When is selection effective?

Simon Gravel, McGill University, Montreal, Canada

October 29, 2014

Abstract

Deleterious alleles are more likely to reach high frequency in small populations because of chance fluctuations in allele frequency. This may lead, over time, to reduced average fitness in the population. In that sense, selection is more 'effective' in larger populations. Many 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 seeks to clarify some terms of the debate by identifying differences in definitions and assumptions used in these studies and providing an intuitive explanation for the observed similarity in genetic load among populations. The intuition is verified through analytical and numerical calculations. First, even though rare variants contribute to load, they contribute little to load differences across populations. Second, the accumulation of non-recessive load after a bottleneck is slow for the weakly deleterious variants that contribute much of the long-term variation among populations. Whereas a bottleneck increases drift instantly, it affects selection only indirectly, so that fitness differences can keep accumulating long after a bottleneck is over. Third, drift and selection tend to have opposite effects on load differentiation under dominance models. Because of this competition, load differences across populations depend sensitively and intricately on past demographic events and on the distribution of fitness effects. A given bottleneck can lead to increased or decreased load for variants with identical fitness effects, depending on the subsequent population history. Because of this sensitivity, both classical population genetic intuition and detailed simulations are required to understand differences in load across 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]. The reduction in diversity has been observed in populations that have undergone strong population bottlenecks: Diversity is much decreased in populations that left Africa, and further decreases with successive founder events [2, 3, 4, 5]. 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 [6]. In humans, fitness is usually estimated through bioinformatic prediction [7, 8]. Lohmueller *et al.* [9] found a higher proportion of predicted delete-

rious variants among Europeans than 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 simulation studies [10, 11] 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. Peischl *et al.* [12], by contrast, have claimed significant differences among populations under range expansion models, and Fu [13] claims a slight excess in the number of deleterious alleles in European-Americans compared to that in African-Americans. These results have sparked a heated debate as to whether the efficacy of selection has indeed been different across human populations [14, 13].

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 relation between the fitness of a variant and its ultimate fate in the population—that is, *natural selection*—holds independently of the biology and the history of the population. However, the rate of increase in frequency for favorable alleles depends on mutation, dominance, linkage, and demography, and can vary across populations. Differences in this rate can be thought of as differences in the efficacy of selection.

The purpose of this article is to provide an intuitive but quantitative understanding of the interplay of mutation, selection, and drift in the variation in genetic load across human populations, and more specifically to discuss how the classical prediction about the reduced efficacy of selection in small populations could be verified in real populations.

1   Models

We will assume that alleles have constant fitness coefficients s and that selection is additive over genes (no epistasis). Given alleles a and A, we assume that genotype aa has fitness 1, aA has fitness $1 + 2sh$, and AA has fitness $1 + 2s$. In a random-mating population, an allele A at frequency x adds an average of $2s\left(2hx + (1 - 2h)x^2\right)$ to individual fitness compared to the aa genotype. Fitness and favorable allele frequency are proportional only for genic selection (no dominance; $h = 1/2$).

I consider natural selection to be *intense* if the average frequency of favorable alleles increases by a large amount. I will say that natural selection is *reliable* if the frequency of favorable alleles increases with high probability, and that it is *thorough* if a large proportion of deleterious mutations are eradicated from the population. In addition, I will say that *fitness increases* if the mean fitness in the population increases. Since the genetic load is the difference between the mean fitness and optimal fitness, a fitness increase corresponds to a reduction in the genetic load. Intensity, reliability, thoroughness, and increase in fitness are all measures of the dynamics of
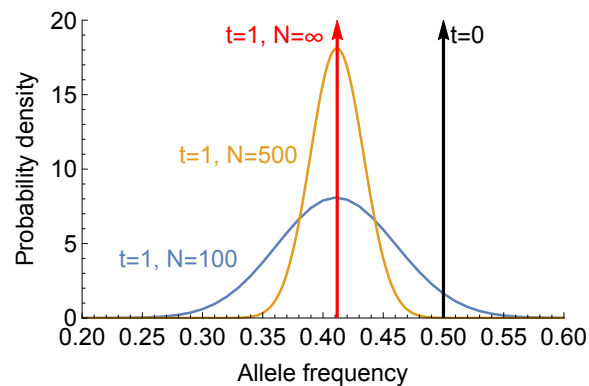

Figure 1: Frequency distribution of a deleterious allele after one generation, assuming population sizes of $2N = 100, 500$, and ∞; an initial frequency of 0.5; a selection coefficient $s = -0.3$; and no dominance ($h = 0.5$). The variance in frequency depends on N, but the mean does not. The probability of a deleterious variant increasing in frequency is larger in the smaller populations, but the genetic load is unchanged.

selected alleles that can be compared across populations, and that have been used to define the efficacy of selection [11, 9, 15].

If we wait for the frequency of alleles to reach 0 or 1, all four definitions are in agreement: In that case, the average frequency of favorable alleles equals the proportion of favorable alleles with frequency 1 and also equals the proportion of favorable alleles whose frequency has increased. However, these concepts are not equivalent if we consider evolution over a shorter time span. In some cases, differences across populations may have little to do with the action of selection. This short-term disagreement among definitions that are equivalent in the long term has led to much confusion (see [14] and references therein).

Figure 1 illustrates the difference between intensity, reliability, and thoroughness for populations with different sizes. During Wright–Fisher reproduction, an allele with parental frequency x, fitness coefficient s, and dominance coefficient h will be drawn with probability $x' \simeq x + 2sx(1 - x)(h + x(1 - 2h))$ in the descending population. Figure 1 shows the re-

sulting binomial distributions in offspring allele frequency for $x = 0.5$, $s = -0.3$, $h = 0.5$, and $2N = 100, 500$, and ∞. The average frequency x' is independent of N, hence the intensity of selection is the same in all populations. By contrast, the variance in allele frequency is much larger in the smaller population, and so selection is much less reliable. Since very little time has elapsed, a vanishingly small proportion of deleterious alleles have been eliminated. The thoroughness of selection over that period is close to 0. Finally, the sign and magnitude of fitness changes depend on the population size as well as on the selection and dominance coefficients.

If we let these populations evolve further, however, we will eventually find that deleterious allele frequencies decrease more slowly in the smaller population. This is because selection requires favorable alleles to outcompete deleterious ones, and has little effect on the frequency of very common deleterious variants. If drift pushes a deleterious variant to fixation, selection can no longer act to reduce its frequency. Mathematically, the average increase in frequency of favorable alleles after one generation, $\Delta x = 2sx(1 - x)(h + x(1 - 2h))$, is a convex function of x when $\frac{1}{3} \le h \le \frac{2}{3}$ (Figure 2). The smaller population, having accumulated more variance in allele frequency, will be able to eliminate fewer deleterious alleles. The intensity of selection does not depend on the current amount of drift in a population, but on the drift accumulated in previous generations. Conversely, drift during one generation can affect the intensity of selection for many future generations. When $0 \le h \le \frac{1}{3}$ or $\frac{2}{3} \le h \le 1$, the increase in favorable allele frequency Δx is not a convex function of x, and drift can lead to increased intensity of selection and eventually to increased fitness, as discussed below.

Figure 2 also suggests that substantial differences in the efficacy of genic selection require deleterious alleles to reach appreciable frequencies. Thus, rare alleles may contribute to load, but their contribution is relatively insensitive to recent demography as long as they are not pushed to high frequency.
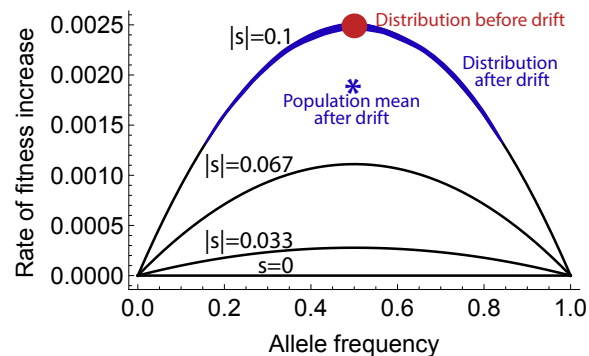

Figure 2: Effect of drift on the rate of fitness increase for $h = 0.5$. We consider two populations with initial allele frequency 0.5 and fitness coefficient $s = -0.1$. The red population does not undergo drift, and the blue population undergoes one generation of neutral drift, leading to increased variance in allele frequency. Even though drift has not affected the mean frequency of alleles in the blue population, the added variance results in a lower average *rate* of fitness increase per generation (blue star). The convexity of the curve means that this holds for arbitrary starting frequency.

2 Asymptotics

To quantify these arguments, we calculate the moments of the allele frequency distribution $\phi(x, t)$ under the diffusion approximation. Specifically, $\phi(x, t)$ represents the number of alleles with frequency x at time t. In a randomly mating population of size $N_e = N_e(t) \gg 1$, it obeys

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

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, with Dirac's

3

delta distribution δ, describes the influx of new mutations. 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 rate of change in allele frequencies $\frac{\partial \mu_1}{\partial t}$, is driven by mutation and selection:

$$\frac{\partial \mu_1}{\partial t} = u + \frac{s\Gamma_{1,h}}{2}, \qquad (2)$$

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 $\pi_{1,0} = \Gamma_{1,1/2}$ (see Appendix for detailed calculations). We can define the contributions of selection and mutation as $\dot{\mu}_{1s} \equiv \frac{s\Gamma_{1,h}}{2}$ and $\dot{\mu}_{1u} \equiv u$. The effect of mutation is constant and independent of population size, but the effect of selection is modulated by $\Gamma_{1,h}$, and therefore depends on the history of the population.

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

$$\dot{F} = \dot{F}_u + \dot{F}_N + \dot{F}_s$$
$$= 2s\Big[2hu + \frac{(1-2h)\pi}{N} \qquad (3)$$
$$+ s(h\Gamma_{1,h} + (1-2h)\Gamma_{2,h})\Big].$$

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

It is therefore natural to define the cumulative effect of selection on load as $F_s = \int_0^t \dot{F}_s dt$, the change in fitness caused directly by selection. Fisher's Fundamental Theorem, for example, equates the effect of selection on fitness change, $\dot{F}_s$, to the additive variance in fitness [16]. The relationship does not hold for total fitness changes $\dot{F}$ [17]: in the presence of dominance and drift, fitness can decrease despite the favorable action of selection. Similarly, a population in a fluctuating environment is constantly adapting, in the sense that $\dot{F}_s > 0$, but may not gain fitness in the long term [17, 18].

Now consider an ancestral population that splits into two randomly mating populations with initial sizes N_1 and N_2. The populations may experience migration and continuous size fluctuations. Using the moments method, the Appendix shows that the difference in fitness $\Delta(t) = F_1(t) - F_2(t)$ between populations 1 and 2 is

$$\Delta(t) = \frac{-s(1-2h)t\pi_{1,0}}{2}\left(\frac{1}{N_2} - \frac{1}{N_1}\right) + O(t^2), \quad (4)$$

where t is the time in generations, $\pi_{1,0}$ is the expected heterozygosity in the source population, and $O(t^2)$ represents terms at least quadratic in t—these will eventually dominate, but are small right after the split. This rapid, linear differentiation is entirely driven by drift coupled with dominance, and is independent of the effect of selection after the split. The smallest population has higher fitness when the heterozygote is at a disadvantage.

By contrast, the effect of selection on load differences $\Delta_s(t)$ grows only quadratically:

$$\Delta_s(t) = \frac{s^2 t^2 \Pi_h}{4}\left(\frac{1}{N_2} - \frac{1}{N_1}\right) + O(t^3), \qquad (5)$$

where Π_h is a moment of the ancestral frequency distribution that also reduces to $\pi_{1,0}$ when $h = 0.5$ (see Appendix). This slower response is the mathematical consequence of the intuition provided by Figures 1 and 2: Differences in drift need to accumulate before differences in the rate of selection can be observed, and differences in the rate of selection need to accumulate to produce differences in fitness.

The effect of new mutations is even slower to appear, since the direct effect of mutation on load is independent of demography [Equation (3)], as is the combined effect of mutation and selection (Table 1). An extra factor of ut accounts for the time necessary for new mutations to accumulate in the population, before drift and selection can induce differences in load:

$$\Delta_{s,new}(t) = \frac{2s^2 t^3 uh(5h-2)}{3}\left(\frac{1}{N_2} - \frac{1}{N_1}\right) + O(t^4). \qquad (6)$$

Finally, even though a bottleneck inexorably leads to increased load when no dominance is present, we
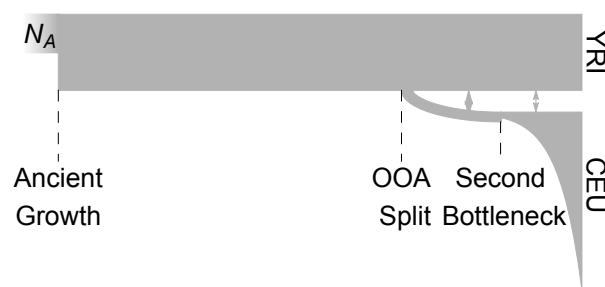
Figure 3: The demographic model used in simulations, where time flows from left to right and population sizes are indicated by line thickness. Vertical arrows represent migrations. The OOA split was inferred at 51 thousand years ago in [20], using a mutation rate of $\mu = 2.35 \times 10^{-8}$ per base pair per generation. Recent estimates of the mutation rate suggest a lower value for μ [21]. Using $\mu = 1.2 \times 10^{-8}$, we get a split at 100kya, which would imply that the first bottleneck started well before the Out-Of-Africa event.

show in the Appendix that the generic intermediate-term effect of a bottleneck is to reduce the genetic load caused by recessive variants. However, we will see in simulations that the familiar short-term increase in recessive load can last hundreds or thousands of generations for weakly deleterious variants.

3   Simulations

Evolution was simulated using $\partial a \partial i$ [19] and the Out-Of-Africa demographic model inferred from synonymous variation in [20] and illustrated in Figure 3. Simulated genetic loads were obtained for all combinations of selection coefficients $\gamma \equiv 2N_r s \in \{0, -0.01, -0.1, -0.3, -1, -3, -10, -30, -100\}$, with $N_r = 7300$, and dominance coefficients $h \in \{0, 0.05, 0.5, 1\}$. The contributions of selection and drift were obtained using Equation (3). Simulations were carried beyond the present time by assuming equal and large population sizes ($N_e = 20N_r$) to emphasize the long-lasting effect of past drift on the efficacy of selection. In all cases, Equations (4), (5),

and (6) capture the initial increase in load (Figures 4B, 5B, 6B, and 7B).

3.1   Genic selection; $h = 1/2$

The models predict differences in load that are small and limited to intermediate-effect variants ($.3 < |\gamma| < 30$, $2 \times 10^{-5} < |s| < 0.002$). Assuming the distribution of fitness effects from Boyko *et al.* [22], the excess load in the OOA population is about 0.27 per Gb of amino-acid–changing variants, compared to a total load of 18.69 per Gb. 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 [23], the model predicts a non-synonymous load difference of 0.01. The total estimated non-synonymous load, excluding mutations fixed in the ancestral state, is 0.7. In this model, the reduced efficacy of selection caused by the OOA bottleneck leads to a relative increase in non-recessive load of 1.4%. Since we did not consider fixed ancestral deleterious alleles in the total load, we expect the relative increase in load due to the bottleneck to be even smaller. The relative increase reaches a maximum of 8% for mutations with $-20 < \gamma < -10$.

The contribution of new mutations is important for very deleterious variants, but these contribute a small fraction of total load differences. The fast response of very deleterious variants to changes in population sizes described in Equation (5) can also be seen at the time of the second bottleneck in Figures 4B and 4C. By contrast, load due to benign additive variants reacts much more slowly: To date, its evolution is almost entirely determined by ancient diversity and the size of the early bottleneck.

3.2   Partial and complete dominance

The picture changes dramatically when we consider recessive deleterious variants ($h = 0$, Figure 5). Reactions to changes in population size are linear rather than quadratic, and they are more substantial than in the additive case (Figure 5). The load due to variants with $\gamma = -100$ almost doubles after 500 generations, excluding fixed ancestral deleterious alleles. This excess load in the OOA population is due entirely to
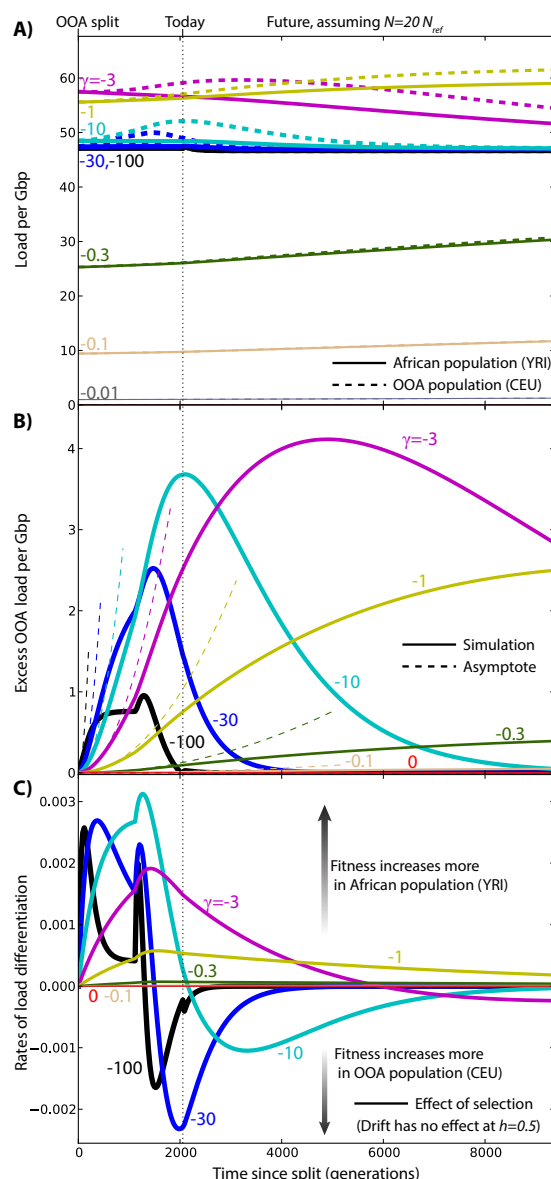
Figure 4: Changes in load after the split between ancestors of Yoruba and CEU populations, assuming no dominance ($h = 0.5$) and $\gamma = 2N_A s$. (A) Overall genetic load evolution. We subtracted the load due to variants fixed in all populations. (B) Difference in load between the two populations. The dashed lines represent the asymptotic results derived in the text. (C) Rate of load differentiation due to selection. Drift does not contribute to load when $h = 0.5$.

drift, and leads to an increased intensity of selection in the OOA population (Figure 5C), since a higher proportion of deleterious alleles are now visible to selection. The difference in load for the most deleterious variants is not sustained. Both the number of very deleterious variants and the associated genetic load eventually becomes higher in the Yoruba population model. In this case, both the early bottleneck and later growth in the European population model act to reduce recessive load from strong-effect variants. By contrast, weak-effect deleterious variants contribute more load in the European population model.

The opposite occurs for dominant deleterious variants (Figure 6). Drift tends to increase fitness by combining more of the deleterious alleles into homozygotes, reducing their average effect on fitness. The difference is much less pronounced and less sustained than in the recessive case. Equation (4) 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 drift. Here again, the population with the highest load depends on the selection coefficient, with a higher load in the European population model for strongly deleterious variants and a higher load in the Yoruba population model for the weakly deleterious variants.

If a deleterious allele is not completely recessive ($h = 0.05$, Figure 7), load differentiation is suppressed for the most deleterious alleles. This effect can be traced back to a reduction in the initial diversity. However, load differentiation is largely unchanged compared to that in the recessive case for weakly deleterious alleles.

4 Discussion

Selection affects evolution in many ways. It tends to increase the frequency of favorable alleles and the overall fitness of a population living in a constant environment, 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.
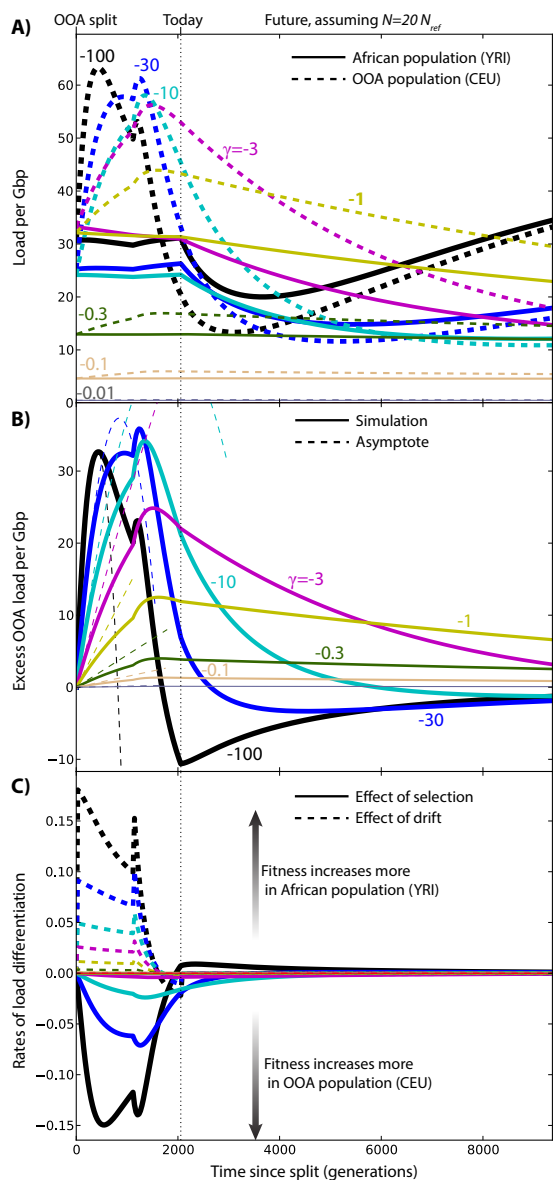
Figure 5: Inferred changes in load after the split between ancestors of the Yoruba and CEU populations, assuming a recessive deleterious variant ($h = 0$) and $\gamma = 2N_A s$. (A) Overall genetic load evolution. The load due to variants that are fixed in all populations is not included. (B) Difference in load between the two populations. The dashed lines represent the asymptotic results derived in the text. (C) Rate of load differentiation due to selection and drift.
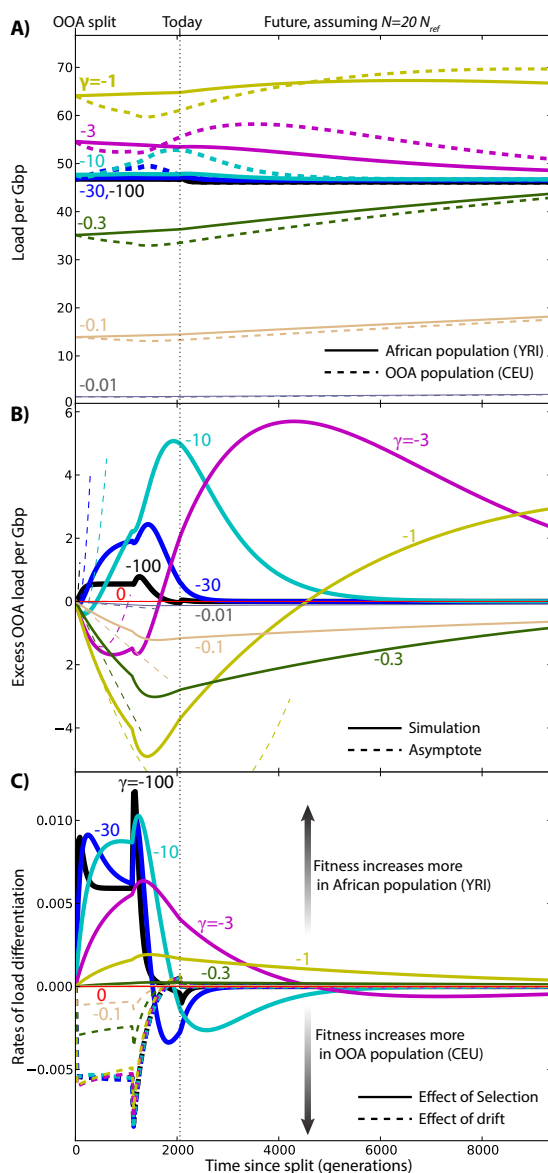
Figure 6: Inferred changes in load after the split between ancestors of the Yoruba and CEU populations, assuming a dominant deleterious variant ($h = 1$ and $\gamma = 2N_A s$). (A) Overall genetic load evolution. The load due to variants that are fixed in all populations is not included. (B) Difference in load between the two populations. The dashed lines represent the asymptotic results derived in the text. (C) Rate of load differentiation due to selection and drift.
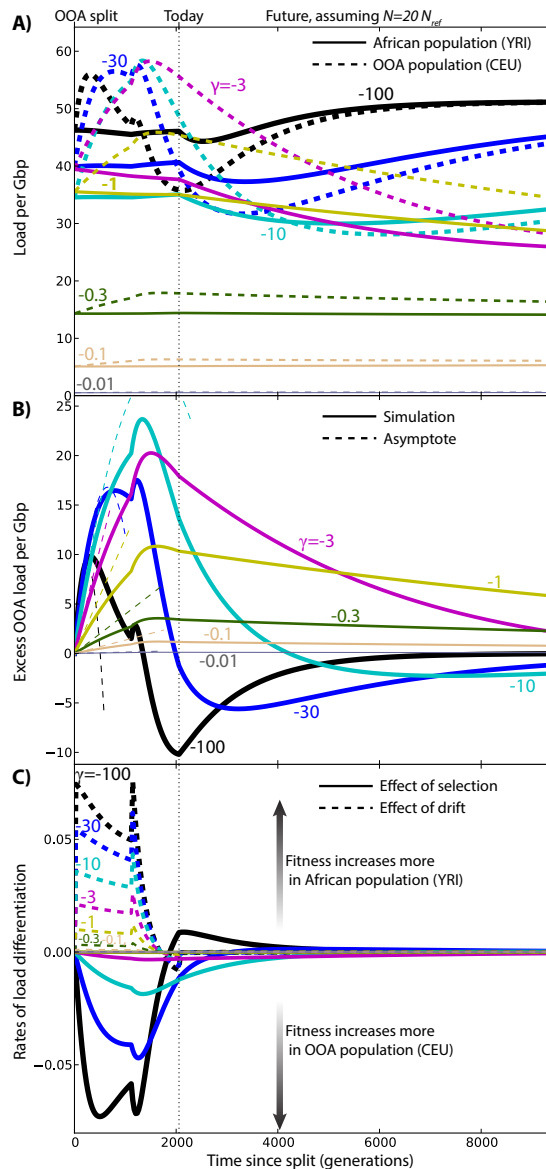
Figure 7: Inferred changes in load after the split between ancestors of the Yoruba and CEU populations, assuming incomplete dominance ($h = 0.05$) and $\gamma = 2N_A s$. (A) Overall genetic load evolution. The load due to variants that are fixed in all populations is not included. (B) Difference in load between the two populations. The dashed lines represent the asymptotic results derived in the text. (C) Rate of load differentiation due to selection and drift.

Defining such an 'efficacy' suggests a purpose to evolution, and unacknowledged disagreement about this purpose can lead to confusion.

One class of definitions focuses on the outcome of the entire evolutionary process: Selection is deemed effective if the final product looks 'selected'—as measured by fitness, by the frequency of deleterious alleles, or by the relative impact of fitness and drift. This approach is convenient, because it enables us to compare the 'efficacy' of selection across populations without having to worry about what exactly led to the current state. On the other hand, because it is not tied to what selection is actually doing, it can lead to paradoxes. For example, considering only fitness differences would force us to conclude that selection is 'effective' even during a period when it is entirely turned off.

The other class of definitions attempts to measure what selection is actually doing. Here we define the efficacy of selection as the instantaneous rate of fitness increase due to selection, $\dot{F}_s$. This is the definition implicitly used by Fisher in the derivation of his fundamental theorem [17], which relates the direct effect of selection on fitness to the additive variance in fitness. Importantly, $\dot{F}_s$ does not correspond to the total rate of fitness increase, because drift and mutation can impact fitness as well. Unfortunately, Fisher's efficacy is difficult to measure even for constant fitness, because it requires dense time-series data or accurate models.

An alternate observable is the intensity of selection, that is, the rate of increase in favorable allele frequency, which has expectation $\dot{\mu}_{1s}$. The intensity of selection can be directly attributed to selection, and it can be compared across populations without the need for time-series data. In addition, because the weight we assign to each variant can be an arbitrary constant, we can use any a priori information about the importance of the variant in our definition. A polyphen-weighted intensity of selection is equally valid a measure of the effects of selection as a gerp-weighted one or a selection-coefficient–weighted one.

There are many ways to assess the role of selection in shaping genetic diversity. The optimal measure ultimately depends on the biological phenomenon we are attempting to model. When seeking to determine

8

whether selection is 'less effective' in a population going through a bottleneck, I believe that the effect that we are trying to demonstrate is the one illustrated in Figures 1 and 2: Drift tends to reduce the number of beneficial alleles that an average deleterious allele has to outcompete. This effect can be measured equally as a change in the efficacy of selection, as in Fisher's Fundamental Theorem, or as a change in the intensity of selection. In cases with dominance, drift initially affects the efficacy of selection by increasing homozygosity. The more subtle action of drift—the reduction in the number of competitors per deleterious allele—can be the dominant long-term effect, but it may be extraordinarily difficult to measure at short time scales given the overbearing effect of drift on homozygosity.

Recessive, fairly deleterious alleles have the most potential to cause substantial differences in load across populations, because drift causes rapid and pronounced increase in recessive load in a bottleneck population. However, this increase in load is tied to an increase in the efficacy and intensity of selection: Drift pushes rare recessive alleles to frequencies where selection is more likely to remove them. These competing effects create complex dynamics. In simulations, the CEU model population showed a reduced load caused by very deleterious recessive variants—and an increased load due to intermediate-effect recessive variants—compared to the YRI model. By contrast, differences in load for additive variants systematically favor the larger population, but are smaller and limited to a narrow range of fitness effects. In both cases, the effect of the bottleneck on the efficacy of selection continues long after the bottleneck is over and is modulated by the later history of the populations: The OOA bottleneck has had a different effect on the efficacy of selection for all the populations that have experienced it.

5   Acknowledgements

References

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

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

[3] 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.

[4] Consortium GP, et al. (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature* 491:56–65.

[5] 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.

[6] Reed D, Frankham R (2003) Correlation between Fitness and Genetic Diversity. *Conservation biology.*

[7] 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.

[8] 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.

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

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

[11] Do R, et al. (2014) No evidence that natural selection has been less effective at removing deleterious mutations in Europeans than in West Africans. *arXiv*.

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

[13] 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.

[14] Lohmueller KE (2014) The impact of population demography and selection on the genetic architecture of complex traits. *PLoS Genet* 10:e1004379.

[15] 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.

[16] Fisher RA (1958) *The genetical theory of natural selection, 2nd edition* (Dover Publications, New York).

[17] Price GR (1972) Fisher's 'fundamental theorem' made clear. *Ann. Hum. Genet.* 36:129–140.

[18] Mustonen V, Lassig M (2010) Fitness flux and ubiquity of adaptive evolution. *Proceedings of the National Academy of Sciences* 107:4248–4253.

[19] 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.

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

[21] Scally A, Durbin R (2012) Revising the human mutation rate: implications for understanding human evolution. *Nat Rev Genet* 13:745–753.

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

[23] 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.

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

6 Appendix

6.1 Background

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

$$
\begin{aligned}
\frac{\partial \phi(x,t)}{\partial t} \simeq &\frac{1}{4N_e}\frac{\partial^2}{\partial x^2}x(1-x)\phi(x,t) \\
&- 2s\frac{\partial}{\partial x}\left(h + (1-2h)x\right)x(1-x)\phi(x,t) \\
&+ 2N_e u\delta(x - \frac{1}{2N_e}).
\end{aligned}
\tag{7}
$$

The mutational model, described using Dirac's δ, is an infinite-sites model that neglects the possibility of back mutations. A complete solution of this problem can be expressed as a superposition of Gegenbauer polynomials [24]. However, here we are looking for simple asymptotic results that will help us understand the dynamics of the problem. We integrate both sides using $\int_{0^+}^{1^-} dx\, x^k$ to obtain the time dependence of the moments of the allele frequency distribution: $\mu_k \equiv \int_{0^+}^{1^-} dx\, x^k\phi(x,t)$. Thus, μ_0 is the number of segregating sites, and μ_1 is the average allele frequency over all sites. Using some integration by parts, we get

$$
\frac{\partial}{\partial t}\mu_0 = -\frac{\phi(0,t) + \phi(1,t)}{4N_e} + 2N_e u,
$$

where $\phi(0,t)$ and $\phi(1,t)$ are defined by continuity from the function values in the open interval $(0,1)$.

That is, we lose segregating sites due to fixation, and we win them due to mutation. Similarly, if $h = 1/2$,

$$
\frac{\partial}{\partial t}\mu_1 = -\frac{\phi(1,t)}{4N_e} + s(\mu_1 - \mu_2) + u.
$$

Segregating alleles can disappear due to fixation, change in frequency due to selection, or appear due to mutation. Of course, alleles that fix at frequency 1 do not really disappear; they just stop segregating. If we keep track of fixed alleles, the first term vanishes. If we account for both segregating and fixed non-ancestral sites, the general term is

$$
\frac{\partial}{\partial t}\mu_k = \frac{k(k-1)}{8N_e}\pi_{k-1} + \frac{sk}{2}\Gamma_{k,h} + \frac{u}{(2N)^{k-1}}, \quad (8)
$$

where π_k and $\Gamma_{k,h}$ are moments of the distribution:

$$
\pi_k = 2(\mu_k - \mu_{k+1})
$$

and

$$
\Gamma_{k,h} = 2\left(h\pi_k + (1-2h)\pi_{k+1}\right).
$$

6.2 Response in allele frequencies

Solving Equation (8) is challenging, but it can be used to calculate the response of allele frequency to a sudden change in demographic or selective conditions. Consider a population of size N_0 that experiences a change in size to N_1 at time $t = 0$. We can expand μ_k for short times:

$$
\mu_k(t) = \mu_{k,0} + \mu_{k,1}t + \mu_{k,2}t^2 + O(t^3),
$$

where $\mu_{k,0}$ is the kth moment prior to the population size change. The other terms can be evaluated by collecting powers of t in Equation (8). For example, we get

$$
\mu_1(t) = \mu_{1,0} + (s\frac{\Gamma_{1,h}}{2} + u)t + O(t^2).
$$

Load increases even if $N_1 = N_0$, since our model assumes a constant supply of irreversible mutations. Since this linear term is independent of N_1, it does not contribute to differences across populations that share a common ancestor. Differences appear at the next order:

$$
\begin{aligned}
\Delta\mu_1(t) =&\frac{st^2}{4}\left((3h-1)\pi_{1,0} + 2(1-2h)\pi_{2,0}\right) \times \\
&\times \left(\frac{1}{N_2} - \frac{1}{N_1}\right) + O(t^3).
\end{aligned}
\tag{9}
$$

6.3 Response in genetic load

To compute the fitness in the diploid case, we write

$$
F = 2s\left(2h\mu_1 + (1-2h)\mu_2\right).
$$

Using (8), we get

$$
\dot{F} = \dot{F}_s + \dot{F}_u + \dot{F}_N,
\tag{10}
$$

where

$$\dot{F}_s = 2s\left(s\left(h\Gamma_{1,h} + (1-2h)\Gamma_{2,h}\right)\right),$$
$$\dot{F}_u = 2s\left(2hu\right),$$
$$\dot{F}_N = 2s\left(\frac{(1-2h)}{4N}\pi_1\right)$$

(11)

are the contributions of selection, mutation, and drift to changes in fitness. The mutation term is constant in time and independent of population size; it will not lead 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:

$$\Delta F = F_1 - F_2 = \frac{-s(1-2h)t\pi_{1,0}}{2}\left(\frac{1}{N_2} - \frac{1}{N_1}\right) + O(t^2),$$

where $\pi_{1,0}$ is the heterozygosity in the ancestral population. This reduction in load is driven purely by drift and dominance, and would happen even if the selection term were removed from the diffusion equation. It does not lead to 'natural selection', in that it does not increase the frequency of the favorable allele.

The changes in load due to selection are quadratic in time:

$$\Delta F_s = \frac{s^2t^2\Pi_h}{4}\left(\frac{1}{N_2} - \frac{1}{N_1}\right) + O(t^3),$$

where Π_h is a moment of the ancestral frequency distribution:

$$\Pi_h = 4h\pi_{1,0}(5h-2) - 12\pi_{2,0}(1-2h)(1-4h)$$
$$+ 24\pi_{3,0}(1-2h)^2,$$

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. If we imagine that the source population has frequency distribution $\phi(x) = \delta(x-f)$, we can compute the efficacy of selection as a function of h and the frequency f (Figure 8).
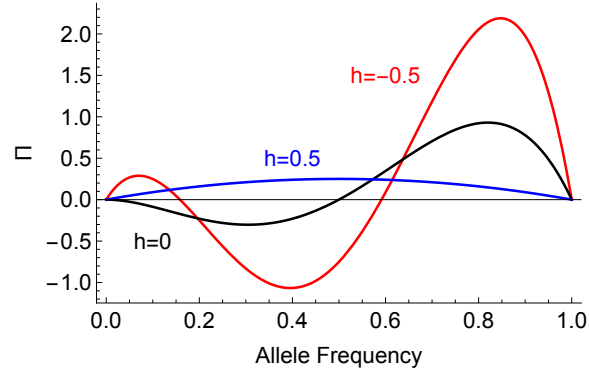

Figure 8: Π_h as a function of initial allele frequency and dominance ratio, representing the difference between the effect in a large population and that in a small population. Negative values indicate that selection has a bigger impact in the smaller population. This happens when drift increases the load, bringing more variants to the attention of selection.

Quantity	Expectation $(t \ll N)$
Frequency of mutation created t generations ago	$\frac{(1-2hs)^t}{N_t}$
Individual fitness reduction due to a mutation created t generations ago	$\frac{2hs(1-2hs)^t}{N_t}$
Number of new mutations created t generations ago	$N_t\mu$
Load from all mutations created t generations ago	$2hs(1-2hs)^t\mu$
Load from all mutations created in last t generations	$(1-(1-2hs)^t)\mu$

Table 1: Expected impact of demography on recent deleterious variation, assuming $t \ll N$. As long as the number of generations t is much smaller than the population size N, copies of the new allele evolve almost independently of one another, and roughly independently of the overall population size N.
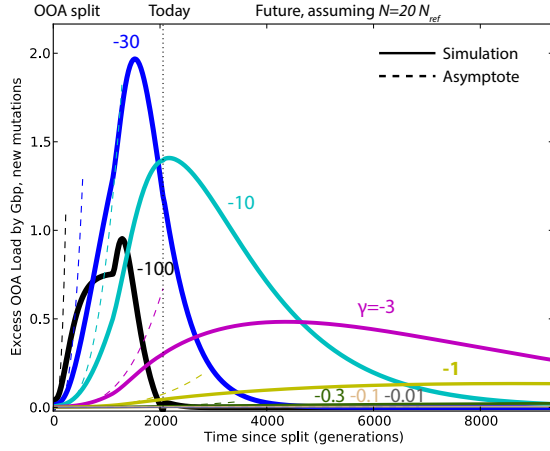
12

Figure 9: Changes in load after the split between ancestors of the Yoruba and CEU populations due to post-split mutations, assuming $h = 0.5$ and $\gamma = 2N_A s$. The dashed lines represent the asymptotic results derived in the text.

6.4 Effect of new mutations

If we set $\pi_{i,0} = 0$ 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:

$$\Delta F_{new} = \frac{-sut^2(1 - 2h)}{2}\left(\frac{1}{N_2} - \frac{1}{N_1}\right),$$

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

$$\Delta F_{s,new} = \frac{2s^2 t^3 uh(5h - 2)}{3}\left(\frac{1}{N_2} - \frac{1}{N_1}\right).$$

6.5 Intermediate and long-term effects of a bottleneck under a completely recessive model

Under recessive models, drift will make deleterious alleles more visible to selection, and therefore selection will become more intense. However, drift also increases the expected genetic load by increasing homozygosity. The short-time asymptotic results tell us that drift acts faster than selection, and that the initial effect of the bottleneck will be an increase in load. Can the increased intensity of selection eventually overcome the increased homozygosity and lead to an overall reduction in load? Kimura found that for $h = 0.05$, smaller populations could have a slightly reduced load compared to larger populations [1], leading us to expect that drift can indeed increase fitness in some cases. In fact, increased fitness to deleterious alleles may be the generic response to a bottleneck for all but the very shortest and very longest times.

To see this in the recessive case ($h = 0$), consider two randomly mating populations with identical frequency spectrum $\phi_0(x)$ at $t = 0$ and identical infinite population sizes for $t > 0$, except for a brief but intense bottleneck in population 1. The expected load contributed by mutations that occur after the bottleneck are identical in the two populations, so differences in load can be traced back to the effect of drift on the initial distribution. Furthermore, drift after the bottleneck can be ignored, and the frequency of an allele with initial frequency x_0 can be approximated by

$$\frac{\partial x}{\partial t} \simeq 2sx^2(1 - x).$$

If $-st \gg 1$, we have

$$x \simeq \frac{1}{-2st + \frac{1}{x_0} + \log\left(\frac{1}{x_0} - 1\right)}.$$

The fitness is simply $F(t) = 2sx(t)^2$, and therefore

$$F(x_0, t) \simeq \frac{2s}{\left(-2st + \frac{1}{x_0} + \log\left(\frac{1}{x_0} - 1\right)\right)^2}. \quad (12)$$

To study the effect of drift on fitness, we can use intuition similar to that suggested by Figure 2: If the genetic load $-F(x_0, t)$ is a concave function of x_0, drift leads to higher load. Conversely, a convex function means that drift leads to reduced load. Figure 10 shows that the function becomes convex after a finite amount of time, at least for variants with low frequency. We have

$$\frac{\partial^2 F}{\partial x_0^2} = \frac{s(2 - 3x_0)}{2(st)^3(x_0 - 1)^2 x_0^3} + sO\left(\frac{1}{(st)^4}\right).$$
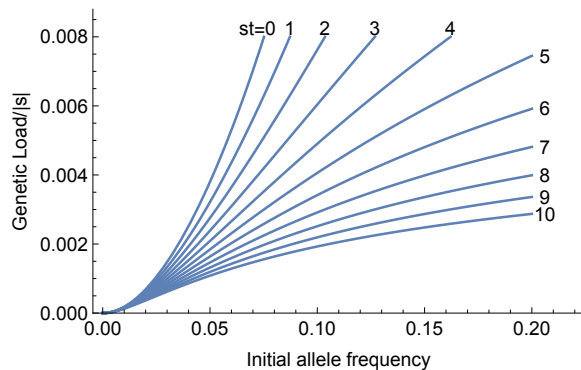
13

Figure 10: Genetic load at time t as a function of initial allele frequency for variants with fitness s, according to equation (12). An early bottleneck spreads the initial allele frequencies, leading to increased load when the function is concave but decreased load when it is convex. See also Figure 2.

For variants with $x_0 < \frac{2}{3}$, a small spread in allele frequency leads to a decrease in load. However, even a slight amount of drift can lead a minuscule number of variants to fixation. After a long enough time, unfixed deleterious variants will be eliminated by selection, and only fixed variants will be kept. It may take a very long time to reach this situation, but this guarantees that the very long-term effect of the bottleneck in the infinite-sites model is a reduction in fitness. However, even though the short- and long-term consequence of a bottleneck is an increase in genetic load, the generic behavior for intermediate times is an increase in fitness.

6.6 Microscopic and macroscopic efficacy and intensity of selection

The main text discusses the evolution of the expected allele frequency distributions by averaging over all possible allelic trajectories. Here we want to measure the efficacy and intensity of selection for individual allelic trajectories. If the allele frequency has trajectory $\{x_t\}_{t=1,\dots,T}$, with t the time in generations, we can write the fitness change ΔF_t at generation t as

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

where $\Delta x_t = x_{t+1} - x_t$ and F' represents the partial derivative of the fitness function with respect to frequency x. In the constant-fitness models discussed above, $F'(x_i) = 4s(h + (1 - 2h)x)$ and $F''(x_i) = 4s(1 - 2h)$. 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. We note that the expectation of $F'(x_t)\Delta x_t$ gives our $\dot{F}_s$, and the expectation of $F''(x_t)\frac{(\Delta x_t)^2}{2}$ gives $\dot{F}_N$ when $|s| \ll 1$. We therefore define the quantities $\sigma_t = F'(x_t)\Delta x_t$ and $\nu_t = F''(x_t)(\Delta x_t)^2$ as microscopic analogs of the macroscopic efficiency of selection $\dot{F}_s$ and drift $\dot{F}_N$. 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. The combination $\sigma_t + \nu_t$ corresponds to the change in fitness of existing alleles, or 'fitness flux' ρ_t (labeled ϕ_t in [18]). The fitness flux definition is almost identical to our definition for σ, namely $\rho_t = \sum_i F'(y_i)\Delta y_i$, where y_i is a dense allele frequency trajectory. Whereas our trajectory $\{x\}_t$ is labeled by the time in generation, the time steps in $\{y_i\}$ are chosen so that $\Delta y_i \ll \frac{1}{N}$. In other words, frequencies y_i are interpolated within generations so that the quadratic terms can be neglected.

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

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

When $0 \leq h \leq 1$ and selection coefficients do not change sign over time, $\frac{F'(x)}{|F'(x)|}$ 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.