nature genetics

The deleterious mutation load is insensitive to recent population history

Yuval B Simons[1,8], Michael C Turchin[2,8], Jonathan K Pritchard[2–5] & Guy Sella[6,7]

Human populations have undergone major changes in population size in the past 100,000 years, including recent rapid growth. How these demographic events have affected the burden of deleterious mutations in individuals and the frequencies of disease mutations in populations remains unclear. We use population genetic models to show that recent human demography has probably had little impact on the average burden of deleterious mutations. This prediction is supported by two exome sequence data sets showing that individuals of west African and European ancestry carry very similar burdens of damaging mutations. We further show that for many diseases, rare alleles are unlikely to contribute a large fraction of the heritable variation, and therefore the impact of recent growth is likely to be modest. However, for those diseases that have a direct impact on fitness, strongly deleterious rare mutations probably do have an important role, and recent growth will have increased their impact.

Recent work has highlighted the impact of demographic history on the distribution of human genetic variation. Deep sequencing studies have identified huge numbers of very rare variants in human populations, which are the consequence of explosive population growth in the past 5,000 years[1–6]. Additionally, Europeans and east Asians have a greater fraction of high-frequency variants compared to Africans, probably because of an ancient bottleneck of non-African populations[5,7–10].

Given these observations, it is natural to ask whether recent demographic history has affected the burden of genetic disease in modern human populations[3,6,11,12]. Keinan and Clark[3] recently hypothesized that "some degree of genetic risk for complex disease may be due to this recent rapid increase in the number of rare variants in the human population." A second important question concerns the relative importance of rare and common variants in causing disease[13–15]. If much of the genetic variation underlying disease is due to rare variants, it could help to explain the so-called 'missing heritability' of complex traits, implying that mapping approaches based on deep sequencing will be essential for the dissection of complex traits[16].

[1]Department of Ecology, Evolution and Behavior, Hebrew University of Jerusalem, Jerusalem, Israel. [2]Department of Human Genetics, University of Chicago, Chicago, Illinois, USA. [3]Howard Hughes Medical Institute, Stanford University, Stanford, California, USA. [4]Department of Biology, Stanford University, Stanford, California, USA. [5]Department of Genetics, Stanford University, Stanford, California, USA. [6]Department of Ecology and Evolution, University of Chicago, Chicago, Illinois, USA. [7]Present address: Department of Biological Sciences, Columbia University, New York, New York, USA. [8]These authors contributed equally to this work. Correspondence should be addressed to J.K.P. (pritch@stanford.edu) or G.S. (gsella@math.huji.ac.il).

Received 27 August 2013; accepted 16 January 2014; published online 9 February 2014; doi:10.1038/ng.2896

RESULTS

The model

To address these questions, we analyzed a theoretical model with a large number of biallelic sites, each of which was subject to two-way mutation, and natural selection against one of the alleles (Online Methods). We studied three types of demographic models thought to be relevant for human populations: (i) a bottleneck; (ii) exponential growth starting from a constant-sized population; and (iii) a complex demographic model for African Americans (including rapid recent growth) and European Americans (including two bottlenecks followed by growth) inferred by Tennessen *et al.*[5]. The main features of the Tennessen model are similar to those of other recent models[9,10,17], but the Tennessen model uses a larger data set for parameter estimation. Our main results focus on selection against semidominant (i.e., additive) alleles in which the three genotypes have fitnesses of 1, $1 - s/2$ and $1 - s$, where s is the selection coefficient, and selection against recessive alleles with genotype fitnesses 1, 1 and $1 - s$. The effects of demography in these two models are qualitatively representative of those over the range of dominance coefficients (**Supplementary Note**). In addition to the simulation results shown here, further results and detailed theoretical analysis for all our key results are provided in the **Supplementary Note**.

The impact of demographic changes on individual load

We focused first on the impact of demographic changes on individual load; that is, we wanted to understand whether demographic history has affected the burden of deleterious variation carried by a typical individual in a population. Individual load is related directly to the number of deleterious alleles carried by an individual or, for recessive mutations, the number of homozygous sites per individual (Online Methods and **Supplementary Note**).

Figure 1 illustrates the impact of a bottleneck and population growth on the numbers of deleterious variants when selection is strong (s = 1%). As we expected, these demographic events have a major impact on the number and frequency spectra of deleterious variants: the bottleneck causes a decrease in the total number of segregating sites in a population largely because of loss of rare variants, whereas the mean frequency of alleles that survive increases. Meanwhile, exponential growth causes a rapid increase in the number of segregating sites because of a major influx of rare variants but also causes a consequent drop in the mean frequency at segregating sites.

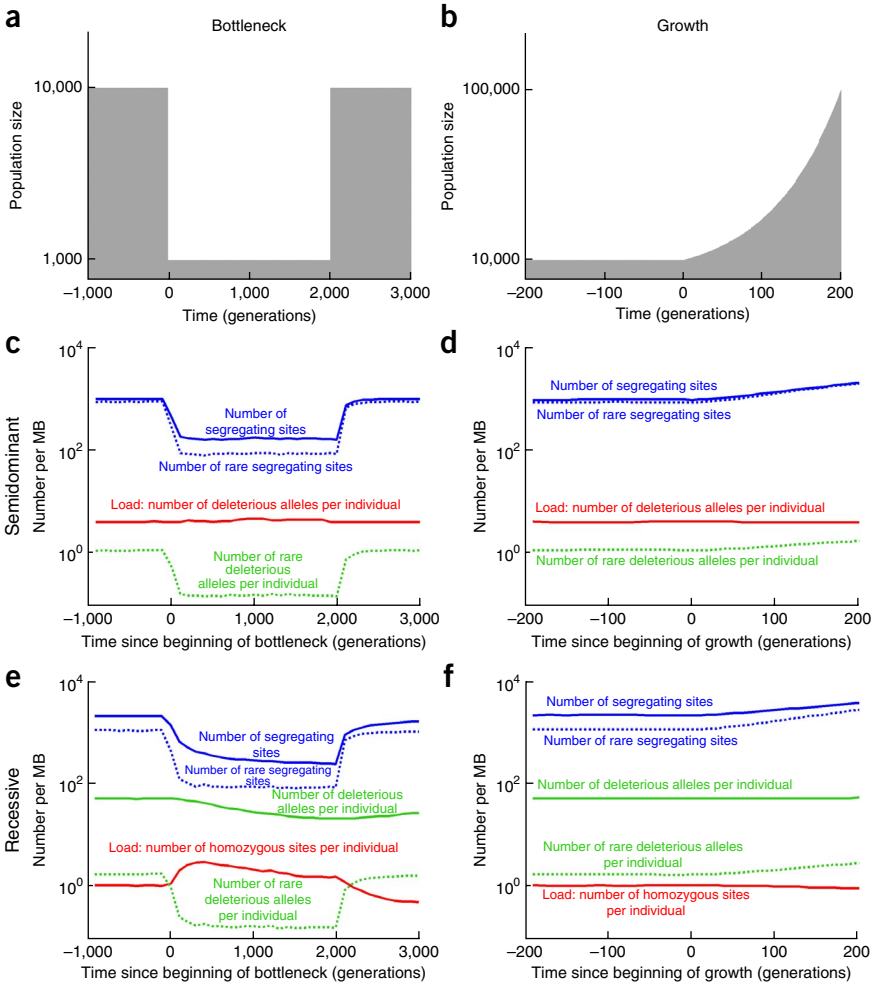

Figure 1 Time course of load and other key aspects of variation through a bottleneck and exponential growth. (**a**,**b**) The bottleneck (**a**) and exponential growth (**b**). (**c**–**f**) The expected number of variants and alleles per MB assuming semidominant mutations (**c**,**d**) or recessive mutations (**e**,**f**) with $s = 1\%$ and a mutation rate per site per generation of 10^{-8}.

But despite these substantial shifts in the overall frequency spectrum, the impact on genetic load—namely, the mean number of deleterious variants per individual and thus the average fitness—is much more subtle.

In the semidominant case, the individual burden is essentially unaffected by these demographic events (**Fig. 1c**,**d**). With growth, the increased number of segregating sites is balanced exactly by a decrease in the mean frequency (with the converse being true for the bottleneck model) so that the number of variants per individual stays constant. This kind of balance is predicted by classic mutation-selection balance models[18] and can be shown to hold for general changes in population size, provided that selection is strong and deleterious alleles are at least partially dominant (**Supplementary Note**).

The behavior of the recessive model is more complicated (**Fig. 1e**,**f**). In the bottleneck model, the mean number of deleterious variants per individual drops by 60% as a result of the bottleneck. This drop is due to the loss of rare alleles. However, during the bottleneck, some deleterious alleles drift to higher frequencies[11,19], contributing disproportionately to the number of homozygotes. This causes a transient increase in the number of deleterious homozygous sites per individual, i.e., the recessive load. Meanwhile, population growth has a less pronounced effect on recessive variation, leaving the mean number of deleterious alleles per individual unchanged but causing a slight decrease in load.

More generally, the manner in which demography affects individual load varies with the degree of dominance and the strength of selection (**Fig. 2**, **Supplementary Note** and **Supplementary Table 1**). The behavior of these models can be classified into three selection regimes: strong, weak and effectively neutral. In the case of strong selection, i.e., where selection is much stronger than drift (approximately $s \geq 10^{-3}$ for semidominant mutations), deleterious variants are extremely unlikely to fix, and virtually all of the genetic load is due to segregating variation. In this range, we infer that human demography has had no impact on semidominant load (and, more generally, for mutations with at least some dominance component) and has had only small effects on recessive load.

The case of weak selection—where drift and selection have comparable effects—is more complex, as fixed alleles may contribute appreciably to load, and the steady-state load depends on population size[20]. However, the approach to the steady state is very slow, being limited by both the time to fixation (on the order of $4N$ generations) and the mutational input (on the order of $1/2NU$ generations, where U is the mutation rate). For both the semidominant and recessive cases, population growth is too recent to have substantially decreased the load. Recent growth increases the input of new deleterious mutations, but this effect is counterbalanced by the fact that the new deleterious mutations are proportionally rarer, as well as by the input of beneficial mutations. The bottleneck in Europeans is estimated to have occurred further in the past and at much lower population sizes[5] (**Supplementary Fig. 1**), thus increasing its effect. In this case, the increase in drift causes segregating deleterious alleles to increase in frequency, sometimes reaching fixation, and results in a slight increase in load (**Supplementary Fig. 2**). The out-of-Africa bottleneck should thus lead to a slight increase of load in Europeans, most notably for recessive sites.

In the effectively neutral range—where selection has negligible effects on the population dynamics—segregating variation contributes negligibly, and hence the load does not change with demography. Thus, across all three selection regimes, recent human demographic history is likely to have had virtually no impact on genetic load at partially dominant sites and only weak effects at recessive sites.

Analysis of exome data

To test these predictions, we analyzed two recent data sets of exome sequences from individuals of west African and European descent. Previous work comparing load in different populations has produced conflicting conclusions depending on the data set, choice of measures and functional annotations used. For example, Lohmueller *et al.*[11] reported that there is "proportionally more deleterious variation in European than in African populations." Similarly, Tennessen *et al.*[5] found that European Americans had more nonreference genotypes when they used a conservative classification of deleterious sites but

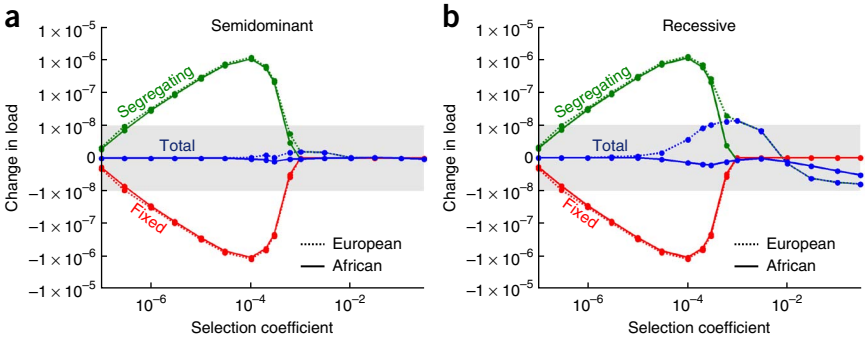

Figure 2 Changes in load due to changes in population size during the histories of European and African Americans. (**a**,**b**) Semidominant (**a**) and recessive (**b**) sites. The blue lines denote the difference in load per base pair of DNA sequence in the present-day population compared to the ancestral (constant) population size as a function of the selection coefficient. The green and red lines show the difference in load due to segregating and fixed variants, respectively. The increase in load due to segregating variation in modern populations approximately cancels out the decrease in load due to fixed sites. The scale on the *y* axis is linear within the gray region and is logarithmic outside this region.

observed the opposite result when using a more liberal classification of sites (both observations were highly significant).

We first analyzed single-nucleotide variant (SNV) frequency data from a recent exome sequencing study of 2,217 African Americans (AAs) and 4,298 European Americans (EAs) sequenced at 15,336 protein-coding genes by Fu *et al.*[6] (the allele frequencies are available from the National Heart, Lung, and Blood Institute (NHLBI) Grand Opportunity (GO) Exome Variant Server). Additionally, we analyzed exome data from 88 Yoruba (YRI) and 81 European (CEU) individuals collected by the 1000 Genomes Project[21].

To test whether there are differences in load between individuals of west African and European descent, we considered the average number of derived alleles per individual at putatively deleterious segregating sites. For this purpose, we considered a site to be segregating if and only if it is variable within the combined sample of both populations. This definition ensures that the derived counts are comparable across populations. Under a semidominant model, the number of derived alleles increases monotonically with the segregating genetic load. Thus, any difference in average load between populations would be apparent as a difference in the mean number of derived alleles per individual. Here we focused on an equivalent measure that also facilitates comparisons across different types of sites, namely, the mean derived allele frequency within functional classes. The mean derived allele frequency is equal simply to the number of derived alleles per individual divided by twice the number of segregating sites in that class, and so any difference in the mean number of derived alleles per individual will also be a difference in the mean derived frequencies. For sites that are either neutral or semidominant, our model predicts that the mean derived allele frequency should be virtually identical in Africans and Europeans (**Supplementary Note** and **Supplementary Fig. 3**). At recessive sites, we expect a slight increase in mean derived frequency in Africans compared to Europeans (**Supplementary Fig. 3**), but overall we expect any differences to be small.

We obtained functional predictions of SNVs from PolyPhen-2, which employs a method that uses sequence conservation and structural information to infer which nonsynonymous changes are most likely to have functional consequences[22] (**Supplementary Table 2** shows similar analyses with other functional prediction methods). When using the functional predictions, we observed a strong bias: SNVs for which the genome reference carries the derived allele are much more likely to be classified as benign than SNVs for which the reference allele is ancestral—this observation was true even when we controlled for the overall population frequency (**Supplementary Fig. 4**). Hence, our analysis incorporates a correction to account for this bias; we obtained very similar results using a separate set of unpublished human-independent PolyPhen scores provided by the Sunyaev lab (**Supplementary Tables 3** and **4**).

Figure 3 summarizes the results from the data of Fu *et al.*[6]. The mean allele frequency declines with increasing functional severity[5] from 2.8% at noncoding SNVs to 0.6% at probably damaging SNVs, implying that there is selection against most SNVs with predicted damaging effects. More striking, however, is the finding that within each of the five functional categories, the mean allele frequencies—and hence the numbers of derived alleles per individual—are essentially identical in the two populations despite the very large size of the data sets ($P > 0.05$ for all five comparisons). Results from the 1000 Genomes Project data are qualitatively similar: we found no significant differences between the YRI and CEU populations in the numbers of derived alleles per individual in any functional category (**Supplementary Table 5**).

In summary, these observations are consistent with our model predictions that load should be very similar in these populations. Our conclusions probably differ from those of previous studies in part because earlier studies used measures that are related to load but are also sensitive to other differences between the populations being compared (for example, the number of neutral segregating sites and the frequency spectrum) and in part because of the reference bias in the functional annotations accounted for here (**Supplementary Note**). We note that D. Reich, S. Sunyaev and colleagues have recently made similar observations regarding load in different populations (personal communication).

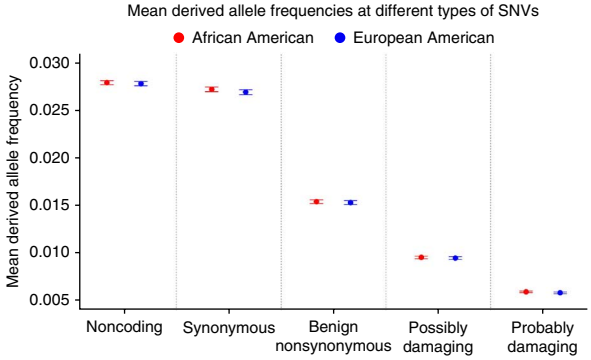

Figure 3 Observed mean allele frequencies in AAs and EAs at various classes of SNVs. The plot shows the mean frequencies in each population (± 2 s.d.) using exome sequence data from Fu *et al.*[6]. Here a site is considered a SNV if it is segregating in the combined AA-EA sample of 6,515 individuals. The functional classifications of sites are from PolyPhen-2 (ref. 22) with bias-correcting modifications. The AA and EA mean frequencies are essentially identical within all five functional categories ($p > 0.05$).

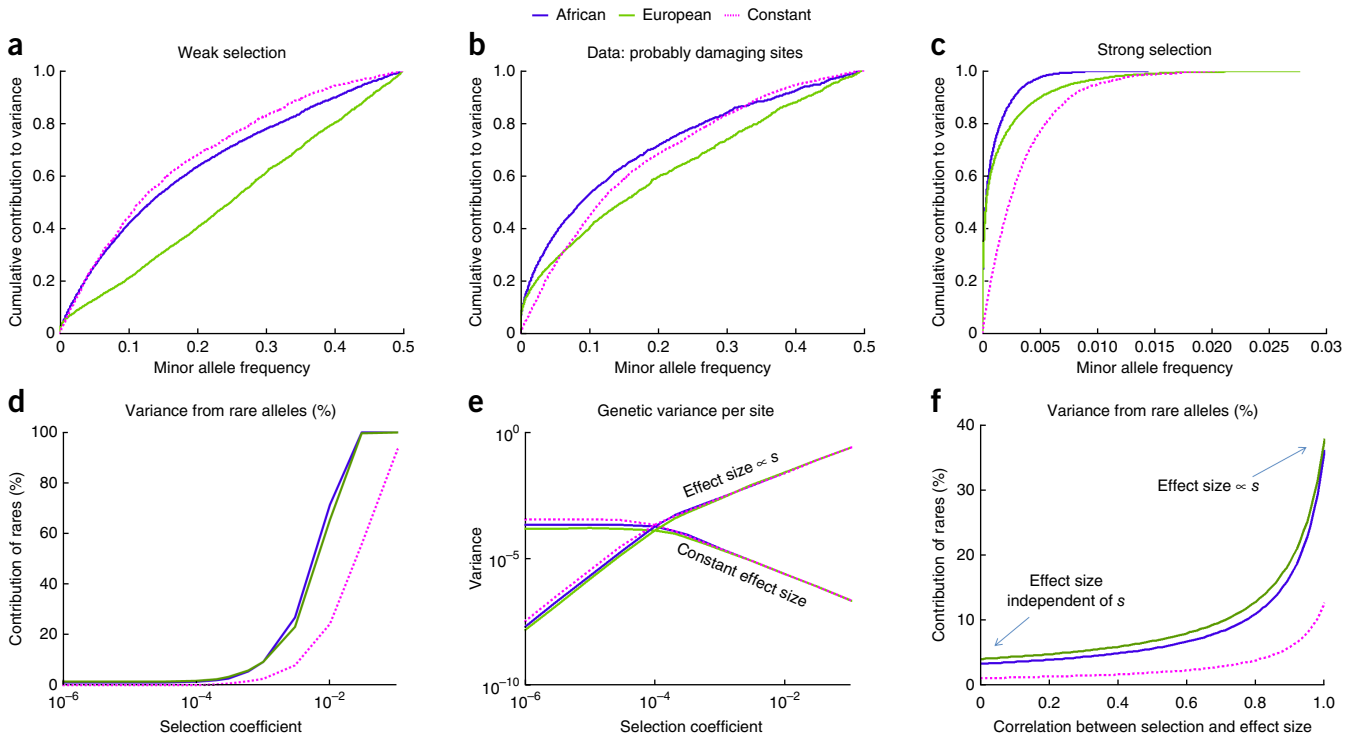

Figure 4 Predicted effect of demography on the genetic architecture of disease risk. All plots (**a**–**f**) assume an additive trait and, with the exception of **b**, are based on simulations with semidominant selection under the Tennessen *et al.*[5] demographic model. Results for the constant population size model are also provided for comparison. The upper plots (**a**–**c**) show the cumulative fractions of genetic variance due to alleles at frequency $<x$ based on simulated data with weak selection ($s = 0.0002$; **a**); assuming the observed frequency spectrum at probably damaging sites[6,22], where a constant population size of 14,474 and selection coefficient of 0.02% are used for comparison (**b**); and simulated data with strong selection ($s = 0.01$; **c**). (**d**–**f**) The fraction of variance due to rare alleles (i.e., <0.1%) as a function of the selection coefficient (**d**); the per-site contribution to variance as a function of the selection coefficient under two extreme models, with effect sizes that are either independent of s (constant) or proportional to s (**e**); and the expected fraction of the variance due to rare variants (i.e., <0.1%) as a function of the correlation between the selection on and the effect sizes of variants (**f**). Further details on the model are provided in the Online Methods.

The impact of demography on genetic architecture

Although changes in population size have had little impact on the average load carried by individuals, growth has greatly increased the number of rare variants in populations. So do rare variants have a greater (and substantial) role in the genetics of disease as a result of recent growth (**Fig. 4**)? Given the differences in population history, do higher-frequency variants have a greater role in Europeans and Asians than in Africans? The answers to these questions are of practical importance because different study designs may be needed to identify rare variants[13,15,16,23].

To study these questions, we computed the contributions of different allele frequencies to the heritable phenotypic variation among individuals in the population, namely $x(1-x)f(x)/2$, where $f(x)$ is the probability that a derived allele is at frequency x given the demographic model and selection coefficient. These distributions show the fraction of genetic variance for a disease that is contributed by alleles below frequency x for the simplest case in which the loci underlying the trait all have the same effect size and selection coefficient and are all semidominant (**Supplementary Note**). In practice, we anticipate that variants underlying a given disease would have a variety of selection coefficients and effect sizes, in which case the overall distribution would be an appropriately weighted mixture of distributions for different selection coefficients. Of note, here we consider the proportional contribution of variants at different frequencies, and thus these results should hold regardless of the number of loci underlying variation in the trait.

Analysis of this model shows several interesting points. For effectively neutral or weakly deleterious sites (**Fig. 4a**), only a small fraction of the total variance comes from very rare alleles: although there are many rare alleles, each one contributes very little to population variance and individual load. The same is true for recessive variation across almost the entire range of selection coefficients (**Supplementary Note** and **Supplementary Fig. 5**). Likewise, if we assume that the frequency density $f(x)$ follows the frequency spectrum observed at all nonsynonymous sites classified as probably damaging[22], then under the same model, it is still only a modest fraction of the genetic variance that is due to rare alleles (**Fig. 4b** and ref. 5). Meanwhile, in all of these cases, the out-of-Africa bottleneck increases the contribution of intermediate-frequency alleles to the genetic variance (**Fig. 4a**–**c**); for example, at probably damaging sites, 62% of the variance in EAs is contributed by alleles with minor allele frequency above 10% as compared to only 49% in AAs.

It is only for the case of strong, dominant selection that very rare variants (<0.1%) become important (**Fig. 4c**,**d**). For example, for a selection coefficient of 1%, most of the variation is due to rare alleles that arose within the recent exponential-growth phase. As a result, the contribution of extremely rare variants is much greater than it would have been in the absence of growth; for example, in AAs and EAs, 80% and 65% of the variance, respectively, is due to alleles below frequency 0.1% compared to just 25% in the constant population model.

In practice, the genetic variants that contribute to a complex trait probably have a range of selection coefficients (s) and a range of effect sizes (a) on the phenotype in question (**Supplementary Note**). When

there is a mixture of selective coefficients and effect sizes, what can we say about the relative importance of rare and common variants? The answer crucially depends on the relationship between a and s[14,24]. To illustrate this dependence, we consider two extreme cases: (i) a is independent of s, namely, the trait itself has little effect on fitness, but specific variants could have fitness consequences because of pleiotropic effects on other phenotypes; and (ii) a is proportional to s, a model that is likely most relevant for traits with a direct impact on fitness, such as early onset diseases or diseases affecting fertility. **Figure 4e** shows the expected contribution of each site to genetic variance as a function of s under these two models. When a is independent of s, we would expect weakly selected mutations to contribute most of the variance because they have the same average effect on the trait but can drift to higher frequencies. But the reverse occurs in the model in which a increases with s: highly deleterious, rare mutations will have a greater contribution to variance because their increased effect size outweighs their lower frequencies.

Many traits presumably lie between these two extreme cases. To study how demography affects genetic architecture across this range, we consider a second model. In this model, we assume that the heritable variance in a trait is due to a mixture of weakly ($s = 0.0002$) and strongly ($s = 0.01$) selected mutations, and we vary the correlation between selection on a variant and its effect on the trait (Online Methods). **Figure 4f** shows how the contribution of rare alleles to genetic variance changes with the correlation between the selection coefficient and effect size. As can be seen in the case with a constant population size, the contribution of rare variants becomes substantial only when the effects of variants on fitness and on the trait are highly correlated (presumably because the trait itself is strongly coupled with fitness). Although growth affects the frequencies of strongly selected alleles regardless of the correlation, it will have a substantial effect on the genetic architecture only for traits in which strongly selected alleles contribute substantially to variance. In this case, we see that recent growth greatly amplifies the contribution of rare alleles to variance. A similar argument implies that the out-of-Africa bottleneck should have substantially increased the contribution of intermediate-frequency alleles to the variance, unless the effects of variants on fitness and on the trait are highly correlated, in which case rare alleles will still dominate.

DISCUSSION

Although recent demographic events have had well-documented effects on the frequency spectrum of SNVs in modern populations, we find that these events have had negligible impact on the average burden of mutations carried by individuals. Moreover, we conclude that although there are large absolute numbers of rare variants, they do not necessarily contribute a large fraction of the genetic variance underlying complex traits. An earlier paper from one of the present authors (Pritchard, 2001 (ref. 13)) also discussed the possible role of allelic heterogeneity and rare variants in disease using a model that is closer to the independent s model examined here. Although the earlier model is not comparable exactly to our present work, the overall results are broadly consistent, as the bulk of the genetic variance was predicted to be due to variants that would not be considered rare by modern standards. To summarize, it is only for diseases that are caused primarily by strongly deleterious mutations that we can expect much of the variance to be due to rare alleles; these will likely be diseases that are tightly coupled to fitness.

URLs. NHLBI GO Exome Variant Server, http://evs.gs.washington.edu/EVS; 1000 Genomes public server, ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/.

METHODS

Methods and any associated references are available in the online version of the paper.

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

ACKNOWLEDGMENTS

This work was supported by grants from the US National Institutes of Health (NIH) (MH084703 to J.K.P. and GM083228 to G.S.), the Israel Science Foundation (grant 1492/10 to G.S.) and the Howard Hughes Medical Institute (J.K.P.). M.C.T. was supported in part by NIH grant T32 GM007197. We thank D. Reich and S. Sunyaev for helpful discussions and generous input regarding the interpretation of PolyPhen-2, I. Adzhubey for human-independent PolyPhen scores, J. Akey for assistance in accessing data, J. Akey, A. Siepel, G. Coop, I. Eshel, R. Hudson and two anonymous reviewers for comments on the manuscript and M. Przeworski for many discussions and comments on the manuscript.

AUTHOR CONTRIBUTIONS

J.K.P. and G.S. conceived and supervised the research. Y.B.S., G.S. and J.K.P. developed theory. Y.B.S. performed simulations. M.C.T. and J.K.P. performed data analysis. J.K.P. and G.S. wrote the manuscript with input from Y.B.S. and M.C.T.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Coventry, A. *et al.* Deep resequencing reveals excess rare recent variants consistent with explosive population growth. *Nat. Commun.* **1**, 131 (2010).
2. Marth, G.T. *et al.* The functional spectrum of low-frequency coding variation. *Genome Biol.* **12**, R84 (2011).
3. Keinan, A. & Clark, A.G. Recent explosive human population growth has resulted in an excess of rare genetic variants. *Science* **336**, 740–743 (2012).
4. Nelson, M.R. *et al.* An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people. *Science* **337**, 100–104 (2012).
5. Tennessen, J.A. *et al.* Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* **337**, 64–69 (2012).
6. Fu, W. *et al.* Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nature* **493**, 216–220 (2013).
7. Keinan, A., Mullikin, J.C., Patterson, N. & Reich, D. Measurement of the human allele frequency spectrum demonstrates greater genetic drift in East Asians than in Europeans. *Nat. Genet.* **39**, 1251–1255 (2007).
8. Wall, J.D. & Przeworski, M. When did the human population size start increasing? *Genetics* **155**, 1865–1874 (2000).
9. Voight, B.F. *et al.* Interrogating multiple aspects of variation in a full resequencing data set to infer human population size changes. *Proc. Natl. Acad. Sci. USA* **102**, 18508–18513 (2005).
10. Gutenkunst, R.N., Hernandez, R.D., Williamson, S.H. & Bustamante, C.D. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet.* **5**, e1000695 (2009).
11. Lohmueller, K.E. *et al.* Proportionally more deleterious genetic variation in European than in African populations. *Nature* **451**, 994–997 (2008).
12. Casals, F. & Bertranpetit, J. Human genetic variation, shared and private. *Science* **337**, 39–40 (2012).
13. Pritchard, J.K. Are rare variants responsible for susceptibility to complex diseases? *Am. J. Hum. Genet.* **69**, 124–137 (2001).
14. Eyre-Walker, A. Genetic architecture of a complex trait and its implications for fitness and genome-wide association studies. *Proc. Natl. Acad. Sci. USA* **107**, 1752–1756 (2010).
15. Gibson, G. Rare and common variants: twenty arguments. *Nat. Rev. Genet.* **13**, 135–145 (2011).
16. Manolio, T.A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747–753 (2009).
17. Schaffner, S.F. *et al.* Calibrating a coalescent simulation of human genome sequence variation. *Genome Res.* **15**, 1576–1583 (2005).
18. Hartl, D.L. *A Primer of Population Genetics* (Sinauer Associates, Inc., 2000).
19. Travis, J.M. *et al.* Deleterious mutations can surf to high densities on the wave front of an expanding population. *Mol. Biol. Evol.* **24**, 2334–2343 (2007).
20. Lynch, M., Conery, J. & Burger, R. Mutational meltdowns in sexual populations. *Evolution* **49**, 1067–1080 (1995).
21. The 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).
22. Adzhubei, I.A. *et al.* A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248–249 (2010).
23. Thornton, K.R., Foran, A.J. & Long, A.D. Properties and modeling of GWAS when complex disease risk is due to non-complementing, deleterious mutations in genes of large effect. *PLoS Genet.* **9**, e1003258 (2013).
24. Johnson, T. & Barton, N. Theoretical models of selection and mutation on quantitative traits. *Phil. Trans. R. Soc. Lond. B* **360**, 1411–1425 (2005).

ONLINE METHODS

Model. Our basic model starts by considering selection at a single site. We use the standard biallelic diploid model with two-way mutation, viability selection, drift and, in some cases, migration[25]. Specifically, we assume there are two possible alleles at each site: normal (N) and deleterious (D). An N allele mutates to the D allele with probability u per gamete, per generation, and the reverse mutation occurs with probability v. Unless noted otherwise, we assume that mutation is symmetric, i.e., $u = v$. The absolute fitness of the three genotypes NN, ND and DD are 1, $1 - hs$ and $1 - s$, respectively, where $s > 0$ and the dominance coefficient $h \geq 0$. We focus on semidominant ($h = 1/2$) and fully recessive ($h = 0$) selection, as these two cases exhibit the full range of qualitative behaviors, with selection acting primarily on heterozygotes when $h = 1/2$ and only on homozygotes when $h = 0$. Allele frequencies in the next generation follow from Wright-Fisher sampling with these viabilities, sometimes with migration, and the population size and migration rates vary according to the demographic scenario considered.

We assume that fitness is multiplicative across sites and that there is linkage equilibrium among sites. Under these assumptions, the evolutionary dynamics at each site are independent from those at all other sites. In practice, linked selection is likely to have negligible effects on differences between populations because, to a first approximation, this reduces the effective population size at a given site by similar proportions regardless of demographic history, and these effects are thought to be modest in humans (compare to ref. 26).

Demographic scenarios. We consider three demographic scenarios. The most detailed is the out-of-Africa demographic model for AAs and EAs estimated by Tennessen *et al.*[5] (**Supplementary Fig. 1a**). The model includes the out-of-Africa split of European ancestors, changes in population size before and after the split (specifically, a severe bottleneck in Europeans after the split and recent rapid growth in both Europeans and Africans) and migration between the populations after the split. In addition, the model includes recent admixture between the populations, which we include in our simulations only when we compare our results to data from AAs.

We also study two simpler demographic scenarios (**Supplementary Fig. 1b**,**c**). To understand the effects of the recent explosive growth of human populations, we use a simple model of exponential growth from a population of constant size, and to investigate the effects of the bottleneck in Europeans at the out-of-Africa split, we consider a simple model of a bottleneck where population size instantaneously changes to a lower value at which it stays constant until it instantaneously reverts back to its original size.

Simulations. For each demographic scenario, we run simulations of a single site for the semidominant and recessive cases and vary the selection coefficient such that the strength of selection ranges from effectively neutral to strong. Each run begins with one of the two alleles fixed, where the proportion of runs that start with each allele is given by the expectation at equilibrium. A burn-in period of $\geq 10N$ generations with constant population size N follows to ensure an equilibrium distribution of segregating sites. The initial state is defined as ancestral, and the other state is defined as derived; the derived and deleterious allele frequencies are recorded at the end of the simulation. The code is written in C++ and is available by request (**Supplementary Note** and **Supplementary Figs. 6–8**).

Load. Genetic load is defined as the relative reduction in average fitness caused by deleterious alleles compared to the maximum absolute fitness[25]. In our model, the maximal absolute fitness equals 1, allowing us to directly consider differences in average fitness in populations with different demographic histories. Given our model, the average fitness function can be written as

$$\bar{W} \approx \exp(-\sum_{j=1}^{M} l(h_j, s_j))$$

where

$$l(h,s) \equiv 2hsE(pq) + sE(q^2) = s(2hE(q) + (1-2h)E(q^2)),$$

relates the quantities at a locus with load, p and q are the beneficial and deleterious allele frequencies at a locus ($p + q = 1$), and h_j and s_j are the dominance and selection coefficient at locus j. For a model with a single site and $s \ll 1$, $l(h, s)$ coincides with the definition of load. For more than one site, load is a simple function of the sum over all $l(h, s)$. For brevity, we therefore refer to $l(h, s)$ as load.

Change in load. To assess whether there has been a change in load due to demography, we consider the difference between the load at the present time and the load before recent demographic events. Specifically, in the exponential and bottleneck models, the reference time is before the change in population size, and in the Tennessen model, the reference time is the split between the African and European populations (**Supplementary Note**, **Supplementary Figs. 2** and **9–20** and **Supplementary Table 1**).

Data analyses. We used exome resequencing data from Fu *et al.*[6] and from the 1000 Genomes Project[21]. Allele frequency estimates from Fu *et al.*[6] are available from the NHLBI GO Exome Variant Server. These data provide estimates of the derived allele frequencies (DAFs) at exonic SNVs in EAs and AAs. 1000 Genomes Project vcf files (phase 1, version 3) were downloaded from the official 1000 Genomes public server. YRI and CEU individuals with (at least) exome sequencing coverage were extracted from the original vcf files (88 YRI individuals and 81 CEU individuals). 7 YRI individuals, chosen at random, were removed to match the sample sizes between the YRI and CEU groups. Variants that were fixed for either allele in both populations were removed. Any variant that was not a SNV or did not contain ancestral allele information was also dropped.

The ANNOVAR suite of scripts[27] was used to obtain functional predictions for each SNP from each of four prediction methods: PolyPhen-2 (ref. 22), SIFT[28], LRT[29] and MutationTaster[30]. We observed a strong reference bias in the functional classifications for all four prediction methods: sites at which the reference genome carries the derived allele are much more likely to be classified as benign than are sites at which the reference is ancestral; this is a very strong effect even when we control for the true population frequency in a very large sample (**Supplementary Fig. 4**) and hence does not simply reflect the tendency for common alleles to be less functional. We therefore treated the functional designations at sites where the genome reference is derived as unreliable. To deal with this problem, we used a simple procedure to estimate the probability that each reference-derived site would have been classified as damaging had the reference allele been ancestral (conditional on the overall population frequency). Specifically, we binned SNVs by overall population frequency in the full sample, and for each bin, we determined the fraction of reference-ancestral sites in each functional category. For SNVs in that bin that are reference derived, we treated those fractions as estimates of the probability that these SNVs would have been in each functional category had they instead been reference ancestral. Next, to estimate the mean DAF for each functional category, we summed across all sites in that category that were reference ancestral and added a contribution from all sites that were reference derived, weighted according to the estimated probability that the site would have been in the relevant functional category if it had been reference ancestral. We also provide supplementary results (**Supplementary Table 3**) in which we used a new unpublished version of PolyPhen's PSIC scores that are calculated in a human-independent (i.e., unbiased) manner and obtained qualitatively similar results. We thank I. Adzhubey and S. Sunyaev for prepublication access to these data.

We calculated mean derived frequencies within functional categories and the corresponding standard errors (calculated as s.d.(DAF)/$\sqrt{}$(number of sites)). Individual-level counts for the 1000 Genomes data simply counted the numbers of derived alleles per individual within a functional class (there are no missing genotypes in this data set, as these data have been imputed by the 1000 Genomes Project). For each population and functional category, we estimated the s.d. of the mean number of derived alleles per individual by bootstrapping across sites. This method is more appropriate than computing the standard error directly from the distribution of the derived allele counts across individuals, as the latter method ignores variation in the genealogical process. Because we are working with mean allele counts or frequencies, these analyses are unaffected by linkage disequilibrium or Hardy-Weinberg disequilibrium (which may affect variances but not means).

Our analysis effectively uses the derived allele count as a proxy for the deleterious allele count. Hence, there will be a low rate of misclassification

at weakly selected sites for which the deleterious allele is ancestral. However, this does not change the qualitative predictions about patterns of differences between populations, and we expect the number of derived alleles to have a monotonic relationship with the number of deleterious alleles. Specifically, for sites that are either neutral or semidominant, we predict that this measure should yield virtually identical counts in AAs and EAs (**Supplementary Note** and **Supplementary Fig. 20**). At recessive sites, our model predicts slight differences (**Supplementary Note**), but overall we expect these differences to be negligibly small. When SNVs are defined within populations, as has been done in some previous papers, these simple predictions do not hold.

Models for variance. We consider how the relationship between the effects of mutations on fitness and a trait affect the genetic architecture. For that purpose, we calculate the expected contribution of mutations to the heritable variation in a trait. We assume an additive trait and that the fitness effects of mutations are semidominant. At a site with selection coefficient s, the expected contribution to the variance from deleterious alleles below frequency ω is therefore

$$V_\omega(s) = \tfrac{1}{2}CE(a^2 \mid s)\int_0^\omega f(x \mid s)x(1-x)dx$$

where $E(a^2|s)$ is the expectation of the squared effect size, $f(x|s)$ is the probability of the deleterious allele being at frequency x (without conditioning of the site being segregating, i.e., including $x = 0$ and $x = 1$) and C is a proportion coefficient (**Supplementary Note**). A site's expected contribution to variance is $V_1(s)$ and the proportional contribution from variants below frequency ω is

$$\Theta_\omega(s) \equiv \frac{V_\omega(s)}{V_1(s)}$$

Note that while $V_1(s)$ depends on the relationship between selection coefficients and effect sizes, $\Theta_\omega(s)$ does not. When all sites are considered jointly, denoting the input of mutations with selection coefficient s by $\mu(s)$, the expected proportion of variance from deleterious alleles below frequency ω is

$$\Theta_\omega = \frac{\int_s \mu(s)V_1(s)\Theta_\omega(s)ds}{\int_s \mu(s)V_1(s)ds}$$

As an illustration, we consider a simple model in which we vary the correlation between selection on variants and their effects on a trait. We assume that half of the newly arising mutations have a weak selection coefficient $s_w = 0.0002$ and half have a strong selection coefficient of $s_s = 0.01$. For strongly selected mutations, the effect size on the trait, a, is chosen to be cs_s with probability $1/2(1 + p)$ and cs_w with probability $1/2(1 - p)$, where c is a positive constant and $0 \le p \le 1$; correspondingly, for weakly selected mutations, the effect size is chosen to be cs_w with probability $1/2(1 + p)$ and cs_s with probability $1/2(1 - p)$. In this model, the marginal distributions of selection coefficients and effect sizes do not depend on p, whereas the correlation between them is equal to p. To obtain **Figure 4f** we therefore varied p between 0 and 1. In **Figure 4e**, we consider the two extremes ($p = 0$ and $p = 1$).

25. Charlesworth, B. & Charlesworth, D. *Elements of Evolutionary Genetics* (Roberts and Company Publishers, 2010).
26. McVicker, G., Gordon, D., Davis, C. & Green, P. Widespread genomic signatures of natural selection in hominid evolution. *PLoS Genet.* **5**, e1000471 (2009).
27. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
28. Kumar, P., Henikoff, S. & Ng, P.C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* **4**, 1073–1081 (2009).
29. Chun, S. & Fay, J.C. Identification of deleterious mutations within three human genomes. *Genome Res.* **19**, 1553–1561 (2009).
30. Schwarz, J.M., Rödelsperger, C., Schuelke, M. & Seelow, D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat. Methods* **7**, 575–576 (2010).

doi:10.1038/ng.2896