1956) An estimate of the mutational damage in man from data on consanguineous marriages. *Proc Natl Acad Sci USA* 42(11):855–863.
- Tabor HK, et al.; NHLBI Exome Sequencing Project (2014) Pathogenic variants for Mendelian and complex traits in exomes of 6,517 European and African Americans: Implications for the return of incidental results. *Am J Hum Genet* 95(2):183–193.
- Ohta T (1973) Slightly deleterious mutant substitutions in evolution. *Nature* 246(5428):96–98.
- Gravel S (2014) When is selection effective? *bioRxiv*, dx.doi.org/10.1101/010934.
- Henn BM, Botigué LR, Bustamante CD, Clark AG, Gravel S (2015) Estimating the mutation load in human genomes. *Nat Rev Genet* 16(6):333–343.
- Lohmueller KE (2014) The distribution of deleterious genetic variation in human populations. *Curr Opin Genet Dev* 29:139–146.
- Peischl S, Excoffier L (2015) Expansion load: Recessive mutations and the role of standing genetic variation. *Mol Ecol* 24(9):2084–2094.
- Casals F, et al. (2013) Whole-exome sequencing reveals a rapid change in the frequency of rare functional variants in a founding population of humans. *PLoS Genet* 9(9):e1003815.
- Kehdy FSG, et al.; Brazilian EPIGEN Project Consortium (2015) Origin and dynamics of admixture in Brazilians and its effect on the pattern of deleterious mutations. *Proc Natl Acad Sci USA* 112(28):8696–8701.
- Lohmueller KE, et al. (2008) Proportionally more deleterious genetic variation in European than in African populations. *Nature* 451(7181):994–997.
- Fu W, Gittelman RM, Bamshad MJ, Akey JM (2014) Characteristics of neutral and deleterious protein-coding variation among individuals and populations. *Am J Hum Genet* 95(4):421–436.
- Simons YB, Turchin MC, Pritchard JK, Sella G (2014) The deleterious mutation load is insensitive to recent population history. *Nat Genet* 46(3):220–224.
- Do R, et al. (2015) No evidence that selection has been less effective at removing deleterious mutations in Europeans than in Africans. *Nat Genet* 47(2):126–131.
- Fu W, et al.; NHLBI Exome Sequencing Project (2013) Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nature* 493(7431):216–220.
- Tennessen JA, et al.; Broad GO; Seattle GO; NHLBI Exome Sequencing Project (2012) Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* 337(6090):64–69.
- Boyko AR, et al. (2008) Assessing the evolutionary impact of amino acid mutations in the human genome. *PLoS Genet* 4(5):e1000083.
- Bittles AH, Black ML (2010) Evolution in health and medicine Sackler colloquium: Consanguinity, human evolution, and complex diseases. *Proc Natl Acad Sci USA* 107(Suppl 1):1779–1786.
- Slatkin M (2004) A population-genetic test of founder effects and implications for Ashkenazi Jewish diseases. *Am J Hum Genet* 75(2):282–293.
- Agrawal AF, Whitlock MC (2011) Inferences about the distribution of dominance drawn from yeast gene knockout data. *Genetics* 187(2):553–566.
- Cann HM, et al. (2002) A human genome diversity cell line panel. *Science* 296(5566):261–262.
- Henn BM, et al. (2011) Hunter-gatherer genomic diversity suggests a southern African origin for modern humans. *Proc Natl Acad Sci USA* 108(13):5154–5162.
- Henn BM, et al. (2012) Genomic ancestry of North Africans supports back-to-Africa migrations. *PLoS Genet* 8(1):e1002397.
- Wang S, et al. (2007) Genetic variation and population structure in native Americans. *PLoS Genet* 3(11):e185.
- Maples BK, Gravel S, Kenny EE, Bustamante CD (2013) RFMix: A discriminative modeling approach for rapid and robust local-ancestry inference. *Am J Hum Genet* 93(2):278–288.
- Ramachandran S, et al. (2005) Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. *Proc Natl Acad Sci USA* 102(44):15942–15947.
- Prugnolle F, Manica A, Balloux F (2005) Geography predicts neutral genetic diversity of human populations. *Curr Biol* 15(5):R159–R160.
- Li H, Durbin R (2011) Inference of human population history from individual whole-genome sequences. *Nature* 475(7357):493–496.
- Meyer M, et al. (2012) A high-coverage genome sequence from an archaic Denisovan individual. *Science* 338(6104):222–226.
- Kidd JM, et al. (2012) Population genetic inference from personal genome data: Impact of ancestry and admixture on human genomic variation. *Am J Hum Genet* 91(4):660–671.
- Cooper GM, et al.; NISC Comparative Sequencing Program (2005) Distribution and intensity of constraint in mammalian genomic sequence. *Genome Res* 15(7):901–913.
- Cooper GM, et al. (2010) Single-nucleotide evolutionary constraint scores highlight disease-causing mutations. *Nat Methods* 7(4):250–251.
- Peischl S, Dupanloup I, Kirkpatrick M, Excoffier L (2013) On the accumulation of deleterious mutations during range expansions. *Mol Ecol* 22(24):5972–5982.
- Goode DL, et al. (2010) Evolutionary constraint facilitates interpretation of genetic variation in resequenced human genomes. *Genome Res* 20(3):301–310.
- 1000 Genomes Project Consortium (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature* 491:56–65.
- Lohmueller KE, et al. (2011) Natural selection affects multiple aspects of genetic variation at putatively neutral sites across the human genome. *PLoS Genet* 7(10):e1002326.
- Kimura M, Maruyama T, Crow JF (1963) The mutation load in small populations. *Genetics* 48:1303–1312.
- Reich DE, Lander ES (2001) On the allelic spectrum of human disease. *Trends Genet* 17(9):502–510.
- Gao Z, Waggoner D, Stephens M, Ober C, Przeworski M (2015) An estimate of the average number of recessive lethal mutations carried by humans. *Genetics* 199(4):1243–1254.
- Lachance J, Tishkoff SA (2014) Biased gene conversion skews allele frequencies in human populations, increasing the disease burden of recessive alleles. *Am J Hum Genet* 95(4):408–420.
- McQuillan R, et al.; ROHgen Consortium (2012) Evidence of inbreeding depression on human height. *PLoS Genet* 8(7):e1002655.
- Hamosh A, Scott AF, Amberger JS, Bocchini CA, McKusick VA (2005) Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res* 33(Database issue):D514–D517.
- Henn BM, et al. (2012) Cryptic distant relatives are common in both isolated and cosmopolitan genetic samples. *PLoS One* 7(4):e34267.
- Mukai T, Chigusa SI, Mettler LE, Crow JF (1972) Mutation rate and dominance of genes affecting viability in *Drosophila melanogaster*. *Genetics* 72(2):335–355.
- Houle D, Hughes KA, Assimacopoulos S, Charlesworth B (1997) The effects of spontaneous mutation on quantitative traits. II. Dominance of mutations with effects on life-history traits. *Genet Res* 70(1):27–34.
- Racimo F, Schraiber JG (2014) Approximation to the distribution of fitness effects across functional categories in human segregating polymorphisms. *PLoS Genet* 10(11):e1004697.
- Henn BM, Cavalli-Sforza LL, Feldman MW (2012) The great human expansion. *Proc Natl Acad Sci USA* 109(44):17758–17764.
- Laval G, Patin E, Barreiro LB, Quintana-Murci L (2010) Formulating a historical and demographic model of recent human evolution based on resequencing data from noncoding regions. *PLoS One* 5(4):e10284.
- Marth GT, Czabarka E, Murvai J, Sherry ST (2004) The allele frequency spectrum in genome-wide human variation data reveals signals of differential demographic history in three large world populations. *Genetics* 166(1):351–372.
- Tabane K, et al. (2010) Plasmodium falciparum accompanied the human expansion out of Africa. *Curr Biol* 20(14):1283–1289.
- Linz B, et al. (2007) An African origin for the intimate association between humans and *Helicobacter pylori*. *Nature* 445(7130):915–918.
- Pollard KS, Hubisz MJ, Rosenblloom KR, Siepel A (2010) Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res* 20(1):110–121.
- Keinan A, Clark AG (2012) Recent explosive human population growth has resulted in an excess of rare genetic variants. *Science* 336(6082):740–743.
- Lohmueller KE (2014) The impact of population demography and selection on the genetic architecture of complex traits. *PLoS Genet* 10(5):e1004379.
- Pennings PS, Kryazhimskiy S, Wakeley J (2014) Loss and recovery of genetic diversity in adapting populations of HIV. *PLoS Genet* 10(1):e1004000.
- Erickson RP, Mitchison NA (2014) The low frequency of recessive disease: Insights from ENU mutagenesis, severity of disease phenotype, GWAS associations, and demography: An analytical review. *J Appl Genet* 55(3):319–327.
- Manna F, Martin G, Lenormand T (2011) Fitness landscapes: An alternative theory for the dominance of mutation. *Genetics* 189(3):923–937.
- Sabeti PC, et al.; International HapMap Consortium (2007) Genome-wide detection and characterization of positive selection in human populations. *Nature* 449(7164):913–918.
- Hernandez RD, et al.; 1000 Genomes Project (2011) Classic selective sweeps were rare in recent human evolution. *Science* 331(6019):920–924.
- Granka JM, et al. (2012) Limited evidence for classic selective sweeps in African populations. *Genetics* 192(3):1049–1064.
- Enard D, Messer PW, Petrov DA (2014) Genome-wide signals of positive selection in human evolution. *Genome Res* 24(6):885–895.
- Grossman SR, et al.; 1000 Genomes Project (2013) Identifying recent adaptations in large-scale genomic data. *Cell* 152(4):703–713.
- Yi X, et al. (2010) Sequencing of 50 human exomes reveals adaptation to high altitude. *Science* 329(5987):75–78.
- Pickrell JK, et al. (2009) Signals of recent positive selection in a worldwide sample of human populations. *Genome Res* 19(5):826–837.
- Scheinfeldt LB, Tishkoff SA (2013) Recent human adaptation: Genomic approaches, interpretation and insights. *Nat Rev Genet* 14(10):692–702.
- Coop G, et al. (2009) The role of geography in human adaptation. *PLoS Genet* 5(6):e1000500.
- Sousa V, Peischl S, Excoffier L (2014) Impact of range expansions on current human genomic diversity. *Curr Opin Genet Dev* 29:22–30.
- Flicek P, et al. (2013) Ensembl 2013. *Nucleic Acids Res* 41(Database issue):D48–D55.

## SUPPLEMENTARY MATERIAL

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

<sup>1</sup> Stony Brook University, SUNY, Department of Ecology and Evolution, Stony Brook, NY 11794

<sup>2</sup> Institute of Ecology and Evolution, University of Berne, Berne, Switzerland 3012.

<sup>3</sup> Stanford University School of Medicine, Department of Genetics, Stanford, CA 94305

<sup>4</sup> Foundation Jean Dausset, Centre d'Etude du Polymorphisme Humain, Paris, France 75010

<sup>5</sup> University of Michigan Medical School, Department of Human Genetics and Department of Computational Medicine and Bioinformatics, Ann Arbor, MI 48109

<sup>6</sup> Swiss Institute of Bioinformatics, Lausanne, Switzerland, 1015

<sup>7</sup> Interfaculty Bioinformatics Unit, University of Berne, Berne, Switzerland 3012

## Table of Contents

<b>Supplementary Methods .....</b>	<b>4</b>
<b>Mapping and Variant Calling.....</b>	<b>4</b>
<b>Contamination and Data Quality Control.....</b>	<b>4</b>
<b>Identifying Single Nucleotide Variants .....</b>	<b>4</b>
<b>Callable Genome Mask for WGS analysis.....</b>	<b>4</b>
<b>Exome Sequence Data Analysis .....</b>	<b>5</b>
<b>Variant Annotation .....</b>	<b>5</b>
<b>Local Ancestry Assignment.....</b>	<b>5</b>
<b>Models of dominance.....</b>	<b>6</b>
Testing for a recessive model of dominance .....	6
Model for the underlying distribution of dominance.....	7
Testing for significance in differences in Load.....	9
<b>Supplementary Results .....</b>	<b>10</b>
<b>PSMC Simulations and demography.....</b>	<b>10</b>
<b>Effect of sample size on mean number of homozygotes.....</b>	<b>10</b>
<b>Effect of sample size on <math>A_i</math> for each functional category .....</b>	<b>10</b>
<b>Extreme alleles (GERP <math>\geq 6</math>) across populations .....</b>	<b>11</b>
<b>Hardy-Weinberg Equilibrium Test .....</b>	<b>11</b>
<b>Inference based on the site frequency spectrum: .....</b>	<b>11</b>
<b>Mutation Load .....</b>	<b>12</b>
<b>Table S1: Genome and exome variant statistics by population after imputation .....</b>	<b>13</b>
<b>Table S2: ANNOVAR functional annotations as compared to GERP .....</b>	<b>14</b>
<b>Figure S1: Assessment of PSMC coverage correction .....</b>	<b>15</b>
<b>Figure S2: Determination of whole genome masks.....</b>	<b>16</b>
<b>Figure S3: Genotype concordance for full genome data .....</b>	<b>17</b>
<b>Figure S4: Contrasting the SFS for ancestral and derived alleles .....</b>	<b>18</b>
<b>Figure S5: Number of heterozygotes per individual genome for 7 populations.....</b>	<b>19</b>
<b>Figure S6: Karyograms of the Mayan individuals reflecting the inferred ancestry .....</b>	<b>20</b>
<b>Figure S7: Simulations of bottleneck length and magnitude as inferred from PSMC.....</b>	<b>21</b>
<b>Figure S8: Comparison of ancestral/derived variants by GERP score.....</b>	<b>22</b>
<b>Figure S9: Distribution of derived variants with conservation scores <math>-2 \leq \text{GERP} \leq 6.5</math> .....</b>	<b>23</b>
<b>Figure S10: Median number of derived variants per individual.....</b>	<b>24</b>
<b>Figure S11: Individual counts of neutral derived variants.....</b>	<b>25</b>

<b>Figure S12: Number of common and rare variants per individual's genome by predicted effect.....</b>	<b>26</b>
<b>Figure S13: Number of homozygotes per population with subsampling .....</b>	<b>27</b>
<b>Figure S14: Site Frequency Spectra (SFS) of neutral and extreme effect variants .....</b>	<b>28</b>
<b>Figure S15: Relative reduction in heterozygosity (RH) .....</b>	<b>29</b>
<b>Figure S16: Heterozygosity under range expansion simulations assuming codominance at selected loci .....</b>	<b>30</b>
<b>Figure S17: Luhya het/hom<sub>der</sub> ratio by effect category.....</b>	<b>31</b>
<b>Figure S18: Testing a recessive model.....</b>	<b>32</b>
<b>Figure S19: Mutational load in 1000 Genomes exome data .....</b>	<b>33</b>
<b>Figure S20: Distribution of highly differentiated variants vs. the genome ..</b>	<b>34</b>
<b>Figure S21: Schematic of the range expansion model .....</b>	<b>35</b>
<b>Figure S22: Sharing of GERP <math>\geq 6</math> variants across populations .....</b>	<b>36</b>
<b>Figure S23: Site frequency spectrum under different selection regimes and locations of the range expansion. .....</b>	<b>37</b>
<b>Figure S24: Testing significance in observed differences in load under the assumed models of dominance.....</b>	<b>38</b>
<b>Figure S25: Relationship between runs of homozygosity and mutation load .....</b>	<b>39</b>
<b>Figure S26: Annotation of variants with PhyloP and correlation with distance from Africa .....</b>	<b>40</b>
<b>Supplementary References:.....</b>	<b>42</b>

## Supplementary Methods

### Mapping and Variant Calling

Read-pairs were mapped onto the human genome reference (GRCh37, with the pseudo-autosomal regions of the Y chromosome masked). Briefly, reads are mapped using the *bwa* mapper (version 0.5.9), an aligner that reports a confidence metric associated with each aligned read. The resulting alignments are then processed to identify PCR duplicates (using Picard, <http://picard.sourceforge.net/>), empirically recalibrate the quality values associated with each base call based on observed rates of differences from the reference, and realignment of all samples together around candidate small insertion-deletion variants (using the Genome Analysis Tool Kit [GATK] version 1.2-65) (1). The output of this process is a set of cleaned, calibrated, and mapped reads from each individual, suitable for subsequent analysis.

### Contamination and Data Quality Control

Sample contamination and data quality issues can compromise the results of large-scale genome sequencing efforts. Contamination was assessed for each individual by comparing the genotypes from Illumina Human660K array SNP data (2) and the Illumina HiSeq data from an initial per-sample call set using *samtools*. A concordance rate was calculated from the number of HiSeq homozygous non-reference calls (HNR) that were also homozygous non-reference on the Illumina Human660K array, divided by the total number of HNR calls from the 660K array. If the concordance dropped below 90%, a new library was made and contamination assessed in a second run.

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

### Identifying Single Nucleotide Variants

Candidate single nucleotide variants were identified based on joint calling across all samples using the Unified Genotyper in the GATK. We applied the Variant Quality Score Recalibration (VQSR) procedure to retain a set of variants such that 99% of variant positions that overlap with HapMap3 SNPs were retained. Refined genotypes for the resulting set of positions were obtained using Beagle v3 (3). Sites were called on the autosomes and the pseudo-autosomal portions of the X chromosome, but only variants on the autosomes were utilized in subsequent analysis.

### Callable Genome Mask for WGS analysis

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

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

## Exome Sequence Data Analysis

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

## Variant Annotation

The putative ancestral state of each variant was annotated following the 1000 Genomes Project (4) based on ancestral sequences determined by Ortheus using multi-species alignments from Ensembl Compara release 59 ([http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/phase1/analysis\\_results/supporting/ancestral\\_alignments/](http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/phase1/analysis_results/supporting/ancestral_alignments/)). Only variants for which the ancestral state was known were kept for downstream analysis.

Recent studies have shown that establishing the damaging potential of a variant is extremely difficult. As an example, one of the commonly used predictive algorithms, Polyphen (5), has been shown to have a strong reference bias, annotating as neutral variants that are represented in the reference genome, regardless of their ancestral state (6). We therefore used two algorithms, one that is a measure of conservation across species (GERP scores) (7), and another that is based on the biological effect of the variant (ANNOVAR) (**Table S2**). Positive GERP (RS) scores reflect a site showing high degree of conservation, based on the inferred number of “rejected substitutions” across the phylogeny. GERP scores were obtained from the UCSC genome browser ([http://hgdownload.cse.ucsc.edu/gbdb/hg19/bbi/All\\_hg19\\_RS.bw](http://hgdownload.cse.ucsc.edu/gbdb/hg19/bbi/All_hg19_RS.bw)) based on an alignment of 35 mammals to human. The allele represented in the human hg19 sequence was not included in the calculation of GERP RS scores. GERP scores from the exome dataset range from -12 to 6.17, though only variants with a GERP score greater than -2 were selected for subsequent analysis, as negative values may be indicative of poor sequence alignments across the phylogeny. Most analysis focus on variants with positive GERP RS scores > 2. Since the range of RS scores is dependent on the depth of the multi-species phylogeny used, we re-annotated GERP scores for the 1000 Genomes data using this procedure. We first examined the distribution of GERP RS scores for both all exome single nucleotide changes and for only nonsynonymous changes (**Figure S8**). We observe an approximately normal distribution for all exome variants between -2 and 6.5, but for nonsynonymous variants, there is sharp decrease in the number of variants greater than GERP score of 4. This difference in the distributions is consistent with the prediction that more conserved nonsynonymous sites are more likely to be functionally important and therefore subject to purifying selection when mutations occur at highly conserved sites. We focus on all exome variants in the analyses that follow. In order to explore a possible ancestral bias we examined the site frequency spectra across effects for the different populations. **Figure S4** shows results for moderate effect variants. Though no evidence of an ancestral bias was detected, we note that excluding variants where the ancestral allele is the alternate allele according to the reference sequence has a very strong effect in the site frequency spectrum overall shape. While the true fitness of these mutations cannot be measured directly, GERP scores are indicative of long-term selection in many species and the severity of mutation effect should be similar in human populations.

## Local Ancestry Assignment

Local ancestry segments in the 8 Mayan samples were inferred using RFMix (8). Two reference panels were constructed, one for Native American ancestry and one for European ancestry. The Native American reference panel was constructed by including all samples from the Maya, Pima, Columbian, Karitiana and Surui populations in HGDP(2). The European reference panel was constructed by using all samples from the Sardinian and French populations. One Mayan sample at a time from the Native American reference panel was removed to form the admixed panel for the initial inference step. RFMix was run in PopPhased mode with the

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

## Individual-based simulations

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

## Models of dominance

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

## Testing for a recessive model of dominance

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

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

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

$$\text{hom}_{Anc}^{HW} = p^2 \times N = (1-q)^2 \times N$$

$$\text{het}^{HW} = 2pq \times N = 2(1-q)q \times N$$

$$\text{hom}_{Der}^{HW} = q^2 \times N$$

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

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

#### *Proportion of deleterious variants in dominant and recessive genes*

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

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

## Model for the underlying distribution of dominance

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

### 1. Boundary conditions

To begin, we make use of a relationship between  $h$  and  $s$  that was previously obtained for mutations in yeast (9):

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

where  $\beta_1$ ,  $\beta_2$ , and  $d$  are some constants. As described in Agrawal and Whitlock and others (9), as selection strength increases, the dominance coefficient tends to zero. In other words, we assume that strongly deleterious mutations are fully recessive:

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

We also make use of a frequent assumption that the dominance coefficient for neutral mutations (i.e. those for which  $s = 0$ ) is equal to  $\frac{1}{2}$ :

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

When we specify  $d = 0$  and  $\beta_1 = \frac{1}{2}$  in equation (1), the dependence of the dominance coefficient on the selection coefficient becomes

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

## 2. Least-squares fit

In order to find the best value for parameter  $\beta_2$  in  $h(s)$  above, we start by defining  $h(s)$  as a function of both  $s$  and  $\beta_2$ :

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

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

We assume that, of the four classes of mutations mentioned above, the one with the smallest selection coefficient has a dominance coefficient that is very close to  $\frac{1}{2}$  and that the class of mutations with the largest selection coefficient has a dominance coefficient that is very close to zero.

One can show that the former requirement, that  $|h(\beta_2, s = s_0) - \frac{1}{2}|$  is minimized, tends to decrease  $\beta_2$ . At the same time, the latter requirement, that  $|h(\beta_2, s = s_3) - 0|$  is minimized tends to increase  $\beta_2$ . If we require that the sum of  $|h(\beta_2, s = s_0) - \frac{1}{2}|$  and  $|h(\beta_2, s = s_3) - 0|$  is minimized – or, similarly, that the sum of the squares of these two expressions is minimized – one obtains an intermediate value of  $\beta_2$  that corresponds to a balance between the two requirements. In other words, we are looking for

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

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

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

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

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

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

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

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

### 3. Dominance coefficient function.

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

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

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

$$\begin{aligned} h(s_0) &= h(0) = 0.46698, \\ h(s_1) &= h(10^{-4}) = 0.292893, \\ h(s_2) &= h(10^{-3}) = 0.0619497 \\ h(s_3) &= h(2 \times 10^{-3}) = 0.0330204. \end{aligned}$$

### 4. Variance of the dominance coefficient

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

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

where  $s_j, \beta_1, \beta_2$  have already been estimated,  $\mu_{h(\text{del}),j}$  is the dominance coefficient for each selection coefficient that has also been calculated,  $d \approx \delta$ , and  $\sigma_h^2$  and  $\delta$  are the variance and the mean of  $Q_G$ , which is a gamma distribution to introduce variance to the dominance coefficient. Values for  $\sigma_h^2$  and  $\delta$  have been taken from (9) and are 0.010 and 0.038, respectively.

## Testing for significance in differences in Load

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

## Supplementary Results

### PSMC Simulations and demography

We constructed profiles of effective population size through time using PSMC method (10). Since this model relies on heterozygous sites within an individual it is not applicable to low or moderate coverage whole genome sequencing. However, if the rate of ‘missing’ heterozygotes is known and uniform, the PSMC curves can be corrected through a rescaling of the mutation rate to an effective rate that incorporates heterozygote false negative rates. We applied this rescaling idea, utilizing Pathan sample HGDP00222, which has 22x coverage, as a test case. We subsampled reads from this sample to lower coverage levels, ran the PSMC calling procedure on the sub sampled read sets, and compare the proportion of heterozygous sites identified at each coverage level. Based on this, we constructed a correction curve relating coverage level with to heterozygote SNP false negative rates. We found that reasonable concordance between down-sampled and original PSMC curves could be obtained for coverage levels >10x. Since all of the samples were sequenced and processing in the same manner, we reasoned that the correction curve constructed for HGDP00222 would be applicable to other samples in this data set. We verified this through comparison of our corrected PSMC curves with PSMC curves constructed from a high coverage San individual and a high coverage Mbuti Pygmy sample obtained from (11) (**Figure S1**).

### Effect of sample size on mean number of homozygotes

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

### Effect of sample size on $A_i$ for each functional category

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

## Extreme alleles (GERP >= 6) across populations

We were interested in looking at the pattern of extreme alleles across populations. Population theory predicts that extreme alleles will be held at low frequencies if their effect is deleterious, and eventually be eliminated. The SFS of extreme variants shows an excess of low frequency variants (namely singletons), compared to the neutral SFS (**Figure S13**). For a given population, no less than 45% of the extreme variants are singletons. We next asked how variants were distributed across the complete dataset (**Figure S20A**). Surprisingly when we consider all the populations 60% of the variants are singletons (514 out of 854). If we focus on variants private to a specific population the percentage increases to 76%. Thus, the vast majority of variants with extreme effect are either new or kept at very low frequencies, being private to a population. Interestingly, few variants (a dozen) are almost fixed in the dataset. This could be due to errors in the assignment of the ancestral allele or evidences of positive selection. When we focus on variants found in homozygosity (**Figure 20B**) we observe as expected an increase in the number of homozygotes, with distance from Africa. Sub-Saharan African populations have more variants in homozygosity that are found only once in the dataset (like “homozygous – singletons”), whereas Out of Africa populations have more homozygous singletons at higher frequencies. Some variants are found at high frequencies in African populations, and are found at lower frequencies elsewhere.

## Hardy-Weinberg Equilibrium Test

We tested for deviations from Hardy-Weinberg equilibrium in a sample of 72 Luhya individuals from 1000 Genomes Nimblgen exome capture (**Figure S16A**). We show that there is an excess of heterozygotes compared to Hardy-Weinberg expectations, particularly when the homozygotes are at low frequency. However, no extreme effect alleles were found to have significantly more heterozygotes than predicted. The bulk of the heterozygotes with a paucity of corresponding derived homozygotes occurs in the neutral and moderate effect categories. We conclude that alleles are either generally additive or moderately recessive such that incomplete penetrance does not cause them to significantly violate Hardy-Weinberg at  $p<0.01$ . Alternatively, we note that the HW model has low power for rare allele frequencies, so if most selection occurs against deleterious recessive variants less than 25% in frequency than this test does not have sufficient power to detect deviations from an additive model. For example, if there is one derived homozygotes in the population then there would need to be more than 37 heterozygotes to deviate from Hardy-Weinberg at  $p<0.01$ , an allele frequency of at least 28%. Interestingly, we also observe many variants that have a deficient number of heterozygotes / excess of homozygotes. This pattern can occur due to haploinsufficiency (12) or false negatives in the next-generation sequencing data (i.e. heterozygotes are more error prone for variant calling software).

## Inference based on the site frequency spectrum:

Although we classify extreme effect mutations as being potentially deleterious, there is also a possibility that these mutations are functionally adaptive, large effect mutations that are under positive selection. We test this hypothesis by considering the site frequency spectrum (SFS) of predicted extreme and neutral effect mutations. For each population, we considered the number of extreme and neutral effect variants in each allele frequency bin, proportional to the total number of mutations in the extreme and neutral category such that the spectra are directly comparable. While the two spectra generally demonstrate an exponential decay, as expected under constant size or low population growth, there is an enrichment of extreme effect mutations in low frequency bins for all populations. This observation is consistent with other studies that have shown an enrichment of deleterious alleles at low frequencies (13, 14). Some populations also display an enrichment of extreme effect variants at intermediate frequencies (e.g. Pathan), potentially indicative of adaptive alleles under balancing selection; such inference would require additional modeling (15). No populations display an enrichment of extreme effect alleles at fixation, suggesting that overall, selective sweeps have not played a dominant role in shaping the

frequencies of extreme effect alleles (16). No such pattern is present for large effect alleles either (**Fig. 3B**).

## Mutation Load

It has also been argued that the relationship between effective population size and load is non-linear for a model with partially, but not completely, recessive mutations (i.e.  $h=0.05$ ) (17). This is because in a population with larger effective size, mutations of equal  $s$  are less likely to be lost by drift and thus recessive deleterious alleles can float to higher frequencies, impacting more individuals when exposed as homozygotes. We do not observe this effect within African populations, which carry fewer weakly deleterious alleles per individual than non-African populations (**Figure 2A**).

It is also interesting to note that there are negligible differences in additive load between western Africans and Europeans. This is in keeping with the fact that western African populations have experienced dramatic population growth over the past 5,000 years (18), which alters the distribution of deleterious alleles within a population (19). There are sharp differences in demography among African populations, and populations with western African ancestry should not be taken to be representative of all of Africa.

**Table S1: Genome and exome variant statistics by population after imputation**

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

<sup>1</sup> Mean population coverage for genomes assessed by sampling each individual for ~650,000 sites on the Illumina Human660K BeadChip SNP platform and counting read depth after quality filtering. Median population coverage for the exomes encompassing all mapped, on target reads.

<sup>2</sup> Mean number of non-reference alleles for an individual in the population (i.e., a non-reference homozygous genotype is counted twice.)

<sup>3</sup> Mean number of heterozygotes for an individual in the population.

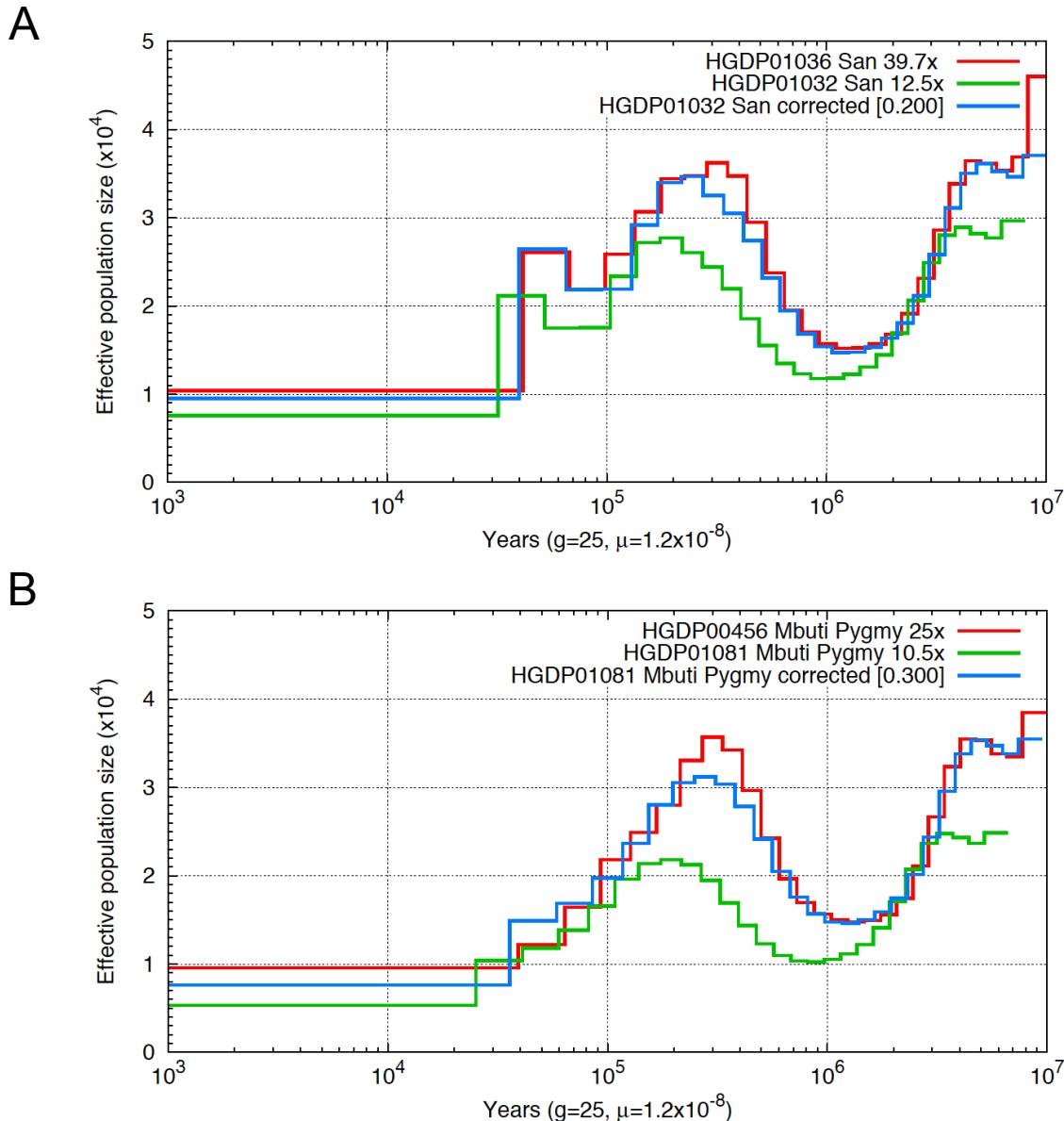
<sup>4</sup> Mean number of singletons for an individual in the population.

<sup>5</sup> Eight individuals were included for exome and genome sequencing; one sample did not pass genome quality control and was excluded from the full genome dataset.

**Table S2: ANNOVAR functional annotations as compared to GERP**

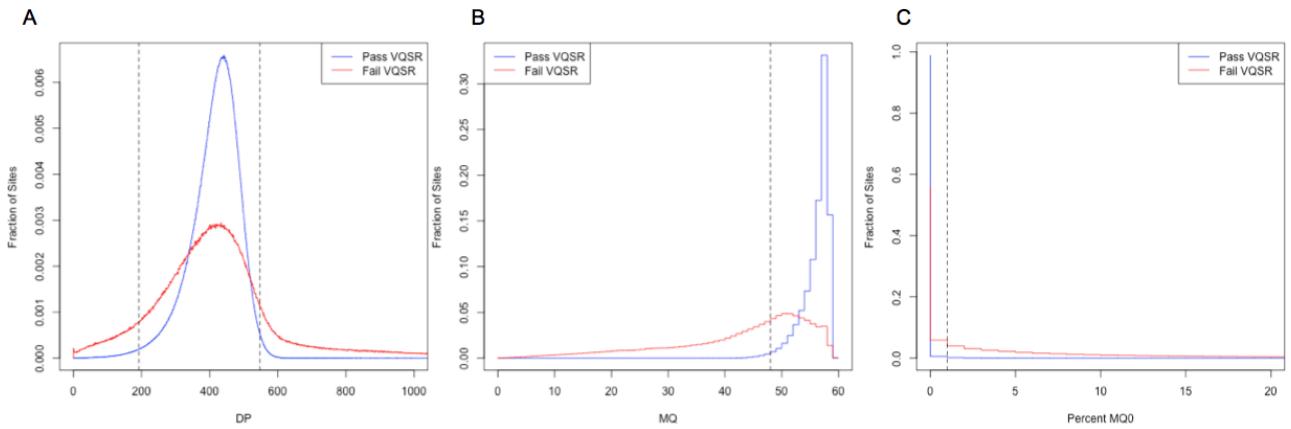
RS Score	Intergenic	Intronic	UTR-5	UTR-3	Missense	Nonsense	Synon.	Total
<b>Neutral: -2,2</b>	5566	28592	1167	1751	14614	187	16883	68760
<b>Moderate: 2,4</b>	1676	8956	540	721	13648	160	9913	35614
<b>Large: 4,6</b>	682	2789	248	314	19645	183	4935	28796
<b>Extreme: &gt;6</b>	11	64	4	8	741	7	88	923
<i>Total:</i>	7935	40401	1959	2794	48648	537	31819	134093

## Figure S1: Assessment of PSMC coverage correction



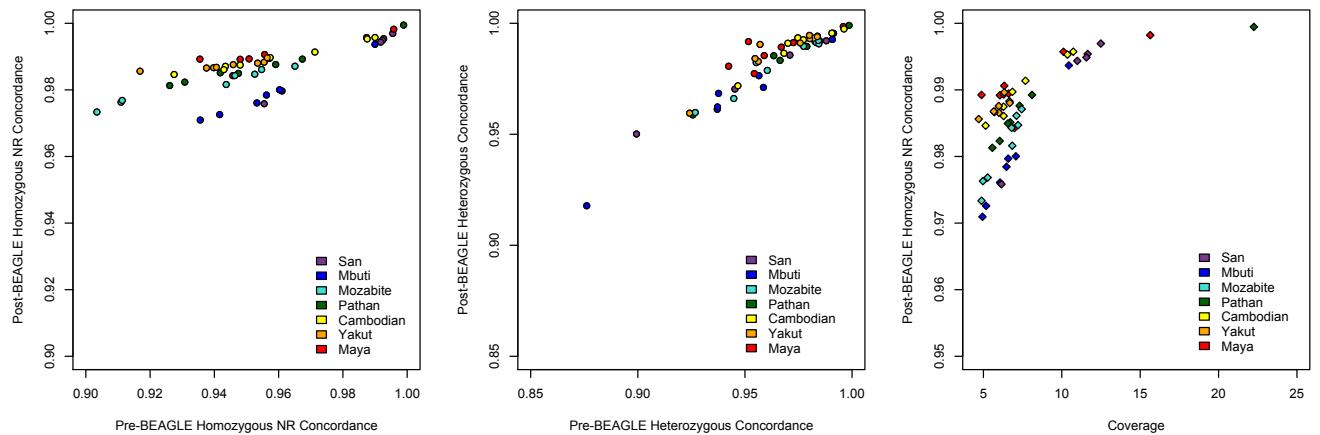
**Figure S1 Assessment of PSMC coverage correction.** **A)** PSMC curves from original moderate coverage data before and after coverage correction are compared with **B)** PSMC curves constructed from high-coverage sequences from the same populations. Strong concordance is observed, with discrepancies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

## Figure S2: Determination of whole genome masks



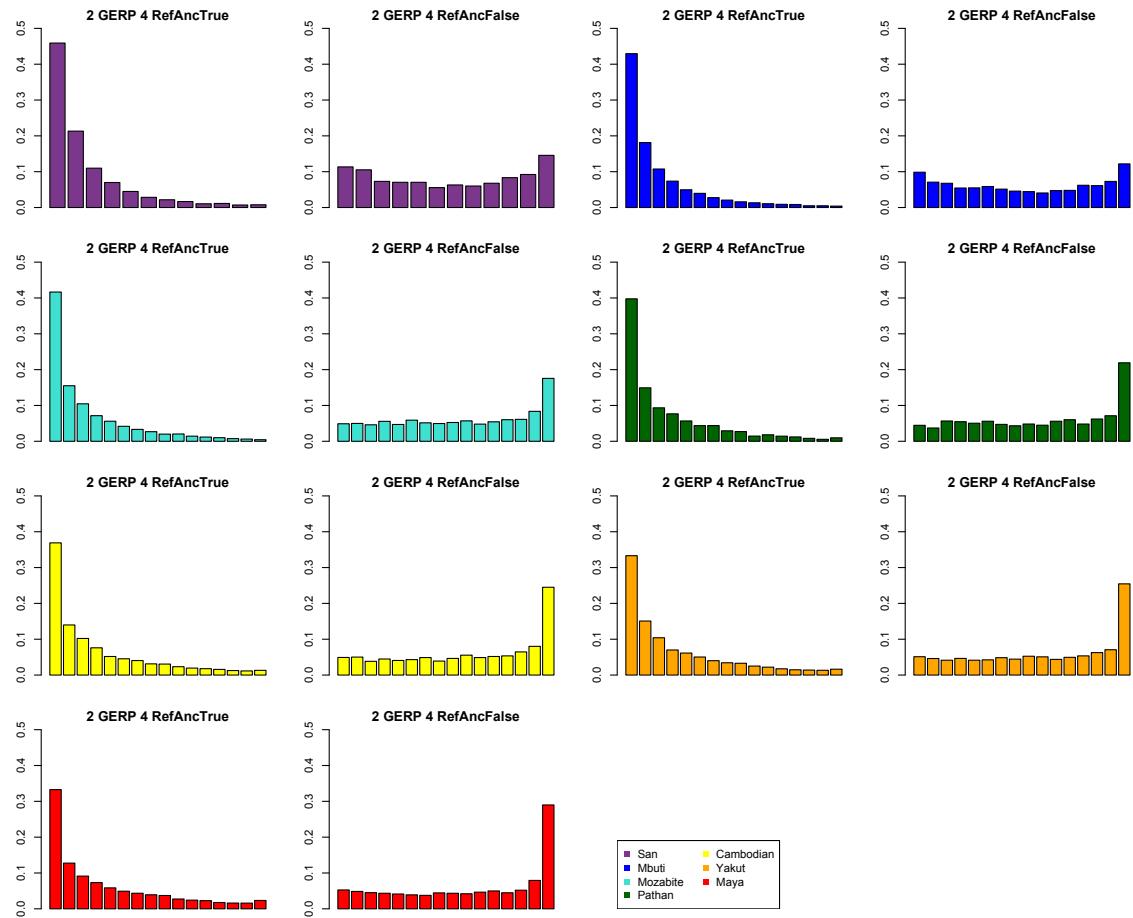
**Figure S2: Determination of whole genome masks.** Distribution of DP, MQ, and MQ0 fraction values for genomic sites that pass (blue) and fail (red) the VQSR procedure are shown. Cutoffs correspond to  $192 \leq DP \leq 547$ ,  $MQ \geq 48$  and  $MQ0 \text{ fraction} \leq 0.01$ .

## Figure S3: Genotype concordance for full genome data



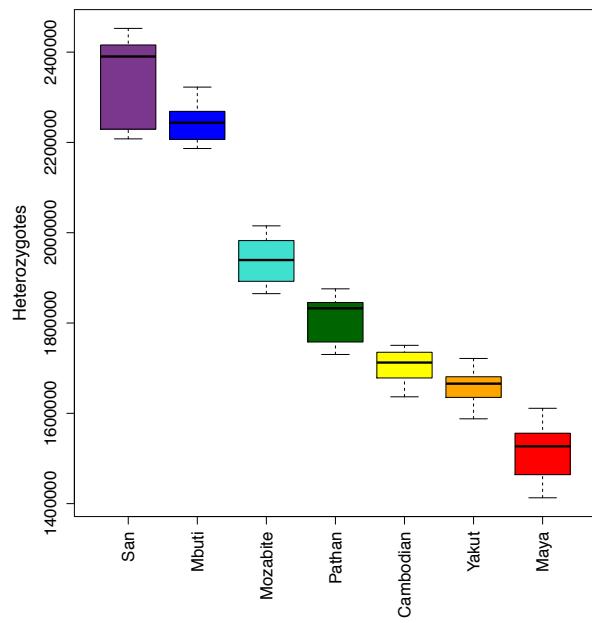
**Figure S3: Genotype Concordance for Full Genome Data.** Phasing and imputation for the full genomes was performed using BEAGLE v3.2. We assessed genotype concordance for SNP calls pre- and post-BEAGLE by comparing genotypes to the Illumina 660K SNP array data for each individual (2). **A)** Concordance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

## Figure S4: Contrasting the SFS for ancestral and derived alleles



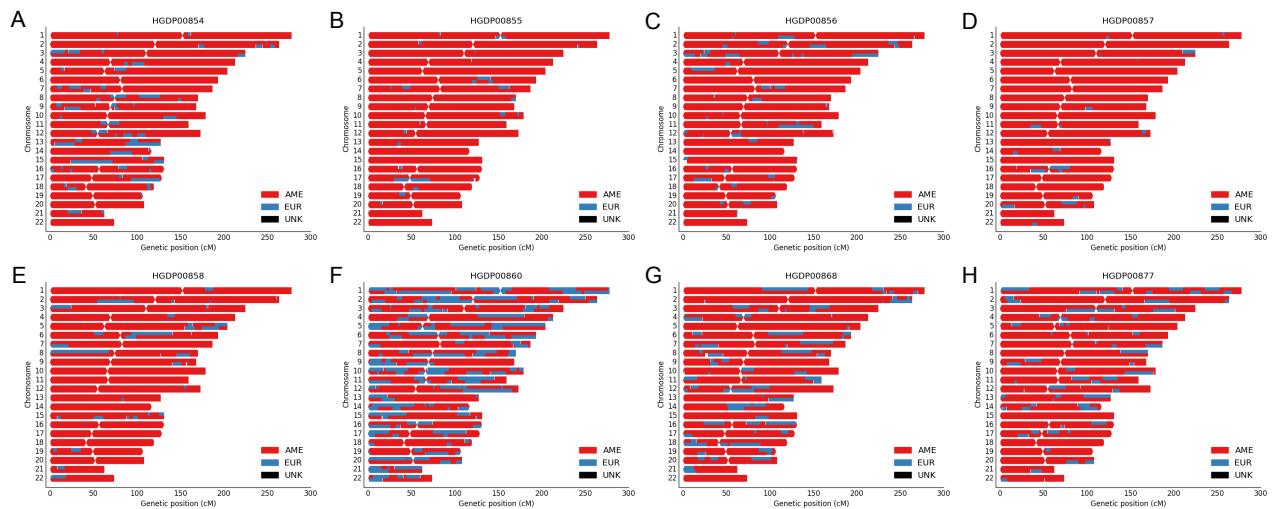
**Figure S4: Contrasting the SFS for Ancestral and Derived Alleles.** We asked whether there were systematic differences in the SFS for ancestral and derived variants, relative to the human reference genome. Shown are moderate effect variants, GERP >2 and <4. The left plot for each population shows the SFS for which the reference allele is ancestral, and thus the non-reference allele is derived. The right hand SFS shows the opposite pattern, where the reference allele is derived and the non-reference allele is ancestral. This pattern has been observed elsewhere (20), and is even expected because alleles that have already been observed once, in the human reference genome, have a higher probability of being observed again when sampling a new population. OOA populations, being more closely related to the human reference genome, have more alleles that have been previously observed in the single human reference sample. African populations have a higher proportion of novel, derived alleles (or conversely fewer derived alleles shared with the reference).

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



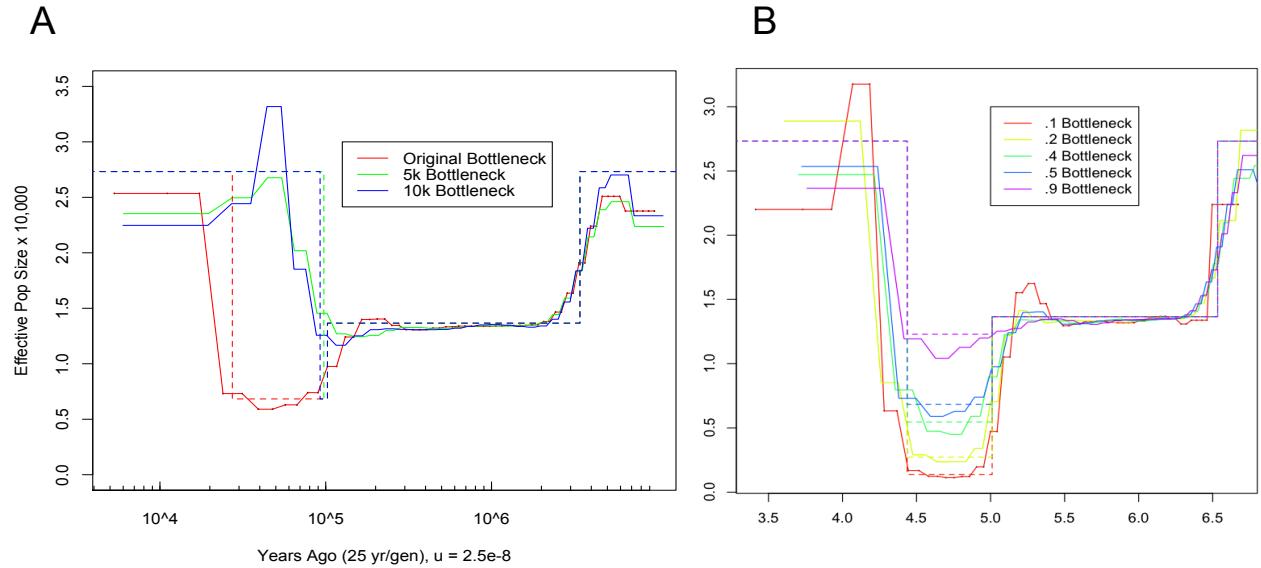
**Figure S5: Number of heterozygotes per individual genome for 7 populations.** Boxplots of number of heterozygotes per individual from the 2.48Gb callable region of the human genome for all 7 seven populations.

## Figure S6: Karyograms of the Mayan individuals reflecting the inferred ancestry



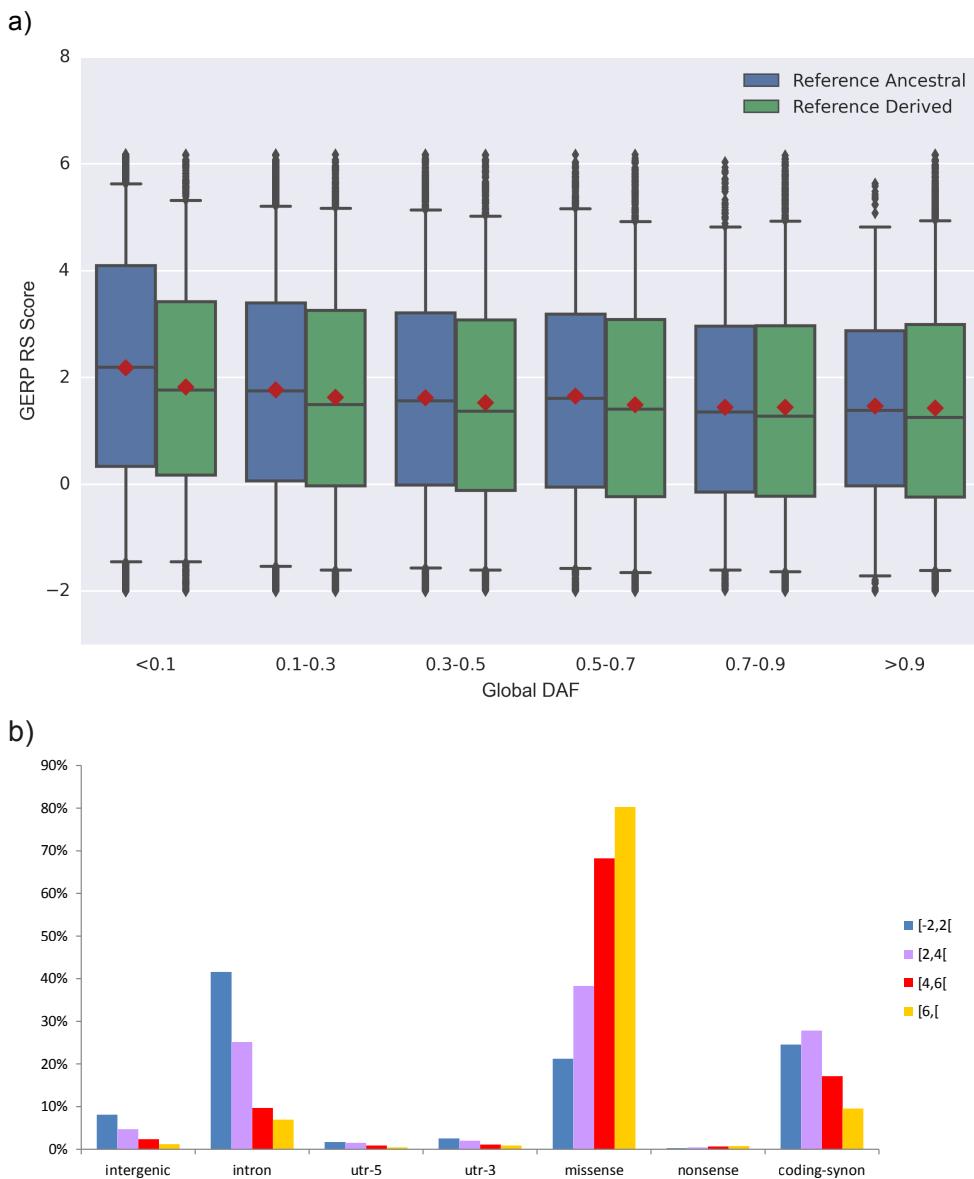
**Figure S6:** Estimates of European (blue) and Native American (red) ancestry at the chromosome level were plotted for every individual (A-G). Every pair of chromosomes is depicted along the Y-axis and the genetic position is reflected on the X-axis. Note that two out of eight individuals (F,G) showed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

## Figure S7: Simulations of bottleneck length and magnitude as inferred from PSMC



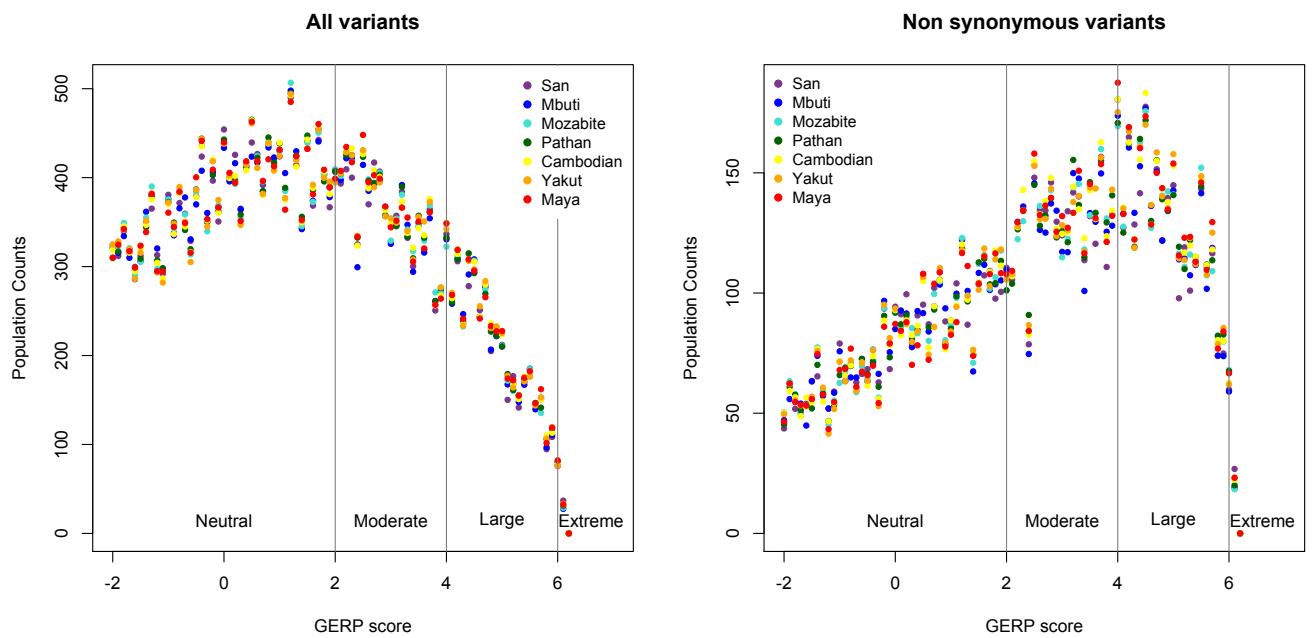
**Figure S7: Simulations of Bottleneck Length and Magnitude as Inferred from PSMC.** We tested whether PSMC was robust to changes in either the duration of a bottleneck or the magnitude of a population bottleneck. **A)** Using a simulation of population history parameters similar to the original paper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

## Figure S8: Comparison of ancestral/derived variants by GERP score



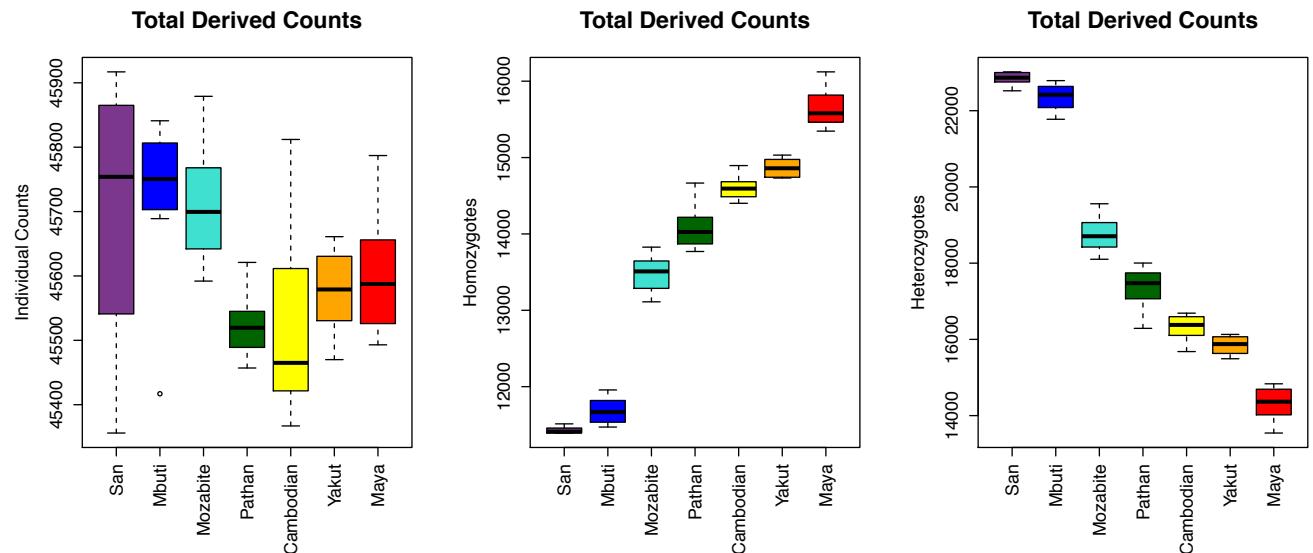
**Figure S8. Comparison of GERP scores across sites where the reference allele is derived or ancestral, and to ANNOVAR classes.** A) Variant sites were binned by derived allele frequency (across all 54 samples from the HGDP collection) and categorized based on the state of the allele represented in the human genome reference assembly. The box plots represent the median and 25th and 75th percentiles of data, the whiskers correspond to the 5th and 95th percentile, and the red diamond indicates the mean. Sites with a GERP RS score < -2 were omitted from analysis. B) Proportion of functional classes of mutations, as defined by ANNOVAR, in different GERP scores categories (see also Table S2).

**Figure S9: Distribution of derived variants with conservation scores  $-2 \leq \text{GERP} \leq 6.5$**



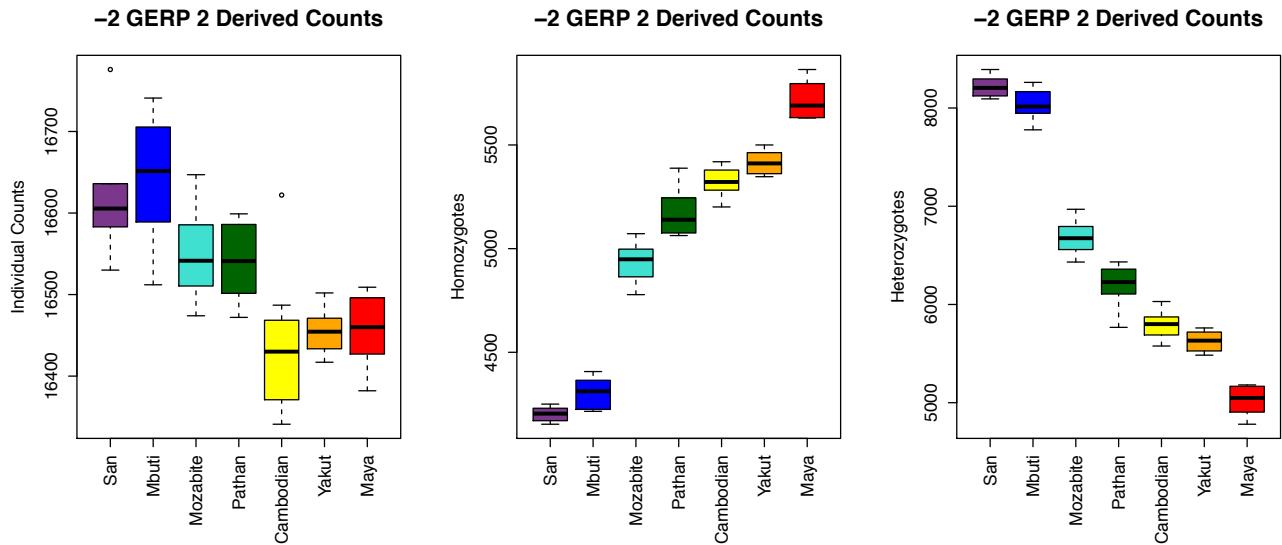
**Figure S9: Distribution of derived variants with conservation scores  $-2 \leq \text{GERP} \leq 6.5$ .** For bin sizes of 0.2, the number of derived variants within each population are plotted according to the prior population color scheme (Figure 1A). Binned counts were standardized by the number of samples per population. GERP scores were divided into four functional categories: neutral (-2 to 2), moderate (2 to 4), large (4 to 6), extreme ( $>6$ ). **A)** Nonsynonymous variants are not normally distributed. **B)** All exome variants conform to a normal distribution. No population had a significant excess or deficit of variants within a particular GERP score range.

## Figure S10: Median number of derived variants per individual



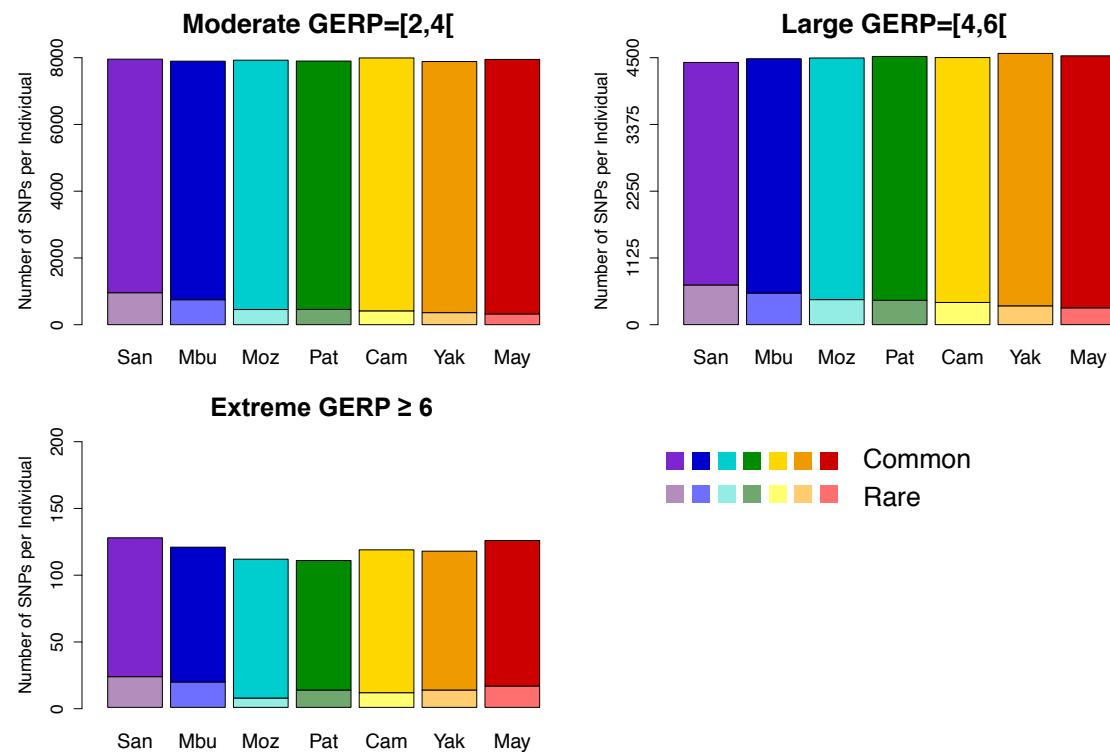
**Figure S10: Median number of derived variants per individual.** For all variants, regardless of GERP score annotation, we tabulated the number of derived variants per individual, heterozygotes and derived homozygotes. Out of Africa populations have roughly equivalent numbers of derived variants per individual. African populations have ~1% fewer derived variants per individual.

**Figure S11: Individual counts of neutral derived variants**



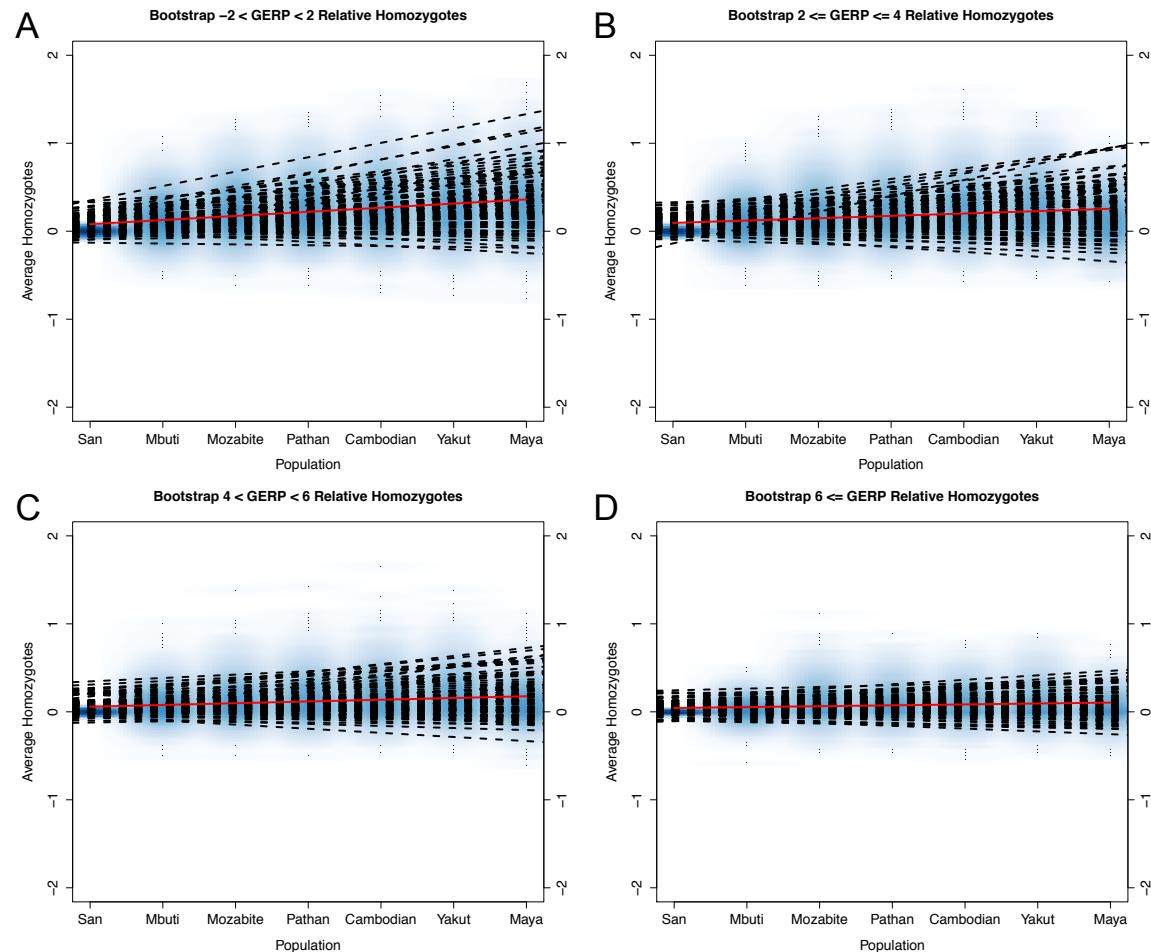
**Figure S11: Individual counts of Neutral derived variants.** For exome variants with GERP score in the -2 to 2 range, we evaluated the average number of **A)** The total number of derived variants (equivalent to number of heterozygotes + twice the number of homozygotes), **B)** derived homozygotes, and **C)** heterozygotes by population.

**Figure S12: Number of common and rare variants per individual's genome by predicted effect**



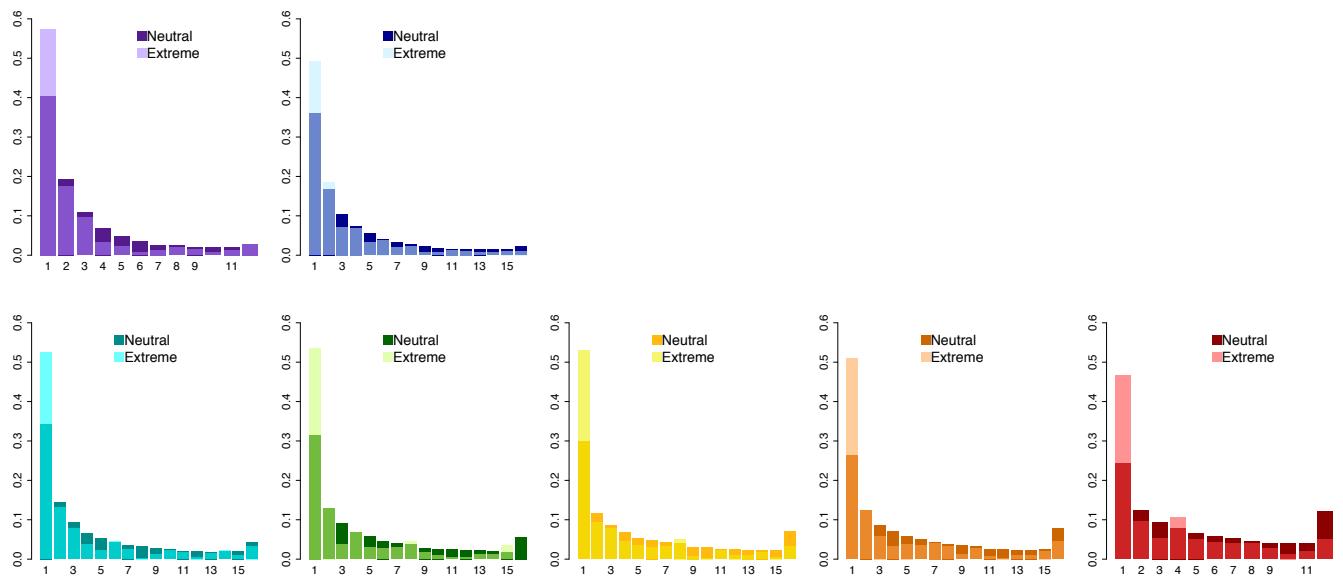
**Figure S12: Number of common and rare variants per individual's genome by predicted effect.** For a given individual, deleterious variants within each predicted effect category were divided into common (>10%, solid colors) and rare (<10%, shaded colors). The contribution of common deleterious variants to an individual's burden is much greater than rare variants. **A)** Moderate, **B)** Large, and **C)** Extreme.

## Figure S13: Number of homozygotes per population with subsampling



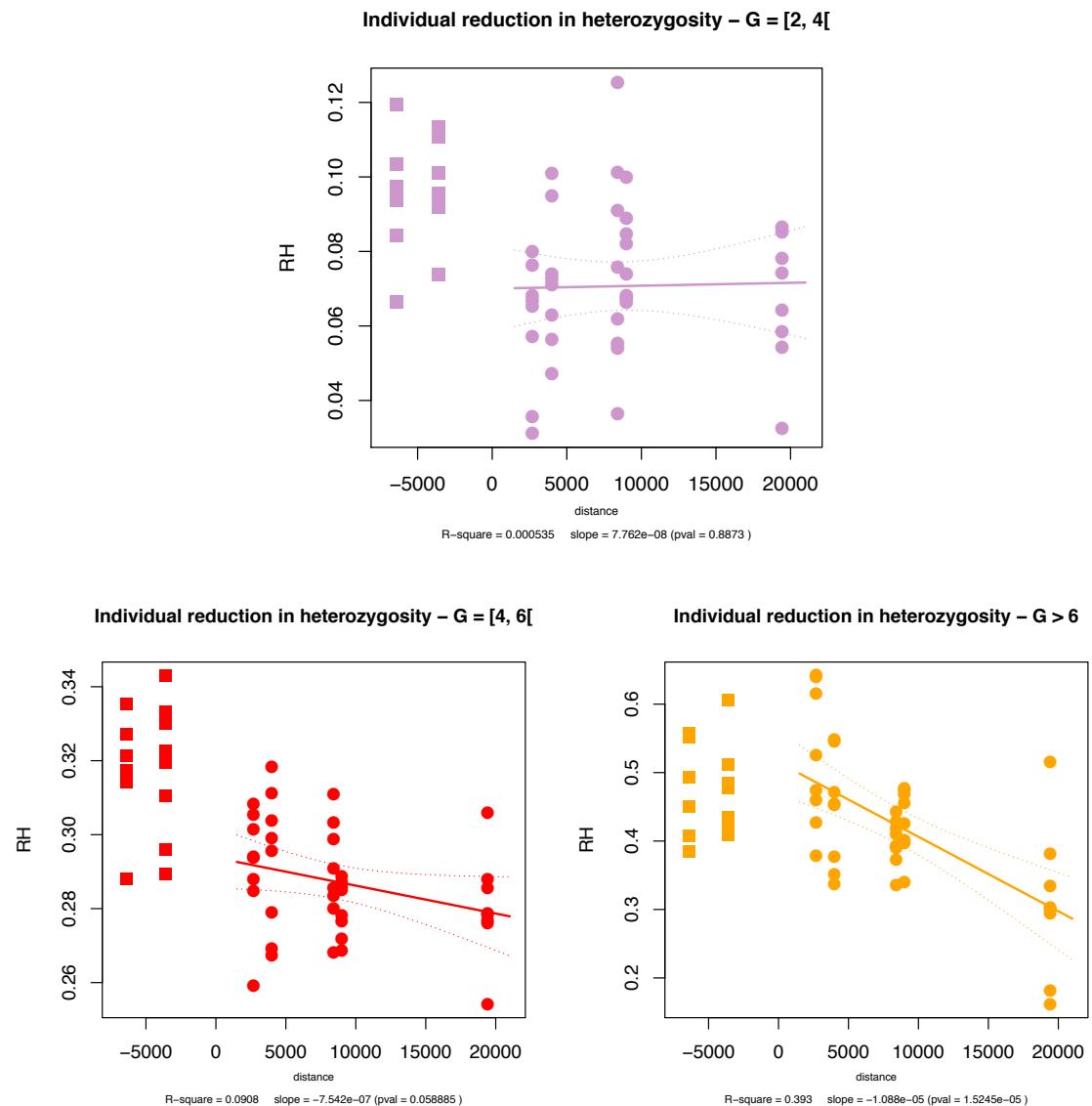
**Figure S13: Number of homozygotes per population with subsampling.** The minimum number extreme homozygotes in a San individual, 26, was used to sub-sample the variants of neutral, moderate, large and extreme effect, and calculate the average number of homozygotes per population. The red line indicates the average number across 10,000 bootstraps, the dashed lines indicate the average number per population for every bootstrap and blue background indicated the individual ranges for every bootstrap.

## Figure S14: Site Frequency Spectra (SFS) of neutral and extreme effect variants



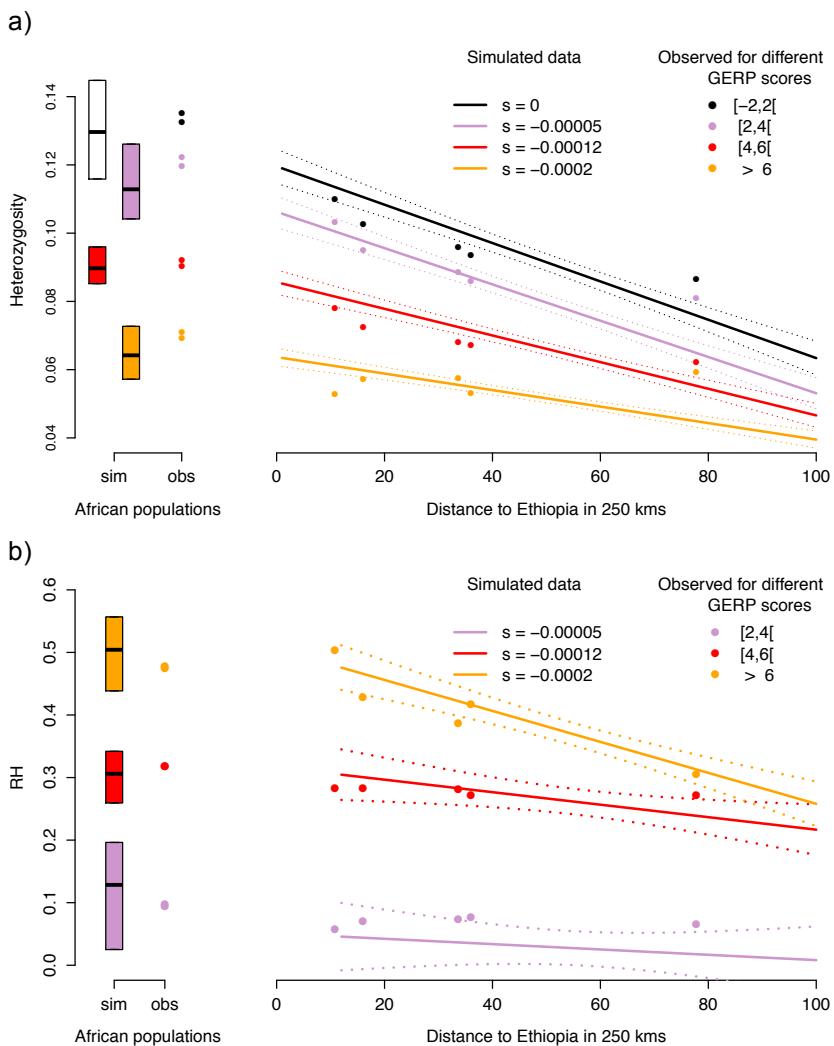
**Figure S14: Site Frequency Spectra of Neutral and Extreme Effect Variants.** We compared the proportion of neutral variants by their frequency class to extreme effect variants by frequency class. The proportion of variants is shown along the Y-axis and each frequency bin is shown along the X-axis. Extreme effect variants are colored translucent. Neutral variants are shaded grey. Overlap between the two categories is opaque.

**Figure S15: Relative reduction in heterozygosity (RH)**



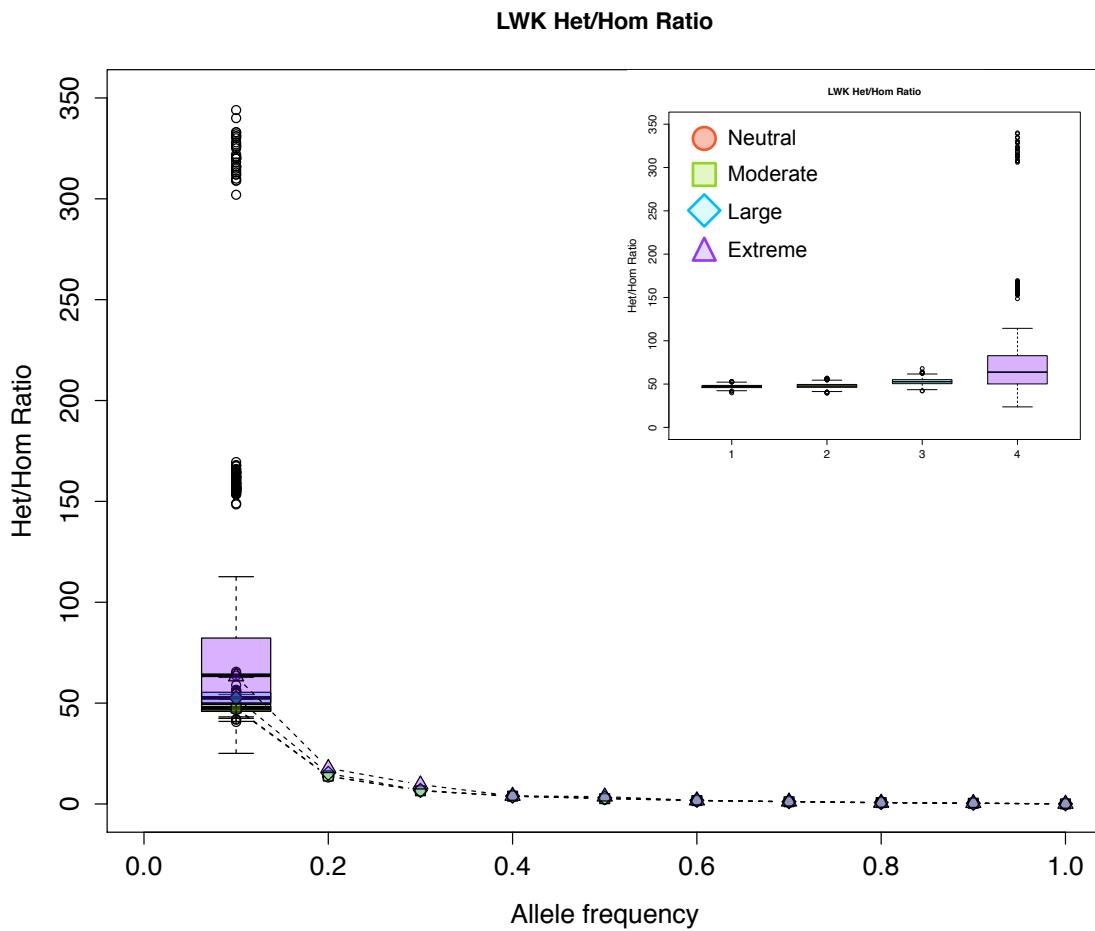
**Figure S15: Relative reduction in heterozygosity (RH) at sites under selection as compared to neutral sites.** Correlation between reduction in heterozygosity, or “RH”, and distance from northeastern Africa for all 41 OOA individuals, separated by GERP score category. African individuals are represented by the squares.

## Figure S16: Heterozygosity under range expansion simulations assuming codominance at selected loci



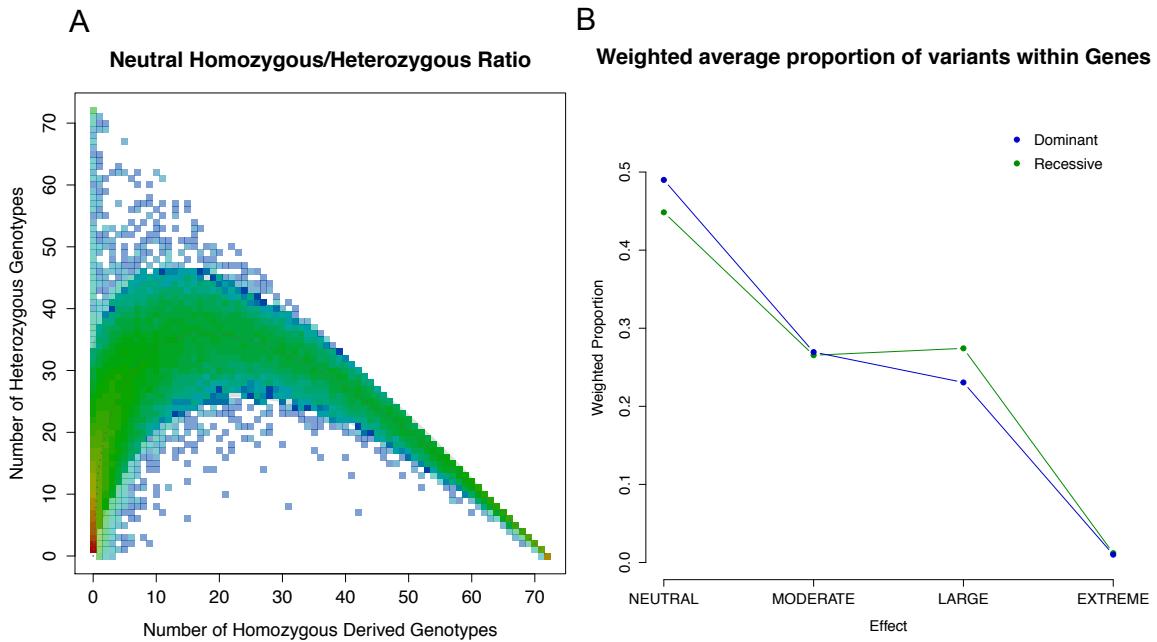
**Figure S16: Heterozygosity under Range Expansion Simulations Assuming Codominance at Selected Loci.** A) Colored circles show average expected heterozygosity for populations with ancestry from the OOA bottleneck. Solid lines show the regression lines obtained from simulations and dashed lines indicate 95% confidence intervals for the regression. The boxplots and colored circles on the left show the simulated heterozygosities in ancestral (i.e. African) populations, and the observed heterozygosity in our African dataset (San / Mbuti), respectively. B) Observed and simulated patterns of the reduction of heterozygosity ( $RH$ ) under an additive model. Selection coefficients used in the simulations are  $s = 0$  (black),  $s = -0.00005$  (lavender),  $s = -0.00012$  (red), and  $s = -0.0002$  (orange).

**Figure S17: Luhya het/hom<sub>der</sub> ratio by effect category**



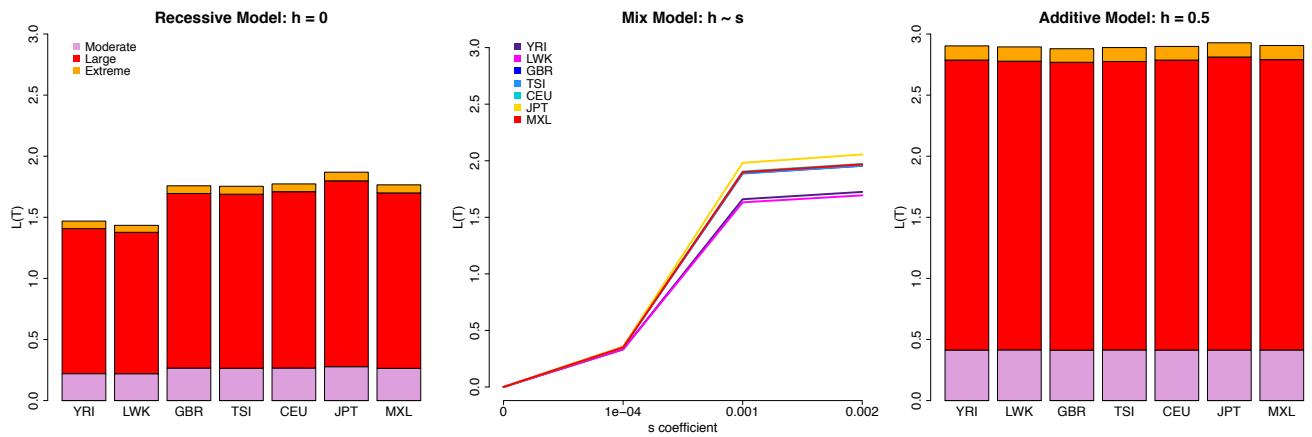
**Figure S17: Luhya (LWK) Het/Hom Ratio by effect category:** Under a recessive model, it is expected that EXTREME effect variants will have an excess of heterozygotes, compared to homozygotes, because of the effect of purifying selection on homozygotes. However, this pattern could also be biased by an excess of low frequency variants with extreme effect, compared to other categories. In order to distinguish between the two processes, we removed singletons for the dataset and calculated the ratio of heterozygotes / homozygotes in the 1000G LWK for all variants within each effect, and plotted the results according to the variants frequency. Results show an excess of heterozygotes in variants of extreme effect for low frequency bins ( $\leq 30\%$ ), being particularly evident for variants between 10% derived allele frequency. The inset shows boxplots for the 10% allele frequency bin along the x-axis.

## Figure S18: Testing a recessive model



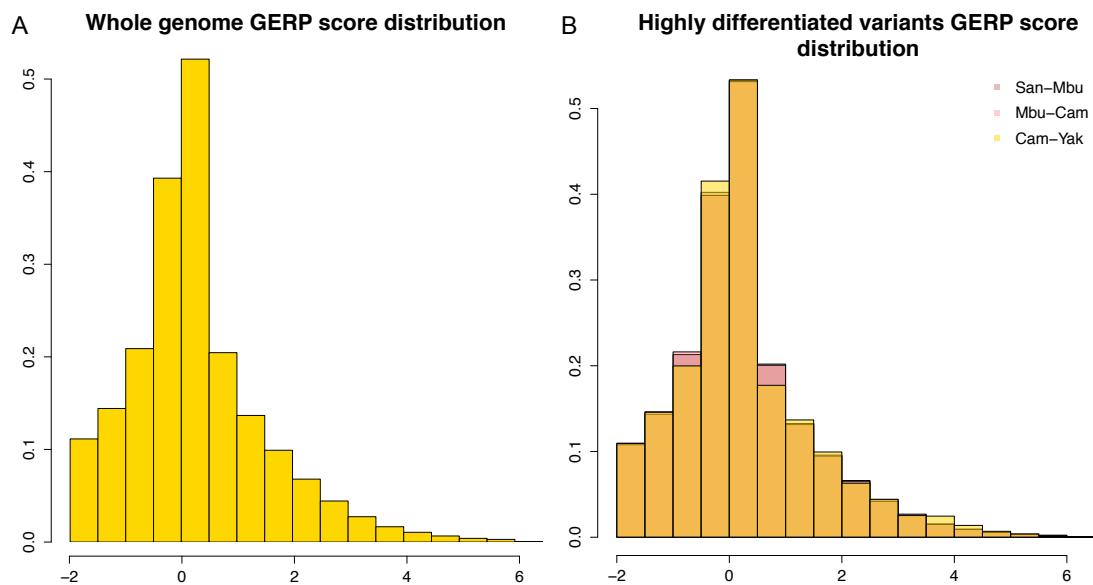
**Figure S18: Testing a recessive model.** **A)** A non-additive model of dominance could lead to deviations from HW equilibrium if the derived variant selection coefficient is strong enough. We compared the observed genotype frequencies with the expected frequencies, considering the observed allele frequencies. Variants are plotted according to the observed number of homozygotes (x-axis) and heterozygotes (y-axis) in the LWK population. Heat map reflect a higher number of variants. The grey dashed line reflects the HW expectation. Colors are shaded when variants significantly deviate from HW expectation ( $p\text{-value} < 0.01$ ). Specifically, variants on the upper left corner represent an excess of heterozygotes compared to what would be expected, compatible with a recessive model. **B)** *Weighted average proportion of variants grouped by effect in recessive (green) vs. dominant (blue) genes.* LARGE effect variants are found, on average, at lower proportions in OMIM annotated dominant genes, compared to OMIM recessive genes, consistent with purifying selection acting more efficiently in dominant genes, where the LARGE effect variants is more likely to be expressed.

## Figure S19: Mutational load in 1000 Genomes exome data



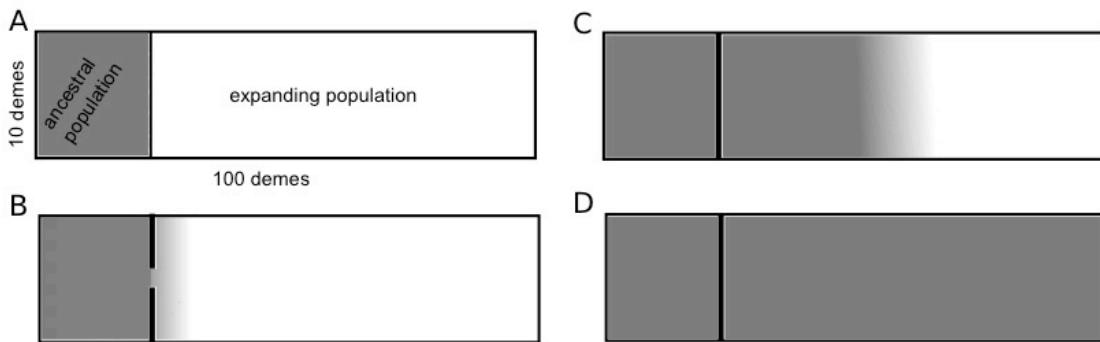
**Figure S19: Differences in Load - 1000 Genomes Dataset.** For each population, load is calculated under a recessive, intermediate and dominant model (as in **Figure 4**), reflecting contributions from variants with moderate, large and extreme effect.

## Figure S20: Distribution of highly differentiated variants vs. the genome



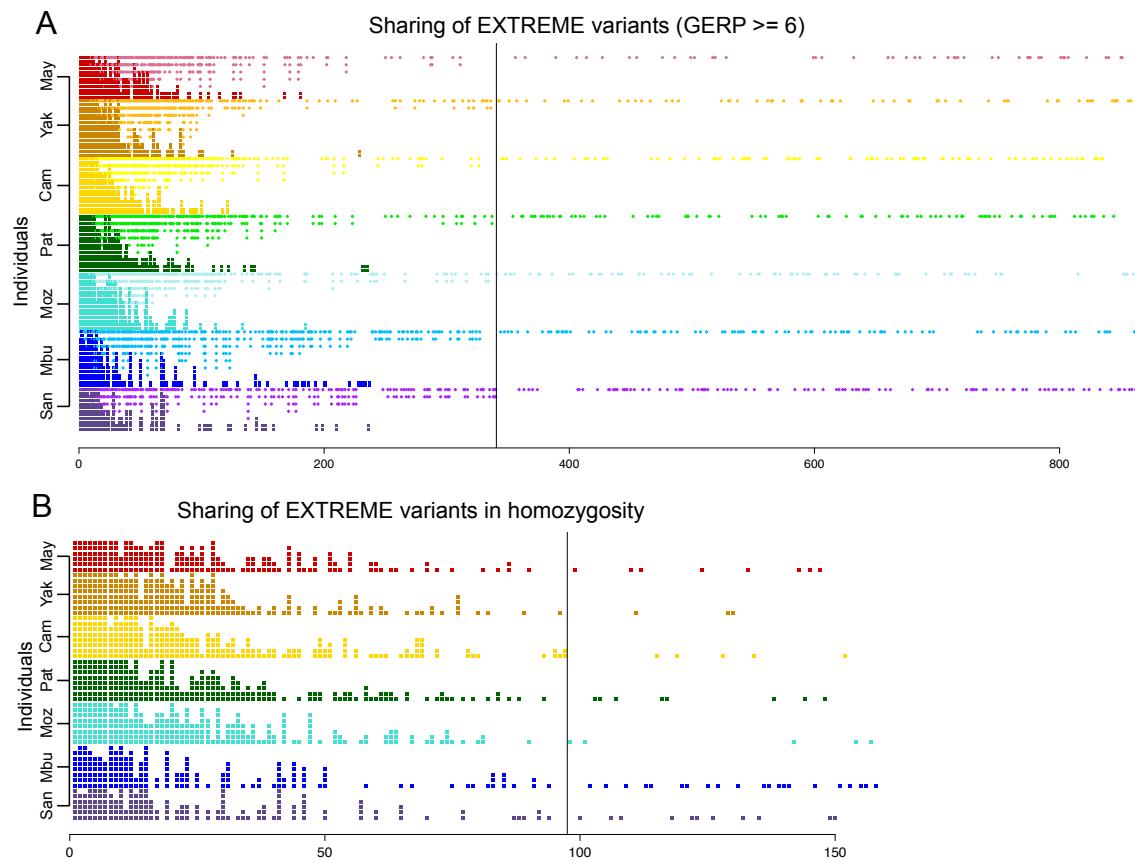
**Figure S20: Distribution of functional alleles in highly differentiated variants vs. the whole Genome.** **A)** Distribution of GERP scores across the Genome **B)** Distribution of GERP scores in highly differentiated variants for different demographically relevant population comparisons: (Afr-Afr), (Afr-OoA), (OoA,OoA). Results show now apparent differences in the distribution of functional variants in those two datasets.

**Figure S21: Schematic of the range expansion model**



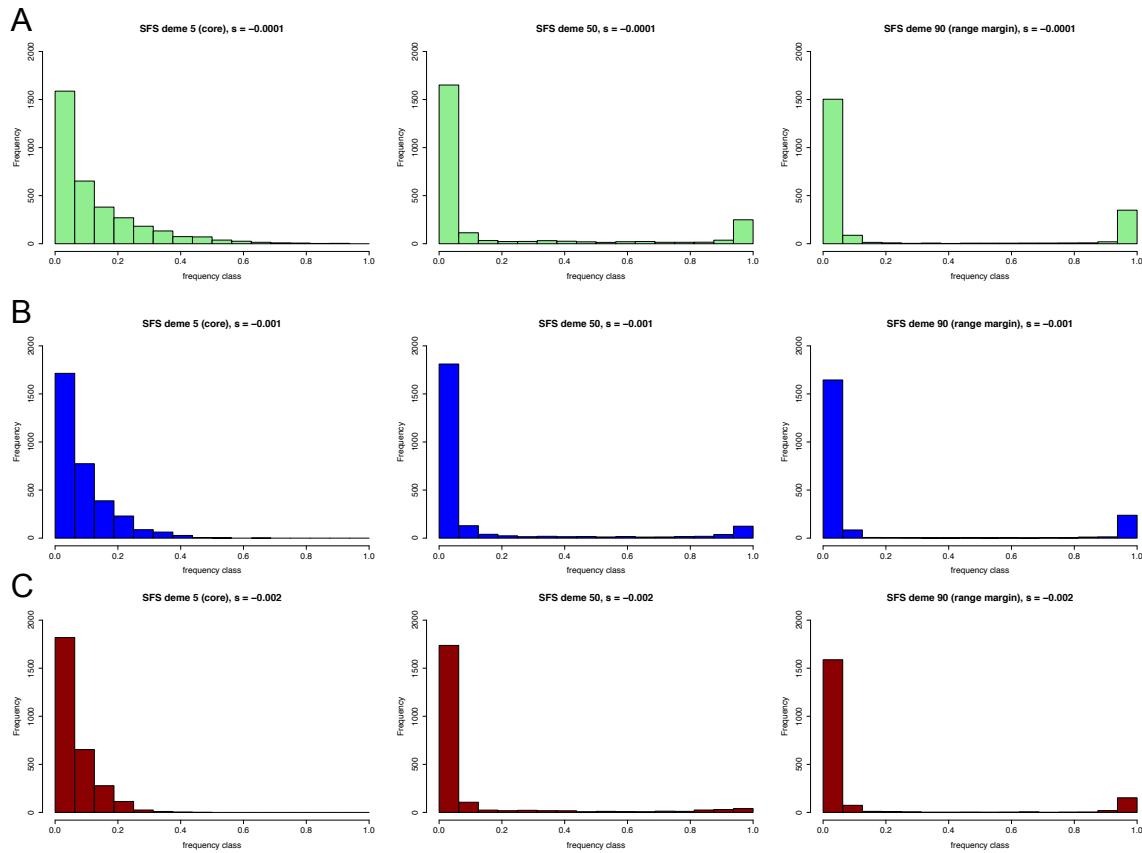
**Figure S21: Schematic of the range expansion model.** The model includes a spatial bottleneck we used to simulate the evolution of heterozygosity during a linear 2D expansion. Panel **A**) shows the ancestral population (gray) separated from the empty habitat by a migration barrier (black line). After a burn-in phase of 20,000 generations, a single deme in the middle of the migration barrier is removed for 5 generations, during which individuals from the ancestral population can migrate into the empty habitat. Panel **B**) shows the onset of the expansion and panel **C**) the colonization of the empty habitat by the expanding population (gray). Panel **D**) shows the whole metapopulation after the colonization is complete. Migration is bidirectional among demes in the simulation. For a similar simulation model, see (21).

## Figure S22: Sharing of GERP $\geq 6$ variants across populations



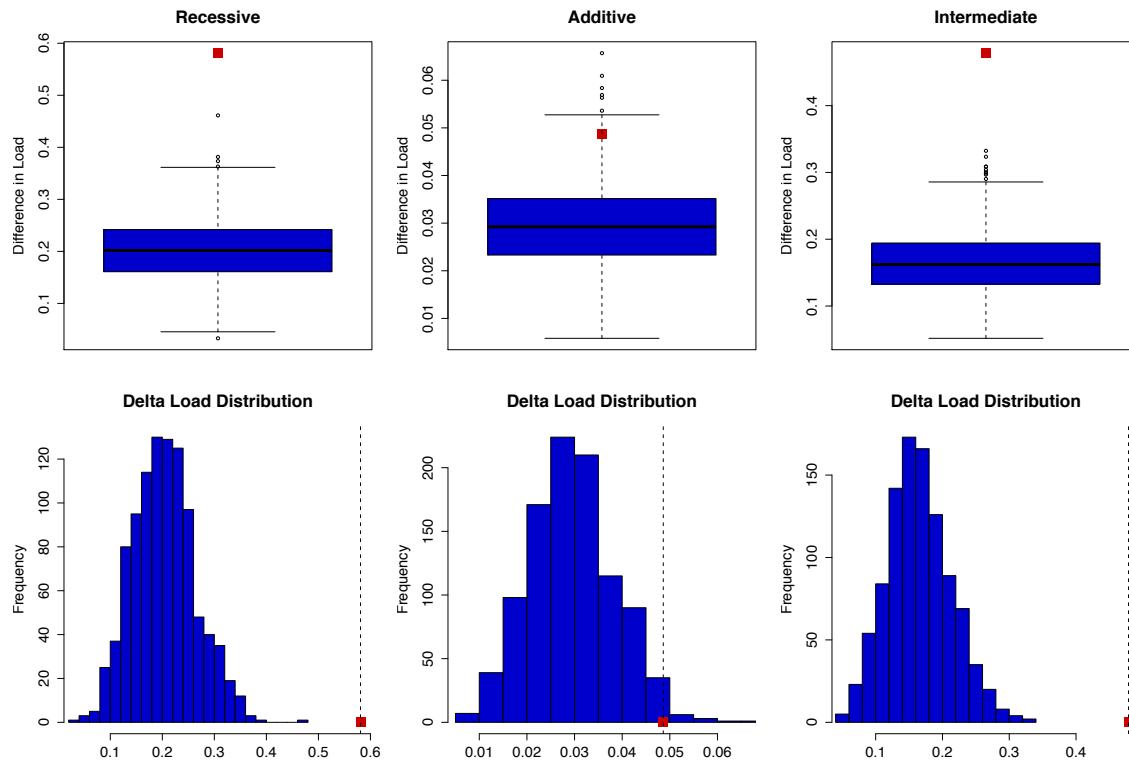
**Figure S22: Sharing of GERP  $>6$  Variants Across Populations.** **A)** For the 800 extreme variants we sorted alleles into homozygote and heterozygote states. Variants are sorted along the X-axis according to their global frequency in the dataset, with common variants on the left and rare variants on the right. In each population, the counts of heterozygotes are ordered in decreasing frequency from top to bottom. Homozygotes are ordered in the opposite fashion, with frequent counts on the bottom row and increasing toward the top within each population. The majority of variants are singletons, indicated to the right of grey line. Out of Africa populations carry more EXTREME variants at higher frequencies and share more EXTREME variants with each other than they share with African populations. Only a small number of GERP  $>6$  variants are fixed in African populations. **B)** A version of the homozygous GERP $>6$  variants is shown in the bottom panel.

## Figure S23: Site frequency spectrum under different selection regimes and locations of the range expansion.



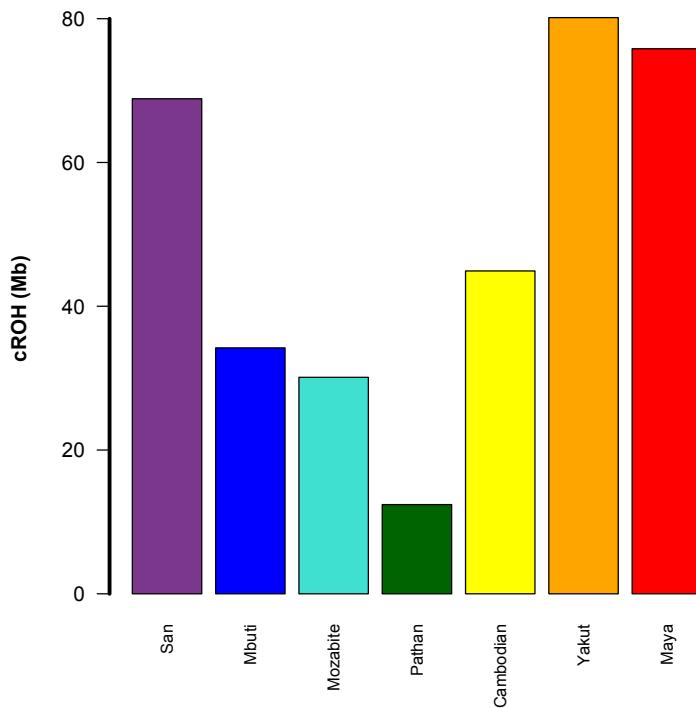
**Figure S23: Site frequency spectrum under different selection regimes and locations of the range expansion.** The site frequency spectrum was plotted for simulated demes from different locations under a range expansion model. Each row represents a different simulated selection coefficient, corresponding to **A**) moderate **B**) large **C**) extreme estimated effect. As the negative selection coefficient increases, the proportion of low frequency variants increases, and as geographic distance between the deme and the ancestral population increases, a greater amount of variants reaches fixation, even for highly deleterious variants.

**Figure S24: Testing significance in observed differences in load under the assumed models of dominance.**



**Figure S24: Testing significance in observed differences in Load under the assumed models of dominance.** A) Under each model, 1,000 iterations were performed where individuals were randomly re-assigned to populations and the maximum difference in mutation load was calculated. The observe difference in load is represented by a red square and the simulated differences are represented via boxplots. Under all three models the observed difference in Load is statistically significant with a p-value < 0.05 (See SI Methods). B) Distribution of the simulated differences in mutation load (blue) and the observed difference in load (red square).

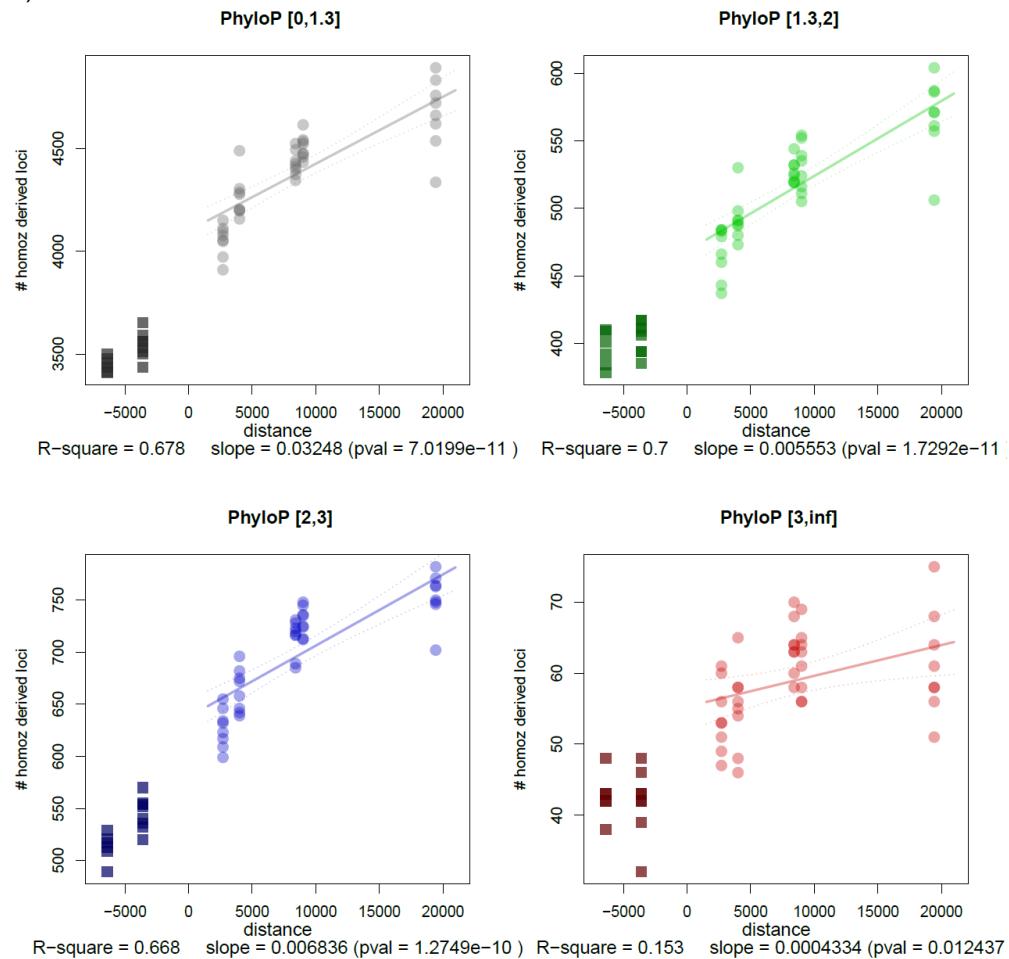
## Figure S25: Relationship between runs of homozygosity and mutation load



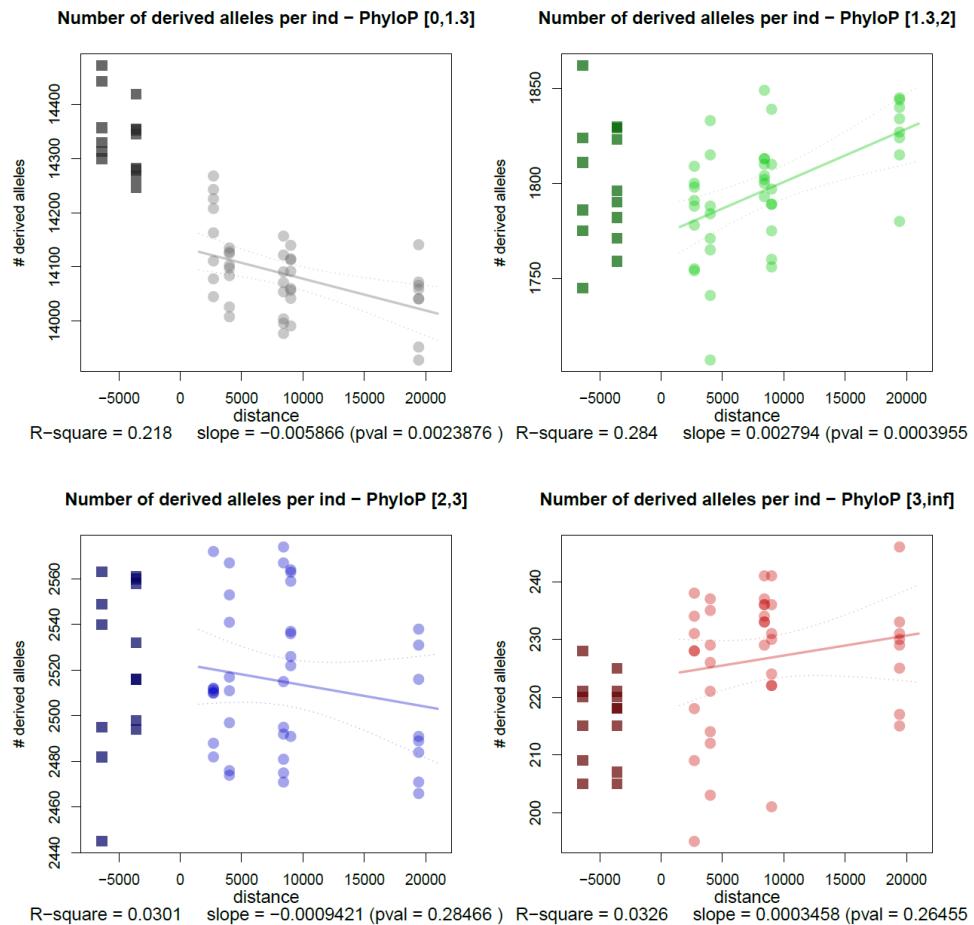
**Figure S25: Median cumulative runs of homozygosity (cROH) for HGDP populations.** The median number of cumulative runs of homozygosity was calculated for each of our seven populations from >600,000 SNPs obtained from SNP array data (2). Each ROH was at least 1 megabase (Mb) and contained a minimum of 25 SNPs. We allowed for 1 missing genotype per window and 2 heterozygotes per Mb in order to account for genotyping error rates. ROHs were calculated in *plink*. Long runs of homozygosity represent segments shared IBD between an individual's parents and represent inbreeding in the population. While strong genetic drift in Native Americans results in long ROH, endogamy in the San has also resulted in a substantial fraction of the genome in ROH (22) and shared IBD among members of the group (23).

**Figure S26: Annotation of variants with PhyloP and correlation with distance from Africa**

a)



b)



**Figure S26: Number of derived homozygous sites per individual and derived deleterious alleles per individual for different PhyloP categories.** A) The PhyloP categories are chosen such that the probability of a site to be neutral is  $>0.05$  (black, Neutral),  $0.05 < p < 0.01$  (green, Moderate),  $0.01 < p < 0.001$  (blue, Large) and  $p < 0.001$  (red, Extreme). B) The PhyloP categories are chosen such that the probability of a site to be neutral is  $>0.05$  (black, Neutral),  $0.05 < p < 0.01$  (green, Moderate),  $0.01 < p < 0.001$  (blue, Large) and  $p < 0.001$  (red, Extreme).

## Supplementary References:

1. DePristo MA et al. (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics* 43:491–498.
2. Li JZ et al. (2008) Worldwide Human Relationships Inferred from Genome-Wide Patterns of Variation. *Science* 319:1100–1104.
3. Browning BL, Yu Z (2009) Simultaneous Genotype Calling and Haplotype Phasing Improves Genotype Accuracy and Reduces False-Positive Associations for Genome-wide Association Studies. *The American Journal of Human Genetics* 85:847–861.
4. 1000 Genomes Project Consortium et al. (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature* 491:56–65.
5. Sunyaev S et al. (2001) Prediction of deleterious human alleles. *Human Molecular Genetics* 10:591–597.
6. Simons YB, Turchin MC, Pritchard JK, Sella G (2014) The deleterious mutation load is insensitive to recent population history. *Nature Genetics* 46:220–224.
7. Cooper GM et al. (2005) Distribution and intensity of constraint in mammalian genomic sequence. *Genome Research* 15:901–913.
8. Maples BK, Gravel S, Kenny EE, Bustamante CD (2013) RFMix: a discriminative modeling approach for rapid and robust local-ancestry inference. *Am J Hum Genet* 93:278–288.
9. Agrawal AF, Whitlock MC (2011) Inferences About the Distribution of Dominance Drawn From Yeast Gene Knockout Data. *Genetics* 187:553–566.
10. Li H, Durbin R (2011) Inference of human population history from individual whole-genome sequences. *Nature* 475:493–496.
11. Prüfer K et al. (2013) The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* 505:43–49.
12. Huang N, Lee I, Marcotte EM, Hurles ME (2010) Characterising and Predicting Haploinsufficiency in the Human Genome. *PLoS Genet* 6:e1001154.
13. Nelson MR et al. (2012) An Abundance of Rare Functional Variants in 202 Drug Target Genes Sequenced in 14,002 People. *Science* 337:100–104.
14. Fu W et al. (2013) Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nature* 493:216–220.
15. Andres AM et al. (2009) Targets of Balancing Selection in the Human Genome. *Mol Biol Evol* 26:2755–2764.
16. Hernandez RD et al. (2011) Classic Selective Sweeps Were Rare in Recent Human Evolution. *Science* 331:920–924.
17. Kimura M, Maruyama T, Crow JF (1963) The Mutation Load In Small Populations. *Genetics* 48:1303–1312.

18. Tennessen JA et al. (2012) Evolution and Functional Impact of Rare Coding Variation from Deep Sequencing of Human Exomes. *Science* 337:64–69.
19. Gazave E, Chang D, Clark AG, Keinan A (2013) Population growth inflates the per-individual number of deleterious mutations and reduces their mean effect. *Genetics* 195:969–978.
20. Chen H, Green RE, Paabo S, Slatkin M (2007) The Joint Allele-Frequency Spectrum in Closely Related Species. *Genetics* 177:387–398.
21. Peischl S, Dupanloup I, Kirkpatrick M, Excoffier L (2013) On the accumulation of deleterious mutations during range expansions. *Molecular Ecology* 22:5972–5982.
22. Szpiech ZA et al. (2013) Long runs of homozygosity are enriched for deleterious variation. *Am J Hum Genet* 93:90–102.
23. Henn BM et al. (2012) Cryptic Distant Relatives Are Common in Both Isolated and Cosmopolitan Genetic Samples. *PLoS ONE* 7:e34267.