bioRxiv preprint doi: https://doi.org/10.1101/010934; this version posted March 11, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

When is selection effective?

Simon Gravel, McGill University, Montreal, Canada

March 11, 2016

Abstract

Deleterious alleles can reach high frequency in small populations because of random fluctuations in allele frequency. This may lead, over time, to reduced average fitness. In that sense, selection is more 'effective' in larger populations. Recent studies have considered whether the different demographic histories across human populations have resulted in differences in the number, distribution, and severity of deleterious variants, leading to an animated debate. This article first seeks to clarify some terms of the debate by identifying differences in definitions and assumptions used in recent studies. We argue that variants of Morton, Crow and Muller's 'total mutational damage' provide the soundest and most practical basis for such comparisons. Using simulations, analytical calculations, and 1000 Genomes data, we provide an intuitive and quantitative explanation for the observed similarity in genetic load across populations. We show that recent demography has likely modulated the effect of selection, and still affects it, but the net result of the accumulated differences is small. Direct observation of differential efficacy of selection for specific allele classes is nevertheless possible with contemporary datasets. By contrast, identifying average genome-wide differences in the efficacy of selection across populations will require many modelling assumptions, and is unlikely to provide much biological insight about human populations.

One of the best-known predictions of population genetics is that smaller populations harbor less diversity at any one time but accumulate a higher number of deleterious variants over time [1]. Considerable subsequent theoretical effort has been devoted to the study of fitness differences at equilibrium in populations of different sizes (e.g., [2]) and in subdivided populations (e.g., [3, 4]). The reduction in diversity has been observed in human populations that have undergone strong population bottlenecks: For example, heterozygosity decreased in populations that left Africa, and further decreased with successive founder events [5, 6, 7, 8]. The effect of demography on the accumulation of deleterious variation has been more elusive in both humans and non-human species. In conservation genetics, where fitness can be measured directly and effective population sizes are small, a modest correlation between population size and fitness was observed [9]. In humans, the first estimates of the fitness cost of deleterious mutations were obtained through the analysis of census data [10], but recent studies have focused on bioinformatic prediction using genomic data [11, 12]. Lohmueller *et al.* [13] found that sites variable among Europeans were more likely to be deleterious than sites variable among African-Americans, and attributed the finding to a reduced efficacy of selection in Europeans because of the Out-of-Africa (OOA) bottleneck. However, recent studies [14, 15] suggest that there has not been enough time for substantial differences in fitness to accumulate in these populations, at least under an additive model of dominance. By contrast Peischl *et al.* [16], and more recently Henn *et al.* [17], have claimed significant differences among populations under range expansion models, and Fu *et al.* [18] claims a slight excess in the number of deleterious alleles in European-Americans compared to that in African-Americans. These apparent contradictions have sparked a heated debate as to whether the efficacy of selection has indeed been different across human populations [19, 18]. Part of the apparent discrepancy stems for disagreement about how we

bioRxiv preprint doi: https://doi.org/10.1101/010934; this version posted March 11, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

should measure the effect of selection.

What does it mean for selection to be 'effective'? Some genetic variants increase the expected number of offspring by carriers. As a result, these variants tend to increase in frequency in the population. This correlation between the fitness of a variant and its fate in the population—that is, *natural selection*—holds independently of the biology and the history of the population. However, the rate at which deleterious alleles are removed from a population depends on mutation, dominance, linkage, and demography, and can vary across populations. Multiple metrics have been proposed to quantify the action of selection in human populations and verify the classical population genetic predictions, leading to apparent discrepancies between studies.

In this article, we first review different metrics used in recent empirical work to quantify the action of selection in human populations. We show that many commonly used metrics implicitly rely on 'steady-state' or 'equilibrium' assumptions, wherein genetic diversity within populations is independent of time. This condition is not met in human populations. We discuss two measures for the efficacy of selection that are appropriate for the study of human populations and other out-of-equilibrium populations.

We then seek to provide an intuitive but quantitative understanding of the effect of mutation, selection, and drift on the efficacy of selection in out-of-equilibrium populations. This is done through a combination of extensive simulation and analytical work describing differentiation between populations after a split from a common ancestor. Using this information, we discuss how the classical predictions concerning the effect of demography on selection could be verified in empirical data from human populations.

1 Measuring selection in out-of-equilibrium populations

We consider large panmictic populations whose size $N(t)$ may change over time, and whose reproduction follows the Wright-Fisher model [20]. Given alleles a and A, we assume that genotype aa has fitness 1, aA has fitness $1 + s_i h_i$, and AA has fitness $1 + s_i$. We suppose that A is the least favourable allele ($s_i < 0$) and that $0 \leq h_i \leq 1$. In a random-mating population, an allele A at frequency x_i adds an average of $\delta w_i = s_i \left(2h_i x_i + (1 - 2h_i)x_i^2\right)$ to individual fitness compared to the optimal genotype. We compute the expected fitness over multiple loci as $w = \prod_i (1 + \delta w_i) \simeq 1 + \sum_i \delta w_i$, under the assumption that the individual selection coefficients s_i are small. Finally, we define the genetic load $L = 1 - w$ as the total relative fitness reduction compared to the optimal genotype, $w_{max} = 1$. This yields

$$L = -\sum_i s_i \left(2h_i x_i + (1 - 2h_i)x_i^2\right).$$

To study the effect of selection over short time spans and in out-of-equilibrium populations, we want to define instantaneous measures of the effect of selection on the genetic load and the frequency of deleterious alleles. In this article, the *rate of adaptation* refers the instantaneous rate of fitness increase (or load decrease) in a population. It has contributions from selection, mutation, and drift. The contribution of selection has been the object of considerable theoretical attention: It is the object of the Fitness Increase Theorem (FIT) (see, e.g., [20]). We will refer to the contribution of selection to the rate of adaptation as the *FIT efficacy of selection.*

We also wish to study the effect of selection on the *frequency* of deleterious alleles. There are multiple ways to combine frequencies across loci to obtain a single, genome-wide metric: Any linear function $L_{lin} = \sum_i \rho_i x_i$ of the allele frequencies, with $\rho_i > 0$ a weight assigned to locus i, provides an equally acceptable metric. A natural option, which weights alleles according to their selection coefficient, is Morton, Crow and Muller's *total mutational damage* [21], which is equivalent to the *additive genetic load* that would be observed if all dominance coefficients were replaced by 1/2, i.e., $L_{add} = -\sum_i s_i x_i$. Mutation and selection systematically affect L_{lin}, but genetic drift does not. We define the *Morton efficacy of selection* as the contribution of selection to $\frac{\partial L_{lin}}{\partial t}$. In simulations, where all alleles have equal fitness, we use $\rho_i = -s$. Another common choice, in empirical studies, is to set $\rho_i = 1$ for all sites annotated as

bioRxiv preprint doi: https://doi.org/10.1101/010934; this version posted March 11, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

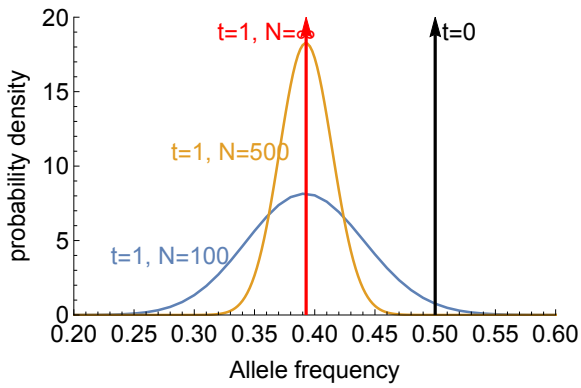

Figure 1: Frequency distribution of deleterious alleles with initial frequency 0.5 after one generation, assuming population sizes of $2N = 100, 500$, and ∞; a selection coefficient $s = -0.6$; and no dominance ($h = 0.5$). The average effect of selection after one generation does not depend on N.

deleterious by a prediction algorithm, and zero otherwise [15, 14, 18]. Since different empirical studies use different ρ_i, direct comparison of the results can be challenging.

Because Morton and FIT efficacies are instantaneous measures of the effect of selection, they can be integrated over time to measure the effect of selection over arbitrary periods. Their integrals over long periods are directly related to classical steady-state metrics such as the rate of fixation of deleterious alleles and the average genetic load in a population.

To understand how genetic drift affects the FIT and Morton efficacy of selection, consider an allele with parental frequency x, selection coefficient $|s| \ll 1$, and no dominance ($h = 0.5$). In the descending population, this allele is drawn with probability $x' \simeq x + sx(1-x)/2$. Figure 1 shows the resulting distribution in offspring allele frequency for $x = 0.5$, $s = -0.6$, and $2N = 100, 500$, and ∞. The average frequency x' is independent of N, hence the expected FIT and Morton efficacies are equal in all populations: Genetic drift does not instantaneously change the effect of selection.

If we let these populations evolve further, however, we will eventually find that deleterious allele frequencies decrease more slowly in smaller populations. This is because natural selection acts on fitness *differences*, and therefore requires genetic variation. By dispersing allele frequencies and reducing diversity, genetic drift also reduces the subsequent effect of selection (see Figure 2). Drift accumulated during

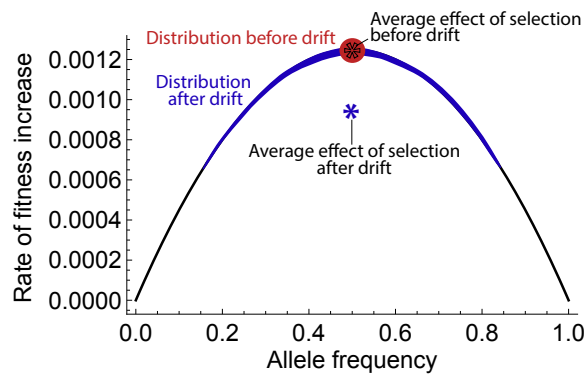

Figure 2: Effect of drift on the rate of adaptation for $h = 0.5$. We consider two populations with initial allele frequency 0.5 and selection coefficient $s = -0.1$. The red population does not undergo drift, and the blue population undergoes one generation of neutral drift, leading to increased variance in allele frequency. The reduced diversity at each locus leads to a lower average *rate* of fitness increase per generation (blue star). Because the rate of fitness increase is a concave function of allele frequency, drift always reduces the future effect of selection when $h = 0.5$.

one generation can change the efficacy of selection for many future generations. Conversely, the current average efficacy of selection depends on the drift accumulated in many previous generations. This delay between the action of drift and its impact on selection can be ignored in steady-state populations but not in out-of-equilibrium populations. For this reason, measures of the effect of selection that have been developed for populations of constant size can be misleading or biased when applied to populations out of equilibrium.

bioRxiv preprint doi: https://doi.org/10.1101/010934; this version posted March 11, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

1.1 Other measures of the effect of selection

The rate at which deleterious mutations are eradicated from a population, for example, is an intuitive metric for the effect of selection that has been recently applied to out-of-equilibrium populations [22]. Over long time scales or in the steady state, this rate of eradication is indeed equivalent to Morton's efficacy of selection. However, in out-of-equilibrium populations, the rate of eradication is a biased measure of the effect of selection. In Figure 1, the smaller population has a higher rate of eradication of deleterious alleles, but this reflects the action of drift rather than the effect of selection. This effect of drift on the rate of eradication of deleterious alleles is short-lived on phylogenetic time scales, but it can be the dominant effect for time-scales relevant to human populations.

Classical work on the efficacy of selection in steady-state populations has emphasized the role of the combined parameter Ns in the dynamics of deleterious alleles. The importance of this combined parameter has led some authors to argue that it should be used as a metric for the efficacy of selection even outside the steady-state [15, 19]. This is problematic for practical and fundamental reasons. On the practical side, the parameter $N(t)s$ is a function of time and does not allow for comparison between populations over finite times: $N(t)s$ is not a rate, and its time integral is meaningless. At a more fundamental level, an instantaneous difference between two populations in the product $N(t)s$ simply indicates a difference in effective population sizes. The interesting biological question is not whether the population sizes are different, but whether these differences lead to differential action of selection by the process illustrated in Figure 2.

More generally, it is commonly proposed that the effect of selection should be measured relative to the effect of drift [19], because the classical parameter Ns is a ratio between a selection term s and a drift term $1/N$. Such a relative measure is not necessary: Morton and FIT efficacies are absolute measures of the effect of selection and they do capture the classical interaction between selection and genetic drift: In populations of constant size, these efficacies do depend on the relative magnitude of selection and drift *coefficients* through the classical parameter Ns. In out-of equilibrium populations, however, they depend on a more complex function of s and $N(t)$. In other words, the classical parameter Ns does not measure the effect of selection *as compared to* the effect of drift; but rather the effect of selection *as modulated by* past genetic drift.

Finally, even though most classical work has focused on the effect of selection on fitness or allele frequency, Henn *et al.* [17] recently proposed to measure the effect of selection on *diversity*, defining a 'reduction in heterozygosity' (RH) statistic that compares the heterozygosity of selected and neutral sites. We show in Section S1.2 that RH is robust to the effect of genetic drift, but it can be biased by recent mutations.

2 Asymptotic results

To study the effect of selection after a population split, we calculate the moments of the expected allele frequency distribution $\phi(x,t)$ under the diffusion approximation. In this formulation, $\phi(x,t)dx$ represents the expected number of alleles with frequency between x and $x+dx$ at time t. In a randomly mating population of size $N = N(t) \gg 1$ and constant s and h, the evolution of $\phi(x,t)$ approximately follows the diffusion equation [20]:

$$\begin{aligned}\frac{\partial \phi(x,t)}{\partial t} \simeq & \frac{1}{4N}\frac{\partial^2}{\partial x^2} x(1-x)\phi(x,t) \\ & - s\frac{\partial}{\partial x}\left(h + (1-2h)x\right) x(1-x)\phi(x,t) \\ & + 2Nu\delta(x - \frac{1}{2N}),\end{aligned} \tag{1}$$

where u is the total mutation rate. The first term describes the effect of drift; the second term, the effect of selection; and the third term describes the influx of new mutations: δ is Dirac's delta distribution. From this equation, we can easily calculate evolution equations for moments of the expected allele frequency distribution $\mu_k = \langle x^k \rangle$. For example, the

bioRxiv preprint doi: https://doi.org/10.1101/010934; this version posted March 11, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

rate of change in allele frequencies $\dot{\mu}_1 \equiv \frac{d\mu_1}{dt}$, is driven by mutation and selection:

$$\dot{\mu}_1 = u + \frac{s\Gamma_{1,h}}{4}, \tag{2}$$

where $\Gamma_{i,h} = 4(\mu_i - \mu_{i+1})h + 4(1-2h)(\mu_{i+1} - \mu_{i+2})$ is a function of the diversity in the population that generalizes the heterozygosity $\pi_1 = \Gamma_{1,1/2}$ (see Appendix for detailed calculations, and References [23, 24] for other applications of the moment approach). We can define the contributions of selection and mutation to changes in allele frequency as $\dot{\mu}_{1s} \equiv \frac{s\Gamma_{1,h}}{4}$ and $\dot{\mu}_{1u} \equiv u$. Morton's efficacy of selection at a locus is simply $s\dot{\mu}_{1s}$. Whereas the effect of mutation is constant and independent of population size, Morton's efficacy depends on the history of the population through $\Gamma_{1,h}$:

$$s\dot{\mu}_{1s} = \frac{s^2\Gamma_{1,h}}{4}, \tag{3}$$

Similarly, changes in the expected fitness W can be decomposed into contributions from mutation, drift, and selection:

$$\begin{aligned} \dot{W} &= \dot{W}_u + \dot{W}_N + \dot{W}_s \\ &= s\Big[2hu + \frac{(1-2h)\pi_1}{4N} \\ &\qquad + s(h\Gamma_{1,h} + (1-2h)\Gamma_{2,h})/2\Big]. \end{aligned} \tag{4}$$

Favorable mutations increase fitness, drift increases fitness when fitness of the heterozygote is below the mean of the homozygotes, and selection always increases average fitness.

The FIT efficacy, $\dot{W}_s$, is therefore

$$\dot{W}_s = s(h\Gamma_{1,h} + (1-2h)\Gamma_{2,h})/2, \tag{5}$$

The right-hand side is the additive variance in fitness, and Equation (5) is an expression of the Fitness Increase Theorem (see, e.g., Equations 1.9 and 1.42 in [20]). Importantly, the FIT efficacy only describes one of three genetic contributions to the rate of adaptation. Interpreting changes in fitness in terms of FIT efficacy requires picking apart the effects of drift and mutation from those of selection. In addition to these genetic effects, changes in the environment can directly affect fitness, introducing a further confounder [25].

Now consider an ancestral population that splits into two isolated randomly mating populations with initial sizes N_A and N_B at time $t = 0$. The populations may experience continuous population size fluctuations. If we expand the moments μ_k of the allele frequency distribution in Taylor series around $t = 0$, we can easily solve the diffusion equation to study the differentiation between the two populations right after the bottleneck. Here we provide an overview of the main results. Detailed derivations are provided in the Appendix.

The difference in fitness between the two populations, $\Delta W(t) = W_A(t) - W_B(t)$ grows linearly in time under dominance:

$$\Delta W(t) = \frac{-s(1-2h)t\pi_{1,o}}{4}\left(\frac{1}{N_B} - \frac{1}{N_A}\right) + O(t^2), \tag{6}$$

Here, t is measured in generations, $\pi_{1,o}$ is the expected heterozygosity in the source population, and $O(t^2)$ represents terms at least quadratic in t. This rapid, linear differentiation is driven by drift coupled with dominance. The smaller population has higher fitness when $h > 0.5$ for $s < 0$: Drift hides dominant, deleterious alleles from the action of selection.

If the source population is large and $h > 0$, we have $\pi_{1,o} \simeq 2\mu_{1,o} \simeq \frac{-2u}{hs}$ [26] and the rate of fitness differentiation is independent of s. This generalizes Haldane's observation that load is insensitive to the selection coefficient in large populations [27]. By contrast to the constant-size population case, however, the observation does not hold when $h = 0$. The initial response to the bottleneck is independent of fitness for $0 < h < .5$ (see Figures S2 and S3), but not for $h = 0$ or $h = 0.5$ (see Figures 3 and S1).

The effect of selection on fitness differences, $\Delta W_s(t)$, grows only quadratically:

$$\Delta W_s(t) = \frac{s^2t^2\Pi_h}{16}\left(\frac{1}{N_B} - \frac{1}{N_A}\right) + O(t^3), \tag{7}$$

where Π_h is a measure of diversity that reduces to $\pi_{1,o}$ when $h = 0.5$ (see Appendix). This slower response

bioRxiv preprint doi: https://doi.org/10.1101/010934; this version posted March 11, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

is the mathematical consequence of the intuition provided by Figures 1 and 2: Right after the split, the fitnesses are identical and the efficacy of selection is the same in both populations. It takes time for drift to increase the variance in allele frequency and cause differences in the efficacy of selection, accounting for a factor $t\left(\frac{1}{N_B}-\frac{1}{N_A}\right)$. It then takes time for differences in the efficacy of selection to accumulate and produce differences in fitness, accounting for an additional factor st.

Combining Equations (6) and (7), we get an asymptotic result for the load differentiation.

$$\Delta W_{asymptotic}(t) = \left(\frac{-s(1-2h)t\pi_{1,o}}{4} + \frac{s^2t^2\Pi_h}{16}\right) \times \left(\frac{1}{N_B}-\frac{1}{N_A}\right). \tag{8}$$

This expression describes the leading differentiation in fitness in all simulations below. It is straightforward to refine this asymptotic result by computing higher-order corrections, however the number of terms in the expansion increases rapidly. Some of these terms are of particular interest, such as the contribution of new mutations. Since the direct effect of mutation on load is independent of demography [Equation (4)], we must wait for mutations to accumulate before load differentiation can begin. This leads an additional factor of ut compared to the case of standing variation. The contribution of drift acting on new recessive mutations is therefore quadratic:

$$\Delta W_{new} = \frac{-sut^2(1-2h)}{4}\left(\frac{1}{N_B}-\frac{1}{N_A}\right) + o(t^3), \tag{9}$$

The effect of selection on new mutations is only cubic in time: We must wait for mutations to appear (contributing a factor of ut), then wait for drift to cause differences on the frequency distribution of the new mutations [contributing a factor of $t\left(\frac{1}{N_B}-\frac{1}{N_A}\right)$], and finally wait for selection to act on these frequency distribution differences (contributing a factor of st.) The leading contribution of selection is therefore

$$\Delta W_{s,new} = \frac{s^2t^3uh(5h-2)}{6}\left(\frac{1}{N_B}-\frac{1}{N_A}\right) + o(t^4).$$

Finally, since drift alone does not produce differences in average allele frequencies, the rate of differentiation in deleterious allele frequencies is always quadratic in time:

$$\Delta\mu_{1s}(t) = \frac{st^2}{8}\left((3h-1)\pi_{1,o} + 3(1-2h)\pi_{2,o}\right) \times \times \left(\frac{1}{N_B}-\frac{1}{N_A}\right) + O(t^3). \tag{10}$$

3 Simulations

We simulated the evolution of $\phi(x,t)$ using ∂a∂i [28] and the Out-Of-Africa (OOA) demographic model illustrated in Figure 3A. This model begins with an ancestral population of size $N_r = 11930$ with frequency distribution following the quasi-stationary distribution of Kimura [29], and features population splits and size changes that were inferred from synonymous polymorphism from the 1000 Genomes dataset [30]. We estimated the probability $\phi_n(i,t)$ that a variant is at frequency i in a finite sample of size $n = 100$ for each population, given a mutation rate of $\mu = 1.44 \times 10^{-8}\text{bp}^{-1}$ generation^{-1} [31] in an infinite genome. We used the finite sample predictions to estimate the expected genetic load and the contributions of drift, selection, and mutation. Finally, to ensure that results were not model-dependent, we repeated each simulation using a different demographic model described in [15], featuring a single deeper but shorter OOA bottleneck.

We simulated all combinations of selection coefficients $2N_rs \in \{0, -0.01, -0.1, -0.3, -1, -3, -10, -30, -100\}$, and dominance coefficients $h \in \{0, 0.05, 0.2, 0.3, 0.5, 1\}$. The contributions of selection and drift were obtained using Equation (4). To emphasize the long-term effects of the OOA bottleneck even after drift is suppressed, simulations were also carried to future times assuming large population sizes ($N = 20N_r$) and no migrations (Figure S6). In all cases, Equations (8), (9), and (10) capture the initial increase in load (Figures 3, S1, S2, and S7).

bioRxiv preprint doi: https://doi.org/10.1101/010934; this version posted March 11, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

3.1 Genic selection; $h = 1/2$

Simulated differences in load are modest and limited to intermediate-effect variants ($.3 < |2N_r s| < 30$, $2.5\times10^{-5} < |s| < 0.0025$). Assuming the distribution of fitness effects inferred from European-American data by Boyko *et al.* [32], the excess load in the OOA population is 0.49 per Gb of amino-acid–changing variants, in addition to a total accumulated load of 24 per Gb in the African population (this accumulated load does not include variation that was fixed at the time of the split). If we consider the 24 Mb of exome covered by the 1000 Genomes project, and assume that 70% of mutations are coding in that region [33], the model predicts a non-synonymous load difference of 0.008. The total estimated non-synonymous load, excluding mutations fixed in the ancestral state, is 0.4 in the African-American population. In this model, the reduced efficacy of selection caused by the OOA bottleneck leads to a relative increase in non-recessive load of 2%. Since we did not consider fixed ancestral deleterious alleles in the total load, this figure is an overestimate of the relative increase in load due to the bottleneck. The relative increase reaches a maximum of 8% for mutations with $-20 < 2N_r s < -10$. The results are similar if we use the distribution of fitness effects inferred from African-American data [32].

Using the simple bottleneck demographic model of Do et al [15], we find very similar load (24 per Gbp) and load differences across populations (2% of the total load).

3.2 Partial and complete dominance

The picture changes dramatically when we consider recessive deleterious variants ($h = 0$). Reactions to changes in population size are linear rather than quadratic, and they are more substantial than in the additive case (Figure S1). The OOA load due to segregating variants with $2N_r s = -100$ almost doubles after 500 generations. This excess load in the OOA population is due entirely to drift, and leads to an increased efficacy of selection in the OOA population, since a higher proportion of deleterious alleles are now visible to selection. The difference in load for the most deleterious variants is therefore not sustained.

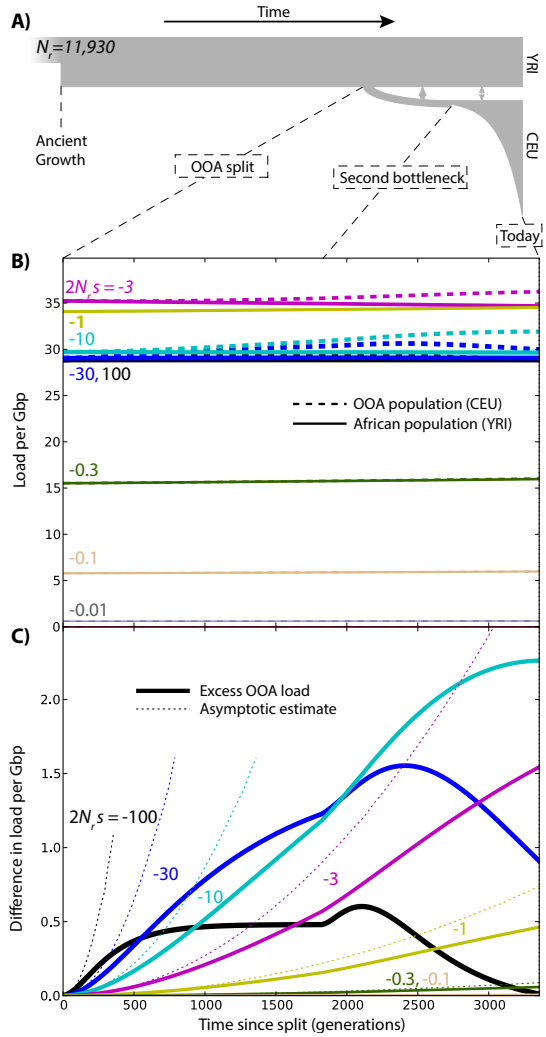

Figure 3: Changes in load after the OOA split model illustrated in (A) as a function of s, given ancestral population size $N_r = 11,930$ and no dominance ($h = 0.5$). (B) Load evolution after the split. We subtracted the load due to variants fixed in all populations. (C) Difference in load between populations. Since $h = 0.5$, this is equivalent to the difference in additive load. Dotted lines show asymptotic results from Equations (8) and (10). Load is given per Gbp of variants at the specified selection coefficient. The total amount of variation under strong selection in the human genome is likely much less than 1 Gbp.

bioRxiv preprint doi: https://doi.org/10.1101/010934; this version posted March 11, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Both the number of very deleterious variants and the associated genetic load eventually becomes higher in the simulated Yoruba population. By contrast, weak-effect deleterious variants contribute more load in the simulated European population.

Even though a bottleneck inexorably leads to increased load when no dominance is present, the additional exposure of recessive variants therefore leads to 'purging', a reduction of the frequency of deleterious alleles (see [2] and references therein). Simulations show that the increase in recessive load can last hundreds or thousands of generations for weakly deleterious variants. Glémin argued that the purging effect is suppressed in constant-sized population when Ns is much less than "2 to 5". [2]. This also holds in a non-equilibrium setting in recessive alleles going through a bottleneck (Figure S1, see also [34]). The time required for purging to compensate the initial fitness loss increases rapidly as the magnitude of the selection coefficients decreases: Whereas our model predicts a reduced load in present-day OOA populations for alleles with $2N_r s = -100$, it would take over $20,000$ generations of continued isolation in large constant-sized population to see purging in alleles with $2N_r s = -3$ (Figure S6).

Opposite effects are observed for dominant deleterious variants (Figure S7). Drift tends to increase fitness by combining more of the deleterious alleles into homozygotes, reducing their average effect on fitness. The difference between populations is much less pronounced and less sustained than in the recessive case. Equation (6) shows that the reduced magnitude is caused by reduced ancestral heterozygosity, $\pi_{1,0}$: Dominant deleterious alleles are much less likely to reach appreciable allele frequencies before the split. Here again, the population with the highest load depends on the selection coefficient, with a higher load in the simulated European population for strongly deleterious variants and a higher load in the simulated Yoruba population for the weakly deleterious variants.

The distribution of dominance coefficients for mutations in humans is largely unknown, but non-human studies suggest that partial recessive may be the norm (see, e.g., [35] and references therein). Under models with $h = 0.2$, we find that the genetic load is elevated in OOA populations for most selection coefficients $N_r s$, whereas the *additive* genetic load is mostly reduced (Figure S3B-C and S4B-C). These simulations suggest that the rate of adaptation was reduced in OOA populations (i.e., the genetic load is higher in OOA population), while the efficacy of selection was higher in the OOA population, whether it is measured by the Morton efficacy or the FIT efficacy (Figure S5). Thus, unless most nearly-neutral variation has $h > 0.20$, we do not expect an overall elevated number of deleterious variants in OOA populations. As we move closer to additive selection, for example at $h = 0.3$, the contributions of alleles with larger and weaker selection coefficient are of comparable magnitude and opposite direction. Because of our limited ability to estimate selection coefficients in humans, this might explain why observing differences in the overall frequency of deleterious alleles between populations has been so difficult. This also suggests that any claim for an across-the-board difference in the efficacy of selection between two populations will have to rely on a number of assumptions about fitness coefficients in human populations.

4 Present-day differences in the efficacy and intensity of selection

The Wright-Fisher predictions for the instantaneous Morton and FIT efficacies of selection, Equations (3) and (5), depend on the present-day allele frequency distribution, on the dominance coefficient h, and on the selection coefficient s. However, s is a multiplicative factor in both equations and cancels out when we consider the relative rate of adaptation across populations. We can therefore use Equations (3) and (5) to estimate differences in the efficacy of selection between populations based on the present-day distribution of allele frequencies. For nearly-neutral alleles, the present-day frequency distribution is similar to the neutral frequency spectrum and largely independent of h. We can therefore use the present-day frequency spectrum for synonymous variation to estimate the relative efficacy of selection for all

bioRxiv preprint doi: https://doi.org/10.1101/010934; this version posted March 11, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

nearly-neutral alleles at different values of h (Figure 4). Figures S10 and S11 show similar results for non-synonymous and predicted deleterious alleles (For the most deleterious classes, the assumption that the present-day frequency spectra depend weakly on h is less accurate).

In the nearly-neutral case, the Luhya population (LWK) shows the highest Morton an FIT efficacy of selection for most dominance parameters and is used as a basis of comparison. The estimated FIT efficacy of selection is higher in African population for all dominance coefficients, as is the Morton efficacy, except for completely recessive alleles. The reduction in Morton's efficacy of selection for nearly-neutral variation in OOA populations is 25% to 39% for dominant variants, 19% to 31% for additive variants, and 6% to 13% for fully recessive variants. The reduction in the FIT efficacy in OOA populations is 29% to 44% for dominant variants, 19% to 31% for additive variants, and 0.2% to 6% for fully recessive variants. This is also consistent with the interpretation of Glémin that purging, the reduction in the frequency of recessive alleles caused by a bottleneck, is not expected for nearly neutral variants. By contrast, estimates using sites with high predicted pathogenicity according to CADD [36] do suggest that purging of deleterious variation by drift is still ongoing in OOA populations (Figures S10 and S11).

Admixed populations from the Americas with the highest African ancestry proportion also show elevated efficacy of selection: African-Americans (75.9% African ancestry [37]), Puerto Rican (14.8% African ancestry [31]), Colombians (7.8% African ancestry [31]), and Mexican-Americans (5.4% African ancestry [31]). The Morton efficacy of selection in admixed populations is much larger than the weighted average of source populations would suggest (Figure 4C, which uses CHB, CEU, and YRI as ancestral population proxies for Native, European, and African ancestries). By averaging out some of the genetic drift experienced by the source populations since their divergence, admixture increases the overall amount of additive variance in the population, and therefore leads to a substantial and rapid increase in the predicted efficacy of selection for nearly neutral alleles.

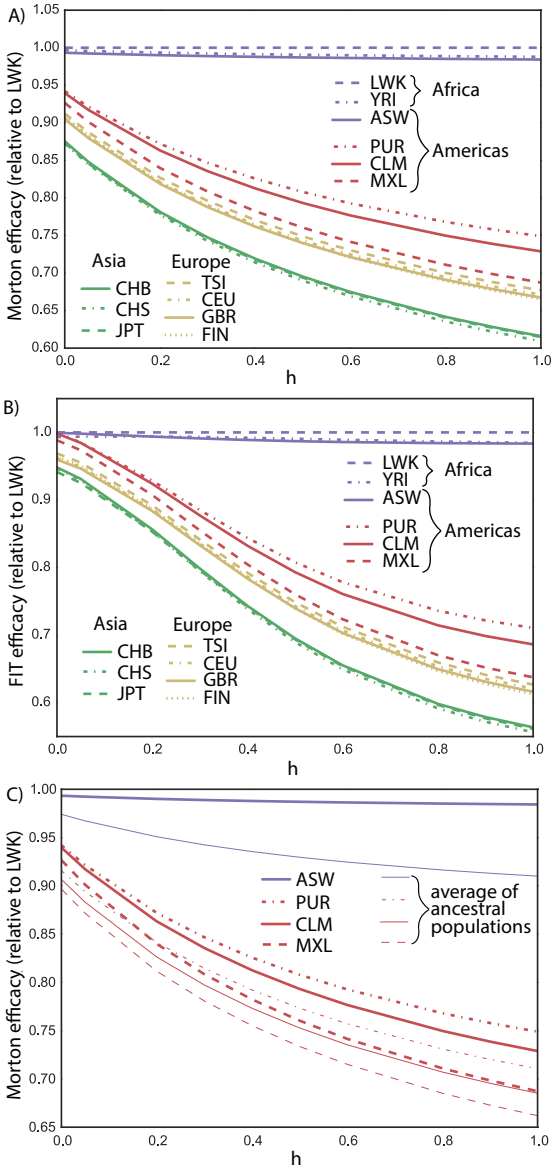

Figure 4: A) Present-day Morton efficacy of selection for nearly neutral variants, estimated from 1000 Genomes synonymous variation. B) FIT efficacy for the same variants, and C) Morton efficacy in admixed populations is increased compared to the average efficacies in ancestral populations.

bioRxiv preprint doi: https://doi.org/10.1101/010934; this version posted March 11, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

5 Discussion

Selection affects evolution in many ways. It tends to increase the frequency of favourable alleles and the overall fitness of a population, and it often reduces diversity. The rates at which it performs these tasks varies across populations, and population geneticists like to frame these differences in terms of the efficacy of selection. The word 'efficacy' implies a measure of achievement, but there are many ways to define achievement for selection. We considered two measure of achievement: the change in deleterious allele frequency (i.e., Morton's efficacy), and the change in load caused by selection (i.e., the FIT efficacy). Even though the two quantities are closely related, and are equal for additive selection, Morton's efficacy is much easier to measure: systematic differences in the frequency of deleterious alleles are robust to drift and to modest changes in the environment. By contrast, the FIT efficacy is impossible to observe directly and requires picking apart the contributions of selection, drift, and the environment. Given the long-standing controversy about how this should be done in the context of Fisher's Fundamental Theorem [38], we would advise against using it.

We have argued that other popular measures for the efficacy of selection [13, 19, 8, 17] are biased in out-of-equilibrium populations studied over short time-scales. Many previous claims that selection acted differentially in human populations [13, 8] could be explained by these biases. Confirming this interpretation, Fu et al. [18] found no differences in the average frequency of deleterious alleles between African-Americans and European-Americans in the ESP 6500 dataset [39]. However, they did report a slight but extremely significant difference in the average number of deleterious alleles per individual for a set of putatively deleterious SNPs. The contrasting results are surprising, since the two statistics are equal up to a multiplicative constant: the average number of deleterious alleles per genome equals the mean frequency of deleterious alleles multiplied by the number of loci. We could reproduce the results from [18], but found that the statistical test used did not account for variability introduced by genetic drift in a finite genome: results remained significant if allele frequencies were randomly permuted between African-Americans and European-Americans (see Section S1.1 for details). This emphasizes that an empirical observation of differences in genetic load must be robust to both finite sample size and finite genome to be attributed to differences in the efficacy of selection.

Figures 4, S10 and S11 strongly suggest that the OOA bottleneck still influences the present-day efficacy of selection. By extension, they also suggest that the efficacy of selection did differ and will differ among populations. Importantly, the differences in frequency distributions across populations that provide this support are not a *consequence* of past differences in the efficacy of selection but a possible *cause* for such differences in the present and future. We have shown that some of the future differences are not inevitable and can be attenuated by demographic processes including admixture. Therefore, measuring actual differences in the efficacy of selection can only be achieved by measuring actual differences in the average frequency or effect of deleterious alleles.

Simulations presented here, together with the results of [14, 15], do suggest that the classical prediction on the differential efficacy of selection in small populations can be verified if only we can accurately isolate variants of specific selective effect and dominance coefficients. By picking apart variants of different selection and dominance coefficients, we should soon be able to convincingly and directly observe the consequences of differences in the efficacy of selection. The recent results of Henn et al. [17], using a version of Morton's efficacy, do suggest such differences for a subset of variants and therefore provide important experimental validation for a classical population genetics prediction. By contrast, the observation of genome-wide differences in the efficacy of selection across populations depends on the cancellation of effects across different variant classes, and can therefore depend sensitively on the particular choice of a metric. For this reason, overall differences in load among populations may not be particularly informative about the fundamental processes governing human evolution.

bioRxiv preprint doi: https://doi.org/10.1101/010934; this version posted March 11, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

6 Acknowledgements

I thank S. Baharian, M. Barakatt, B. Henn, D. Nelson, and S. Lessard for useful comments on this manuscript, and W. Fu and J. Akey for help in reproducing their results. This research was undertaken, in part, thanks to funding from the Canada Research Chairs program and a Sloan Research Fellowship.

References

[1] Kimura M, Maruyama T, Crow JF (1963) The mutation load in small populations. *Genetics* 48:1303.

[2] Glémin S (2003) How are deleterious mutations purged? Drift versus nonrandom mating. *Evolution* 57:2678–2687.

[3] Roze D, Rousset F (2003) Selection and drift in subdivided populations: a straightforward method for deriving diffusion approximations and applications involving dominance, selfing and local extinctions. *Genetics* 165:2153–2166.

[4] Whitlock MC (2002) Selection, load and inbreeding depression in a large metapopulation. *Genetics* 160:1191–1202.

[5] Tishkoff SA, et al. (1996) Global patterns of linkage disequilibrium at the CD4 locus and modern human origins. *Science* 271:1380–1387.

[6] 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* 102:15942–15947.

[7] The 1000 Genomes Project Consortium (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature* 491:56–65.

[8] 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:e1003815.

[9] Reed D, Frankham R (2003) Correlation between Fitness and Genetic Diversity. *Conservation biology* 17:230–237.

[10] Crow JF (1958) Some possibilities for measuring selection intensities in man. *Hum Biol* 30:1–13.

[11] Davydov EV, et al. (2010) Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Comput Biol* 6:e1001025.

[12] Adzhubei I, Jordan DM, Sunyaev SR (2013) Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet* Chapter 7:Unit7.20.

[13] Lohmueller KE, et al. (2008) Proportionally more deleterious genetic variation in European than in African populations. *Nature* 451:994–997.

[14] Simons YB, Turchin MC, Pritchard JK, Sella G (2014) The deleterious mutation load is insensitive to recent population history. *Nat. Genet.* 46:220–224.

[15] Do R, et al. (2014) No evidence that selection has been less effective at removing deleterious mutations in Europeans than in Africans. *Nat. Genet.* 47:126–131.

[16] Peischl S, Dupanloup I, Kirkpatrick M, Excoffier L (2013) On the accumulation of deleterious mutations during range expansions. *Mol Ecol* 22:5972–5982.

[17] Henn BM, et al. (2016) Distance from sub-Saharan Africa predicts mutational load in diverse human genomes. *Proc Natl Acad Sci USA* 113:E440–9.

[18] 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* 95:421–436.

[19] Lohmueller KE (2014) The distribution of deleterious genetic variation in human populations.

bioRxiv preprint doi: https://doi.org/10.1101/010934; this version posted March 11, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Current Opinion in Genetics & Development 29:139–146.

[20] Ewens WJ (2012) *Mathematical Population Genetics 1*, Theoretical Introduction (Springer Science & Business Media).

[21] Morton NE, Crow JF, Muller HJ (1956) An estimate of the mutational damage in man from data on consanguineaous marriages. *Proc Natl Acad Sci USA* 42:855.

[22] 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.

[23] Evans SN, Shvets Y, Slatkin M (2007) Non-equilibrium theory of the allele frequency spectrum. *Theor Popul Biol* 71:109–119.

[24] Balick DJ, Do R, Cassa CA, Reich D, Sunyaev SR (2015) Dominance of Deleterious Alleles Controls the Response to a Population Bottleneck. *PLoS Genet* 11:e1005436.

[25] Mustonen V, Lassig M (2010) Fitness flux and ubiquity of adaptive evolution. *Proc Natl Acad Sci USA* 107:4248–4253.

[26] Crow JF, Kimura M (2009) An Introduction to Population Genetics Theory. (Blackburn Press).

[27] Haldane J (1937) The Effect of Variation of Fitness on fitness. *American Naturalist* p 337.

[28] Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD (2009) Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet* 5:e1000695.

[29] Kimura M (1964) Diffusion models in population genetics. *Journal of Applied Probability* 1:177.

[30] Gravel S, et al. (2011) Demographic history and rare allele sharing among human populations. *Proc Natl Acad Sci USA* 108:11983–11988.

[31] Gravel S, et al. (2013) Reconstructing Native American migrations from whole-genome and whole-exome data. *PLoS Genet* 9:e1004023.

[32] Boyko AR, et al. (2008) Assessing the evolutionary impact of amino acid mutations in the human genome. *PLoS Genet* 4:e1000083.

[33] Hwang DG, Green P (2004) Bayesian Markov chain Monte Carlo sequence analysis reveals varying neutral substitution patterns in mammalian evolution. *Proc Natl Acad Sci USA* 101:13994–14001.

[34] Wang J, Hill WG, Charlesworth D, Charlesworth B (1999) Dynamics of inbreeding depression due to deleterious mutations in small populations: mutation parameters and inbreeding rate. *Genet. Res.* 74:165–178.

[35] Henn BM, Botigué LR, Bustamante CD, Clark AG, Gravel S (2015) Estimating the mutation load in human genomes. *Nat Rev Genet* 16:333–343.

[36] Kircher M, et al. (2014) A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* 46:310–315.

[37] Baharian S, Barakatt M, Gignoux CR, Shringarpure S (2015) The Great Migration and African-American genomic diversity. *bioRxiv*.

[38] Ewens WJ, Lessard S (2015) On the interpretation and relevance of the Fundamental Theorem of Natural Selection. *Theor Popul Biol* 104:59–67.

[39] National Heart, Lung, and Blood Institute (NHLBI) GO Exome Sequencing Project (2012) Exome variant server, Retrieved Oct. 2015.

A1 Appendix

A1.1 Background

To derive the asymptotic results in the text, we start with the diffusion approximation for the distribution

bioRxiv preprint doi: https://doi.org/10.1101/010934; this version posted March 11, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

$\phi(x,t)$ of allele frequencies x over time t in an infinite-sites model (see [26], section 8.6):

$$\begin{aligned}\frac{\partial\phi(x,t)}{\partial t} \simeq & \frac{1}{4N}\frac{\partial^2}{\partial x^2}x(1-x)\phi(x,t) \\ & - s\frac{\partial}{\partial x}\left(h+(1-2h)x\right)x(1-x)\phi(x,t) \\ & + 2Nu\delta(x-\frac{1}{2N}),\end{aligned} \tag{A1}$$

where N is the effective population size, h is the dominance coefficient, s is the selection coefficient, and u is the mutation rate. In this model, new mutations are constantly added via Dirac's delta function δ. Because there are no back mutations in this model, the proportion of fixed mutations increases over time without bound. Because we are only interested in population differences accumulating over a short time span, however, we can simply ignore the (infinite) number of deleterious alleles that fixed before the population split. The time-scales that we will consider are short enough that back-mutations and multiple mutations contribute little to changes in allele frequencies.

A complete solution of this problem can be expressed as a superposition of Gegenbauer polynomials [29]. However, here we are looking for simple asymptotic results that will help us understand the dynamics of the problem. We will consider the evolution of moments of the allele frequency distribution: $\mu_k \equiv \int_0^1 dx x^k \phi(x,t)$. Similar moment approaches have been used in [23, 24]. Because there is a possibly infinite number of fixed sites at frequencies 0 and 1, it is often convenient to distinguish contributions from segregating sites and fixed sites:

$$\mu_k = \int_{0^+}^{1^-} dx x^k \phi(x,t) + K_0\delta_{k,0} + K_1,$$

where K_0 is the number of sites fixed at frequency 0, K_1 is the number of sites fixed at frequency 1, and $\delta_{k,0}$ is Kronecker's delta. Both K_0 and K_1 can be infinite in this model, but this will not be a problem since we will ultimately consider only differences or rates of change in the moments, and these remain finite. In this notation, μ_0 is the (possibly infinite) number of sites, and μ_1 is the expected number of alternate alleles per haploid genome.

To obtain evolution equations for the moments, we integrate both sides of equation (A1) using $\int_{0^+}^{1^-} dx x^k$. The left-hand side gives

$$\begin{aligned}\int_{0^+}^{1^-} dx x^k \frac{\partial\phi(x,t)}{\partial t} &= \frac{\partial \int_{0^+}^{1^-} dx x^k \phi(x,t)}{\partial t} \\ &= \dot{\mu}_k - \dot{K}_0\delta_{k,0} - \dot{K}_1,\end{aligned} \tag{A2}$$

and the right-hand side can be integrated by parts. For $k=0$, this yields

$$\dot{\mu}_0 = 2Nu - \frac{\phi(0,t)+\phi(1,t)}{4N} + \dot{K}_0 + \dot{K}_1,$$

where $\phi(0,t)$ and $\phi(1,t)$ are defined by continuity from the open interval $(0,1)$ and do not include fixed sites. Because the number of sites is constant ($\dot{\mu}_0 = 0$) and the diffusion equation is continuous, we require

$$\begin{aligned}\dot{K}_0 &= -2Nu + \frac{\phi(0,t)}{4N}, \\ \dot{K}_1 &= \frac{\phi(1,t)}{4N}.\end{aligned} \tag{A3}$$

These equations are equivalent to Equations 3.18 and 3.19 in [29].

To obtain an evolution equation for μ_k at arbitrary k, we return to the integration of Equation (A1) with $\int_{0^+}^{1^-} dx x^k$. We use the left-hand-side expression obtained in Equation (A2), and once again integrate the right-hand side by parts. This yields

$$\dot{\mu}_k = \frac{k(k-1)}{8N}\pi_{k-1} + \frac{sk}{4}\Gamma_{k,h} + \frac{u}{(2N)^{k-1}}, \tag{A4}$$

where

$$\begin{aligned}\pi_k &= 2(\mu_k - \mu_{k+1}) \\ \Gamma_{k,h} &= 2\left(h\pi_k + (1-2h)\pi_{k+1}\right).\end{aligned} \tag{A5}$$

These are functions of the moments μ and can therefore be thought of as measures of the shape of the frequency distribution ϕ.

The first term in (A4) represents the effect of drift, the second term the effect of selection, and the third term the effect of mutation.

bioRxiv preprint doi: https://doi.org/10.1101/010934; this version posted March 11, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

For example, if $k = 1$ and $h = 1/2$, we get

$$\dot{\mu}_1 = s(\mu_1 - \mu_2)/2 + u$$

The frequency of damaging alleles can decrease because of selection, or increase because of mutation.

A1.2 Response in allele frequencies

Solving Equation (A4) in general is challenging, because $\dot{\mu}_k$ can depend on μ_{k+1} and μ_{k+2}, leading to an infinite number of coupled equations. However, it can be used to calculate the response in allele frequency to a sudden change in demographic or selective conditions. Consider a population of size N_o that experiences a change in size to N_A at time $t = 0$. We can expand μ_k for short times:

$$\mu_k(t) = \mu_{k,o} + \mu_{k,1}t + \mu_{k,2}t^2 + O(t^3), \tag{A6}$$

where $\mu_{k,o}$ is the kth moment prior to the population size change and $O(t^3)$ represent terms at least cubic in t. The coefficients can be evaluated by expanding both sides of Equation (A4) using Equation (A6), then collecting powers of t. For example, we get

$$\mu_1(t) = \mu_{1,o} + \left(\frac{s\,\Gamma_{1,h}|_{t=0}}{4} + u\right)t + O(t^2).$$

The frequency of variants can increase even in a steady-state regime with $N_A = N_o$, since our model assumes a constant supply of irreversible mutations. However, this linear term is independent of N_A and does not contribute to differences across populations that share a common ancestor. Differences $\Delta\mu_1(t)$ in the number of segregating sites between two populations with sizes N_A and N_B appear at the next order in t. Computations are elementary but a bit more cumbersome. Matching terms linear in t in Equation (A4), we find equation (10):

$$\begin{aligned}\Delta\mu_1(t) =& \frac{st^2}{8}\left((3h-1)\pi_{1,o} + 3(1-2h)\pi_{2,o}\right)\times \\ &\times\left(\frac{1}{N_B} - \frac{1}{N_A}\right) + O(t^3),\end{aligned}$$

where $\pi_{i,o}$ is the moment π_i computed for the common ancestral population.

A1.3 Response in genetic load

To compute the fitness in the diploid case, we write

$$W = s\left(2h\mu_1 + (1-2h)\mu_2\right). \tag{A7}$$

Using (A4), we get

$$\dot{W} = \dot{W}_s + \dot{W}_u + \dot{W}_N, \tag{A8}$$

where

$$\begin{aligned}\dot{W}_s &= s\left(\frac{s}{2}\left(h\Gamma_{1,h} + (1-2h)\Gamma_{2,h}\right)\right), \\ \dot{W}_u &= s\left(2hu\right), \\ \dot{W}_N &= s\left(\frac{(1-2h)}{4N}\pi_1\right)\end{aligned} \tag{A9}$$

are the instantaneous contributions of selection, mutation, and drift to changes in fitness. The mutation term is constant in time and independent of population size; it does not directly contribute to differences across populations. The drift term, by contrast, has an explicit dependence on the population size; this leads to differentiation between populations that grows linearly in time. To see this, we compute the load using Equation (A7) and the time dependence computed in Equation (A6), as in section A1.2:

$$\begin{aligned}\Delta W = W_1 - W_2 =& \frac{-s(1-2h)t\pi_{1,o}}{4}\left(\frac{1}{N_B} - \frac{1}{N_A}\right) \\ &+ O(t^2).\end{aligned} \tag{A10}$$

This reduction in load is driven by drift, i.e., the third term in equation (A9). It is not caused by selection, in the sense that it does not result from differential reproductive success between individuals. As expected, the contribution of drift vanishes for additive variants ($h = 1/2$).

For arbitrary h, the change in fitness caused by selection is

$$\Delta W_s = \frac{s^2t^2\Pi_h}{16}\left(\frac{1}{N_B} - \frac{1}{N_A}\right) + O(t^3),$$

bioRxiv preprint doi: https://doi.org/10.1101/010934; this version posted March 11, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

where Π_h is a statistic of the ancestral frequency distribution:

$$\begin{aligned}\Pi_h =& 4h\pi_{1,o}(5h-2) - 12\pi_{2,o}(1-2h)(1-4h) \\ &+ 24\pi_{3,o}(1-2h)^2,\end{aligned} \tag{A11}$$

which reduces to the heterozygosity $\pi_{1,0}$ when $h = 1/2$. The statistic Π_h depends only on the ancestral frequency distribution and the dominance coefficient.

Genetic drift also contributes to the changes in load at second order in t through $\dot{W}_N$. In addition to the linear term from Equation (A10), we find three quadratic contributions that vanish when $h = 1/2$: a second-order contribution of genetic drift, a contribution from the rate of change in population size and drift, and a contribution from new mutations and drift. Even though these terms can be comparable in magnitude to the contribution of selection in Equation (A11) when $h \neq 1/2$, they are sub-dominant to Equation (A10). We only consider the contribution of new mutations in some detail, as this contribution tells us whether population differentiation in the genetic load is due to old, shared variation or to new, population specific variation.

A1.4 Effect of new mutations

If we set $\pi_{i,o} = 0$ in the equations above, we can calculate the impact of new mutations on the genetic load. The leading term is again due to drift and dominance:

$$\Delta W_{new} = \frac{-sut^2(1-2h)}{4}\left(\frac{1}{N_B} - \frac{1}{N_A}\right) + o(t^3), \tag{A12}$$

while the leading term describing the efficacy of selection is now cubic in t:

$$\Delta W_{s,new} = \frac{s^2t^3uh(5h-2)}{6}\left(\frac{1}{N_B} - \frac{1}{N_A}\right) + o(t^4).$$

When $h \neq 1/2$, drift also contributes t^3 terms to ΔW_{new}. These are reasonably straightforward to compute, but are sub-dominant to Equation (A12).

We therefore use the asymptotic result:

$$\begin{aligned}\Delta W_{new,asymptotic} =& \frac{-sut^2(1-2h)}{4}\left(\frac{1}{N_B} - \frac{1}{N_A}\right) \\ &+ \frac{s^2t^3uh(5h-2)}{6}\left(\frac{1}{N_B} - \frac{1}{N_A}\right).\end{aligned} \tag{A13}$$

Comparisons with simulated data are shown on Figure S9.

bioRxiv preprint doi: https://doi.org/10.1101/010934; this version posted March 11, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

S1 Supplementary Material

S1.1 Analysis of the Fu et al. result

Fu et al presents two different tests for differences in the number of deleterious alleles between individuals. In the first test, a t-test, they compare the average frequency of deleterious alleles between the two populations and find no significant differences ($p = 0.82$). In the second test (a Mann-Whitney test), they compared the number of deleterious alleles per individual, and found an extremely significant difference ($p < 10^{-15}$). We could reproduce these results using the publicly available ESP 6500 data [39] using a permutation test. Since the average frequency of deleterious alleles and the number of deleterious alleles per individual are related by a proportionality constant (i.e., the length of the genome), we suspected that the significant difference results from a difference in the underlying null models of each test.

To verify this, we produced simulated datasets in which the allele frequency at each site was randomly permuted between European-Americans and African-Americans, creating a dataset for which there is no meaningful evolutionary difference between the two populations, except for randomly assigned differences in allele frequency. We nevertheless found very significant differences between the average number of deleterious alleles carried by individuals of the two populations for most simulations. This effect can even be reproduced by analyzing a single SNP. This is the Mann-Whitney test used in Fu et al. conditions on the set of SNPs used: it does not account for the fact that drift can have affected different SNPs differently. The difference in deleterious allele count observed in Fu et al is therefore real, but it only applies to a particular set of SNPs and it is consistent with the action of genetic drift acting on neutral variation. It does not indicate differences in the action of selection, nor systematic differences in fitness across populations. To assign differences in deleterious allele frequency to the systematic action of selection, one must show that the difference is robust to both finite sample size and finite genome effects. Both can be tested through simple resampling strategies. In the ESP6500 example, resampling over SNPs by bootstrap led to no significant differences in the number of deleterious alleles per individual.

S1.2 Reduction in heterozygosity (RH) statistic

The reduction in heterozygosity (RH) statistic was recently introduced [17] as a tool to measure the effect of selection on population diversity as an addition to analyses based on Morton's efficacy. In this section, we show that even though RH is an interesting measure of diversity in a population, its interpretation in terms of the effect of selection can be challenging: First, RH can be affected by recent mutation and, second, differences in the effect of selection on RH may reflect rather mundane normalization issues rather than interesting biology.

RH is defined as

$$RH \equiv 1 - \frac{\pi_{|s}}{\pi_{|n}},$$

where π is the average heterozygosity, $|s$ indicates a quantity measured at selected sites, and $|n$ indicates a quantity measured at neutral sites. The rate of change in RH is therefore

$$\dot{RH} = -\frac{\dot{\pi}_{|s}}{\pi_{|n}} + \frac{\pi_{|s}\dot{\pi}_{|n}}{\pi_{|n}^2} \tag{S2}$$

Using equation (A4) and (A5) to leading order in $\frac{1}{N}$, we find

$$\begin{aligned} \dot{\pi}_{|n} &= \frac{-\pi_{|n}}{2N} + 2u_{|n} \\ \dot{\pi}_{|s} &= \frac{-\pi_{|s}}{2N} + 2u_{|s} + \frac{s\Gamma_{1,h|s}}{2} - s\Gamma_{2,h|s}. \end{aligned} \tag{S3}$$

where u is the mutation rate and $\Gamma_{i,h}$ is defined in Equation (A5). The first terms describe the action of drift, the second terms describe the action of mutation, and the last two terms in the second line describe the action of selection. The contributions of drift cancel out in Equation (S2), so that $\dot{RH}$ only depends on contributions from selection and mutation:

$$\dot{RH} = s\frac{2\Gamma_{2,h|s} - \Gamma_{1,h|s}}{2\pi_{|n}} + 2u_{|n}\frac{\pi_{|s}}{\pi_{|n}^2} - \frac{2u_{|s}}{\pi_{|n}}. \tag{S4}$$

bioRxiv preprint doi: https://doi.org/10.1101/010934; this version posted March 11, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Now consider the rate of change in the difference ΔRH between populations A and B:

$$\begin{aligned}\Delta\dot{RH} =& s\Delta\frac{2\Gamma_{2,h|s}-\Gamma_{1,h|s}}{2\pi_{|n}}+2\Delta\left(u_{|n}\frac{\pi_{|s}}{\pi_{|n}^2}-u_{|s}\frac{1}{\pi_{|n}}\right)\\ =& s\Delta\frac{2\Gamma_{2,h|s}-\Gamma_{1,h|s}}{2\pi_{|n}}\\ &+2\Delta\left(\frac{1}{\pi_{|n}}\left(u_{|n}(1-RH)-u_{|s}\right)\right),\end{aligned} \tag{S5}$$

By contrast with Morton's efficacy, the mutation term does not cancel out for ΔRH.

To show that the selection term can be substantial, we consider the case of strong selection, where $1-RH \ll 1$. In this case, the contribution of mutation to $\Delta\dot{RH}$ is

$$-2u_{|s}\Delta\frac{1}{\pi_{|n}},$$

while the contribution of selection to $\Delta\dot{RH}$ is

$$s\Delta\frac{2\Gamma_{2,h|s}-\Gamma_{1,h|s}}{2\pi_{|n}}\simeq -2hs\Delta\frac{\mu_{1|s}}{\pi_{|n}}\simeq 2u_{|s}\Delta\frac{1}{\pi_{|n}} \tag{S6}$$

where the first approximation uses Equation (A5) and assumes that higher moments are sub-dominant under strong selection because allele frequencies are small. The second approximation uses the mutation-selection balance relation $\mu_{1|s}\simeq\frac{-u_{|s}}{hs}$ (e.g, equation 6.2.9 in [26]). The effects of mutation and selection on ΔRH cancel out: average frequencies of very deleterious alleles are governed by the mutation-selection balance and are unaffected by population size differences. Mutation tends to increase RH in the less diverse population.

Identifying the effect of selection on diversity therefore requires picking apart the effects of selection and mutation (if measuring ΔRH), or the effect of selection and drift (if measuring $\Delta\pi_s$). As in the case of fitness, knowing only the present-day distribution of allele frequency is not enough to identify the effect of selection unambiguously, unless the alternate effect can be shown to be weak.

S1.3 Example of two populations with identical start and end frequency distributions, but different measures of FIT efficacy

Imagine that all alleles all have dominance coefficient $0<h<1$, selection coefficient s and dominance h, and consider two identical populations with exactly L segregating alleles, all at frequency x_o at time t_o. The initial fitness in both populations is $W_o = Ls\left(2hx_o+(1-2h)x_o^2\right)$. Assume no mutation. In the second population, a short bottleneck occurs that increases the variance in allele frequency before selection had time to change mean allele frequencies. The difference in fitness between the two populations after the short bottleneck is $\delta W_{\text{bottleneck}} = Ls(1-2h)\sigma^2$, where σ^2 is the variance in allele frequency in the second population. The change in fitness in population 2 up to this point is entirely attributed to drift.

After the bottleneck, the population sizes increase to a very large value, so that subsequent effects of drift can be neglected. Selection is left to act until we have approached the maximum fitness state, which has fitness $W=Ls$ if $s>0$ and $W=0$ if $s<0$. The change in fitness $W-W_o$ is equal in both populations, since the initial and final states are exactly the same. The FIT 'effect of selection' is $W-W_o$ in population 1, since all changes are due to selection. In population 2, the effect of selection is $W-W_o-\delta W_{\text{bottleneck}}$, because a change $\delta W_{\text{bottleneck}}$ was caused by drift rather than by selection.

S1.4 Microscopic and macroscopic efficacy and intensity of selection

The main text discusses the effect of selection by averaging over all possible allelic trajectories. A natural question is whether individual allelic trajectories can be interpreted in a similar manner. Of course, it is not possible in general to attribute specific changes in frequency to the effect of drift or to selection, just as it is impossible to attribute the precise number of offspring borne by an individual to drift or fitness alone—these attributions can be made only in an average sense. However, individual trajectories can con-

bioRxiv preprint doi: https://doi.org/10.1101/010934; this version posted March 11, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

tain more information than allele frequency distributions, for example when selection coefficients fluctuate. Recent work has defined a microscopic 'rate of adaptation' [25] as a measure of the effect of selection on individual trajectories. Recent work has defined a microscopic 'rate of adaptation' [25] as a measure of the effect of selection on individual trajectories. Here we show how the different metrics of the impact of selection can be applied to individual trajectories, and how they relate to Mustonens rate of adaptation.

Consider an individual allelic trajectories. If the allele frequency has trajectory $\{x_t\}_{t=1,\ldots,T}$, with t the time in generations, we can write the fitness change ΔW_t at generation t as

$$W(x_t+\Delta x_t)-W(x_t) \simeq W'(x_t)\Delta x_t+W''(x_t)\frac{(\Delta x_t)^2}{2},$$

where $\Delta x_t = x_{t+1} - x_t$ and W' represents the partial derivative of the fitness function with respect to frequency x. In the constant-fitness models discussed above, $W'(x_i) = 2s(h + (1 - 2h)x)$ and $W''(x_i) = 2s(1 - 2h)$.

The expectation of $W'(x_t)\Delta x_t$ gives our $\dot{W}_s$, and the expectation of $W''(x_t)\frac{(\Delta x_t)^2}{2}$ gives $\dot{W}_N$ when $|s| \ll 1$. We therefore define the quantities $\sigma_t = W'(x_t)\Delta x_t$ and $\nu_t = W''(x_t)(\Delta x_t)^2$ as microscopic analogs of the macroscopic efficiency of selection $\dot{W}_s$ and drift $\dot{W}_N$. These are not the only possible analogs—for example, we could consider the expectation of the linear term $\overline{\sigma}_t$ as the microscopic effect of selection, and $\nu_t + \sigma_t - \overline{\sigma_t}$ as the microscopic effect of drift, without changing the expected values.

Mustonen and Lassig define a 'fitness flux' ϕ_t in [25] as a measure of the rate of adaptation. The fitness flux definition appears identical to our definition for σ, namely $\phi_t = \sum_i W'(y_i)\Delta y_i$, where y_i is an allele frequency trajectory sampled densely in time. However, it is not equal but corresponds to the total change in fitness, $\nu_t + \sigma_t$. Whereas our trajectory $\{x\}_t$ is labeled by the time in generation, the time steps in $\{y_i\}$ are chosen so that $\Delta y_i \ll \frac{1}{N}$. Because of this constraint, time steps in [25] must be finer than one generation, and the y_i must be interpolated within generations. While integrating over this smoothed trajectory, quadratic terms in Δy_t can be neglected: The integrated fitness flux yields the total rate of fitness change, whether it is due to drift or selection.

The intensity of selection i measures the rate of increase in frequency of favorable alleles. At a microscopic level, the intensity of selection is simply

$$i(t) = \frac{W'(x)}{|W'(x)|}\Delta x.$$

When $0 \leq h \leq 1$ and selection coefficients do not change sign over time, $\frac{W'(x)}{|W'(x)|}$ is 1 for favorable alleles, and -1 for deleterious ones: The intensity of selection is simply the average change in frequency of favorable alleles. If $h < 0$ or $h > 1$, we have overdominance, and the favored allele is frequency dependent. In that case, the intensity of selection is also independent of the trajectory followed, and is given by $I = -\operatorname{sgn}(sh)\left(|x_f - \bar{x}| - |x_i - \bar{x}|\right)$, where $\bar{x} = \frac{-h}{1-2h}$ represents the frequency of optimal fitness. As discussed in the main text, there are many ways to weight the intensity of selection when the effect is to be measured across multiple sites. As long as the weighing scheme is not frequency or time-dependent, it remains possible to directly compare the intensity of selection across populations without the need for detailed modeling.

bioRxiv preprint doi: https://doi.org/10.1101/010934; this version posted March 11, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

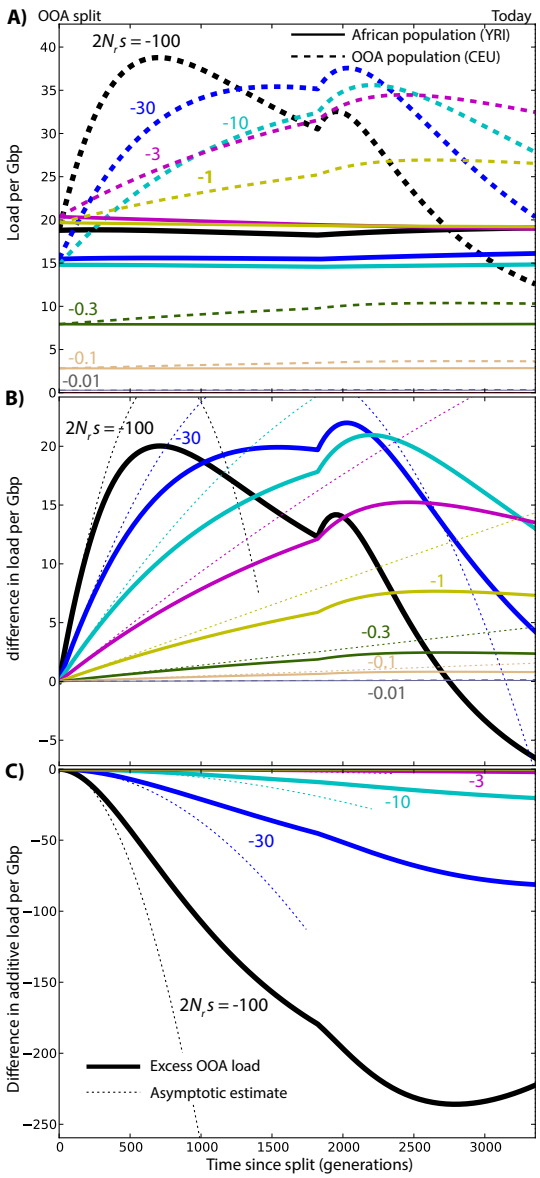

Figure S1: Changes in load after the OOA split model illustrated in Figure 3A, assuming a recessive variant ($h = 0.0$), as a function of s and given ancestral population size $N_r = 11,930$. (A) Overall genetic load evolution. The load due to variants that are fixed in all populations is not included. (B) Difference in load between the two populations. (C) Differences in additive load. The dashed lines represent the asymptotic results from Equations (8) and (10).

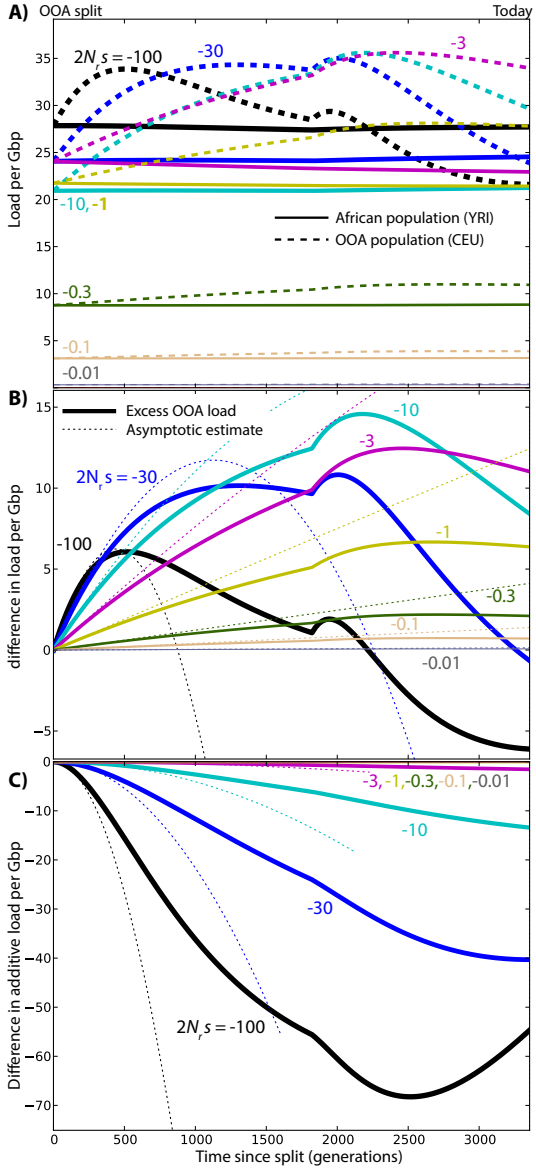

Figure S2: Changes in load after the OOA split model illustrated in Figure 3A, assuming partial recessive variants ($h = 0.05$) as a function of s and given ancestral population size $N_r = 11,930$. (A) Overall genetic load evolution. The load due to variants that are fixed in all populations is not included. (B) Difference in load between the two populations. (C) Differences in additive load. The dashed lines represent the asymptotic results from Equations (8) and (10).

bioRxiv preprint doi: https://doi.org/10.1101/010934; this version posted March 11, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

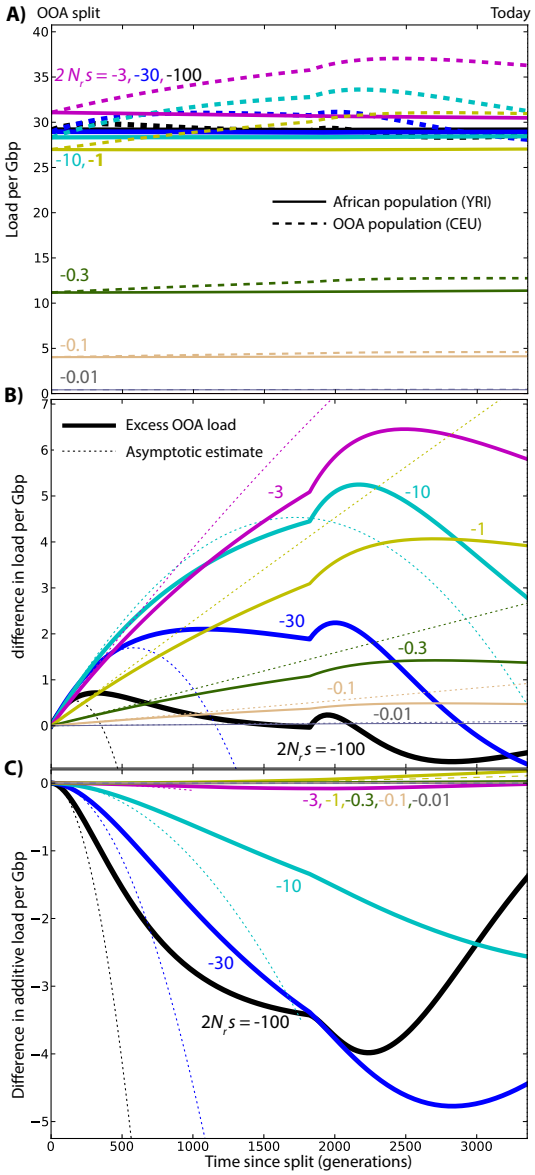

Figure S3: Changes in load after the OOA split model illustrated in Figure 3A, assuming partial recessive variants ($h = 0.2$) as a function of s and given ancestral population size $N_r = 11,930$. (A) Overall genetic load evolution. The load due to variants that are fixed in all populations is not included. (B) Difference in load between the two populations. (C) Differences in additive load. The dashed lines represent the asymptotic results from Equations (8) and (10).

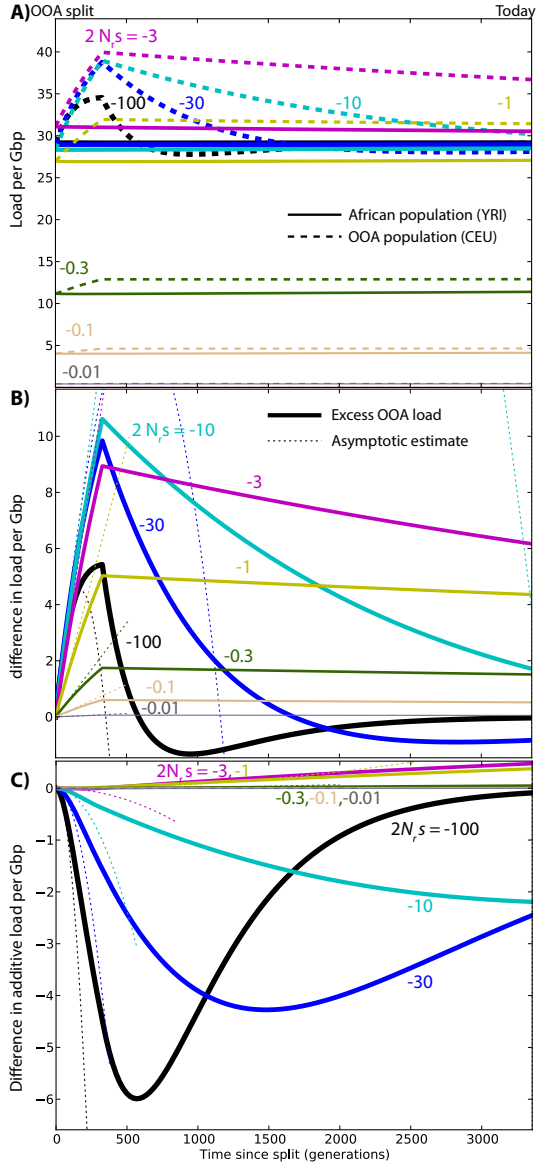

Figure S4: Changes in load after the simple OOA bottleneck model from [15], assuming partial recessive variants ($h = 0.2$) as a function of s and given ancestral population size $N_r = 11,930$. (A) Overall genetic load evolution. The load due to variants that are fixed in all populations is not included. (B) Difference in load between the two populations. (C) Differences in additive load. The dashed lines represent the asymptotic results from Equations (8) and (10).

bioRxiv preprint doi: https://doi.org/10.1101/010934; this version posted March 11, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

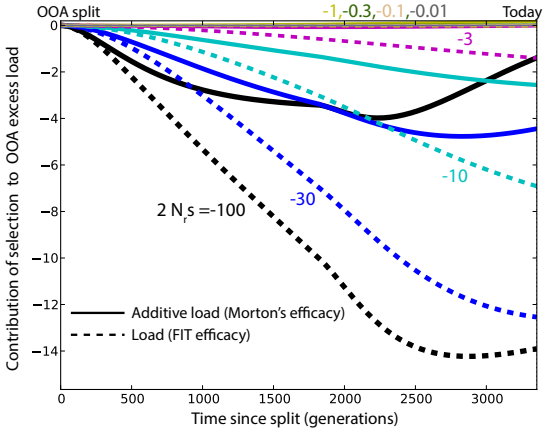

Figure S5: Integrated contributions of selection to load differentiation according to FIT (total load) and Morton (additive load) for partially recessive alleles ($h = 0.2$). Both show an overall increased effect of selection in the OOA population. However, genetic load remains higher in the OOA population because of the contribution of drift (Figure S3).

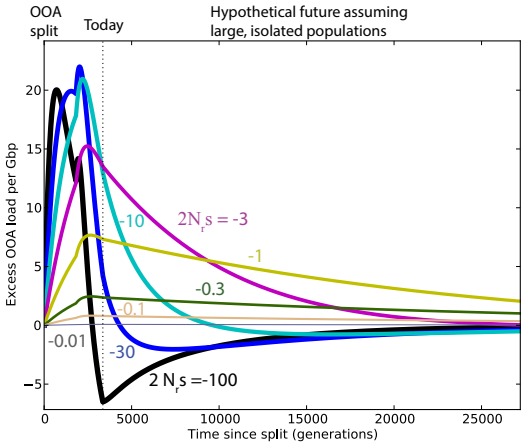

Figure S6: Excess recessive load in OOA populations, assuming continued future isolation between large populations ($N = 20N_r$). This serves to illustrate that the 'purging' effect of the bottleneck on deleterious variants is observed for all alleles with $N_r s < -3$, but that it would require much more time to compensate for the initial loss in fitness for mildly deleterious alleles with $N_r s > -30$.

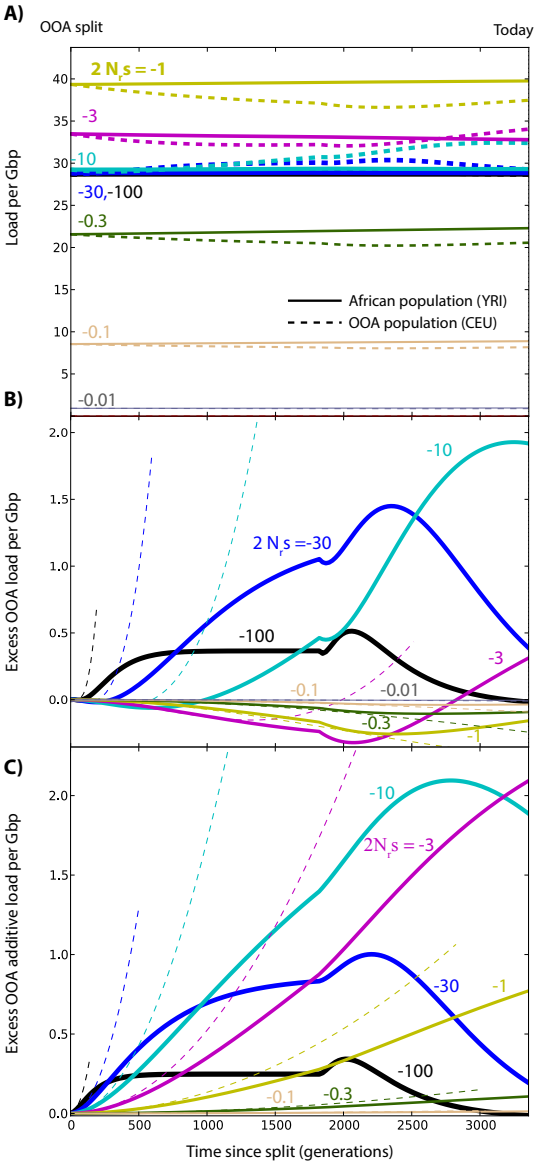

Figure S7: Changes in load after the OOA split model illustrated in Figure 3A, assuming dominant deleterious variants ($h = 1$), as a function of s and given ancestral population size $N_r = 11,930$. (A) Overall genetic load evolution. The load due to variants that are fixed in all populations is not included. (B) Difference in load between the two populations. (C) Differences in additive load. The dashed lines represent the asymptotic results from Equations (8) and (10).

bioRxiv preprint doi: https://doi.org/10.1101/010934; this version posted March 11, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

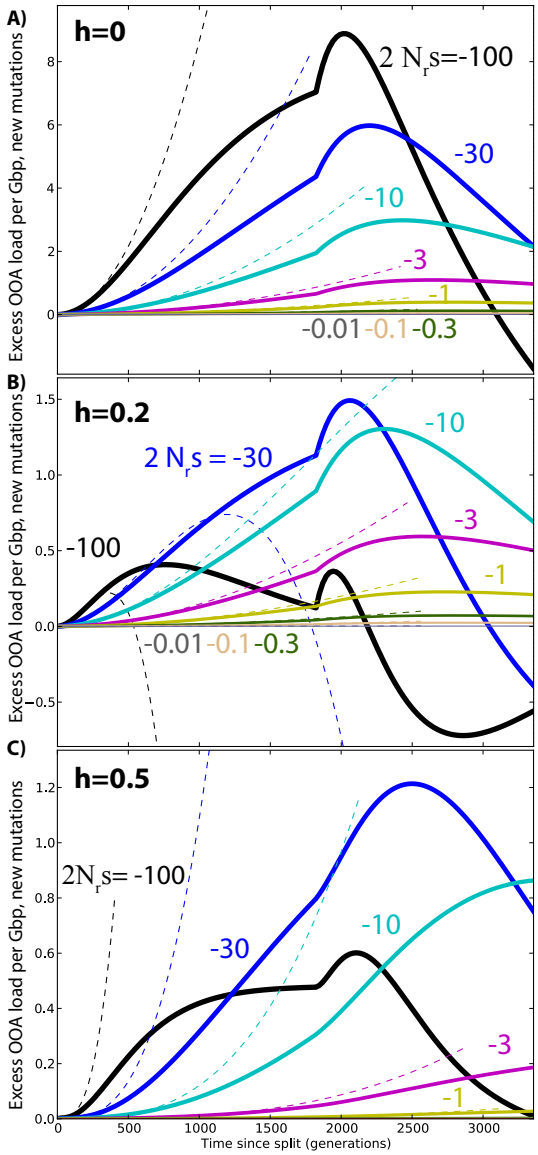

Figure S8: Contributions of recent mutations to load differentiation across populations, for different dominance coefficients. Here, 'new mutations' are mutations that occurred after the OOA split. Solid lines are the results of simulations, and dotted lines represent asymptotic values from Equation (A13).

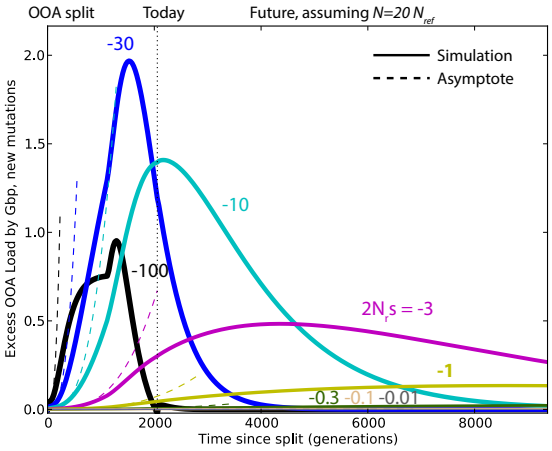

Figure S9: Changes in load caused by new mutations after the OOA split model illustrated in Figure 3A, given $h = 0.5$ and ancestral population size $N_r = 11,930$. The dashed lines represent the asymptotic results from Equation (A13).

bioRxiv preprint doi: https://doi.org/10.1101/010934; this version posted March 11, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

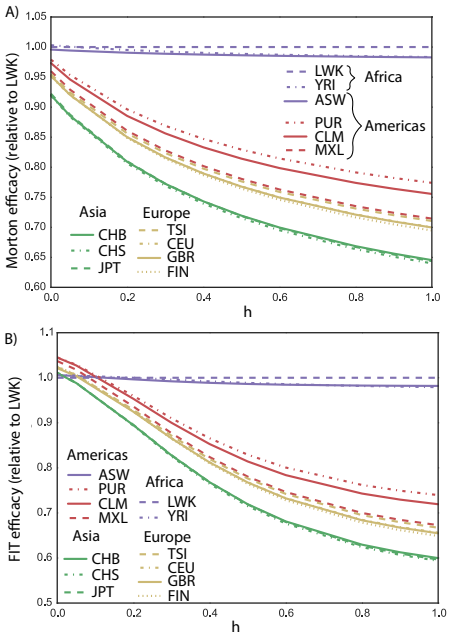

Figure S10: A) Present-day Morton efficacy of selection for nearly neutral variants, estimated from 1000 Genomes non-synonymous variation, assuming different dominance coefficients h B) FIT efficacy for the same variants.

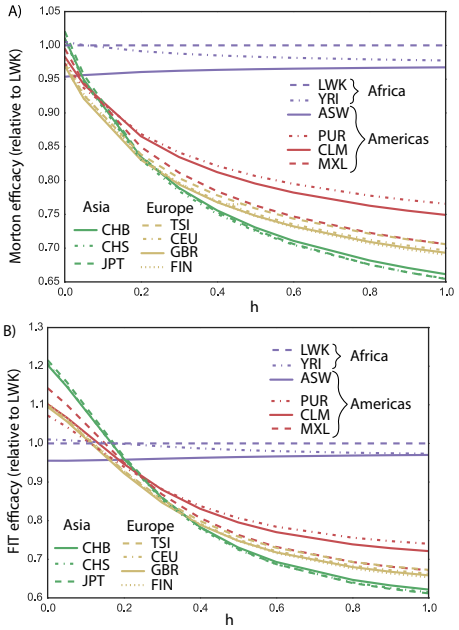

Figure S11: A) Present-day Morton efficacy of selection for nearly neutral variants, estimated from 1000 Genomes variants with CADD score above 2, assuming different dominance coefficients h. B) FIT efficacy for the same variants.